Escola Tècnica Superior d'Enginyeria de Telecomunicació de Barcelona

UNIVERSITAT POLITÈCNICA DE CATALUNYA

GALLUS

A Video Database for Analyzing Affective Physiological Responses

Bachelor's Thesis
Telecommunications Engineering
Audiovisual Systems

Author: Marcel Granero Moya
Advisors: Xavier Giró-i-Nieto (UPC)
Damian Borth (ICS-HSG)
Barbara Weber (ICS-HSG)

Universitat Politècnica de Catalunya (UPC)
University of St.Gallen (HSG)
2018 - 2019

Abstract

Affective computing, leveraged by machine learning techniques, is advancing rapidly in the task of affect recognition in videos. However, there is a need for more annotated data. Several studies have built huge video datasets with emotions annotations. Others have collected music videos or film scenes datasets with physiological signals. Yet, none of them approached a solution with both physiological signals and user-generated videos.

The work introduced here presents *GALLUS*, a novel database of user-generated videos with affective physiological responses. The database is composed of 775 videos that have been previously annotated through an online crowdsourcing platform. Physiological responses such as electroencephalography, electrocardiography, galvanic skin response, facial emotion recognition, and eye-gaze have been collected from 30 participants while they watched the stimuli. Our dataset will be made public to foster research in affect recognition.

Keywords: User-generated videos, physiological signals, affective computing, crowdsourcing.

Resum

La computació afectiva aprofita de les tècniques d'aprenentatge automàtic per avançar ràpidament en la tasca del reconeixement d'emocions en vídeos. Tanmateix, calen més dades anotades. Diversos estudis han construït grans bases de dades de vídeos amb anotacions d'emocions. Altres han recopilat bases de dades de vídeos musicals o escenes de pel.lícules conjuntament amb senyals fisiològiques. Però, cap d'aquests treballs ha abordat una solució tant amb senyals fisiològics com amb vídeos generats per usuaris.

En aquest treball presentem *GALLUS*, una nova base de dades de vídeos generats per usuaris amb respostes fisiològiques afectives. La base de dades es compon de 775 vídeos que s'han anotat prèviament a través d'una plataforma de crowdsourcing en línia. Les respostes fisiològiques com l'electroencefalografia, l'electrocardiografia, l'activitat electrodèrmica, el reconeixement de les emocions facials i el seguiment de mirada s'han recollit de 30 participants mentre observaven els estímuls. La nostra base de dades es farà pública per fomentar la investigació en el reconeixement d'emocions.

Resumen

La computación afectiva aprovecha de las técnicas de aprendizaje automático para avanzar rápidamente en la tarea del reconocimiento de emociones en vídeos. Sin embargo, se necesitan más datos anotados. Varios estudios han construido grandes bases de datos de videos con anotaciones de emociones. Otros han recopilado bases de datos de vídeos musicales o escenas de películas conjuntamente con señales fisiológicas. Pero, ninguno de estos trabajos ha abordado una solución tanto con señales fisiológicas como con vídeos generados por usuarios.

En este trabajo presentamos *GALLUS*, una nueva base de datos de vídeos generados por usuarios con respuestas fisiológicas afectivas. La base de datos se compone de 775 vídeos que se han anotado previamente a través de una plataforma de crowdsourcing en línea. Las respuestas fisiológicas como la electroencefalografía, la electrocardiografía, la actividad electrodérmica, el reconocimiento de las emociones faciales y el seguimiento de mirada se han recogido de 30 participantes mientras observaban los estímulos. Nuestra base de datos se hará pública para fomentar la investigación en el reconocimiento de emociones.

Acknowledgements

First of all, I must thank my tutor, Xavier Giró-i-Nieto, for granting me the opportunity to work on such an interesting project and putting me in touch with Damian and Àgata. Moreover, he has guided and helped me throughout the project; suggesting me either the solutions or where to find the answer.

I would like to thank Damian Borth for hosting me in the brand new Institute of Computer Science from the University of St.Gallen. Damian has granted me the resources that have made this work possible. He has also introduced me to the research group which has received me with open arms.

I want to acknowledge Barbara Weber for joining the project the day I met her. She has always been willing to help and to participate in the study. Her experience in physiological measurements has made a difference since the beginning.

I also thank Àgata Lapedriza for sharing innovative methods and ideas applied in MIT Media Lab. I deeply appreciate that she could review the project due to her expertise in affective computing. Moreover, she has mentored and assisted me in choosing my future.

Apart from my advisors, I would like to acknowledge Juho and Evgenii, personnel from the Behavioral Lab, for supporting me while using their facilities, and for helping me to reach participants for the experiment.

I would like to thank the Agency for Management of University and Research Grants (AGAUR) for conceding me the MOBINT scholarship.

Finally, I have to express my gratitude for the volunteers that participated in the pilot experiments, and Andrea Calafell for sharing her thesis template.

Revision history and approval record

Revision	Date	Purpose
0	07/06/2019	Document creation
1	22/06/2019	Document revision
2	24/06/2015	Document revision
3	25/06/2015	Document approbation

DOCUMENT DISTRIBUTION LIST

Name	e-mail
Marcel Granero Moya	marcel.graneromoya@gmail.com
Xavier Giró i Nieto	xavier.giro@upc.edu
Damian Borth	damian.borth@unisg.ch

Written by:		Reviewed and approved by:		Reviewed and approved by:	
Date	8/07/2015	**Date**	10/07/2015	**Date**	10/07/2015
Name	Marcel Granero Moya	**Name**	Xavier Giró i Nieto	**Name**	Damian Borth
Position	Project Author	**Position**	Project Supervisor	**Position**	Project Supervisor

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Project Overview

The project was carried out at the University of St. Gallen, in the Institute for Computer Science (ICS-HSG). This work was developed as a bachelor thesis under a mobility program.

The project consisted on building a dataset of videos and physiological signals. Our challenge was to use user-generated videos that had to be extracted from YFCC100M database [1], a previous work from my advisor in the University of St.Gallen. Moreover, we aimed to collect a bigger dataset than state-of-the-art projects.

We performed the physiological experiments in the Behavioral Lab, a research infrastructure of the University of St. Gallen. Lab's resources were available free of charge to all HSG faculty for experimental studies of human behavior in a controlled environment.

The dataset is named *GALLUS* in honor of St.Gallen's founder.

1.2 Statement of purpose

The question is not whether intelligent machines can have any emotions, but whether machines can be intelligent without any emotions. — Marvin Minsky (1927–2016)

The ability to recognize emotion is one of the special indications of emotional intelligence, an aspect of human intelligence that has been argued to be even more important than mathematical and verbal intelligence (Picard et al. [2]). There is also significant evidence that rational learning in humans is dependent on emotions.

This findings have led to the emerging fields of **Affective Computing** [3] and sentiment analysis, which leverage human-computer interaction, information retrieval, and multimodal signal processing for recognizing people's sentiments.

Despite the ever-growing amount of online social data, and the attempt of social networks to stimulate users to tag their posts (with friends, content or even emotions), affective information in user-generated content is often wrong. The reason is that tags tend to be sarcastic, personal, and deviated from research purposes.

Therefore, large datasets in affective user-generated content, like ours, are critical for research and exploration as data is required for performing experiments, validating hypotheses, analyzing designs, and building applications. Over the years affective datasets have been put together for research and development using images, songs, film scenes or music videos as stimuli, yet, none of them contained user-generated videos. For this reason, our approach is a significant and novel work that will benefit the community.

1.3 Requirements and specifications

This dataset has been created as a tool that aims to be used for other researchers in the future. It still has to be analyzed.

The requirements of this project are the following:

1. Collect a dataset that should allow the future development of deep learning models capable of predicting the affective response of a human to videos, especially to video with a high emotional content.
2. The dataset should extend the current state of the art in this type of video affective datasets.
3. Publish the dataset with the necessary documentation to be used by third parties.

The specifications are the following:

1. Extract the stimuli from *YFMCC100M*[4] database.
2. Database size of 500 to 1000 videos.
3. Use the software platform *iMotions (2018)* [5] to design the data collection experiment.
4. Use the physiological sensors available in the Behavioral Lab.
5. Develop scripts for Data Management in Python[1] and Jupyter Notebook[2].

1.4 Work Plan

Being this a research-oriented project, it was difficult to define a detailed work plan ahead of time: the next steps to take were generally conditioned on the latest obtained results. Besides, due to the difficulties explained in Section 1.5, we have had several incidents and modifications (Section 1.6) in the work plan. Thus, a work plan with added and removed Work Packages is shown below.

1.4.1 Work Packages

- WP 0: Documentation
- WP 1: Affective Computing Research
- WP 2: Experiment Design
- WP 3: Pilot Experiment
- WP 4: Crowdsourcig Annotations

[1] `https://www.python.org/`
[2] `https://jupyter.org/`

- WP 5: Data Collection
- WP 6: Dataset Creation
- WP 7: Defense Preparation
- WP 8: Dataset Analysis and Training of a Neural Network (Future work)

A couple of Gantt diagrams were created throughout the project: One in the project proposal and another one in the project critical review. Due to the constant incidences the plan changed right after Gantts were delivered. However, a final Gantt is illustrated in Appendix A.

1.5 Challenges

Before explaining the incidents and modifications, we have to bear in mind three aspects that made this project more difficult.

First, the Institute of Computer Science of the University of St.Gallen was founded four months before this project began. This has been a big problem as the servers were still under construction. Moreover, the administration was hired in my first week in St.Gallen. Therefore, dealing with administrative documents was always a struggle.

Second, the Behavioral Lab is a recent acquisition from the University of St.Gallen. Its employees are still building up the lab and being trained to use lab's tools. Also, its facilities are highly-demanded and a reservation is needed before its use.

Third, topics such as Affective Computing or physiological data collection have not been covered during my studies in ETSETB. It was an opportunity to gain knowledge, yet a challenge.

1.6 Incidents and Modifications

As a comment for the reader, this section will be better understood after reading Chapter 3.

The first significant incident was the delayed arrival of the YFCC100M dataset. The dataset was copied to several hard disks in DFKI (Kaiserslautern) and shipped to St.Gallen. It was available to be used on the 1st of May. This drawback prevented us from handling any video beforehand.

Technical problems with the ABM headset delayed the beginning of the experiment. The electroencephalography tool was not sampling data at the desired frequency; therefore, we spent a week solving this problem with the help of an iMotions technician. Finally, it was solved on the 2nd of May.

Due to the previous incidences, we decided not to train a deep neural network after the dataset collection. Therefore, the original Work Package 8 was deleted from our plan.

Further preparation was added to our schedule as we realized the complexity and laboriousness of the experiment. First, we added a Pilot Experiment to evaluate the feasibility, time, cost, adverse events, and improve upon the study design prior to performance of the full-scale experiment. It resulted to be good practice to get experience in using the tools and interiorizing the experiment's procedure (as explained in Section 3.3).

After the pilot test, we noticed that stimuli were neutral and usually evoked no emotions. Given the importance of stimuli, we preferred to spend more time and resources in a more consistent selection of videos. This decision led us to Amazon Mechanical Turk. Using this crowd-sourcing platform we would be able to collect previous annotations of the videos to make a wiser selection of stimuli. These annotations would be the Valence-Arousal ratings of each video.

Two major reasons delayed the final experiment. First, the Behavioral Lab was booked by other researchers from the 15th of May until the 7th of June. Second, my first contract ended on the 31st of May and a second contract could only start from the 11th of June due to administration procedures.

Further incidents occurred at the beginning of the full-scale experiment. They are explained in section 3.4 as they only affected Work Package 6.

Chapter 2

State of the art

We will start discussing the issue of emotion classification before moving on to the affective computing definition. Then, we will mention recent and renowned works in the field of sentiment analysis in relation to images, videos, and physiological responses. Finally, comparable works to ours will be described.

In order to explain what "emotions" are, we will draw on results and discussions from psychology. There is a long tradition of research on emotions, and yet it has not been possible so far to produce a unified, exact definition for the concept of emotion. Researchers have approached the classification of emotions from one of two fundamental viewpoints. First, emotions are discrete and fundamentally different constructs; e.g., Ekman's six basic emotions [6] (anger, disgust, fear, happiness, sadness, and surprise). Second, that emotions can be characterized on a dimensional basis in groupings. Russell's circumplex model of affect [7] and the pleasure-arousal-dominance model [8] are some examples. Nevertheless, Robert Plutchik [9] offers a three-dimensional model that is a hybrid of both basic-complex categories and dimensional theories.

Affective Computing is "Computing that relates to, arises from, or deliberately influences emotion or other affective phenomena" - Picard, MIT Press 1997 [10]. The motivation for the research is the ability to simulate empathy. The machine should interpret the emotional state of humans and adapt its behavior to them, giving an appropriate response to those emotions.

The computational inference of emotions in images has been studied extensively, partly stimulated by the availability of the International Affective Picture System (IAPS) database[11]. And sentiment Analysis can also be performed in a multimodal approach. One of the first approaches in this direction is SentiBank [12] utilizing an adjective-noun pair representation of visual content.

The human face plays a prodigious role in automatic recognition of emotion in the field of identification of human emotion and the interaction between human and computer. Researchers often use Paul Ekman's Facial Action Coding System [13] as a guide. Facial expressions databases [14][15] have led machine learning models [16] to improvement in emotion classification performance.

SEMAINE [17] is a large audiovisual database of face recordings as a part of an iterative approach to building Sensitive Artificial Listener (SAL) agents. High-quality recordings total of 150 participants, for a total of 959 conversations with SAL characters, lasting approximately 5 minutes each.

Physiological monitoring is still in relative infancy as there seem to be more efforts towards affect recognition through facial inputs, as mentioned above. However, some studies claim that physiological responses give more consistent emotional data. For instance, Lang et al. [18] showed colored photographic pictures that varied widely across the valence-arousal dimensions and measured facial electromyographic and visceral (heart rate and skin conductance) reactions. Moreover, Ringeval et al. presented RECOLA [19], a multimodal corpus of spontaneous collaborative and affective interactions. Participants were recorded in pairs during a video conference while completing a task requiring collaboration. Different multimodal data, i.e., audio, video, ECG and EDA, were recorded continuously and synchronously.

We have to mention two more works that can be related to our project but still not fully comparable. For the sake of our knowledge, Jiang et al. [20] present the first database relating user-generated videos and emotions tagging. They introduce a dataset collected from YouTube and Flickr with eight manually annotated emotions from the Plutchik's wheel. They also compute and evaluate a large set of audio-visual features, and introduce the use of semantic attributes for emotion prediction.

The HUMAINE [21] project was to provide the community with examples of the diverse data types that are potentially relevant to affective computing, and the kinds of labeling scheme that address the data. Thus, it provides a Database of naturalistic clips which record forms of feeling, expression, and action that color most of human life.

2.1 Similar works

The following works to be described are datasets comparable to our approach; table 4.3 illustrates a comparison between datasets. All of them are publicly available.

DREAMER [22] is a multimodal database consisting of electroencephalogram (EEG) and electrocardiogram (ECG) signals recorded during affect elicitation by means of movie videos. Signals from participants were recorded along with self-assessment of their affective state after each stimulus, in terms of valence, arousal, and dominance. All the signals were captured using portable, wearable, wireless, low-cost, and off-the-shelf equipment that has the potential to allow the use of affective computing methods in everyday applications.

DECAF [23] is a multimodal data set for decoding user physiological responses to affective multimedia content. Different from other data sets because brain signals are acquired using the Magnetoencephalogram (MEG) sensor, which requires little physical contact with the user's scalp and consequently facilitates naturalistic affective response.

MAHNOB-HCI [24] presents a multimodal database recorded in response to film clips with the goal of emotion recognition. Synchronized recording of face videos, audio signals, eye gaze data, and peripheral/central nervous system physiological signals was captured while participants watched emotional videos.

DEAP [25] is a multimodal data set for the analysis of human affective states. The electroencephalogram (EEG), face video and peripheral physiological signals were recorded as participants watched one-minute long excerpts of music videos. Moreover, participants rated each video in terms of the levels of arousal, valence, like/dislike, dominance, and familiarity.

Last but not least, there is a very recent work that may benefit our approach. As researchers want emotion recognition systems to work across datasets, Gideon et al. focused on adversarial methods to find more generalized representations of emotional speech following an easier to train "meet in the middle" approach. The model iteratively moves representations learned for each dataset closer to one another, improving cross-dataset generalization. Their experiments focus on cross-corpus training for speech emotion recognition, but, they suggest that these methods could be used to remove unwanted factors of variation in other settings.

Chapter 3

Methodology

In this chapter, the methodology followed to build the database will be explained. Firstly, the way we selected the stimuli for the experiment will be narrated. Secondly, a definition of which user data were chosen for the experiment will be discussed. Thirdly, a description of the pilot experiment will be presented as a first practical approach to the real experiment. Lastly, details about the final experiment will be described.

3.1 Video Stimuli

Experiments in human cognitive-behavior research typically involve some kind of stimulation used to evoke a reaction from respondents. The two most crucial stimulus-related questions are: Which stimuli do we need? In which sequence shall we present the stimuli?

Regarding the first question, we will use stimulating user-generated videos. This kind of videos tend to be neutral and usually they are evoking no emotions. Thus, we require a proper stimuli selection (see Section 3.1).

Presenting stimuli in the same sequence to all respondents bears the risk of sequential effects. Respondents might rate the first stimulus always higher because they are still motivated, engaged and curious. After two long hours at the lab, exhaustion might take over, so ratings might be low even if the assessed stimulus exceeds all previous expectations. This will be avoided by presenting stimuli in random order.

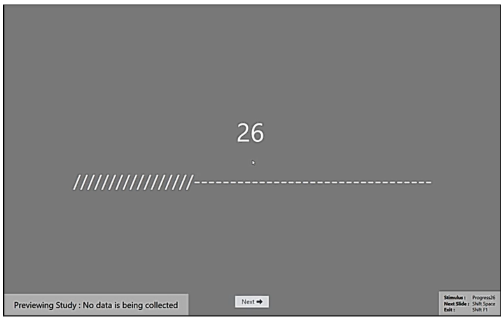
Furthermore, stimuli will be presented as illustrated in figure 3.1. Given the time spent in screens between stimulus, participants will get calmed to the default emotional state.

3.1.1 YFCC100M Database

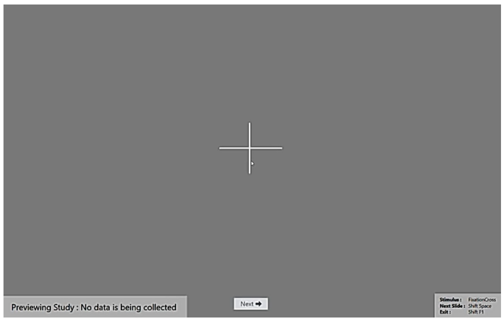
According to requirement 1 (section 1.3), we chose the YFCC100M dataset created by Thomee et al. [1] because of its large size and previous experience on it by Prof. Dr. Borth, one of its co-authors. Thus, he could make it available easily, and he would be able to support throughout the project.

Yahoo Flickr Creative Commons 100 Million Dataset (YFCC100M) is part of the Yahoo Webscope program. This dataset is the largest public multimedia collection that has ever been released, comprising a total of 100 million media objects, of which approximately 99.2 million are photos and 0.8 million are videos, all of which have been uploaded to Flickr between 2004 and 2014 and published under a Creative Commons commercial or non-commercial license.

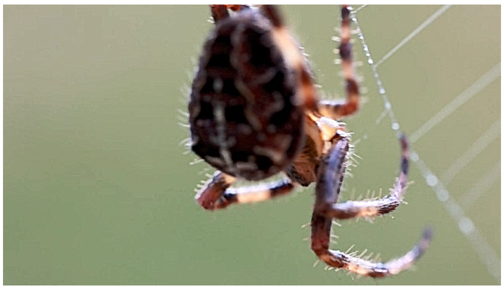
There are 4 characteristics that made this database suitable for our project:



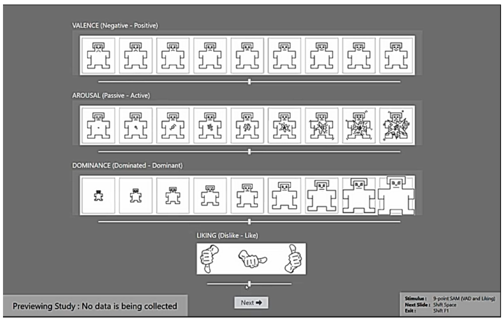
(a) Progress slide. Gives information about the progress of the experiment: Number of stimuli and progress bar.



(b) Fixation Cross Slide. Participants look at the cross so that all stimuli start with the eye-gaze in a central position.



(c) Stimulus. Video with a length between 12 and 25 seconds. Long enough to collect physiological data properly, and short enough to evoke only one emotion



(d) SAM and liking. Users must assess the previous stimulus; moving all sliders is required to proceed to the next stimulus. Further explanation in 3.2.3

Figure 3.1: Stimuli loop. Each stimulus is shown in the form of Progress screen (a), Fixation Cross (b), Video Stimulus (c) and Self-Assessment Manikin (d).

- Content was generated by users. This would make our approach different from the state-of-the-art.
- Meta-data such as title, tags and description is included in each media object.
- It contains a huge amount of videos, more than 800k videos.
- Data is available to be shared and used for research because of the Creative Commons licenses. This allows us to create a free-to-use database.

In addition, YFCC100M has a browser that enables easy and quick access for the type of queries, which define a specific subset of the YFCC100M dataset. It is designed to filter and explore the entire dataset of 100 million images and videos in real-time. Subsets of the complete dataset can be retrieved by a straightforward keyword search and reviewed directly. Despite the datasets vast size, the response time of the search-engine is fast enough to view query results in matters of seconds, enabling a fluid browsing experience.

As it is shown in 3.2, given a user query the browser retrieves the subset of videos matching the query and provides previews of videos in form of thumbnails. Each item is linked to its associated Flickr page, where it can be played and downloaded. In addition, a set of statistics for the retrieved subset is generated dynamically, including a tag-cloud, the global distribution, user participation and time-line of items. With this very vital information it is possible to get a first

Figure 3.2: YFCC100M Browser Illustration. Source: YFCC100M browser website

overview of the subsets as defined by a user query and identify biases or get a quick impression of the quality of the associated videos.

3.1.2 Keyword-based video downloading

Each media object included in the YFCC100M dataset is represented by its metadata in the form of its Flickr identifier, the user that created it, the camera that took it, the time at which it was taken and when it was uploaded, the location where it was taken (if available), and the CC license it was published under. In addition, the title, description and tags are also available, as well as direct links to its page and its content on Flickr.

We got access to the videos metadata text file. From there, we selected the videos according to emotional tags. The goal was to obtain a balanced set of videos in the circumplex model[7] with valence-arousal[1] quadrants as shown in Figure 3.3. Inspired by the work of Laurier et. al. [26], in which they analyzed how people tag music by mood, we made the analogy to video tagging. This set of emotional tags was a good starting point.

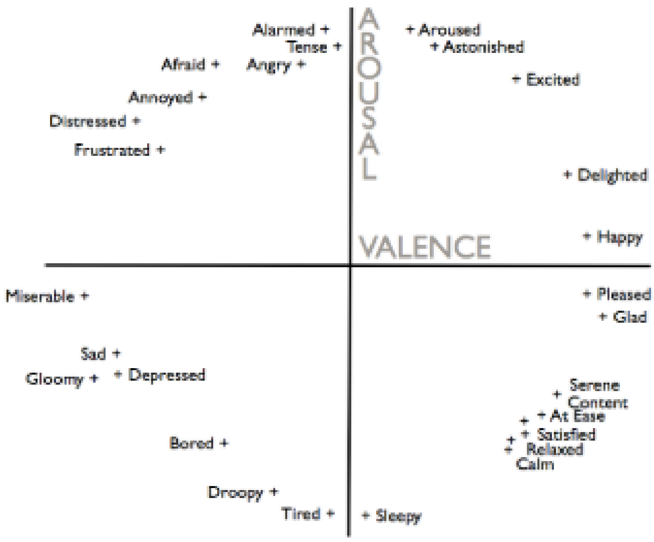


Figure 3.3: Emotional tags in Russell's circumplex model of affect with arousal and valence dimensions.

[1]Valence measures how positive or pleasant an emotion is whereas arousal measures the agitation level of the person, ranging from calm to aroused. See subsection 3.2.3 for further explanation of Valence-Arousal.

After a few searches in YFCC100M browser, a list of emotional tags per cluster was reached, presented in Table 3.1. Regarding Russell's circumplex model, cluster's are distributed in each of its quadrants, e. g., cluster 1 is the low valence high arousal quadrant. Selected videos had the emotional tag in their description, title or tags.

Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 4**
angry	sad	tender	happy
aggressive	bittersweet	soothing	joyous
visceral	sentimental	sleepy	bright
rousing	tragic	tranquil	cheerful
intense	depressing	good natured	happiness
confident	sadness	quiet	humorous
anger	spooky	calm	gay
exciting	gloomy	serene	amiable
martial	sweet	relax	merry
tense	mysterious	dreamy	rollicking
anxious	mournful	delicate	campy
passionate	poignant	longing	light
quirky	lyrical	spiritual	silly
wry	miserable	wistful	boisterous
fiery	yearning	relaxed	fun

Table 3.1: Emotional tags.

Apart from tag-filtering, videos should fulfill the list of requirements shown in table 3.2. First, duration must be long enough to collect proper physiological signals, and it has to be short enough to contain only one predominating emotion. Second, experiment participants would speak English. Third, both YFCC100M dataset and participants would be biased towards western cultures.

Video's requirements	
Duration	Between 12 and 25 seconds
Language	English
Culture	Western

Table 3.2: Video's requirements ensure a consistent set of stimuli.

The two first requirements were tackled through a Python script that read video's ffmpeg data and removed videos that did not fulfill the requirements. The cultural bias could not be tackled through any machine task, it had to be reviewed manually.

As a result of the meta-data filtering, a list of 3000 candidate videos was obtained, in the 4 circumplex quadrants.

3.1.3 Manual Video Filtering

After keyword-based downloading, videos still needed further filtering. First, because there were videos that did not fulfill the requirements: they were flipped, in another language (ffmpeg language data was wrong), or low-quality recordings. And second, because we had to ensure that an event was happening in the video, so that an emotional response would be triggered.

Manual filtering was the only way to achieve a good selection of videos regarding the problems mentioned above.

Then, we started watching videos from the meta-data filtered clusters. Filtering by hand is a tedious task that requires continuous attention. Videos were classified in different folders regarding their emotional quality. One per every three videos was approved because candidate videos were unstimulating.

After two days of tedious manual work, a selection of 400 videos was achieved.

3.1.4 Crowdsourced Annotation

In order to get a more consistent and reliable selection of stimuli, we wanted to annotate our videos in terms of Valence and Arousal. Therefore, we decided to use a crowdsourcing platform to make the assessments faster, more consistent and cheaper. We chose the most popular crowdsourcing platform for Machine Learning.

Amazon Mechanical Turk or MTurk is an Internet crowdsourcing marketplace enabling individuals and businesses (known as Requesters) to coordinate human labor to perform tasks that computers are currently unable to do. It is operated under Amazon Web Services. Requesters post jobs known as Human Intelligence Tasks (HITs), such as identifying specific content in an image or video, writing product descriptions, or answering questions, among others. Workers, colloquially known as Turkers or crowdworkers, browse among existing jobs and complete them in exchange for a rate set by the employer. MTurk can be used to obtain high-quality data inexpensively and rapidly as stated in the work from Buhrmester et al. [27].

First of all, we unsuccessfully tried to register as a worker to have a glance at the HITs that were published by other researchers. Amazon informed us that we would not be permitted to work on Mechanical Turk. Their account review criteria are proprietary and they could not disclose the reason why an invitation to complete registration was denied. Conversely, registering as a Requester was a matter of minutes.

Based on this tutorial *Mechanical Turk Tips* [28], the following tips were introduced in the task definition:

- Turkers have to understand the task and be interested in it. Therefore, the task has to be clearly explained, not too long and not too boring.
- Past performance is a good indicator of future performance. It is recommended to limit HITs to those workers with at least a 95% approval rate.
- Experience is valuable, so a 500-approved-HITs limit was established.
- Giving workers more time than they need because the task closes after it.
- Regarding the payment, it is recommended to reward workers with 3 cents for a one-minute-task.
- Automatically approve HITS within 3 days so that workers get paid fast and do not send emails asking about the payment.

- Blocking workers with bad performance and rejecting their tasks.

MTurk Requester site allows publishing HTML HITs. HTML only supports mp4, webm and ogg video file formats. Therefore, videos were converted into mp4 because it was the format accepted by iMotions as well. Videos were stored in a cloud storage provider (Dropbox). The public rendering links from each video were extracted and uploaded to MTurk through a csv file.

Figure 3.4: Videos prepartion for MTurk.

Figure 3.4 illustrates the process step by step. First, videos were converted from their original video format to mp4. Then, MP4 videos were uploaded to shared folders from a cloud storage platform. After that, by means of a web-browser, the html source of the shared folder was extracted. Later, a Jupyter Notebook analyzed the html source and extracted the public links for rendering videos. Finally, the same notebook printed out a csv batch with one video link per line. The csv file was the input file for MTurk; every HIT used a different video link.

A first testing HIT was published as presented in Figure 3.5. At the top, there are brief and clear instructions about the task; they may be read in the preview of the task. Below the instructions, workers can play the video by using its reproduction controls. At the bottom, valence and arousal 9-point sliders allow respondent's ratings.

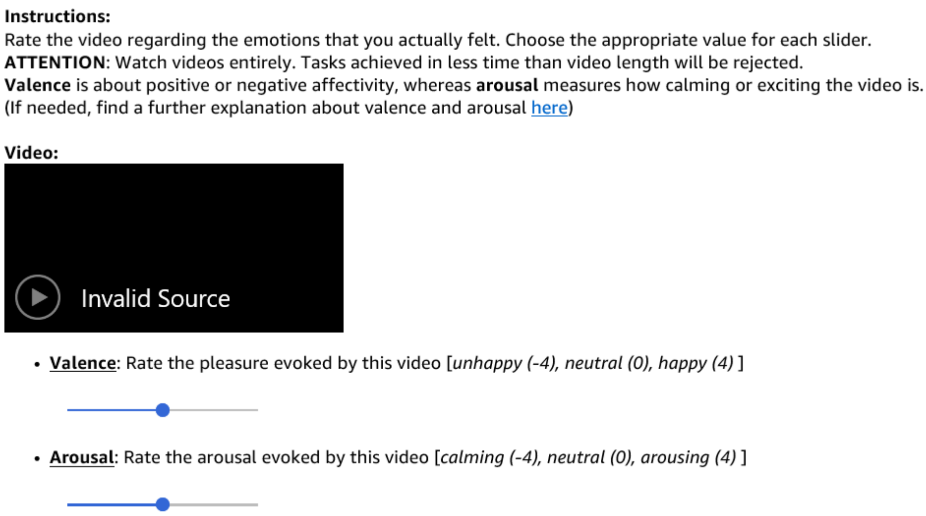


Figure 3.5: First published HIT in MTurk.

3.1.5 Manual vs. Metadata-filtered stimuli

At this point in the project, there were 400 videos manually selected from the metadata-filtered set. We had to choose whether to spend more time selecting stimuli by hand or just choosing the stimuli with the best video definition. We also wondered if the manual selection we did was personal and unreliable.

An AMT experiment was designed using the same HIT from Figure 3.5. The goals were to test MTurk by launching a paid experiment, and compare assessments between a set of manually selected videos and a set of metadata filtered videos.

273 videos were published: 133 manually filtered and 140 that had not been filtered. The HIT was launched following the MTurk recommendations explained previously. Moreover, each stimulus would have 3 assessments from different workers.

65 workers completed the 819 tasks ($273 videos \cdot 3 assessments$) 15 minutes after the batch was uploaded. One of our concerns was if workers would watch videos entirely; they (probably) did it. Videos length is between 12s and 25s, and the average time of the task was 25 seconds; so workers spent the time to watch the video and a bit more for the assessment.

Manually selected videos had less variance in the crowd assessments. Apart from that, videos from the preselected set, i.e. only filtered by its metadata, may not fulfill the requirements: flipped, poor recording or audio quality, or not in English. Thus, only manually selected videos would take part in the next assessments and in the database.

These results led us to the tedious task of filtering videos manually, again. We spent two more days classifying videos in stimulating or dull. Finally, we managed to select over 400 videos more; making a total corpus of over 800 videos.

3.1.6 MTurk assessment to the entire database

Once that over 800 videos had been manually selected, we aimed to assess our candidate videos for the database by gathering a valence-arousal rating for each video. These evaluations would help us to choose the best corpus by removing videos with the highest variance, and informing us about the corpus bias regarding the Valence-Arousal map.

A few improvements were made to the previous published HIT as it is shown in figure 3.6. The most significant change was adding SAM images to help with the rating. Thus, making the evaluation closer to the physiological experiment. Besides, we ensured that all videos would be rendered in the original size, and added a longer description in the HIT's preview. Regarding task's specifications, the number of assessments per HIT increased to 5, and the experiment was launched with 781 manually selected videos.

All 4056 assignments were completed by 180 workers in two hours. We made clear in the instructions that any work achieved in less time than video's length would be rejected; then, the average time per assignment increased to 31 seconds.

As illustrated in the figure 3.7, the corpus results to be biased towards high values of valence and arousal. Therefore, a dozen videos with lower valence were selected in addition to the previous ones. This time, the YFCC100M Browser was used (see Figure 3.2) to look for low-valence related tags such as (car) accident, (building) fire, dead, tsunami and earthquake. These videos were uploaded to Mechanical Turk to get their respective SAM assessments. In order to balance this set of low valence stimuli, 18 videos from the high valence clusters and the neutral one were added. The majority of these videos resulted to be annotated in low-valence clusters.

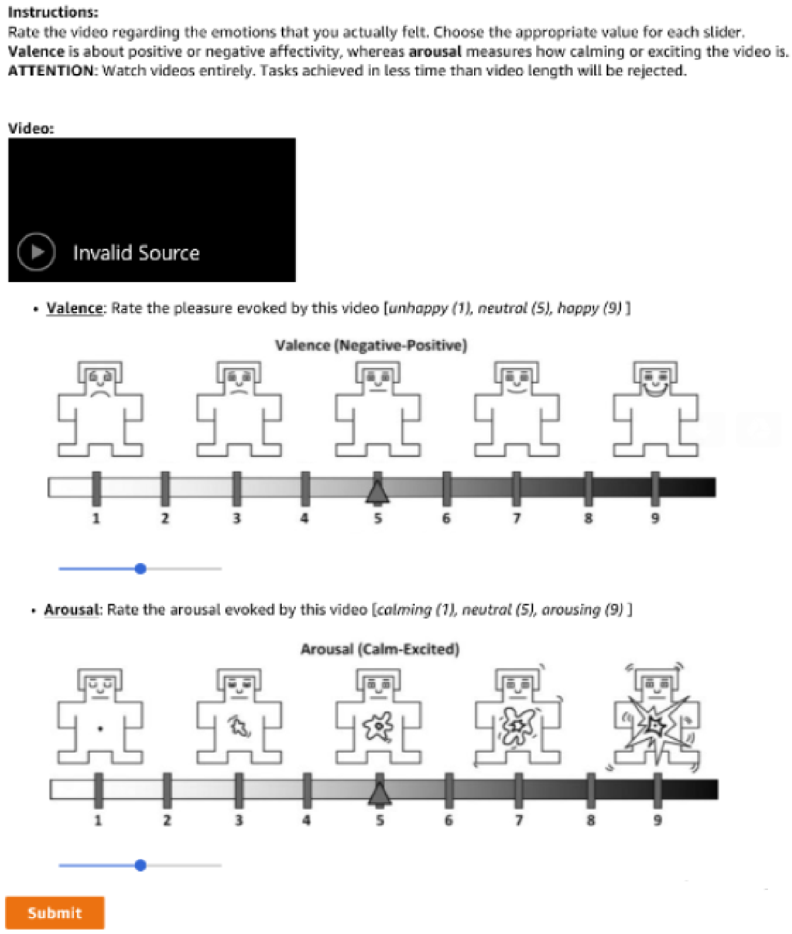
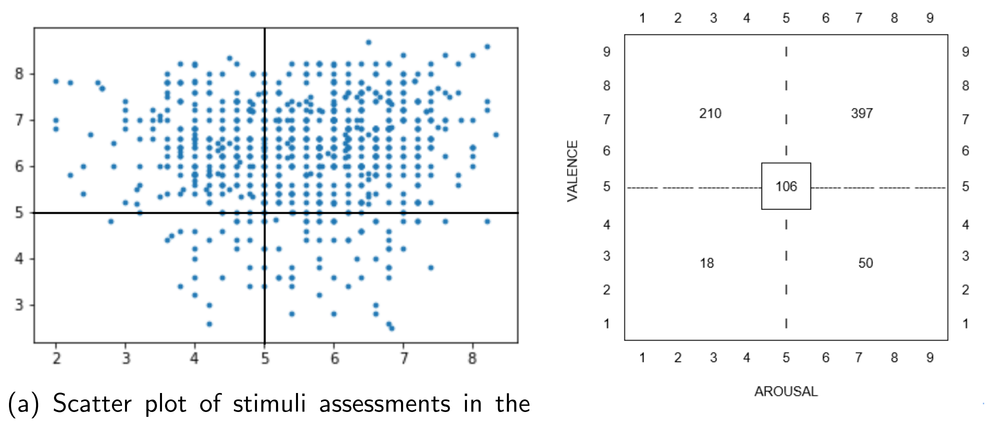


Figure 3.6: Improved HIT used for the biggest experiment in MTurk.



(a) Scatter plot of stimuli assessments in the circumplex model. X-axis Arousal, Y-axis Valence

(b) Cluster Classification

Figure 3.7: Results after the stimuli assessment explained in Subsection 3.1.6.

3.1.7 Final corpus

Once all the videos were assessed, it was possible to create the stimuli corpus. First, around 50 videos that had the highest variances across turkers annotations were removed. Then, the 5 most representative videos from each of the 5 clusters were chosen. For the outer clusters, the furthest videos from the neutral point were selected. Whereas for the neutral cluster the selected videos were the ones with less variance.

Therefore, the final stimuli corpus consisted of 775 videos: 25 cluster-representative videos and 750 more. All of them with a first valence-arousal evaluation with at least 5 assessments.

3.2 User Data

In this section, we will report how the study was designed to collect the aimed video database. Each element from the experiment will be discussed and a founded decision will be formulated.

3.2.1 Respondent group

A group of people will be recruited to participate in the study as respondents. It is necessary to specify the respondent group characteristics to exclude side effects that could alter the outcomes of the experimental data collection. We should make sure that demographic characteristics such as age, gender, education level, income, marital status, occupation, etc. are consistent across the respondent pool. However, as the experiment was developed in the Behavioral Lab, a biased group of respondents was recruited: young HSG students, from Bachelor to Ph.D., mainly from European countries. Thus, we could only make sure that genders were balanced among the group.

Responses to stimuli and self-assessments are subjective and may depend on one's personality. Thus, we propose to collect personality information by means of the Big Five personality test [29]. It is the most scientifically sound way of classifying personality differences and is the most widely used among research psychologists. The Big Five is named so because the model proposes that human personality can be measured along five major dimensions: Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. We will make use of a free, 10-minute, online personality test [30] that will give us precise numerical scores for the five personality traits.

3.2.2 Biometric sensors

The Behavioral Lab offers state-of-the-art software and hardware for collecting affective physiological responses. iMotions [5] is the Biometric Research Platform that we will use. It allows to record data from biometric sensors without having to manually piece everything together. iMotions integrates and synchronizes all sensors with the stimuli, as illustrated in figure 3.8. iMotions integrates the acquisition of physiological signals, the collection subjective responses, render of stimuli and data post-processing. In our experiment, the physiological signals highlighted with yellow boxes will be used; the specific devices for each signal are written at the left of each box. The Self-Assessment Manikin will be shown as a survey after each video stimulus.

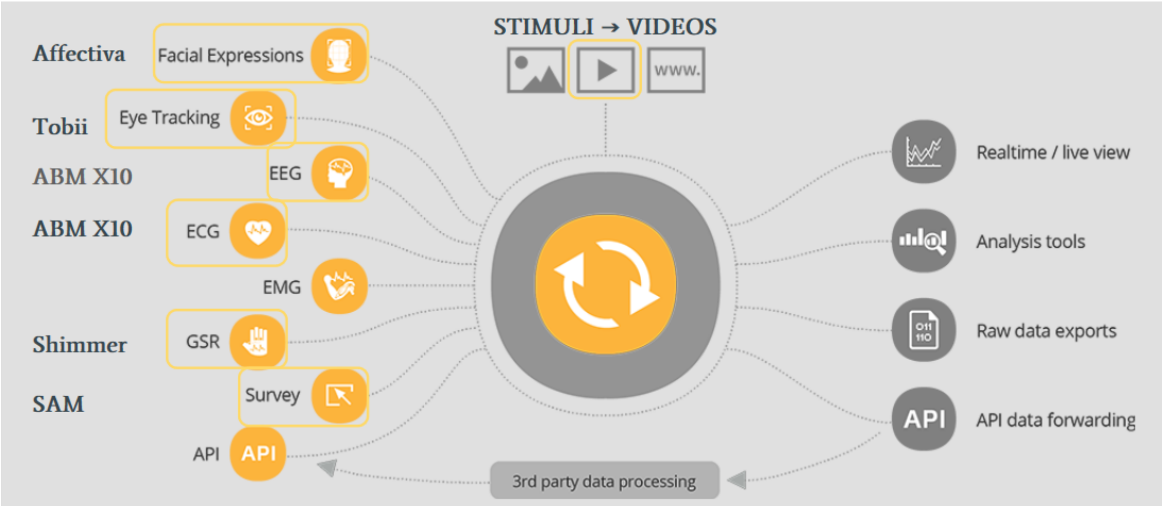


Figure 3.8: Physiological measurements and its devices integrated in iMotions. Source: iMotions Website

The captured physiological signals are the following:

Eye tracking We will collect information on gaze position and pupil dilation with a Tobii[2] 60 Hz screen-based eye-tracker. As we present visual stimuli on screen, we will collect eye-tracking data to know where respondents are directing their gaze to and how this is affecting cognitive processing. Second, monitoring pupil dilation can give valuable insights into arousal and stress levels of a respondent. As pupil dilation is an autonomic process, it cannot be controlled consciously.

GSR Galvanic skin response (GSR) or electrodermal activity (EDA) reflects the amount of sweat secretion from sweat glands in our skin. Increased sweating results in higher skin conductivity. When exposed to emotional content, we sweat emotionally. It offers tremendous insights into the unfiltered, unbiased emotional arousal of a respondent. We will collect GSR data through a Shimmer3[3] unit.

FER[4] Affectiva[5] software will use the Facial Action Coding System (FACS [13]) to classify facial expressions. Combinations of these facial expressions are then mapped to emotions. It is a non-intrusive method to assess head position and orientation, microexpressions (such as lifting of the eyebrows or opening of the mouth) and global facial expressions of basic emotions (joy, anger, surprise etc.) using a webcam placed in front of the respondent. Facial data is extremely helpful to monitor engagement, frustration or drowsiness.

ECG & EP [6]Monitoring heart activity with ECG electrodes (B-Alert X10[7]) attached to the chest and an optical heart rate sensor (Shimmer3) attached to finger tips allows us to track respondents' physical state, their anxiety and stress levels (arousal), and how changes in physiological state relate to stimuli.

[2] https://www.tobiipro.com/
[3] www.shimmersensing.com/products/gsr-optical-pulse-development-kit
[5] https://www.affectiva.com/product/individual-product-page-imotions/
[6] EP stands for Electro Photoplethysmography; to optically detect blood volume changes in the microvascular bed of tissue.
[7] https://www.advancedbrainmonitoring.com/xseries/x10/

EEG Electroencephalography (EEG) is a neuroimaging technique measuring electrical activity generated by the brain from the scalp surface using portable sensors and amplifier systems. It is our means to assess brain activity associated with perception, cognitive behavior, and emotional processes. EEG reveals substantial insights into sub-second brain dynamics of engagement, motivation, frustration, cognitive workload, and further metrics associated with stimulus processing, action preparation, and execution. EEG tells which parts of the brain are active while respondents are exposed to certain stimulus.

3.2.3 Manual Annotation

Apart from the biometric measures, we will ask users to provide a self-assessment of emotions experienced.

First of all, we had to choose which classification of emotions we wanted to use as labels. Given the wide variety of emotions classifications, we investigated state-of-the-art datasets and chose the Valence-Arousal-Dominance (VAD) approach. Valence (V) measures how positive or pleasant an emotion is, ranging from negative to positive. Arousal (A) measures the agitation level of the person, ranging from non-active/in calm to agitated/ready to act. Dominance (D) measures the control level of the situation by the person, ranging from submissive/non-control to dominant/in-control. VAD can be naturally related to the physiological signals, as valence and arousal manifest in our body independently due to the sympathetic (related to fight and flight response) and parasympathetic (related to rest and digest response) nervous systems. We also chose the Valence-Arousal-Dominance classification so that our assessed videos could be represented in Russell's 2-dimensional model.

Two techniques to assess valence, arousal and dominance are discussed in M. Bradley et al. work [31]. First, the Semantic Differential is explained. It is a method in which respondents have to choose one of each bipolar adjective pair, being a total of 18 pairs. Second, the Self-Assessment Manikin (SAM) is compared. As described in [32], SAM is an easy to administer, non-verbal method for quickly assessing VAD. Judgments regarding the amount of pleasure and arousal experienced when viewing a picture using SAM correlated highly with ratings obtained using the verbal, more lengthy semantic differential scale. Besides, it can be employed with a variety of subject populations, including non-English speaking subjects, children, people with language disorders, and of course all clinical syndromes. Differences obtained in judgments of dominance suggest that SAM might be more accurate in tracking the subject's - rather than the stimulus' - feelings of control.

Hence, we chose to use the SAM as well as state-of-the-art studies. Respondents will answer a 9-point SAM and a liking scale after each video as illustrated in Figure 3.9.

3.3 Pilot Experiment

In this section, we will describe the steps to acquire the proper technical skills for the final experiment. Unfortunately, we had an issue with the EEG headset that delayed the whole project one week as mentioned in 1.6.

First of all, we had two online training sessions in both iMotions software and EEG. The first

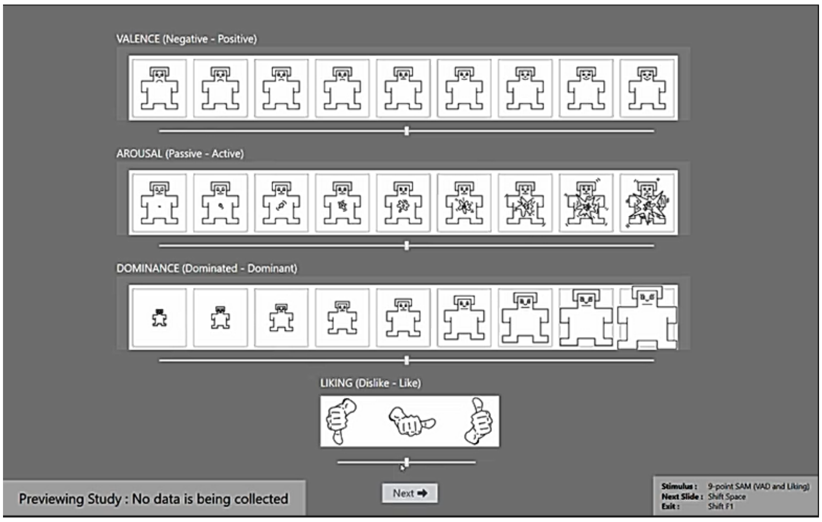


Figure 3.9: 9-point Self-Assessment Manikin with a liking rating that was used in our study. Valence, Arousal and Dominance states are illustrated in manikins. The liking scale comprises three answers: dislike, neutral, like.

one tackled scientific background and experimental design thorough iMotions platform. Whereas in the second one, we learned how to mount the EEG headset using ABM's material, how to conduct impedance and Benchmark tests, and do the post-processing of EEG data.

Figure 3.10: EEG headset mounted for the first time as a result from the online training.

After solving the EEG incident, the pilot experiment was conducted to evaluate the feasibility, time, cost, adverse events, and improve upon the study design prior to performance of the full-scale experiment. The experiment was carried out from the 9th to the 15th of May. Researchers from the Institute of Computer Science volunteered as participants. Overall, 2 females and 4 males attended the experiment.

The pilot experiment was extremely useful because we gained experience in mounting the EEG headset and setting up the other sensors. We had video codec issues that we could solve by converting videos into mp4 format.

(a) Unmounting the headset.

(b) Getting insights from a volunteer.

Figure 3.11: Pictures from the pilot experiment.

Moreover, we got insights from each volunteer. They reported that certain videos were too dull; a better selection of stimuli was needed. By then, only 400 videos were selected and we were about to launch the MTurk experiments (see subsection 3.1.5). We also evaluated the time needed for the experiment and figured out tasks order. In order to remember all the tasks, a checklist was written down. Last but not least, we familiarized with the Behavioral Lab hardware and its facilities, so that the final test could be designed as depicted in the following chapter.

3.4 Final Experiment

In this section, the full-scale experiment that was carried out in the Behavioral Lab will be explained. Firstly, the set-up of the rooms will be described with its hardware and communication systems. Secondly, the scheme followed in every session will be specified. Thirdly, the performance and the incidences that happened throughout the experiment will be reported.

3.4.1 Launching the experiment

Before starting with the experiment, the Behavioral Lab needed a research proposal document (App. B) to approve our experiment. Moreover, the Ethical Committee from the University of St.Gallen had to be informed about the study and and had to give its consent. Once approved, participants recruitment was managed through the Sona Systems software. Two analog studies were created, one for each gender. Potential participants could see the experiment information as illustrated in Appendix C.

Multiple 2-hour time slots were created from the 13th until the 21st of June. 4 to 5 sessions per day were assigned. A regular experiment day had the following schedule: male 9-11h, female 11-13h, male 14-16h, female 16-18h. Males were before females because they often are short-haired; this means that it is easier to mount the headset and probably no delay will affect the following session.

An email to reach out participants was sent on the morning of the 12th of June. After two

(a) Instructor's room

(b) Respondents' room

Figure 3.12: Instructor and respondents rooms are contiguous.

hours all slots were assigned. It was a complete success given that exchange students were gone and regular students were in the exams period. Besides, whenever there was a cancellation, students signed up automatically.

3.4.2 Experiment setup

The experiment needed two contiguous rooms, one for the participant and another one for the instructor. The respondents' room (Fig. 3.12b) had the required hardware to capture physiological data (Tobii eye-tracker and a web-cam for the facial emotion recognition), reproduce stimuli (screen and high-quality speakers), and get participants feedback (mouse and keyboard). Moreover, it had a chair without wheels so that participants could not move during the test, and an adjustable table to make sure that the participant was at a proper height for the eye-tracker. The window was covered with a curtain to maintain a constant light.

The instructor's room (Fig. 3.12a) had the equipment to monitor the experiment. One screen was mirroring the participant's screen[8], and the other one was monitoring the participant and the captured data in real time. Regarding the communication throughout the test, a tablet was placed in each room with an intercom application. The instructor could always hear the participant's room, and whenever it was needed the instructor could activate the tablet's microphone and give instructions to the participant.

3.4.3 Experiment's Schedule

This subsection explains the methodology followed in every session. The schedule was tight and time management was a key factor. The usual duration of each task is indicated in table 3.3. The total duration of the experiment is of approximately 1h 55'.

Preparation - Before the participant comes, the material for the EEG has to be prepared.

[8]The mirroring screen had to be removed because it interfered with videos reproduction.

The neoprene head strap is attached to the headset, and foam sensors are stuck and filled with synapse gel.

Sensors Setup - Once, the participant arrives to the instructor's room. The mounting of the EEG headset starts, which takes 20 minutes, and ECG and GSR sensors are set up while the EEG impedances are checked. Meanwhile, the participant has to fill and sign two documents; the consent form (Appendix D) and the payment form (Appendix E). Then, she/he answers the online personality test with a tablet as illustrated in figure 3.13a.

Benchmark Test - Once in the participant's room, the Benchmark Test is prepared by setting the table to the proper height for the eye-tracker, and telling the participant not to move the left hand nor the jaw. The participant will remain sat until the end of the experiment. After that, the instructor returns to the other room and the participant completes the 9-minute Benchmark Test.

Experiment Explanation and Instructions - The instructor comes back to the respondent's room and explains the experiment to the participant, pointing out their prohibitions. All this information is in a whiteboard, as illustrated in Fig. 3.13b, so that the same instructions is indicated. The whiteboard is behind the participants so that she or he will not look at it during the experiment. Finally, the instructor goes back to the the monitoring room and will enter the respondent's room during breaks.

Beginning of the experiment - Participants do an eye-gaze calibration following a point all over the screen. After that, the experiment begins with three screens to describe the Self-Assessment Manikin.

Stimuli: 1st Part - 33 videos are shown with their respective SAM. While they are watching ad assessing stimuli, their performance (Fig. 3.13c) and the signals quality (3.13d are monitored.

First break - The respondent gets the gaze out of the screen to prevent eyes fatigue and can move the hand with the GSR sensor to prevent numbness. Meanwhile, personal information is asked. This data is collected anonymously in the personal data sheet (Appendix F).

Stimuli: 2nd Part - 33 more videos are shown.

Second break - The participant gets comfortable again and a glass of water and biscuits without arousing ingredients like sugar are offered.

Stimuli: 3rd Part - The participant assesses the 34 remaining videos.

End of the experiment - Sensors are removed from the participant and the headset is unmounted.

Table 3.3 shows the experiment schedule with the accumulated time after each task. The preparation for the experiment has no accumulated time because it can be carried out before the respondent comes to the lab. Participants spend more time in the first part of the stimuli because they are not used to the assessing scale yet.

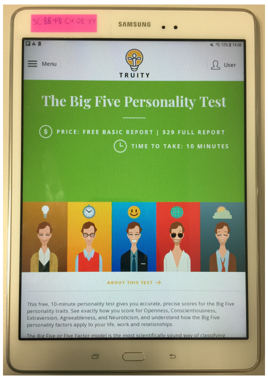
#	Task	Duration	Acc. time
0	*Preparation*	10'	-
1	*Sensors Setup*	25'	25'
2	*Benchmark Test*	15'	40'
3	*Instructions*	5'	45'
4	*Beginning experiment*	5'	50'
5	*Stimuli - 1st Part*	20'	1h 10'
6	*First break*	5'	1h 15'
7	*Stimuli - 2nd Part*	15'	1h 30'
8	*Second break*	5'	1h 35'
9	*Stimuli - 3rd Part*	15'	1h 50'
10	*End of experiment*	5'	1h 55'

Table 3.3: Experiment schedule.

3.4.4 Incidents during data collection

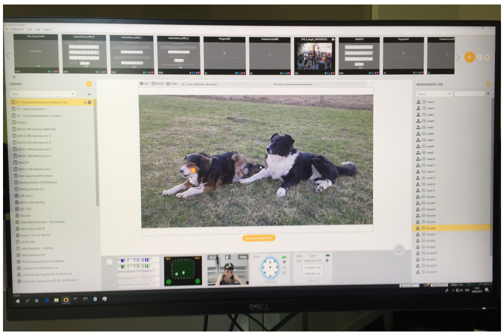
At the beginning of the experiment, we had issues with stimuli reproduction. Videos got frozen the first two seconds and continued with a black frame for a couple of seconds more. It had never happened before. We got in contact with iMotions support and checked codecs, videos format, software issues and drivers; but nothing helped. Almost accidentally, the screen that was mirroring the participant's screen was removed and the issue resulted to be solved. This problem ruined the data from the first two sessions (Male1 and Male2), so we had to assess their sessions again with other participants. Nevertheless, we took it as an extension of the Pilot Experiment.

Another problem arose with Female5; iMotions crashed at the 90th stimulus. No data was saved, so the experiment had to be assessed again by another female participant. As it was a software-related issue, the Behavioral Lab covered the expenses.

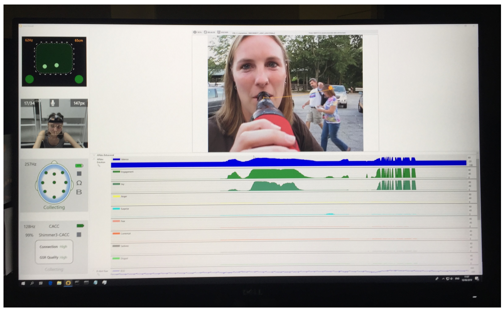


(a) Personality Test

(b) Instructions explanation



(c) Stimuli monitoring



(d) Signals monitoring

Figure 3.13: Experiment explanatory pictures.

Chapter 4

Results

This chapter presents the results obtained with the techniques described in Chapter 3. First, database characteristics are described. Second, the results from stimuli assessments in Mturk are provided and interpreted. Third, a description of the data collected from physiological sensors is presented.

4.1 Dataset

The final result of this project has been the *GALLUS* database. Table 4.1 summarizes its characteristics.

GALLUS - A Video Database for Analyzing Affective Physiological Responses		
Respondents	*Amount*	15 females, 15 males
	Age	Between 19 and 32
	Nationality	Mainly center european
	Studies	Bachelor to Ph.D. students of a Business University (HSG)
Stimuli	*Type*	User-generated videos extracted from YFCC100M Database
	Length	12 - 25 seconds
	Amount	750 assessed 3 times, 25 assessed 30 times
Data	*Personality*	Big Five scores
	Physiological signals	EEG (9 channels), ECG, GSR, electro photoplethysmography eye-gaze, pupillary response, and facial emotion recognition
	Subjective response	9-point Self-Assessment Manikin and a 3-scale liking score

Table 4.1: Final database characteristics.

4.2 Stimuli Corpus

The final corpus of stimuli is described in this section. It is the result of the entire process of selecting and filtering stimuli which is described in section 3.1. The corpus consists of 775 videos extracted from YFCC100M Database [1] which collected images and videos from Flickr.

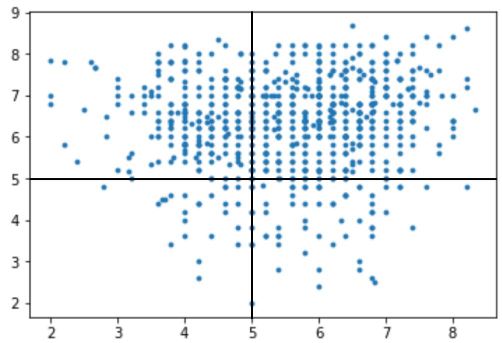
As described in subsection 3.1.6, the whole stimuli corpus was assessed by workers from a crowdsourcing platform. The assessment consisted of ratings of valence and arousal using the 9-point SAM scale; ratings are between 1 and 9, being 5 the neutral punctuation. Each stimulus had at least 5 assessments.

In Table 4.2, an overview of the MTurk assessment to the full corpus is presented. Valence and Arousal mean values are higher than the neutral punctuation. Regarding their variance, both valence and arousal result to be high-variant and therefore subjective.

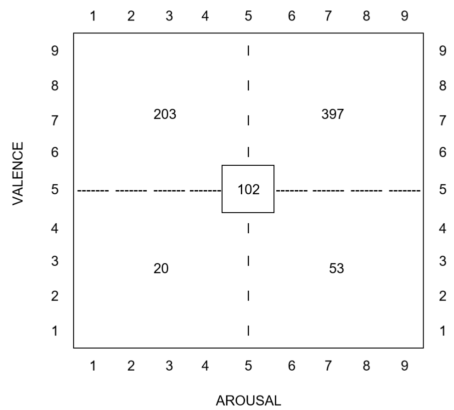
	Mean Value	Mean Variance
Valence	6.30	2.99
Arousal	5.57	4.21

Table 4.2: Results for the stimuli corpus: mean value and mean variance for valence and arousal.

The number of stimuli per cluster is depicted in figure 4.1b. The Stimuli corpus is biased towards high values of both valence and arousal; in fact, cluster 4 contains half of the stimuli. The reason is that user-generated videos posted on social networks like Flickr are biased to happy, high-valence videos.



(a) Scatter plot of stimuli assessments in the circumplex model. X-axis Arousal, Y-axis Valence



(b) Cluster Classification

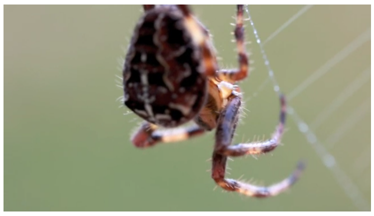
Figure 4.1: Assessment results for the entire stimuli corpus.

Lastly, frames from the representative videos are presented in figure 4.2.

(a) Baby - Cluster 3 (High Val - Low Ar)

(b) Concert - Cluster 4 (High Val - High Ar)

(c) Guns - Cluster 2 (Low Val - High Ar)



(d) Spider - Cluster 1 (Low Val - High Ar)

Figure 4.2: Frames of one representative video per no-neutral clusters. These videos are included in the 25 videos assessed by all respondents.

4.3 Sensors Data

Physiological data were collected by means of iMotions platform [5]. Sensors data include electrocardiography (ECG) with electrodes placed in both sides of the collarbone, 9-channel electroencephalography (EEG), electro-photoplethysmography (EP) applied in one finger to measure the hear-rate, Galvanic Skin Response (GSR), facial emotion recognition, eye-gaze, and pupillary response. The following two figures give details about the signals and how data can be represented.

Figure 4.3 is a capture extracted from the pilot experiment. At the top left, the respondent is shown with the facial box. At his right side, the current frame of the video is presented with eye-gaze information. The yellow point shows where the respondent is directing his gaze, and the lines show its last movements. His facial expressions are analyzed and mapped into facial emotions. Temporal data regarding facial emotions is plotted in below the images. We can see that disgust is detected from the respondent's face in the current frame.

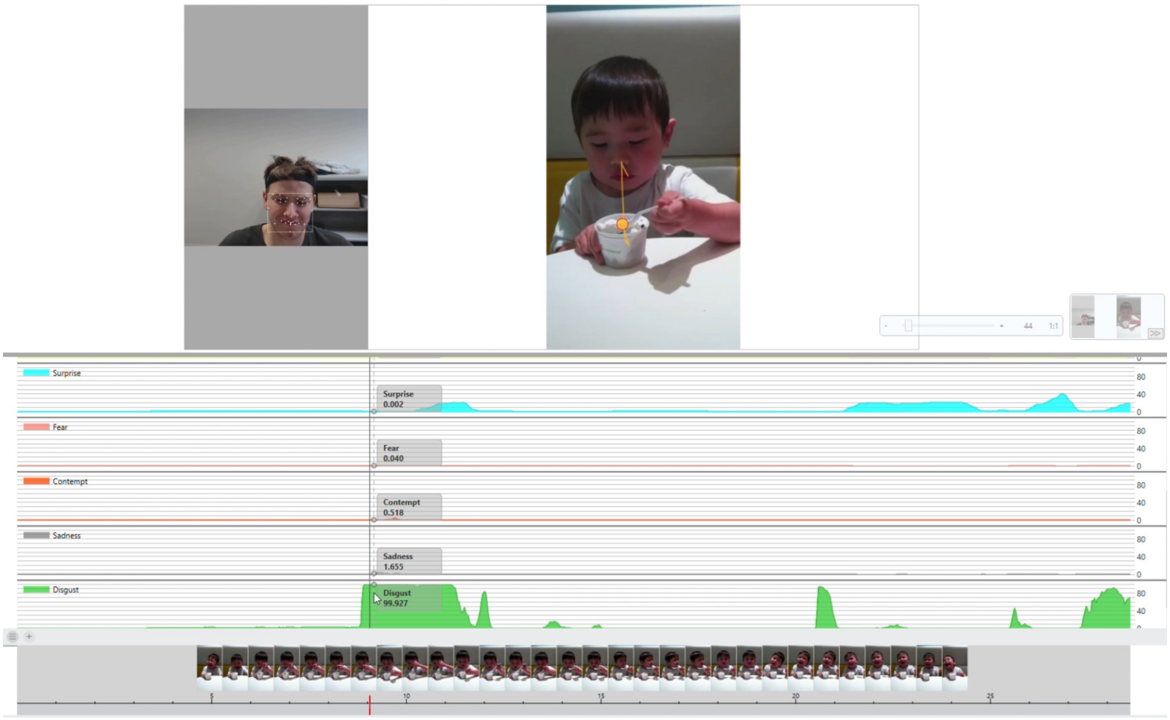


Figure 4.3: Facial emotion recognition and eye-gaze monitoring through iMotions.

Figure 4.4 is a capture extracted from Male2 data while watching the video of the spider, one of the most stimulating videos in the dataset. This participant shows clear physiological signals regarding the arousal. We will review signals from the bottom to the top. First, pupillary response signal (filled in blue) shows a clear dilatation of the pupils once the respondent has recognized the spider, he keeps the same dilatation during the rest of the video. Gaps in this signal correspond to blinks. Galvanic Skin Response is reported to have a physiological delay; human hands need around 4 seconds to start sweating. The participant's GSR signal is clear in this sense, after a certain delay his hands start sweating and the GSR values increase corresponding to his aroused state. Finally, this figure presents two more signals: electro-photoplethysmography (EP) and electrocardiography (ECG). Both signals measure the heart-rate in different ways; EP detects heartbeats which are shown as picks whereas ECG collects the whole electric signal. Still, there is a high correlation between these signals.

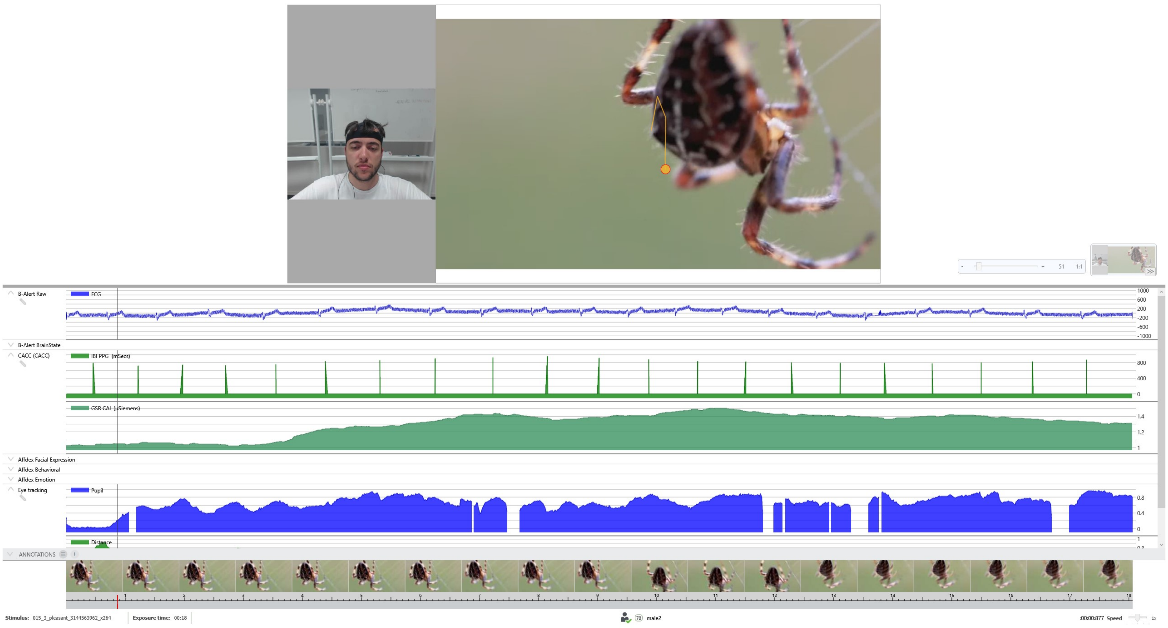


Figure 4.4: Electrocardiography, electro photoplethysmography, galvanic skin response and pupillary response monitored in iMotions.

4.4 Datasets comparison

Finally, a dataset comparison table is presented. Our dataset, *GALLUS*, is compared to state-of-the-art works in terms of stimuli, amount of participants and multimedia content, and physiological signals acquired among others.

Name	Year	Stimuli	Subj.	Exmp.	Affect (Label)	Physiological Signals									PT
						FER	EEG	GSR	ECG	ST	RP	EMG	EG	PR	
DEAP	2012	music videos	32	40	VAD & liking	X	32	X	X	X	X	X			
DREAMER	2017	movie videos	23	18	5-VAD		14		X						
MAHNOB-HCI	2011	movie videos	27	20	VAD, emotion tag	X	32	X	X	X	X		X	X	
DECAF	2015	music videos	30	40	VA	X	MEG		X			X			
GALLUS	2019	user videos	30	100*	VAD & liking	X	9	X	X				X	X	X

Table 4.3: Datasets comparison table. All studies were carried out in a controlled-lab environment. Columns acronyms stand for subjects, examples, facial emotion recognition, electroencephalography, galvanic skin response electrocardiography (Heart rate variability and electro photoplethysmography included), Skin Temperature, respiration pattern, electromyography, eye-gaze, pupillary response, personality test. DECAF uses magnetoencephalogram (MEG) sensor instead of the EEG. * To make our study comparable in term of examples, 100 is the amount of examples shown to each participant.

Chapter 5

Budget

This project has been developed using the resources provided by the Artificial Intelligence and Machine Learning group from the Institute of Computer Science of the University of St.Gallen.

The main costs of this project come from my wages as a research intern at the University of St.Gallen. I have worked full-time for three months and a half receiving a compensation of 1200CHF per month.

	Starting date	**Ending date**	**Months**	**Monthly wage**	**Total**
Period 1	March 1	May 31	3	1200 CHF/month	3600 CHF
Period 2	June 11	June 25	0.5	1200 CHF/month	600 CHF
				Total	4200 CHF

Table 5.1: Wages in the University of St.Gallen

I also received funding from the MOBINT scholarship which had two payments of 400€.

MOBINT scholarship: $2 payments \cdot 400 euros/payment = 800 euros$

Apart from my wages, ICS granted me an office during my three first months of stay. Given a rental fee of 200 CHF, the total office costs are 600 CHF.

Office space: $200CHF/month \cdot 3months = 600CHF$

The Behavioral Lab hardware, software licenses and further materials have an approximate cost of 25.000 CHF per year. As it was used during one entire month, it adds 2.080 CHF to project's budget.

Behavioral Lab Resources: $25.000CHF/year \cdot 1month/12months/year = 2.080CHF$

We have to add the budget for experiments; both in Amazon Mechanical Turk and in the final experiment. Regarding Mechanical Turk, 4 different experiments were launched as explained in Chapter 3. Each assignment for a HIT (Human Intelligent Task) had a reward of \$0.03 for workers plus \$0.01 fees to Mechanical Turk. As calculated in table 5.2, the total cost of AMT

#	**HITs**	**Asgmt/HIT**	**Asgmts**	**Cost/Asgmt**	**Total**
1	11	3	33	\$0.04	\$0.33
2	273	3	819	\$0.04	\$32.76
3	792	5	3960	\$0.04	\$158.40
4	30	5	150	\$0.04	\$6.00
				Total	\$197.49

Table 5.2: Costs of Amazon Mechanical Turk experiments

experiments is of $197.49 .

The videos used for the AMT experiments required an online video storage. As the corpus size exceeded the free storage limit, the Dropbox account was upgraded to Plus. The cost for one month was of 11.99€.

The final experiment, which was developed in the Behavioral Lab, had a cost of 25CHF per hour and participant. Each of the 33 participants attended a two-hours-long session. Therefore, the total amount spent in the final experiment was of 1600 CHF.

Final experiment: $33 participants \cdot 2 hours \cdot 25 CHF/hour \cdot participant = 1650 CHF$

Furthermore, advisors spent part of their working hours assessing the project. Estimating that advisors spent 15 hours per month during the 4 months of the project, and at an average cost of 40€per hour.

$15 h/month \cdot 4 months \cdot 40 euros/h = 2400 euros$

In order to compute the final budget in Euros, each currency was converted with the exchange rates of June 15th. Thus, the total budget for the project is around **11.000€**.

	Amount	**Exchange rate**	**Conversion**
Research intern wages	4200 CHF	0.89 €/CHF	3738.00 €
MOBINT Scholarship	800 €	-	800.00 €
Office Space	600 CHF	0.89 €/CHF	534.00 €
Behavioral Lab Resources	2080 CHF	0.89 €/CHF	1851.20 €
Mechanical Turk experiments	197.49 $	0.89 €/$	175.77 €
Dropbox online storage	11.99 €	-	11.99 €
Behavioral lab experiment	1650 CHF	0.89 €/CHF	1468.50 €
Advisors	2400 €	-	2400.00 €
		Total	**10979.46 €**

Table 5.3: Final Budget

Chapter 6

Conclusions & Future Work

The main objective of this project has been fulfilled; we have collected a dataset that will allow the future development of deep learning models capable of predicting the affective response of a human to videos. Our approach gathers 3000 responses to videos in both physiological signals and subjective annotations.

Moreover, we aimed to extend the current state of the art in this type of video affective datasets. In order to assess the accomplishment of this goal, we will refer to the datasets comparison table 4.3. Despite that our database falls behind DEAP and MAHNOB-HCI in the number of channels for the electroencephalogram and the collection of skin temperature and the respiration pattern signals, GALLUS outperforms the state of the art databases in dimension. Moreover, introduces a novel type of stimuli for affective analysis of physiological responses; user-generated content is our major difference and challenge with respect to the other projects. Another contribution from our work is the handling of the big five personality test.

The presented results motivate further steps regarding this project. First, writing a paper to submit it to ArXiv[1]. After that, the database may be analyzed before making it public. We would make it available to the community uploading it to Kaggle datasets[2] because of its popularity and its possibilities to collaborate and analyze own projects. Thereafter, we aim to publish and present an article to workshops from affective computing conferences.

[1] https://arxiv.org/

[2] https://www.kaggle.com/datasets

Appendices

Appendix A

Gantt Diagram

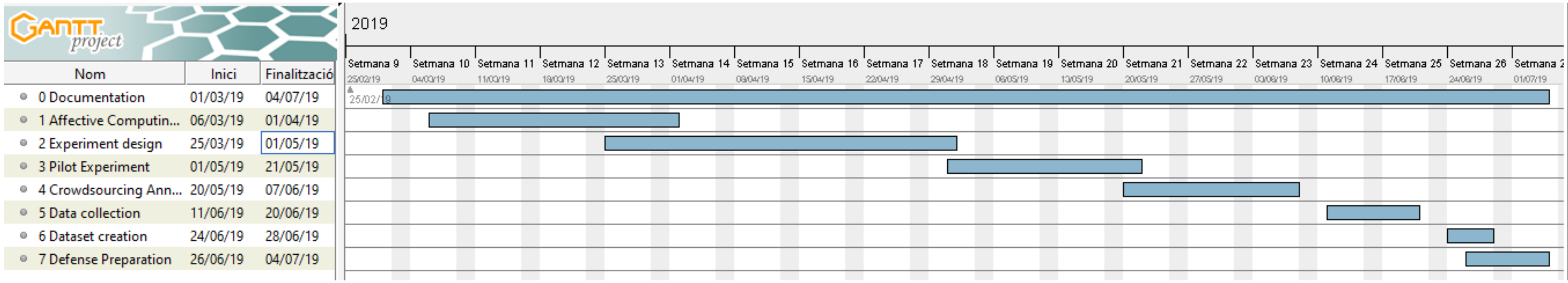


Figure A.1: Project's Gantt Diagram

Appendix B

Behavioral Lab Research Proposal

This document was delivered to the Behavioral Lab so that we could carry out the experiment in their facilities. The Behavioral Lab responsible had to approve it before we could start recruiting participants and launch the final experiment.

The experimental design written in this proposal changed slightly for the actual experiment. The number of participants increased to 30.

Behavioral Lab
Universität St.Gallen

Behavioral Lab (BL-HSG)
Tellstrasse 2
CH-9000 St.Gallen
behaviorlab@unisg.ch
http://www.behaviorallab.unisg.ch

Behavioral Lab Research Proposal

Please send completed proposal form to behaviorlab@unisg.ch

TO BE FILLED BY THE BEHAVIORAL LAB	
Study number:	
Lead:	
Status:	☐ approved ☐ more information is needed ☐ rejected ☐ Recommendation Letter received (only for Bachelor and Master)
Date:	

1. Researcher Details	
Name of main contact:	Marcel Granero Moya
Affiliation: (Institute/Chair)	Institut for Computer Science (ICS-HSG)
Email address:	97granero@gmail.com
Phone number:	(+41) 762676564
Other researchers involved in the project:	Damian Borth (damian.borth@unisg.ch) Barbara Weber (barbara.weber@unisg.ch)
Have you ever conducted an experiment before?	No
Supervisor: (only Bachelor, Master and PhD students)	Damian Borth (damian.borth@unisg.ch)
Major/Program: (only Bachelor, Master and PhD students)	Bachelor's thesis

Behavioral Lab (BL-HSG)
Tellstrasse 2
CH-9000 St.Gallen
behaviorlab@unisg.ch
http://www.behaviorallab.unisg.ch

2. Overview of the research [1]	
Title of the study:	Affective response to videos
Abstract: (purpose of the study, research problem you investigate, research question)	The goal of this project is to create a multimodal dataset by measuring the human affective response to videos. Stimuli will be extracted from YFCC100M dataset which contains user-generated content from Yahoo and Flickr. We will collect several biometric signals (eye-tracking, pupillary response, heart rate, galvanic skin response, facial emotion recognition, and electroencephalography) and a quantitative feedback (self-assessment manikin) for each video. Participants will also answer the Big Five personality test. The final purpose is to publish our dataset under a creative commons license so that the community can benefit from it.

[1] Please note that the title of the project, a short abstract and the researcher's name will be published on our website. If you do not wish to have the information published due to confidentiality reasons, please contact the Behavioral Lab.

Behavioral Lab (BL-HSG)
Tellstrasse 2
CH-9000 St.Gallen
behaviorlab@unisg.ch
http://www.behaviorallab.unisg.ch

Experimental design:	We will collect the neuro-physiological measures explained in the abstract as well as subjective measures from each participant. We will then use these measures to classify the videos in terms of affect that is evoked, i.e., we will allocate these affective states within the valence-arousal plane of the circumplex model of affect. The objective is to obtain a balanced and spread distribution among the Valence-Arousal plane. PROCEDURE Pilot Experiment: We will develop a pilot experiment with researchers from our department. Its goal is to validate both measurements and stimuli that will take place in the final experiment. Final Experiment: A session will follow the schedule below: 1. Gathering participant's anonymous data (age, gender, academic level, country of origin) 2. Setting up biometric sensors while explaining instructions 3. Big Five personality test and Impedance tests for EEG 4. Signing consent form 5. EEG Benchmark test 6. Test Measurement to ensure that sensors are working and the participant understood the instructions 7. 1st part of the measurements (40 trials - 20 to 25 minutes) 8. 5-minute break 9. 2nd part of the measurements (same as the first one) 10. Filling the payment form Biometric signals will be collected during the stimulus visualisation, and after each video we will ask for the user's self-assessment.

Behavioral Lab (BL-HSG)
Tellstrasse 2
CH-9000 St.Gallen
behaviorlab@unisg.ch
http://www.behaviorallab.unisg.ch

3. REQUIREMENTS FROM THE LAB	
Participants	
How many participants are needed? (number per session and total)	We will need 20 participants (10 men and 10 women), each one attending two sessions.
What requirements do the participants have to fulfill? (demographic or other specific attributes)	- english speakers - no psychological diseases or syndromes - no heart pacemaker (ECG) - no glasses (eye-tracker) - no makeup (facial emotion recognition)
Time	
How much laboratory time is required for the experiment? (number of sessions, length of each session)	Study setup, technical problems solving and first measurements: 40 hours (2 weeks) Pilot test: 20 hours (1 week) Experiment: 30 sessions (120 minutes each)
What is your timeline? (set up the study, pre-test, conduct the study)	Set up the study (22/04 - 8/05) Pilot experiment (9/05 - 16/05) Experiment (11/06 - 25/06)
What dates do you prefer?	11/06 to 25/06
Resources	
What lab resources and additional materials are needed? (equipment, software, etc.)	We will need a room for researchers and an adjacent room for the participants. We plan to use all the biometric sensors available. Biometric sensors integrated in iMotions software: EEG, Shimmer (GSR & ECG), Tobii (eye gaze & pupillary response), and facial emotion recognition with Affectiva. Two tablets will be used as intercom; one of these tablets will be used to answer the personality test as well.

Behavioral Lab (BL-HSG)
Tellstrasse 2
CH-9000 St.Gallen
behaviorlab@unisg.ch
http://www.behaviorallab.unisg.ch

4. PAYMENT	
What is the payment structure and amounts for the participants?	We plan to pay 25 CHF/h, up to 50 CHF for each 2-hour session. Participants will fill a payment form with their bank account details.
Source of payment: (e.g. your institute/chair, mini research grant, external source)	The Institute for Computer Science will support the experiment economically.

Appendix C

Study Information in Sona Systems

The Behavioral Lab uses Sona Systems to recruit participants.

Sona Systems is a cloud-based participant management software. It eliminates the need for paper-based methods by integrating every function of the research administration process online. The platform enables researchers to manage schedules and view who has signed up for their study and track a participant's activity.

The following document shows the participants view of the Female study that we launched. An analog study was created for male participants.

Universität St.Gallen

Behavioral Lab (BL-HSG)

Study Information

Study Name	Emotional Response to Videos (Females)
Study Type	**Standard (lab) study** This is a standard lab study. To participate, sign up, and go to the specified location at the chosen time.
Pay	50 CHF via bank transfer
Duration	120 minutes
Abstract	Watch videos while we record your physiological signals (heart rate, electroencephalography, ...) and ask you to rate the emotions you felt during watching videos.
Description	1 hour of preparation for setting up sensors and answering a brief personality test followed by 1 hour of watching and rating videos while we collect your physiological data. **ATTENTION**: Be prepared to get messy hair because of the electroencephalography headset!
Preparation	No makeup! Come 5 minutes before the study begins Fill in the payment form you will receive on your email if you want to get money faster
Eligibility Requirements	English speaker, Female
Researcher	Marcel Granero
Principal Investigator	Principal Investigator

Deadlines	Sign-Up: 6 hour(s) before the appointment Cancellation: 24 hour(s) before the appointment

(18:58)

Appendix D

Consent Form

According to General Data Protection Regulation, participants have to be aware of the data that is being collected, their rights and risks regarding the experiment. Therefore, we created the following form which helped us to get the approval from the Ethics Committee. The Consent Form was signed by all participants before starting the experiment.

University of St.Gallen

Institute for Computer Science

Title of project: Emotional Response to Videos

Principal Investigator: Marcel Granero Moya (marcel.graneromoya@unisg.ch)

Project Summary:

Watch videos while we record your physiological signals (heart rate, electroencephalography, ...) and ask you to rate the emotions you felt while watching videos.

Procedure:

You will answer a personality test while the instructor sets up the electroencephalography headset and the galvanic skin response sensors. You will be asked to take part in the Benchmark test for the electroencephalography and the calibration procedure of the eye tracker, then the experiment will start. You will watch videos from Flickr users and rate how you felt while watching each of them. During the experiment, the GSR signal, EEG, heart rate, eye tracking, pupillary response and facial emotion recognition signals will be recorded.

Video Recording:

As part of this study, you will be video recorded while interacting with the smart device. The video record will be used to analyze facial expressions and recognize emotions.

Confidentiality and Data Security:

Any information that is obtained in connection with this study and that can be identified with you will remain confidential and will be disclosed only with your permission or as required by law. In addition, your information may be reviewed by authorized HSG representatives to ensure compliance with HSG policies and procedures.

All information you provide is considered completely confidential. Participants will be referred to as Female 1, Female2, ... (or Male1, Male2, ...). Data collected during this study will be retained indefinitely in locked cabinets or on password-protected desktop computers in a locked office.

Risks:

A common risk is associated with the infrared light used in the eye-tracking system. The infrared light is emitted from the eye tracker at very low amperage and causes no damage to the eye. This kind of infrared eye tracking has been used for many years at many universities and no negative consequences have been reported.
The use of EEG and GSR devices might cause minor irritation from electrode pads over prolonged periods of time. In case you feel irritation during the experiment, please inform the instructor.
The EEG headset setup may mess up your hair with synapse gel.

Consent Statement and Signature:

I have read this consent form and understand the information that has been provided above. I have had the opportunity to have my questions answered to my satisfaction and understand my rights as a participant. I voluntarily agree to participate in this study. I certify the following:

☐ I allow my data to be stored and used for research purposes.

☐ I allow my data to be used for educational & related non-profit purposes such as conference publications & presentations.

______________________________ ____________________

Participant Signature **Date**

Appendix E

Payment Form

This document is the HSG payment form for guest speakers that was used in our experiment to get the bank acount details from our participants. In fact, our study was the first one to pay participants via bank transfer, and may become a model for the next experiments in the Behavioral Lab.

Universität St.Gallen

Personaladministration
Dufourstrasse 50, CH-9000 St.Gallen, Tel. +41 71 224 39 46, hrm@unisg.ch

HR Master Data Sheet

for faculty and guest speakers

Accounting Section/Institute: Select one element.

Information about gainful employment

***Explanation concerning the State Old-Age and Survivors' Insurance (OASI)**: The fees paid by the University and the institutes are liable to AHV contributions. Members of the faculty must declare them directly to the AHV as income derived from gainful employment. If they are only remuneration for guest lectures, the fees are part of income derived from self-employment provided that the guest speaker is recorded as self-employed by the relevant Compensation Office. The University does not have to declare the fees if the fees are paid to employers of faculty members or into the account of a business partnership.*

Personal data

☐ Ms ☐ Mr

Surname	....	Title, address	....
First name(s)	....	Acad. title	....
Address	....	Acad. degree	....
Postcode/town	....	Nationality	....
Country	....	Date of birth	....
CH: pl. of origin	....	Foreign birthplace	....
Mobile no.	....		

Marital status ☐ Single ☐ Married since Date ☐ Divorced since Date

Denomination ☐ Roman Catholic ☐ Protestant ☐ Other ☐ None

Choice of the applicable variant

☐ A) I am an employee and therefore not affiliated to a Compensation Office as a self-employed person. I therefore request admission to the St.Gallen Pension Fund (sgpk) of the University of St.Gallen (from a gross annual salary of CHF 14,100).

→ In this case, please fill in p. 2 completely.

☐ B) Faculty/guest speakers as a secondary occupation
Since all my lecturing activities at the University of St.Gallen and its institutes are exclusively a secondary occupation and since I am affiliated to my main employer's pension scheme (compulsory occupational pension scheme according to the Swiss Federal Occupational Retirement, Survivors' and Disability Pensions Act (BVG), I waive the right to be accepted into the St.Gallen Pension Fund (sgpk) of the University of St.Gallen.
Name and address of main employer ....

→ In this case, please fill in p. 2 completely

☐ C) Faculty/guest speakers with employer
The fees paid for my lecture(s) are not received by me personally but go directly to an employer or into the account of a business partnership of which I am a partner.

→ Please send us an invoice made out by your employer. In this case, the second page will not be applicable.

☐ D) Faculty/guest speakers who are self-employed
I am affiliated to the following Compensation Office as a self-employed person:
Compensation Office ....
Industry designation ....
Personal insurance number
....

Universität St.Gallen
Personaladministration
Dufourstrasse 50, CH-9000 St.Gallen, Tel. +41 71 224 39 46, hrm@unisg.ch

Please send us an invoice made out in your company's name. Please enclose confirmation by the Compensation Office with your invoice. EU/EFTA nationals have to submit certification by the tax office or the social insurer of their municipality of residence. In this case, the second page will not be applicable.

Salary payments/social insurance

Bank/Post Office ....

Address .... Postcode/town ....

IBAN no. ....

SWIFT code .... Acc. no. ....

SWIFT code and acc. no. need only be indicated for foreign bank accounts.

SI no. 756...

Questions for foreign faculty and guest speakers

Do you have a work/residence permit issued by the immigration authorities?

☐ Yes *It is imperative that you enclose a copy*.

☐ No The work permit will be applied for by the relevant institute.

Starting work without a permit is prohibited and a criminal offence!

Are you in gainful employment outside the University of St.Gallen? ☐ Yes ☐ No

If yes: workload …… %

☐ Other occupation in Switzerland
☐ Other occupation abroad
☐ Other occupation in Switzerland and abroad

Is your employment by the University of St.Gallen a secondary occupation? ☐ Yes ☐ No

Spouse

Surname, 1st name .... Date of birth ....

Nationality .... Employer/town ....

Gainfully employed ☐ No ☐ Yes, since .... Secondary occup. ☐ Yes ☐ No

Country of gainful employment...

Canton of gainful employment (if CH)...

Permit, spouse CH ☐ Permit B ☐ C ☐ G ☐ L ☐

Children ☐ No ☐ Yes, number ....

I hereby confirm the completeness and correctness of the information supplied.

Place/date: .... Signature: ....

Please note

Universität St.Gallen
Personaladministration
Dufourstrasse 50, CH-9000 St.Gallen, Tel. +41 71 224 39 46, hrm@unisg.ch

Please send the completed and signed form to the relevant institute.

Appendix F

Personal Data Sheet

The following sheet was used during the experiment to get participants' personal data anonymously. The *Respondent* cell was filled with the iMotions participant name, e. g. *Female3*, *Male15*. Also, gender, age and nationality were collected. The last data were the numerical results from the Personality Test .

Personal Data

Respondent					
Gender					
Age					
Nationality					
OCEAN					

Respondent					
Gender					
Age					
Nationality					
OCEAN					

Respondent					
Gender					
Age					
Nationality					
OCEAN					

Respondent					
Gender					
Age					
Nationality					
OCEAN					

Respondent					
Gender					
Age					
Nationality					
OCEAN					

Respondent					
Gender					
Age					
Nationality					
OCEAN					

Respondent					
Gender					
Age					
Nationality					
OCEAN					

Respondent					
Gender					
Age					
Nationality					
OCEAN					

Respondent					
Gender					
Age					
Nationality					
OCEAN					

Respondent					
Gender					
Age					
Nationality					
OCEAN					

Respondent					
Gender					
Age					
Nationality					
OCEAN					

Respondent					
Gender					
Age					
Nationality					
OCEAN					

Page out of

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

telecom BCN

Bibliography

[1] Bart Thomee, David A Shamma, Gerald Friedland, Benjamin Elizalde, Karl Ni, Douglas Poland, Damian Borth, and Li-Jia Li. Yfcc100m: The new data in multimedia research. *arXiv preprint arXiv:1503.01817*, 2015.

[2] Rosalind W. Picard, Elias Vyzas, and Jennifer Healey. Toward machine emotional intelligence: Analysis of affective physiological state. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (10):1175–1191, 2001.

[3] Rosalind W Picard. *Affective computing*. MIT press, 2000.

[4] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrel. Caffe: Convolutional architecture for fast feature embedding. In *ACM Multimedia Open Source Competition*, 2014.

[5] iMotions A/S. iMotions Biometric Research Platform 7.1. `https://imotions.com/`, 2018.

[6] Paul Ekman. An argument for basic emotions. *Cognition & emotion*, 6(3-4):169–200, 1992.

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

[8] Albert Mehrabian. Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament. *Current Psychology*, 14(4):261–292, 1996.

[9] Robert Plutchik. Emotion. *A psychoevolutionary synthesis*, 1980.

[10] Rosalind W Picard. Affective computing mit press. *Cambridge, Massachsusetts*, 1997.

[11] Peter J Lang, Margaret M Bradley, and Bruce N Cuthbert. International affective picture system (iaps): Technical manual and affective ratings. *NIMH Center for the Study of Emotion and Attention*, 1:39–58, 1997.

[12] Damian Borth, Tao Chen, Rongrong Ji, and Shih-Fu Chang. Sentibank: large-scale ontology and classifiers for detecting sentiment and emotions in visual content. In *Proceedings of the 21st ACM international conference on Multimedia*, pages 459–460. ACM, 2013.

[13] Rosenberg Ekman. *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)*. Oxford University Press, USA, 1997.

[14] Takeo Kanade, Jeffrey F Cohn, and Yingli Tian. Comprehensive database for facial expression analysis. In *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580)*, pages 46–53. IEEE, 2000.

[15] Maja Pantic, Michel Valstar, Ron Rademaker, and Ludo Maat. Web-based database for facial expression analysis. In *2005 IEEE international conference on multimedia and Expo*, pages 5–pp. IEEE, 2005.

[16] Marian Stewart Bartlett, Gwen Littlewort, Mark Frank, Claudia Lainscsek, Ian Fasel, and Javier Movellan. Recognizing facial expression: machine learning and application to spontaneous behavior. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 2, pages 568–573. IEEE, 2005.

[17] Gary McKeown, Michel Valstar, Roddy Cowie, Maja Pantic, and Marc Schroder. The semaine database: Annotated multimodal records of emotionally colored conversations between a person and a limited agent. *IEEE Transactions on Affective Computing*, 3(1):5–17, 2011.

[18] Peter J Lang, Mark K Greenwald, Margaret M Bradley, and Alfons O Hamm. Looking at pictures: Affective, facial, visceral, and behavioral reactions. *Psychophysiology*, 30(3):261–273, 1993.

[19] Fabien Ringeval, Andreas Sonderegger, Juergen Sauer, and Denis Lalanne. Introducing the recola multimodal corpus of remote collaborative and affective interactions. In *2013 10th IEEE international conference and workshops on automatic face and gesture recognition (FG)*, pages 1–8. IEEE, 2013.

[20] Yu-Gang Jiang, Baohan Xu, and Xiangyang Xue. Predicting emotions in user-generated videos. In *Twenty-Eighth AAAI Conference on Artificial Intelligence*, 2014.

[21] Ellen Douglas-Cowie, Roddy Cowie, Ian Sneddon, Cate Cox, Orla Lowry, Margaret Mcrorie, Jean-Claude Martin, Laurence Devillers, Sarkis Abrilian, Anton Batliner, et al. The humaine database: Addressing the collection and annotation of naturalistic and induced emotional data. In *International conference on affective computing and intelligent interaction*, pages 488–500. Springer, 2007.

[22] Stamos Katsigiannis and Naeem Ramzan. Dreamer: A database for emotion recognition through eeg and ecg signals from wireless low-cost off-the-shelf devices. *IEEE journal of biomedical and health informatics*, 22(1):98–107, 2017.

[23] Mojtaba Khomami Abadi, Ramanathan Subramanian, Seyed Mostafa Kia, Paolo Avesani, Ioannis Patras, and Nicu Sebe. Decaf: Meg-based multimodal database for decoding affective physiological responses. *IEEE Transactions on Affective Computing*, 6(3):209–222, 2015.

[24] Mohammad Soleymani, Jeroen Lichtenauer, Thierry Pun, and Maja Pantic. A multimodal database for affect recognition and implicit tagging. *IEEE transactions on affective computing*, 3(1):42–55, 2011.

[25] Sander Koelstra, Christian Muhl, Mohammad Soleymani, Jong-Seok Lee, Ashkan Yazdani, Touradj Ebrahimi, Thierry Pun, Anton Nijholt, and Ioannis Patras. Deap: A database for emotion analysis; using physiological signals. *IEEE transactions on affective computing*, 3(1):18–31, 2011.

[26] Cyril Laurier, Mohamed Sordo, Joan Serra, and Perfecto Herrera. Music mood representations from social tags. In *ISMIR*, pages 381–386, 2009.

[27] Michael Buhrmester, Tracy Kwang, and Samuel D Gosling. Amazon's mechanical turk: A new source of inexpensive, yet high-quality, data? *Perspectives on psychological science*, 6(1):3–5, 2011.

[28] AJ Kohn. Mechanical turk tips. `http://www.blindfiveyearold.com/mechanical-turk-tips`. Accessed: 2019-06-04.

[29] Murray R Barrick and Michael K Mount. The big five personality dimensions and job performance: a meta-analysis. *Personnel psychology*, 44(1):1–26, 1991.

[30] Truity Psychometrics LLC. The Big Five Personality Test. https://www.truity.com/test/big-five-personality-test, 2019.

[31] 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.

[32] Robert L Hodes, Edwin W Cook III, and Peter J Lang. Individual differences in autonomic response: conditioned association or conditioned fear? *Psychophysiology*, 22(5):545–560, 1985.