Twarz

Tell me What Affects youR buZz

Group work		
Maria Vitali	229036	maria.vitali@studenti.unitn.it
Teresa Rizzi	228813	teresa.rizzi-1@studenti.unitn.it
Francesco Coppola	229005	francesco.coppola@studenti.unitn.it

1. Introduction

"Emotions are a process, a particular kind of automatic appraisal influenced by our evolutionary and personal past, in which we sense that something important to our welfare is occurring, and a set of psychological changes and emotional behaviors begins to deal with the situation."

Paul Ekman, PhD

The problem of emotion recognition has been addressed for years now. Emotions are a very "human" feature, which can give a lot of information on the people experiencing them. Many studies were conducted on the topic, highlighting the importance of (often involuntary) nonverbal behaviours when it comes to emotions. In order to study emotions elicitation and reactions, scientists need a lot of data to observe, compare and analyse.

Twarz is a data collection system that allows researchers to easily collect data in order to build specific datasets for research purposes involving emotion recognition through facial expressions (1).

There already exists some emotional expressions datasets, which are often related to specific problems and contexts. However, the need for new datasets is constantly emphasised, in order to address new issues and research questions. Moreover, the ever-growing development of new technologies and tools opens many possibilities to faster and more precise techniques which aim to capture and analyse a wide range of nonverbal behaviours and emotions.

The aim of this specific project is to provide a new emotional expression dataset by using an easy-to-use data collection tool that, through the smartphone camera and the use of machine learning algorithms, allows to collect and correctly label facial expressions connected to different emotions.

Twarz is composed of a mobile application that offers a clear and fast interface to the user and collects visual data through the smartphone's camera. The image is processed in real time by a machine learning algorithm, which, by combining Action Units detected on the face of the participant, is able to recognize, with a certain probability, the emotion experienced at that specific moment.

The emotion theory used for this project was Ekman's Theory of Discrete Emotions, which recognizes six basic emotions: sadness, happiness, anger, fear, disgust, and surprise. To these, a brand new emotion was added: cringe.

Twarz's aim is to help researchers to collect data easily, taking advantage of the smartphone camera and of Deep Learning models: a whole automated process that supports the entire .

This report presents the motivations and the background theory behind the choices made in the design and implementation process in *Section 2*. Subsequently, *Section 3* describes the application architecture and actual implementation. Then, in *Sections 4* and *5*, the experiment performed to test the system, the data collection and analysis are illustrated. To conclude, *Section 6* provides some final remarks and an evaluation of the results.

2. Getting Ready: Background and Motivations

Twarz uses different theories and studies to support the different choices made during its development.

Choosing the emotion theory

To train the machine learning algorithm, Ekman's Theory of Discrete Emotions was taken into account. Paul Ekman is an American psychologist, a pioneer in the study of emotions and their relation to facial expressions. With his research, considering also previous studies which supported the same idea, he aims to demonstrate the universality of some emotions, also related to facial expressions and nonverbal behaviour (2). His emotion theory is discrete, meaning that he suggests that emotions can be classified according to specific and categorical labels. After the research conducted in Papua Nuova Guinea, Ekman's findings provided evidence that there are at least six basic emotions, associated with their specific facial expressions, that are universal.

The six basic discrete emotions, according to Ekman's theory, are: happiness, sadness, anger, disgust, fear and surprise. However, it is specified that no emotion exists as a single affective or psychological state. Instead, emotions are comprised of a family of related emotional states which are variations on a shared theme (3).

Ekman also conducted important studies on facial expressions, stating that they are both universal and culture-specific. The scientist discovered strong evidence of universality of some facial expressions of emotion as well as why expressions may appear differently across cultures (4).

Ekman's theory is one of several discrete theories of emotions, which consider a small number of basic emotions. Every emotion has its own profile, composed by a specific action tendency, a physiological response pattern, a motor expression and a feeling state, activated by different appraisals. More complex emotions are considered to be combinations of the basic ones.

Another kind of emotion theory is the dimensional model. Here each emotion is defined by one or more dimensions. Scientists proposed different models, also changing the number of dimensions taken into account. One of the most well-known is the one proposed by Russel: a two-dimensional theory that measures valence and arousal (5). Each emotion is a linear combination of these two variables. It could be interesting, as a future development, to take into account also the dimensional aspect of emotions, making Twarz capable of recognizing the level of arousal or valence of each recognized emotion.

Currently, the system is equipped to provide a **percentage of probability**: the higher the percentage, the more likely it is that the emotion stated is the actual emotion shown by the facial expression examined.

In the end, we chose Ekman's discrete model of emotions to narrow our focus and to test how this kind of approach to data collection and analysis might work. To the 6 basic emotions we added one of our own: cringe. Nowadays, the word "Cringe" (6) is associated with a specific reaction or emotion, defined as a mixture of embarrassment and awkwardness. If someone or something is "cringe", it makes you feel uncomfortable in an awkward way.

This new emotion was added to the basic ones proposed by Ekman, in order to try to define a new encoding and to test the app also on unusual contexts.

On facial expressions

Facial expressions are believed to convey a lot of information about people's emotional state. It is a very overt form of nonverbal communication and behaviour.

For decades, scientists have examined facial expressions, studied them, tried to classify them and outline a comprehensive way to encode them. In 1978 (7), Ekman and Friesen proposed FACS, the Facial Action Coding System, later updated in 2002. FACS is a system to taxonomize human facial movements: it encodes any visible change and secondary movements on the face. Single "actions" across the face are called Action Units (AUs), each of which defines specific movements of face muscles (8).

There are more than 80 distinct Action Units. AUs can be recognized both by humans and by automated systems, and this is one of the reasons why FACS was chosen for this specific project, in order for Twarz to automatically detect emotions through facial expressions. By combining different AUs, emotions can be distinguished and encoded. *Table 1* shows a reference for the basic emotions Action Units combinations.

Emotion	**Action Units**
Happiness	6, 12
Sadness	1, 4, 15
Surprise	1, 2, 5B, 26
Fear	1, 2, 4, 5, 7, 20, 26
Anger	4, 5, 7, 23
Disgust	9, 15, 17

Table 1

Since Twarz's emotion recognition uses AUs and facial expressions, a proper encoding for the "Cringe" emotion had to be defined. At first, we tried to analyse its different characteristics by considering our perception of it, so how we picture it, what it elicits in us and how it is conveyed through facial expressions.
We then compared our ideas with online content (pictures and videos) and with 3 external people.

At this point, we outlined an initial composition of Action Units that could represent our idea of experiencing "Cringe" feelings.

To test our hypothesis, we asked 4 people to show us how they would react to something cringe. We recorded short videos (3-4 seconds), after asking for their consent (*Attachments > Cringe Coding > Videos*). By using OpenFace (9), we were able to measure Action Units involved in every "cringe face" performance. Having all the data collected and transferred in Excel (*Attachments > Cringe Coding > Data*), we could compare the various AUs.

In the end, we could identify some recurring patterns and combinations of AUs. Many of them coincided with our initial list of AUs for Cringe. On the other hand, AU1, AU13 and AU15 were added after carefully analysing the videos and the data gathered. *Table 2* shows the selected AUs for the Cringe emotion, with a small description.

Emotion	**Action Units**
Cringe 	AU**4** - Brow lowerer AU**6** - Cheek riser AU**7** - Lid Tightener AU**10** - Upper Lip raiser AU**12** - Lip corner puller AU**14** - Dimpler AU**25** - Lips part

Table 2

Our choice to use FACS as a reference to better analyse facial expression is due to the possibility of using it also in automated systems. It provides a fairly accurate coding of all the different facial movements and features.

The idea is to have a new facial expression emotion dataset, obtained by exploiting new technological devices, such as the smartphone.

There are many other datasets of facial expressions, which can be used for research and which were built using different kinds of methods. For instance, "AffectNet" (10) is a database of facial expressions in the wild, which contains images collected from the

Internet. Images were manually annotated for the presence of seven discrete facial expressions and the intensity of valence and arousal (dimensional model).

We created our own small dataset. We used our own pictures where we tried to stress specific basic emotions and trained the Machine Learning with them. Ideally, the more pictures, the more the model will be trained and the more accurate it should be.

Emotion elicitation

To test the system, some short videos from the Internet were chosen. Eliciting emotions through videos and movies has been widely studied and discussed in different contexts.

Emotion elicitation through media is issued in various fields: from research in affecting computing or signal processing, to psychology, to business and marketing, to advertising and entertainment. Guidelines to elicit emotional reactions when creating videos can be found online (11).

Different research has been conducted on the importance of emotions when it comes to health, and on how to elicit such emotions in patients within the field of e-health monitoring systems (12); others tackled the issue of age difference when it comes to selecting videos which could elicit feelings and emotions in the viewers (13); others focused on differences between elicitation methods, such as between videos and pictures (14).

The 12 videos we chose (see *Attachment 1*) were mostly based on our subjective emotions. We tried to understand which feelings were elicited in us and how people could react after the presentation of the videos for the first time.

After some research we came out with:

- 2 videos which (supposedly) elicited happiness
- 3 videos which (supposedly) elicited sadness
- 1 video which (supposedly) elicited anger
- 1 video which (supposedly) elicited fear
- 1 video which (supposedly) elicited disgust
- 3 videos which (supposedly) elicited surprise
- 1 video which (supposedly) elicited cringe

The research of the videos was performed mostly on Youtube and Reddit.

We then experimented and collected actual reactions from participants.

Face processing

Twarz uses AI's Face Recognition techniques (15) (*Section 3* of this document).

Nowadays, research on Artificial Intelligence, Machine Learning and Deep Learning is progressing really fast. Facial Recognition employs machine learning algorithms which find, capture, store and analyse facial features in order to match them with images of individuals in a pre-existing database.

When dealing with Emotion Recognition, automated systems need to be able to (a) detect faces, i.e. to detect the location of the face in any input image or frame; (b) be able to measure and extract features, that will allow the algorithm to match the face to other faces in its database; and finally (c) be able to detect and classify emotions by analysing specific features and movements and matching them against known faces and labelled emotions in a database.

Twarz has a small training set of images composed of our pictures. We tried to train the model by giving our own representation of happiness, sadness, anger, fear, disgust and surprise.

The data collected from our participants was analysed by the app, which tried to match facial expressions to the aforementioned new dataset. This way, creating datasets and training the deep learning algorithm can be straightforward. The system can be easily adapted to allow labelling new pictures, so that it facilitates an emotional datasets building process.

3. Architecture design and Implementation: the application and data processing

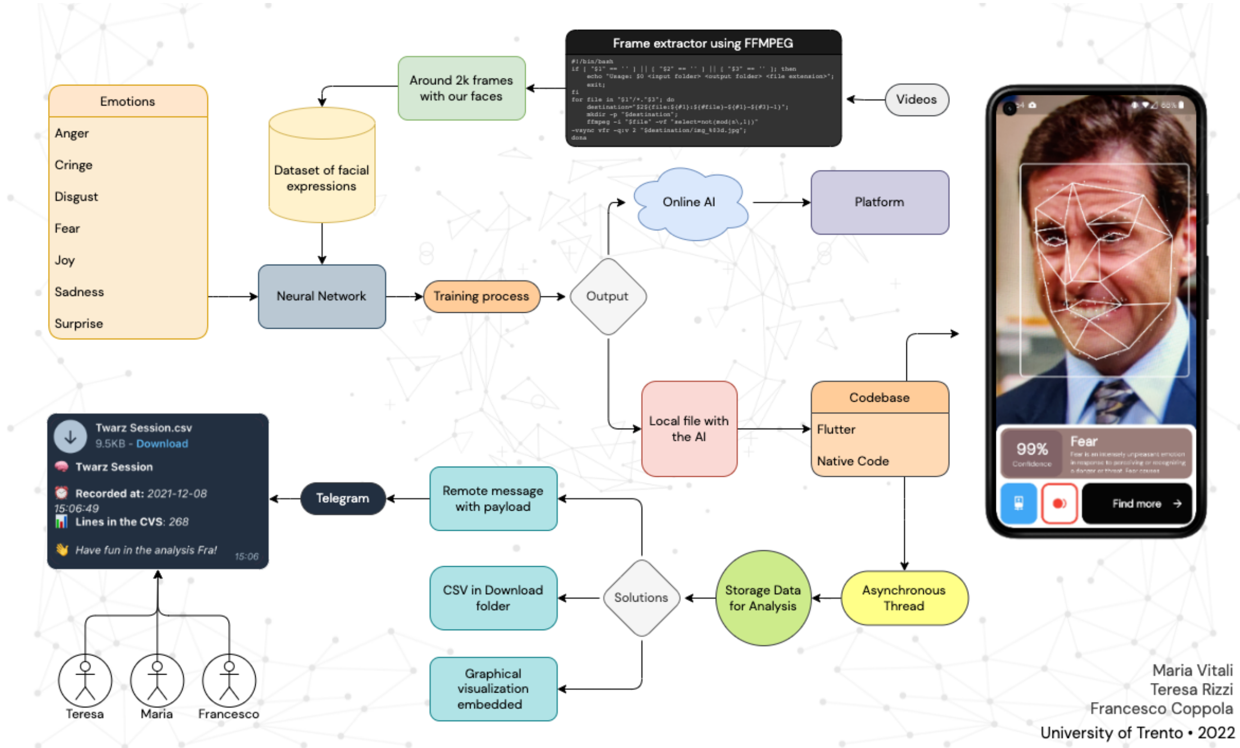


Figure 1

Twarz is able to easily create a connection between the experimenter, the experiment participant and the automated system behind the whole recognition process. As shown in *Figure 1*, before presenting the user with the actual mobile application through its UI, the Neural Network had to be trained. Through some videos we shot, expressing our own ideas of the six Ekman's basic emotions and the newly added cringe feeling, and after having extracted multiple frames, a small dataset of facial expressions was created. This was used to train the Neural Network.

In the meantime, the mobile application was designed and implemented.

After opening the app, users can immediately visualize their smartphone camera image. The system is able to recognize and outline facial features and shapes, in order to help people understand that it is going to analyse those specific characteristics.

Simply by tapping the recording button in the bottom selection menu, the app will start to record data. The moment the recording ends, all the data gathered during that session, will be collected and stored in a CSV file and sent to a Download folder in the experimenter device. In addition, a bot Telegram is provided, to immediately notify the experimenters that a new data file has been created and is ready to be analysed. This way, the experimenters can easily manage the whole data collection process.

Technologies stack

- **TensorFlow**
    - Free and open-source software library for machine learning and artificial intelligence. It can be used across a range of tasks but has a particular focus on training and inference of deep neural networks
- **Teachable Machine**
    - Tool that makes it fast and easy to create machine learning models for your projects. Train a computer to recognize your images, sounds, and poses, then export your model for your application
- **FFMPEG**
    - A complete, cross-platform solution to record, convert and stream audio and video
- **Flutter**
    - Main framework for the building of the front-end applications
- **Telegram API**
    - TwarzBot
    - Dart interface to connect the Telegram API with the native application in an asynchronous thread

The mobile application: challenges

The developed application was quite challenging for three main aspects:

- Accessing the camera's API through the native code implementation in Kotlin
- Using the TensorFlow plugin taking in consideration the performance of the device that might be not that good
- Recognising in real time the face of the user

Flutter, an UI framework written in Dart, was the main tool used for the creation of Twarz.

It gave us the possibility to build a really appealing and easy-to-understand interface thanks to its paradigm based on the reuse of `widgets`. This unique feature helped us in reducing the computational cost in order to have a less impact on the RAM usage of the smartphone.

Furthermore, having a predefined set of components that we designed allowed us to keep a coherence style and a shared UX between all the screens.

Since the camera was the main characteristic of the whole technologic stack we spent a lot of time in binding the native API of that with Flutter. Finally we were able to merge the Kotlin code with the Dart one. Once we completed this part we moved on facial recognition, and to do that, we stuck with an external package called Google ML Vision, that out of the box could simply recognise a face from a video stream. Basically we connected the raw stream of pixels coming from the camera to the package just described.

At this point we knew in which position, in terms of coordinates along the screen, the face and its details were. At this point the following workflow was pretty easy, we just needed to draw some lines along all these discovered pixels and thankfully for us Flutter provides an object called `Painter` that was able to do this task pretty flawlessly.

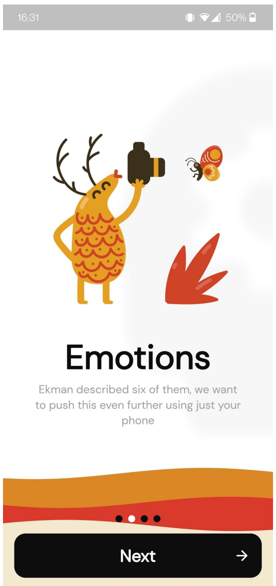
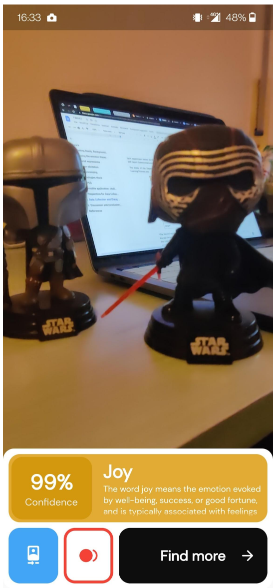
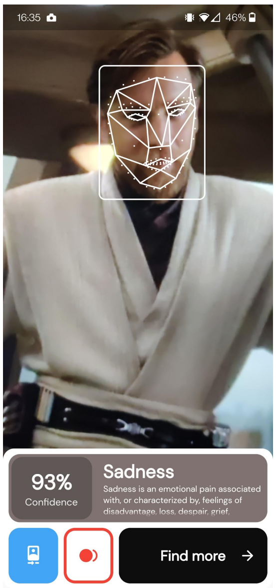
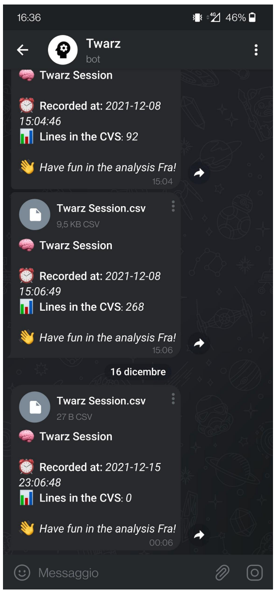
The last missing feature was the emotions recognition. Indeed, after that we trained our neural network using Python and TensorFlow we got back a really tiny file *(in the order of a couple of megabytes)* with a `tflite` extension. We added it in the `assets` folder of the project and we were almost ready to go at that point.

Using the official `TensorFlow Lite` plugin in fact made everything quite accessible, above all for the internal manager of resources that was able to take up and free computational power at its best without any effort from our side.

Finally, we simply mapped the outcomes of the trained neural network with our emotions labels while showing them on the device's screen at each update of the facial expression.

Moreover the whole codebase can be found on `GitHub` and is shared with the `MIT License`.

Application and Bot screenshots

<table>
<tr><th colspan="4">Screenshots</th></tr>
<tr>
<td>
</td>
<td>
</td>
<td>
</td>
<td>
</td>
</tr>
</table>

4. Preparation for Data Collection

Our specific hypothesis leans on the possibility that the app will be able to identify/detect different emotions based on the Action Units that we described before (learned through a Machine Learning process). Simply, we want to explore what type of emotions the Twarz app would identify during the presentation of the videos we found online.

The subjects of our experiment are students from HCI master course from the University of Trento, chosen with a convenience sampling.

Our methodology foresees a within-subjects experimental design with one independent variable and one dependent variable. At this point, the I.V. is the video we present (List in *Attachment 1*) and the D.V. is the identification of the emotions and the "dimension", in terms of percentage of probability, of the recognizable emotion by the Twarz app. Our focus is not on the "correct" or "incorrect" expression of emotions by the subject in relation to the videos, but rather on the reliability of labelling made by the app and humans.

As previously explained in Section 2, the only theory we relied on is a discrete theory, we chose to follow Paul Ekman's Theory of Emotions with his six labels (happiness, sadness, surprise, fear, anger and disgust plus one coded by us using OpenFace, the cringe face expression).

We start with the idea of having 20/25 subjects where every participant should go under every condition.

Our experiment should last 20 to 30 minutes because every video should be seen in its entirety and only once. We plan to organise the experiment at San Bartolomeo's student residence where we have the possibility to use quiet and small rooms to make participants more comfortable.

When a subject comes into our laboratory, we ask him/her to sit in front of a table where our device is ready to start the recording and the analysis of the emotions. Before the beginning of the registration, we present an informational sheet and a consent form (*Attachment 2*) where participants can clearly read our set of instructions for both the task evaluation and the questionnaire.

The instructions are:

- Sit, relax and watch the videos we will propose in the device in front of you.
- We are here to learn and discover what type of emotions the videos are eliciting in you.

- After every video we will ask you to complete a few questions about what you felt in that moment.
- After the experiment you can ask more information about the experiment and the app we are using.
- Enjoy the experience!

When he/she completes the consent form, we present the first video (all the videos are presented in random order). After that, we ask to complete the part of the questionnaire related to the first video (*Attachment 3*). This procedure is repeated with every video.

The data we gathered includes one or more of the seven emotions Labels (0, 1, 2, 3, 4, 5, 6).

The facial expressions of our experimental group is evaluated both by the app and by three different observers (Francesco, Maria and Teresa, *Attachments* > Data Analysis > *Observer Data*) at the same time; plus, we will collect the participant's point of view through the questionnaire.

After the collection we analyse the data. We want to find how much inter-rater reliability there is between the app's detection and the experimenters' point of view (Cohen's Kappa App-Experimenters).

Then we confront the emotions detected by the app with the ones declared by the participants.

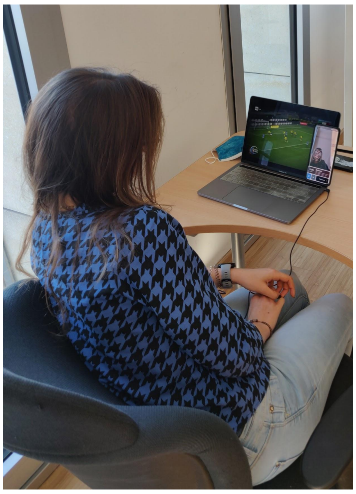
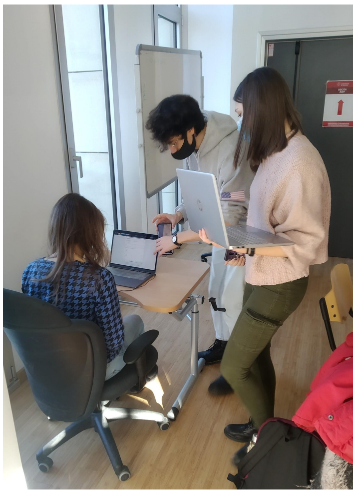
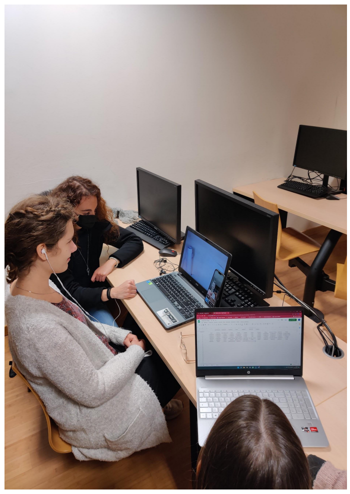
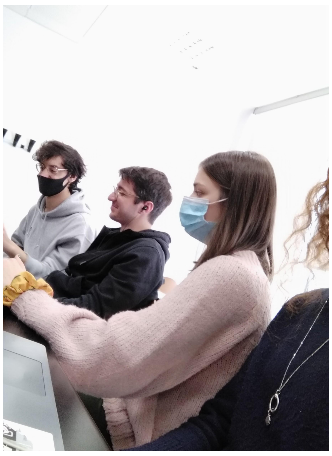
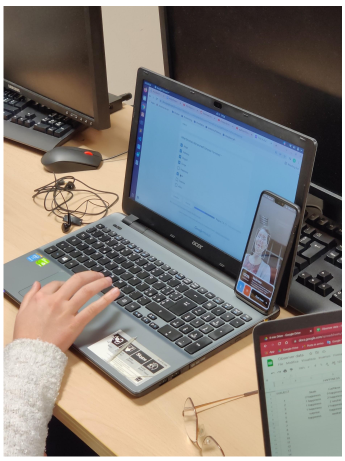
If our Hypothesis is correct, the app should be able to detect the same emotion as the human with a substantial agreement with an external observer.

5. Data Collection and (Data) Analysis

On Tuesday December 7th, we ran the experiments. We found 7 subjects, less than expected, but we agreed on stopping the investigations because we were realizing that the app was not responding the way we hoped.

The experiment was not conducted at San Bartolomeo’s student residence because no one was available the day we proposed, so we took advantage of some empty hours in the university where we were sure our colleagues would have participated.

During the experiment we asked the participants to take off their masks in order to allow the app to better analyze their expressions. This procedure was done in a safe environment as we made sure that the place was empty and the participants were securely distant from us **(we always had our masks on)**.

Pictures		
		
		

Each experiment lasted 20/25 minutes. We collected three types of data: **App Data**, **self-report Questionnaire** and **Observer Data** *(our external point of view)*.

The levels of the Dependent Variable, linked to the code we used in the Machine Learning Process, are:

Level of the D.V.	**Identifiers**
Anger	0
Cringe	1
Disgust	2
Fear	3
Happiness	4
Sadness	5
Surprise	6
Neutral*	7
Other*	8

> **The Neutral expression and Other expressions (7, 8) are not coded in the app because we only considered Ekman's emotions plus one. During the collection and analysis of data we discovered that most of the time people rest their face in a more "neutral" expression (not explicitly considered in Ekman's Theory). With "other" expressions we wanted to stress the possibility of having a wider range of emotions, harder to code using the FACS system (AUs).*

We analysed our data with Jamovi, a free software for statistical analysis.

First, in every CSV file *(sent via Telegram)* we filtered out the Sadness emotion `(NOT(emotion==5))` because our data was full of "5" (Sadness).

This happened because we didn't code the "neutral" expression in the app. At the start of every video the app misidentified "sadness" (5) just because people were in fact "neutral".

Then, we analysed the mode of every video for every person and we added these results in a new Jamovi file (*Attachments > Data analysis > Resume analysis.omv*).

In this file we also added our data collection *(external p.o.v.)* and the coded questionnaire. The problem here is that both from an internal perspective and from an empathic vision, we ended up having data with mixed emotions. That was not easy to analyse because of the quantity of raw data we had and the absence of SPSS. For these reasons we were obliged to select just one main emotion for each video both in the questionnaire data and our observer excel.

The main analysis was inter-rater reliability. We performed three types of inter-rater R. one for each video: we confronted app data vs questionnaire, app data vs our excel and excel data vs questionnaire. Everything was calculated with Cohen's Kappa.

Unfortunately, most of the results were statistically not significant. Most of our P-values are greater than our α (0.05). For this reason we concluded that the appraiser agreement is due to chance *(Fail to reject H0)* because we don't have enough evidence to conclude that the appraiser agreement is different from what would be achieved by chance.
Only two results are statistically significant but they do not add significance to our research: these are the i-r reliability between our excel and the questionnaire of the third and the fourth videos.

Finally, we analyzed the participant's emotions through a qualitative ethnography of the questionnaire we proposed.

We noticed that every video has elicited more than one emotion.

We analyzed each reaction for every video (*Attachments > Data analysis > Questionnaire Data)*:

1. The majority (4 out of 7) selected "happiness". 3 people felt both "surprised" and "cringe". Only 2 choose "fear"
2. Most of the subject (5/7) felt "happy". Other emotions have only one selection: Sadness, Disgust. Someone described "other" emotions: *Cuteness, Calmness, Wholesomeness*.
3. 5 out of 7 selected "Sadness". 3 people also felt "happy". Also "cringe" was elected. "Other" emotions were: *Calmness, Chillness*.
4. The majority (4 out of 7) selected "cringe". 3 people also felt "happy" and "sad" at the same time. "Surprise" and "disgust" were chosen by only 1 person. "Other" emotions were: *Fun, Indifference, Empathy*.
5. 4 people out of 7 felt both "happy" and "surprised". 2 people also selected "sadness" and "cringe". "Fear" and "Disgust" were chosen by 1 person. "Other" emotions were: *Fun, Anxiety*.
6. "Anger" and "Cringe" received 4 votes each. 3 people also selected "Disgust". "Surprise" and "Fear" were chosen 2 times. Only one person selected "Happiness" and/or "sadness". "Other" emotions were: *Frustration, Hate*.

7. Nearly everyone (6 out of 7) selected “Surprise”. 5 people also felt “fear”. The other emotions had 1 vote each, except for “happiness” that has none. “Other” emotions were: *Fun, Suspense, Confusion*.
8. The majority of participants (5 of 7) chose “Disgust”. 3 people selected “Sadness”. “Surprise” and “Fear” received 2 votes each. 1 participant selected “Happiness”.
9. Many participants (5 of 7) claimed they experienced “Happiness”. 4 out of 7 selected “Surprise”. Others (3/7) specified “Cringe” as one of the emotions felt. “Other” emotion: *Funny*.
10. 4 out of 7 participants felt “Happiness” and “Surprise”. 3 of 7 selected “Sadness”. The emotions “Fear” and “Cringe” received 1 vote each. “Other” emotions were: *Satisfying* and *Empathy*.
11. The majority of the people (6 of 7) selected “Sadness”. 3 of 7 people chose “Surprise”. “Happiness” and “Anger” had 2 votes each. In the “other” category: *Empathy* and *Frustration*.
12. Nearly everyone (6 of 7) selected “cringe”. Many people (5 of 7) also selected “disgust”. 3 people claimed they felt “happy” and 3 people claimed they felt “sad”. “Fear” also received 1 vote. The only “other” emotion expressed was: *Unexpected*.

Finally, we conclude that all of the participants have experienced mixed and *(sometimes)* opposite emotions *(taking into account positive or negative valence)*.

6. Discussion and conclusions

From our analysis we discovered that our app *is not really working as we expected*.

We saw some changes in expression recognition during the presentation of the videos but, after the inter-rater reliability analysis, we can not conclude any statistical significance.

With the ethnography we made on the questionnaire, we also became aware of the reality of mixed emotions. The majority of participants always chose more than one emotion for each video and sometimes they declared having another type of emotion, far from the six primary ones described by Ekman.

We tried to analyse our flaws in the whole experiment: firstly we noticed that our dataset was not trained enough for male features **(the only male subject we had was Francesco, so in percentage we covered just a small part of the whole infrastructure)**. Then, we perceived some mis-recognition in expression when the subject had glasses on (we asked to remove them if possible).

Statistically speaking we are aware of two main problems in our research: one of them is the lack of participants. It seemed that our results were due to chance (P-value > α) and if we have had more subjects maybe this type of outcome could have occured.

Secondly, we have reduced both the data of the app and the one of the questionnaires: as said before, we came across mixed emotions. We chose to use the mode from each video (most frequent face expression) losing the other detected emotions.

Moreover, we did not have any type of "neutral" expression, so the app recognized the relaxed face as "sadness" polluting the data.

In addition, we felt obliged to use only one selected emotion from the multiple choice questionnaire because with Jamovi we found out that it is not possible to assign more values to one item *(e.g. "In the first video I felt both happy and sad")*.

Apart from the problems described above, we are happy with our experiment and general research.

Future developments

It could be interesting, as a future development, to take into account also the dimensional aspect of emotions, making Twarz capable of recognizing the level of arousal or valence of each recognized emotion.

Moreover, we think that it would be really interesting to focus on the following aspects of the application in order to have a more solid and stable product:

- Having a more robust codebase, tested and deeply controlled can improve the battery drain that we noticed. Performance at the actual state of the art is really good, *but the battery doesn't think the same*.
    - Moving the analysis of the raw data coming from the video stream to a central backend instead of computing everything inside the CPU and RAM of the smartphone would be a huge improvement.
- Adding a social aspect to the application could increase the usage and engagement of people with the whole system, and, as a result, more participants could contribute to our dataset.
- Monitoring heartbeat through external devices (e.g. smart bands) could provide additional information, resulting in a more clever emotion recognition process.
- User-profiling could help developers to better understand how the application is used, in order to design an improved interaction and experience for the users, thanks to the provided feedback.
- Combining the whole body posture with the facial expressions could increase the accuracy of the results and the general understanding of nonverbal behaviors.
- A bigger and more complex dataset would be needed.

Overall, we believe this system could be systematically improved to better fit the actual needs of users, being them researchers or experiment-participants.

References

(1) https://edps.europa.eu/data-protection/our-work/publications/techdispatch/techdispatch-12021-facial-emotion-recognition_en

(2) Ekman, P. (1970). Universal Facial Expressions of Emotions. California Mental Health Research Digest, 8(4), 151-158.

(3) EKMAN, PAUL and FRIESEN, WALLACE V.. "The Repertoire of Nonverbal Behaviour: Categories, Origins, Usage, and Coding" Semiotica, vol. 1, no. 1, 1969, pp. 49-98. https://doi.org/10.1515/semi.1969.1.1.49

(4) https://www.paulekman.com/universal-emotions/

(5) Russell, James A. "A circumplex model of affect." Journal of personality and social psychology 39.6 (1980): 1161.

(6) https://www.urbandictionary.com/define.php?term=Cringe

(7) Ekman P, Friesen W (1978). Facial Action Coding System: A Technique for the Measurement of Facial Movement. Palo Alto: Consulting Psychologists Press

(8) https://imotions.com/blog/facial-action-coding-system/

(9) Baltrusaitis, Tadas & Robinson, Peter & Morency, Louis-Philippe. (2016). OpenFace: An open source facial behaviour analysis toolkit. 1-10. 10.1109/WACV.2016.7477553

(10) Ali Mollahosseini, Behzad Hasani, and Mohammad H. Mahoor, "AffectNet: A New Database for Facial Expression, Valence, and Arousal Computation in the Wild", IEEE Transactions on Affective Computing, 2017.

(11) https://www.vyond.com/resources/how-to-create-videos-that-evoke-emotion/

(12) L. D. Rumpa, A. D. Wibawa, M. H. Purnomo and H. Tulak, "Validating video stimulus for eliciting human emotion: A preliminary study for e-health monitoring system," 2015 4th International Conference on Instrumentation, Communications, Information Technology, and Biomedical Engineering (ICICI-BME), 2015, pp. 208-213, doi: 10.1109/ICICI-BME.2015.7401364.

(13) Hazer D., Ma X., Rukavina S., Gruss S., Walter S., Traue H.C. (2015) Emotion Elicitation Using Film Clips: Effect of Age Groups on Movie Choice and Emotion Rating. In: Stephanidis C. (eds) HCI International 2015 - Posters' Extended Abstracts. HCI 2015. Communications in Computer and Information Science, vol 528. Springer, Cham. https://doi.org/10.1007/978-3-319-21380-4_20

(14) Uhrig MK, Trautmann N, Baumgärtner U, Treede R-D, Henrich F, Hiller W and Marschall S (2016) Emotion Elicitation: A Comparison of Pictures and Films. Front. Psychol. 7:180. doi: 10.3389/fpsyg.2016.00180

(15) V. Jacintha, J. Simon, S. Tamilarasu, R. Thamizhmani, K. Thanga yogesh and J. Nagarajan, "A Review on Facial Emotion Recognition Techniques," 2019 International Conference on Communication and Signal Processing (ICCSP), 2019, pp. 0517-0521, doi: 10.1109/ICCSP.2019.8698067.