Emotion Star: The Facial Expression Recognition Game

EMIL JOSEFSSON*, FLORIAN HILD*, MATTIAS HALLIN*, LUDVIG ANDERSSON*, and JESPER LINDSTRÖM*, Chalmers University of Technology, Sweden

To evoke interest in technology among teenagers, an innovative game related to facial recognition technology was created. The project was conducted in the context of an educational project in interaction design. Furthermore, the project explored how facial expressions can be used as a method of interacting with the UI. A concept was designed to be suitable to the context of teenagers in the science centre environment. The final prototype is inspired by the other games in the musical genre, but in the sense that the players need to make one of five particular facial expressions at the correct times. Finally the players are presented with statistics about their facial expression performance. To conclude, user testing revealed that the game was considered to be fun and innovative. Furthermore, the interaction was found to be functional to some degree.

Additional Key Words and Phrases: Facial, expression, recognition, collaborative, multiplayer, educational, game

ACM Reference Format:
Emil Josefsson, Florian Hild, Mattias Hallin, Ludvig Andersson, and Jesper Lindström. 2020. Emotion Star: The Facial Expression Recognition Game. In *CHI '21: ACM CHI Conference on Human Factors in Computing Systems, May 08–21, 2021, Yokohama, Japan.* ACM, New York, NY, USA, 8 pages.

1 INTRODUCTION

According to a Swedish study by Teknikföretagen and Ungdomsbarometern [8], the interest in technology and engineering among the youth in Sweden is gradually increasing. However, the study notes that there is still room for improvement, especially among girls. The context of this project is Universeum, the national science centre of Sweden. The project was conducted over eight weeks as a part of a course at the Interaction Design and Technologies master program at Chalmers University of Technology. This project aims to explore new ways of interaction to create an innovative game with interactions based on facial recognition technology. The game should be fun and interactive while simultaneously sparking interest in programming and data science. Furthermore, this project explores how facial expressions can be used as the method of interaction, to control a computer game interface entirely through a webcam instead of using the traditional WIMP interfaces. [1] .

2 BACKGROUND

This section describes the context of the science center Universeum, the goal of engaging teenagers within the field of technology and how we decided to work with facial expression recognition.

*All authors contributed equally to this research.
[1](WIMP is short for; windows, icons, menus, pointers)

Manuscript submitted to ACM

2.1 Context: Universeum

Universeum is a science centre in Gothenburg, Sweden, primarily targeting the age span of 12 to 16. As a result of the Covid-19 pandemic, Universeum created a digital platform, Digitala Universeum, in an effort to be accessible online. The context of this project is a combination of the online platform and the physical exhibition.

2.2 Goal: Engage Teenagers in Technology

The study by Teknikföretagen and Ungdomsbarometern [8] lists several recommendations on the topic of evoking interest. Two of these are to connect the initiative to the interests of the target audience (such as gaming), as well as creating a "wow-factor". This is also supported by Skolinspektionen (the Swedish Schools Inspectorate), which recommends that the underlying technology should be visible to increase interest [7].

2.3 Technology: Facial Expression Recognition

Today a lot of technology is taken for granted, generation z has grown up with smart devices and intelligent technology. For example, applying an expression sensitive filter in Snapchat [2] has been around for years. Facial expression recognition was chosen as a core interaction to explore new ways for innovative interaction in games but also in an attempt to quantify what the software sees and in order to spark interest in data science or technology in general.

3 METHODOLOGY

Throughout the project, various methods were used, either common within interaction design, or game design. In this section, the most prominent ones are listed.

3.1 Context Immersion

Before initialising the ideation process we visited Universeum to understand the context. The science center Universeum have multiple exhibitions with different themes such as the rainforest, the ocean, space, health, a tech hub and a chemistry lab. The different exhibitions were investigated to understand how they spark engagement and interest among the visitors in technology and science. One important takeaway from the visit was to keep a low interaction threshold in the exhibition feature to make it interesting for the audience, not only the person/persons currently interacting.

3.2 Ideation

The ideation process began with a brainstorming [3] session, listing ideas within the stated context of the project. The ideas were then sorted into themes; facial expressions and emotions, measurements and estimates, productivity and reactions, collaboration. These themes served as the basis for the next brainstorming session, where we further explored diverse themes. Conducting dot voting among the themes resulted in a democratic decision to focus on facial expressions and emotions but the other themes would also play a part in the final result. A third session was then conducted to explore the possibilities, requirements, and constraints of facial expressions in the context of this project. Three ideas within the theme were presented to peers and supervisors and a final concept was mutually decided based on the collected feedback.

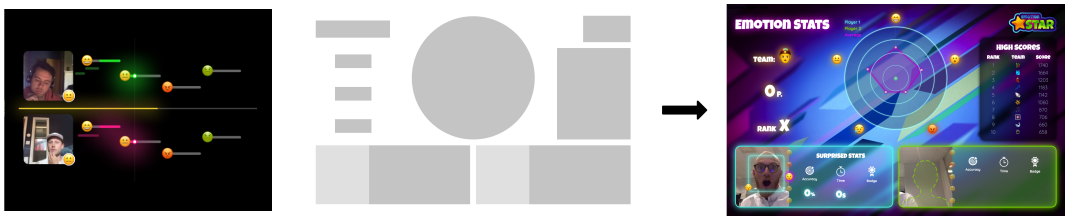
[2]Snapchat is a multimedia messaging app letting people communicate with pictures

3.3 Concept Refinement

The final idea came to be a game so the refinement of the concept was influenced by game development methodology. The concept had similarities with existing musical games such as Singstar and Guitar Hero [3] by making the players perform certain expressions at certain moments as the "notes" slides towards you (See fig. 1a). To understand these games, a rough MDA [2] was conducted for each of the games, followed by constructing an MDA for the project concept. The main aesthetics we decided to use is; Social, fun, frustration, embarrassment, and unfairness. When the basics of the concept were defined, the MoSCoW-method [4] was used to define the priorities between different aspects. During brainstorming and iterations, there were a lot of possibilities and ideas that had to be prioritised to fit within the time-frame of the project. The notes in our concept are visualised by emojis that represents each expression as it fits the intended aesthetics of our game.

3.4 Sketching and Wireframing

To allow sketching and designing simultaneously, Figma, a collaborative interface design tool, was decided to be the main tool for solution sketches [1] and wireframes. By this, all project members had access and could iterate on each other's ideas before deciding which option to choose. Certain components, for example, the logo, were constructed using Adobe illustrator (a vector graphics design software) to allow for more design possibilities. This project used a 12x6 grid for layout, 12 columns is an industry-standard and 6 rows allow dividing it both by two and three. Iterations of grey box wireframes were used to explore different arrangements and alignments to the grid. Fig. 1b shows an example from the process of developing the final screen.



(a) Early concept sketch

(b) Greybox iteration compared to prototype

Fig. 1

3.5 Implementation

The front end implementation was done using *ReactJS* [4]. Furthermore, the *ReactJS*-framework *NextJS* [5] was used to provide convenience and speed through pre-configuration and pre-rendering pages. Additionally, *TailwindCSS* [6] is used in the project for layout. For the back end implementation, two different frameworks were used. For tracking high scores, *Firebase* [7] was used to provide a database implementation. During the gameplay, statistics about the game are stored locally. When the game is finished statistics from the game are sent to *Firebase*. Another framework, *face-api.js* acts as the core of the game which is the facial detection and tracking of expressions. The face detection API sends

[3] Singstar and Guitar Hero are music rhythm games where the player tries to reach correct pitch or hit the right note on time to get points
[4] ReactJS: JavaScript library for building user interfaces. https://reactjs.org/
[5] NextJS: Front end React development framework. https://nextjs.org/
[6] TailwindCSS: utility based CSS tool with pre-defined classes. https://tailwindcss.com/
[7] Firebase: Web application platform service

a constant stream of data regarding the probability of the current faces and expressions it detects. The expressions available are surprised, disgusted, neutral, fearful, sad, angry, and happy [5]. To use this data the list was sorted and the expression with the highest probability was extracted.

3.6 Evaluation

At an early stage of the process, a rough mock-up of the facial recognition application was prepared to evaluate users' ability to perform the different expressions. The application showed the webcam of the user and their current expression. After exploring the application the test subjects answered a survey regarding which expressions was considered the easiest or hardest to achieve. We had a hypothesis that two of the seven possible expressions were much harder to achieve. After conducting ten tests, six on site and four via an online form, we felt comfortable in eliminating the expressions; disgust and fearful, since they were considered too hard to execute by the test subjects. After further application development, a second user test was conducted. In the second evaluation, all parts of the game were evaluated through a test session followed by another survey. Five test sessions with a total of eight participants, three pairs and two single users, were conducted in a controlled environment using group rooms at the University. Some parts of the game were developed enough to be used on their own while other parts had to be simulated using the "Wizard of Oz"-technique [6]. In this test, both the look and feel as well as the interaction of the game were evaluated. The evaluation survey was divided according to the different parts of the interaction. Using pictures from each part to remind the participants of how it looked, minimising the risk of the test subjects mixing them up afterwards. The survey investigated both the overall impression as well as the specific tasks they had to perform. The results from the evaluation were then used to refine the concept through further iterations.

4 RESULTS

The project resulted in an interactive game prototype solely controlled by the users' facial expressions. The project was presented at a combined online and physical exhibition with a physical prototype being available for visitors to the Visual Arena in Gothenburg [8] on October 23rd, 2020.

4.1 Game Concept

The final game concept is a two player competitive and collaborative facial expression game.

4.1.1 Expanding Collaboration. With this game we wanted to explore new ways of interaction. With the context of a digital science center we wanted to rethink and expand the possibilities of collaboration. Our prototype is made to visualise our core concept at a physical exhibition but it does not showcase the intention of letting people play with someone somewhere else. It is intended to be a link between the digital science center and the physical by letting visitors of the physical science center play with people visiting the digital one.

4.1.2 Layers of Competitiveness. This game utilises a combination between collaborative effort and individual effort. During the game the players collectively collect points to a shared score. If the final score is high enough it will enter the leader board. The individual effort is also measured but presented in a different way. Accuracy, achievements and other comparable stats are presented to the players when the game is over. The players can in that way compare their

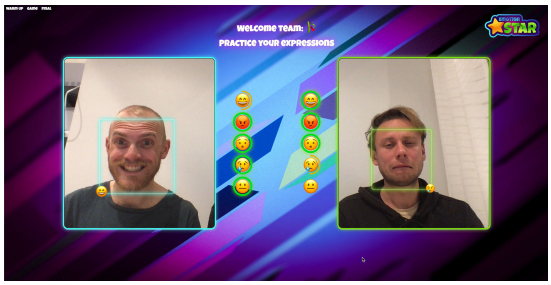
[8]The Visual Arena Lindholmen is an open collaboration platform. With a live streaming studio and multi-functional lobby it makes a perfect location for visual and virtual innovation.

contribution to the shared score. This results in an interesting gameplay that urges the players to create new team constellations in order to reach the leader board

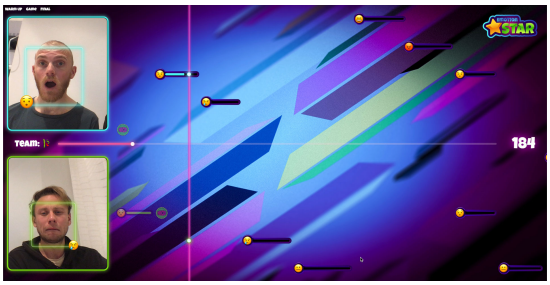
4.2 Final Prototype

The final prototype consists of three parts; a warm-up exercise, the actual game, and exploratory and dynamic statistics.

4.2.1 Warm-up exercise. The warm-up exercise reveals when two persons are detected by the webcam in the play area. The players can practice their expressions and gain an understanding of how the application works (see figure 2a). Each player needs to achieve and hold each expression for one second. The time criteria was a deliberate design choice grounded in the fact that the face recognition technology can, in certain indecisive situations, quickly swap between multiple expressions. Once all facial expressions for both players are successfully achieved, the application automatically proceeds to the game. The amount of time required for each expression affects the difficulty of completing the warm-up, as this is only a warm-up and not the actual game different times were evaluated to balance the required performance.



(a) Screenshot of the warm-up screen



(b) Screenshot of the game

Fig. 2

4.2.2 The game. The game screen is horizontally split into two play areas, one for each player. A screenshot can be seen in figure 2b. The play area consists of a set of notes, moving from right to left, with an attached expression represented by an emoji. Thus, the goal for the players is to perform and hold the expression that the note represents for the length of the note. Completing this generates points that are added to the team total. A song is playing while the game is running. Using a low pass filter, the sound gets muffled if the players' performance completing the notes is too low. If a player misses a note, the sound gets more muffled, and only recovers when a player starts hitting notes again. This audible feedback is designed to indicate how well the players are performing. When the game is finished, a transition is displayed, showing a screen of animated emojis along with the team score that the players achieved together, before transitioning to the statistics screen.

4.2.3 Statistics screen. During the game, the following statistics are captured:

- The time spent making each expression.
- The percentage of notes completed.
- The team score, which is calculated based on performance and completion of notes

(a) Screenshot of the final screen

(b) IDXPO 2020

Fig. 3

These statistics are displayed in the final statistics screen (see figure 3a). The team's score is displayed in plain text and compared to other teams using a ranking and a high score list. A radar chart displays a general overview of how accurate each player was at certain expressions allows comparisons between each other and the global average. To get detailed statistics (individually), each player needs to make the expression of the desired statistics. The interaction includes a time threshold to allow for better control. The following statistics are displayed: accuracy of hitting the notes, time spent making that expression, and possibly a badge. The badge assigned is based on the player's accuracy, and can be one of gold level (90% or higher), silver level (70% to 90%), or bronze level (50% to 70%). The badges are included as a reward to the players, to make the statistics more visually interesting.

4.3 Exhibition

The prototype was available for the audience at IDXPO 2020 at Visual Arena Lindholmen (see fig. 3b on October 23rd, 2020 [9]. At the exhibition, the prototype was tested almost exclusively by students between the ages of 20 to 30 years old. The main result from the participants at the exhibition was that a majority liked both the general idea of the game as well as the looks of it. The game was considered to be both fun and interesting although they experienced some setbacks. Similar to our previous tests some users' struggled to accomplish all emotions. It was also found that a lot of participants were surprised when they were presented with a combined team score rather than individual scores.

5 DISCUSSION

This chapter will discuss considerations, decisions, pros and cons of the process and result.

5.1 Adjustments

During the process adjustments were made to the game.

- Warm-up Timer: The timer for each expression on the warm-up screen was lowered from two seconds to one second due to some people during the exhibition struggled to make it into the actual game
- Excluded expressions: Based on the evaluation results, the disgusted and fearful expressions were removed from the game, as they were too difficult to achieve for many participants.

[9]The prototype is accesible online through https://emotion-star.vercel.app/ but works best in full screen using Chrome from a Macbook

5.2 Pros and Cons

This section discusses the results and offers insights about the pros and cons. In particular, the game idea itself and the ethical aspects of personal data are strong points for the project.

5.2.1 Pros.

- *Game Idea:* Using facial recognition and facial expressions to control a game is something we find unique and which our users found to be both *"cool"* and *"innovative"*. This does not only apply to the game itself, but it is also the only way of navigating within the user interface.
- *The balance of privacy and personalization:* Ethical concerns of gathering personal data have been taken into account throughout the entire project. One goal of the original project description was to think "outside the box" regarding quantifiable data and strive to present it in a fun and challenging way without sharing sensitive information. The teams are assigned a random emoji as "team name" instead of providing their name(s), this is partly an amusing feature but also a way to reduce the amount of personal data to zero. The camera will record your face in real-time, but this data is only available on the local system in real-time, during the game session, and is not stored. The only information being saved is the performance statistics and high score linked to the randomly assigned emoji.

5.2.2 Cons.

- *Interaction Limitations:* Only allowing interaction through facial expressions comes with some difficulties. It limits the number of possible actions since the prototype only recognised a set amount of expressions. Each user can only perform one expression at a time and also have to concentrate to keep the expression which complicates performing certain actions. Furthermore, as noted during the user testing, certain facial expressions are more difficult than others for the model to interpret as intended. Some emotions also require facial parts to move in similar ways, which occasionally caused problems for participants to succeed in the game-play since expressions were too alike. Which expression participants struggled with varied from person to person but most had at least one which were conveyed as harder than the others limiting their interaction and performance in the game. While this can be a great advantage as it affords a very enticing simplicity to the game, it can also discourage many users, particularly those with accessibility concerns.
- *Lack of On-boarding:* The aim was to create a collaborative game but attendees at both playtesting and the physical exhibition showed a lack of understanding that the game was a collaborative effort when they played for the first time. A decision was made to limit the screens and interactions, in order to fit the context and lower the threshold of initiating the game. Initially the ”warm-up screen” was a placeholder loading screen between the start screen and the actual game. The warm-up screen was made more complex which rendered the start screen unneeded. This resulted in some features being harder to understand.
- *Imperfect Model:* The face-detection API only returns a *probability* that a certain face is achieved. This probability depends on the training data that the model is trained on, meaning that it is only as good as its training. If the model is not sufficiently trained it can ruin the gameplay and in the worst case even be harmful. For example, if the model only is trained on images of white males, it can accidentally become racist and/or sexist. As an existing API was used during the project, additional training could not be performed which led to struggles for some participants to do certain expressions, regardless of gender and skin tone.

5.3 Future Improvements

Due to time restrictions and other factors, some features did not make it to the final prototype. In the iterative process, new ideas of possible features were born. These ideas of features are possible implementations in future iterations of the game.

5.3.1 Performance and Scalability. To make the game more responsive and smooth, it would have to be re-written with a game library that excels at handling game logic and game loops for performance reasons. The game runs quite slow on older computers which ruins the experience and also affects the scalability of the game. A reason for this is the fact that the game was built with a library that is designed to construct and render user interfaces, not full games.

5.3.2 Tracking Position of the Users. Using facial expressions as the only controls within the application limits the number of interactions. Enabling tracking of the users' position within the screen would enable more ways of controlling the interaction. Positioning yourself at a certain side or corner of the screen could be used as a way of performing actions, combining this with expressions would enable even more interaction possibilities

6 CONCLUSION

Based on the testing and evaluation of the project, it can be concluded that; the objective of constructing a fun interactive game has been accomplished. The goal of engaging teenagers in technology is reached by attempting to reveal the underlying technology and the underlying statistics behind the final score. The project also leads to raised interest in technology in general as well as highlighting opportunities within facial recognition and programming in particular. Our project explores possibilities to use facial expressions to control a UI. Contrary to the design convention of WIMP interfaces, the project's UI contains neither a pointer or menus. Instead, it makes use of facial expression recognition combined with timers to navigate the interface. To a very small extent, it can be concluded that this way of navigation did not cause major problems in navigating the interface. Yet it can be identified that the navigation through facial expression has to be tied to timers, to prevent users from accidentally triggering a certain navigational option. Some suggestions on how to improve and allow for more diversity is stated in the discussion.

ACKNOWLEDGMENTS

Thanks to our supervisor Mafalda Gamboa for the great support and for inspiring us step outside our comfort zones. Thanks to our examiner Josef Wideström for providing the context and planning the IDXPO 2020. Also a big thanks to Vincent Mühler for providing the open-source face-api.js face recognition model.

REFERENCES

[1] Google. 2020. Solution sketch by Google Design Sprint Kit. https://designsprintkit.withgoogle.com/methodology/phase3-sketch/solution-sketch
[2] Robin Hunicke, Marc LeBlanc, and Robert Zubek. 2004. MDA: A Formal Approach to Game Design and Game Research. https://users.cs.northwestern.edu/~hunicke/pubs/MDA.pdf
[3] Design Kit. 2020. Brainstorm by DesignKit.org. https://www.designkit.org/methods/brainstorm
[4] Eduardo Miranda. 2011. Time boxing planning: buffered moscow rules. *ACM SIGSOFT Software Engineering Notes* 36 (11 2011), 1–5. https://doi.org/10.1145/2047414.2047428
[5] Vincent Mühler. 2020. JavaScript API for face detection and face recognition in the browser and nodejs with tensorflow.js. https://github.com/justadudewhohacks/face-api.js/
[6] Digital Society School. 2020. Wizard of Oz. https://toolkits.dss.cloud/design/method-card/wizard-of-oz-2/
[7] Skolinspektionen. 2014. Teknik - gör det osynliga synligt. *Kvalitetsgranskning* (2014).
[8] Teknikföretagen. 2017. Att få ännu fler att välja teknik. *En guide till ungas intresse för teknikutbildningar* (2017).