

TDK-thesis

Gyöngyvér Ferencz

Imre Molnár

FaceGym: Toward overcoming special needs related to facial expressions

EÖTVÖS LORÁND UNIVERSITY

FACULTY OF INFORMATICS

DEPT. OF ARTIFICIAL INTELLIGENCE



Authors:

Gyöngyvér Ferencz

Computer Science MSc, Artificial
Intelligence Specialisation

Imre Molnár

Computer Science MSc, Artificial
Intelligence Specialisation

Supervisors:

Dr habil Lőrincz András

senior research fellow

Budapest, 2024

Contents

Abstract	3
1 Introduction	4
2 Related Works	6
3 Methods	11
3.1 AI Technologies	11
3.1.1 OpenGraphAU	11
3.1.2 Py-Feat (ResMaskNet)	11
3.1.3 Exordium (EmoNet)	12
3.1.4 RetinaFace	12
3.1.5 PyTorch	12
3.2 Technologies	13
3.2.1 React	13
3.2.2 Flask	13
3.3 Facial Action Coding System - FACS	14
3.4 Personality	14
3.4.1 Big Five Inventory 2 (BFI-2)	17
3.4.2 HEXACO	18
3.4.3 Energy Level	19
3.5 Specification and Design	19
3.5.1 Exercises	19
3.5.2 Architecture	30
4 Experiment	34
4.1 Setup	34
4.2 Participants	35
4.3 Protocol of the second experiment	36

4.3.1	Questionary part	36
4.3.2	Game part	37
4.4	Data Pre-Processing	38
4.4.1	Measurements and Post-Processing	39
4.4.2	Evaluation the AU maximization game	40
4.4.3	Evaluation the Drawing game	40
5	Results	41
5.1	Experiment	41
5.1.1	The distribution of personality traits	41
5.1.2	Energy level	44
5.1.3	Drawing with emotions task	45
5.1.4	Action Unit maximization task	45
5.1.5	Imitation task	46
5.2	Evaluating and interpreting the results	46
5.2.1	Analysis of improvement as a result of AU practice	46
5.2.2	Who struggled with anger and sadness?	47
5.2.3	Correlation between facial emotions and personality	48
5.2.4	Personality and emotions considering the impact of learning	50
5.2.5	AU Maximization task and Personality	51
6	Discussion	72
7	Conclusion and Future Work	74
	Acknowledgements	75
	Bibliography	76
	List of Figures	81
	List of Tables	87

Abstract

'FaceGym' is an innovative development tool incorporating Artificial Emotional Intelligence technologies to teach individuals to imitate, recognize, and generate human emotional and social expressions. In this game we integrated state-of-the-art facial emotion and Action Unit detectors, which are the building blocks of the Facial Acting Code system, to construct a unique self-training platform. The module has promises for a wide range of applications, for example, it can help in training and the quantitative assessment of the facial muscles in the case of physical disorders such as facial paralysis or ALS, as well as can help to overcome impairments originated by developmental disorders, like autism. We worked with neurotypical adolescent subjects. For this case, we could investigate two of the three options listed above, i.e., if FaceGym improves their abilities to generate and imitate facial emotions, especially when we include a practice session between exercises. We have demonstrated that the models utilized in our application can provide real-time feedback for measuring emotional and action unit values. Our experiment indicates – in agreement with the literature – that adolescent girls had the most difficulty in expressing anger out of the six basic emotional expressions. Given our studies, we see ways to improve the experimental design. We present our observations and findings using statistical and machine learning methods such as correlation tests and PCA to connect the user-produced results with their personality traits. These preliminary studies justify the increase of the population to raise the level of significance further.

Chapter 1

Introduction

Human communication is a multifaceted phenomenon, encompassing verbal and non-verbal cues that play key roles in conveying thoughts, emotions, and intentions. In recent years, research endeavors have delved deeper into understanding the intricacies of non-verbal communication, particularly focusing on facial expressions and gestures, both in human-human collaboration and human-machine interaction. This study aims to elucidate these domains, shedding light on the mechanisms underlying human expression and its implications for fields such as psychology or human-computer interaction.

Facial expression, a powerful tool for non-verbal communication, is a cornerstone of human-human and human-computer interactions. It enables us to convey a myriad of emotions without uttering a single word. Notably, fundamental facial expressions such as sadness, anger, surprise, fear, happiness, contempt, and disgust are universally understood, transcending cultural barriers. However, some people may experience difficulties in this area, significantly affecting their ability to interact with others. People with autism spectrum disorder (ASD) may have challenges in recognizing and displaying facial expressions due to differences in social communication and interaction skills. Individuals who have experienced a stroke or facial paralysis may have difficulty controlling facial muscles that are limited or altered to express facial expressions. With the expansion of AI technologies, it is possible to develop new methodologies and technologies to understand the expression of emotions better and help people, particularly children, in self-training to imitate, recognize, and generate human emotional and social expressions or basic acting skills.

Firstly, we introduce the available methods to measure nonverbal communication, focusing only on the face by detecting facial expressions and facial action unit actuation

levels based on the Facial Action Coding System (FACS). By leveraging cutting-edge AI tools and methodologies, the study endeavors to develop a specialized tool capable of facilitating and training the acquisition and assessment of intricate facial expressions. The research aims to unravel the nuances of expression acquisition and its potential applications across diverse domains through meticulous analysis and evaluation of learning trajectories.

Secondly, it delves into the realm of human-machine collaboration, explicitly focusing on the correlation between human personality traits and proficiency in expressing oneself nonverbally. The study uses established personality frameworks and innovative experimentation methodologies to discern patterns and correlations between individual personality traits and nonverbal communication capabilities. By integrating personality tests with nonverbal expression imitation tasks, the research explores the underlying mechanisms governing human expression and its implications for human-robot interaction.

The goal of the project is to develop an application which can measure in real-time the facial expressions and examine the following claims:

1. The different state-of-the-art AI modules that can detect facial expressions (like Action Units or emotions) can send feedback to the user in real time. Can we create a real-time application, a self-training tool to measure, recognize, and generate the user's facial expressions?
2. Is there a correlation between personality types and emotional expressions?
3. Adolescent girls have problems expressing Anger. Can the developed tool confirm this statement and develop it as a self-educational tool?

Chapter 2

Related Works

There are myriad ways to express our emotions and sentiments, encompassing spoken and unspoken communication forms. Non-verbal communication, such as body language, facial expressions, and gestures, are significant in conveying information. Our faces are a powerful medium for transmitting details about gender, age, ethnicity, or social status. It's a common practice to recognize individuals primarily by their facial features. We tend to associate facial expressions (FE) with aspects of a person's character, as well as their immediate emotional state and underlying intentions. FEs are essential to emotional signals and allow people to understand social motivation [1]. Facial expressions are universal and culture-specific, and many apparent differences across cultures are due to context [2]. Paul Ekman describes this phenomenon with the display rules, which basically means that we learn while growing up how, when, and to whom it is convenient to show our emotions. Ekman defines seven universal basic facial emotions: anger, contempt, disgust, enjoyment, fear, sadness, and surprise. Emotions are not singular, isolated affective or psychological states; they encompass a spectrum of related emotional states that share a common theme [3]. Enjoyment is the most easily detectable in the face, whereas fear and disgust are the least identifiable and often mistaken for surprise and anger, respectively [4].

Emotional cues convey valuable insights into the intentions and emotions of individuals [5]. While extensive research has explored children's comprehension of facial expressions (FEs), there has been limited investigation into the production and mimicry/imitation of emotional expressions. We learn in our childhood how to form emotional expressions during social interactions with our relatives or peers, and we acquire emotional competence [6]. However, how we master and produce FEs has not been assessed. It's

important to explore the different factors that could influence the production of FEs in childhood or adolescence, such as age, gender, or emotion subtypes like sadness, anger, joy, and neutral [7]. In the literature, FE production is paid less attention than FE recognition. We knew only three different methods to measure FE imitation [7]. The most commonly used system is the Facial Acting Coding System (FACS) created by Paul Ekman in 1978 (revised in 2022) [2], where they identified more than forty different units, so-called Action Units, which different combinations from the various expressions, emotions. Decoding the FACS precisely requires a professional rater or FACS coder if we don't use any emotional artificial intelligence technology. The second most commonly used method is a judgment approach [8] introduced by Charles Darwin, which is based on the idea that everyone can rate an emotion and doesn't need any specific knowledge. Usually, researchers record videos of different emotions, and annotators have two different tasks: to conclude the emotion and rate its intensity. Using this method, there are well-known corpora like the GEMEP [9] or the CEED [10] dataset. The third solution is when we use Emotion Artificial Intelligence Technology, where FE analysis relies on algorithmic automatic assessments trained on extensive datasets, which offer standardized facial expression materials [11]. Most of the available datasets contain static 2D images and captures of adults.

Artificial intelligence encompasses affecting computing [12] or emotional AI, which measures, understands, simulates, and reacts to human emotions. Machines can analyze images and detect microexpressions on human faces. This feature enables a more seamless and organic interaction between humans and machines, similar to human-human interaction. The technology has undergone such a breakthrough that emotion AI models learned 'the language of emotions' and are precise enough to be integrated into development or training methods. This aptitude will allow humans and machines to interact much more naturally, similar to how human-to-human interaction works. [13]

Emotional Expression Recognition (EER) and Facial Expression Recognition (FER) represent dynamic domains of investigation within the field of affective computing. This field focuses on analyzing human emotions, identifying facial expressions, and conducting sentiment analysis [14]. FER is the detection of human emotional states through facial muscle movements that are triggered by changes in emotional states [15]. While most research on emotion recognition has focused on photographs of expressions intended to depict the apex of the emotional display, these studies have their limitations [16]. They provide important insights into how emotions are perceived in the face, but they neces-

sarily overlook the role of dynamic information, leaving a significant gap for further exploration [16]. Estimating the intensity of action units can provide valuable information, especially for analyzing complex facial behaviors. This approach can help assess the intensity of affective states and distinguish between genuine and posed behavior [17]. AI's ability to analyze facial expressions goes beyond identifying basic emotions. Some AI platforms can predict personality traits and behavioral tendencies based on facial features and expressions. There are a lot of popular Facial Expression Recognition applications to detect facial expressions by using different approaches. We can focus on detecting facial landmarks, or use Ekman's FACS system and detect the different Action Unit intensities, eye-gaze estimation, or deep neural networks.

OpenFace [18] is one of the most commonly used open-source tool for facial behavior analysis. The tool provides facial landmark detection, head pose estimation, facial action unit recognition, and eye-gaze estimation. OpenFace enables seamless integration with other applications and devices using a lightweight messaging system.

OpenGraphAU [19] is another solution to detect Action Units and can be used in real-time applications and uses multi-dimensional edge features and AU relation graphs. Creates a unique graph using deep learning to explicitly describe the relationship between each pair of AUs in the target facial display. The system achieves state-of-the-art AU recognition results and can be incorporated into various backbones.

Facetorch[20] is a Python library that uses neural networks to detect faces and analyze facial features. The project goal is to collect the most suitable open license pre-trained facial analysis models optimize them for TorchScript and combine them into a facial analysis toolbox. The system can be used for face detection, face verification, facial expression recognition, deepfake detection, or 3D face alignment.

Py-Feat (ResMaskNet) [21] is an open-source Python toolbox for analyzing and visualizing facial expressions. The package aims to connect model developers creating algorithms for detecting faces, facial landmarks, action units, and emotions with end users who want to use these models for research purposes.

EmoNet [22] is a deep neural network model that was developed to accurately analyze facial expressions in natural conditions. The network can estimate both categorical and continuous emotions simultaneously while conducting face alignment in real-time, this property makes it suitable for use in our development as one of the promising emotion detection models.

There is a very promising new approach which was published early this year. **Mexca**

[23] - Multimodal Emotion Expression Capture Amsterdam - is an open-source Python package that aims to capture human emotion expressions from videos in a single pipeline. The software combines existing unimodal systems to extract facial muscle movements, vocalizations, and speech content. Additionally, it can track faces and speakers in videos. It has the ability to automatically capture emotional expressions on a large scale, across various individuals and contexts.

An exciting area to explore when examining facial emotions is whether there are any differences between the genders in the correct expression of emotions. Gender differences are mostly analyzed only from facial recognition cues, and females are typically better at facial emotion processing (FEP) [24] than males, which can be associated somehow with the female 'brain type' theory [25]. We hypothesized that girls would outperform boys based on the findings of studies investigating emotion recognition from facial expressions [24]. We can define small but significant gender differences in emotional expressions between genders; girls show greater positive emotional expressions than boys, and this gender difference became apparent during adolescence [26]. In addition, girls tend to express more internalizing emotions such as sadness, fear, sympathy, and shame than boys, particularly in negative situations [26].

During the literature research, we found very few articles by setting up a similar experiment that we designed and focused on child or adolescent emotion expression recognition and production. In experiments focused on emotion imitation, researchers typically seek to gauge a participant's emotional response by measuring the movement of their facial muscles. This is usually done by using traditional medical devices like EMG [27]. However, it's important to note that these experiments are meant to measure reactions rather than the ability to express emotions, and they do not necessarily focus on the educational aspect of the process. These experiments measure the reaction and not the capabilities of the emotion expression and don't focus on the educational part. A similar study [7] aimed to investigate various factors like age, gender, and specific emotion types, such as sadness, anger, joy, and neutrality, that may impact the production of facial expressions in childhood. They designed two different tasks, an imitation, and a request type. In the imitation game, they created a 3D avatar and a concise video showing an emotion (facial and vocal). The task was to imitate both to map the participant's audiovisual and visual modality. This task is similar to what we designed in our acting game. The result showed that as children grew older, their performance in producing facial expressions improved. Positive emotions were found to be more easily produced than negative emotions (At

age 10, girls were better than boys at imitating joy, but there were no significant differences in imitating neutral faces; boys were slightly better at imitating sadness, and boys were significantly better at imitating anger). Children demonstrated better facial expression production when prompted (as opposed to imitating expressions). Another similar study [28] where designed an emotion recognition and imitation task for ASC and NT individuals. In our case, the experiment protocol is interesting. They designed an imitation task, where they played an actor's facial expression to the participants, and the task was to move the muscle similarly; at the end, the participants had to select from two possible options what the short video displayed. This exercise is similar to our Imitation task. They analyzed the Action Unit values by using OpenFace. They designed an imitation imprecision score IIS, which indicates the absolute deviation of the participant's facial expressions from the actors and cosine similarity to compare the dominant Action Units to each other. They concluded that individuals with autism were, on average, capable of imitating facial expressions when instructed. Computer-based face analysis can measure and quantify qualitative differences in imitation.

Establishing the connections between personality and facial expressions of emotion helps us understand how traits and disorders are perceived and responded to by others, and more broadly, how personality and psychopathology contribute to shaping the social environment [29]. We can define a relation between facial expressions and the Big Five OCEAN personality traits [30]. Extraversion is positively correlated with increased 'Duchenne smiles' of enjoyment and amusement which refers to the basic emotion of happiness, because extraverted people are more enthusiastic. Extraversion is also correlated with increased facial expressions of sadness [29, 31]. Neuroticism is correlated with increased facial expressions of anger, contempt, and fear [29], we can predict negative emotions with this personality trait. Agreeableness is negatively correlated with anger and disgust. The Agreeableness trait is also negatively correlated with anger [29, 31]. Conscientiousness, is correlated with reduced facial expressions of negative emotion, and increased 'Duchenne laughter' [31]. Openness can show a negative correlations for anger and fear [31].

Chapter 3

Methods

3.1 AI Technologies

3.1.1 OpenGraphAU

OpenGraphAU is a state-of-the-art Facial Action Unit (AU) detection model. This novel approach leverages multi-dimensional edge features and AU relation graphs. Traditional methods for AU recognition often face challenges in capturing subtle facial expressions accurately. The developed method combines edge-based features extracted from multiple dimensions of facial images and constructs a graph to model the relationships between AUs. This aims to improve the accuracy and robustness of AU recognition by considering both local and global facial information. The method excels in capturing relationship cues essential for recognizing AUs and seamlessly integrates with different backbone architectures. The OpenGraphAU [19] yields the probability of activation for all 41 action units.

3.1.2 Py-Feat (ResMaskNet)

Py-Feat [21] is an open-source Python toolbox for analyzing facial expressions. This package provides a variety of tools such as facial landmarks, action units, emotion, and head pose detection for both images and videos. It has built-in preprocessing and analysis capabilities. For our purposes, we were only interested in the emotion expression detection feature. The model can output the probability distribution of the seven emotional expressions: anger, disgust, fear, happiness, sadness, surprise, and neutral.

We used this package only in the first pilot version of the software because it was easy to integrate into the project, and no extra steps were required. Based on the documentation, it only provides a function that returns the face bounding box, the emotion expression probabilities, and the action unit values. The action unit value estimation should be excluded to speed up the process, so only the relevant parts had to be included in the source code.

3.1.3 Exordium (EmoNet)

Exordium [32] is an open-source Python toolbox for emotional expression analysis. The package includes EmoNet [33] which is a categorical and dimensional emotion estimation model, that can output the probability of eight emotional expressions: neutral, happiness, sadness, surprise, fear, disgust, anger, and contempt. It can also output the valence and the arousal level (how calm or agitated you feel) of the given facial expression, which is another way of describing an expression.

For our purposes, only the seven universal emotional expressions (neutral, happiness, sadness, surprise, fear, disgust, and anger) were required, so we made a small change to the output of the model. The extra logits have been cut off from the output, and softmax has been used to calculate the new values.

3.1.4 RetinaFace

RetinaFace is a deep learning cutting-edge face detector for Python that can discover facial landmarks [34]. RetinaFace utilizes pixel-wise face localization across different scales, leveraging joint extra-supervised and self-supervised multi-task learning techniques. RetinaFace exhibits robust detection capabilities, even within dense crowds. Its notable advantage lies in its ability to achieve real-time performance on a single CPU core for VGA-resolution images. This tool is also included in Exordium.

3.1.5 PyTorch

PyTorch [35] is an open-source machine learning library developed by Facebook's AI Research lab. It provides a flexible and intuitive framework for building and training neural networks and is currently the most popular deep-learning framework for scientific research and industrial applications. It supports GPU acceleration and enables fast com-

putation and training of large models. It also offers a rich ecosystem of tools and libraries for computer vision and natural language processing, making it a versatile tool for various AI projects. Its simple syntax and extensive community support have also contributed to its popularity among researchers and developers.

In some cases, minor modifications were required to the models. In the case of OpenGraphAU, the provided code only implemented inference for a single image, which is not a problem when we are sending one image at a time for feedback for users, however, when it comes to running a batched inference on the saved images it had to be addressed.

3.2 Technologies

3.2.1 React

React [36] is one of the most popular JavaScript libraries for developing interactive user interfaces, particularly suited for single-page applications where smooth, responsive user experiences are essential. Developers can create web applications that update dynamically with changing data without reloading the page. React focuses on creating reusable UI components, allowing encapsulated components to manage their state and assemble them to create complex interfaces. It's part of a larger ecosystem, with tools and extensions like React Native for mobile app development and Redux for state management, enhancing its capabilities and application.

The main reason for selecting React was that it made it easier to create reusable parts by keeping a functional approach for the solution. This reduces the need to write boilerplate code and increases the productivity. The other reason was that if the software had to be ported to mobile devices then React Native would make it easier.

3.2.2 Flask

Flask [37] is a lightweight, easy-to-use, and highly extensible web framework for Python. It is designed to simplify web application development while offering great flexibility for more complex applications. Flask supports extensions that can add application features, such as form validation, file uploading, and authentication technologies. It is ideal for small- to medium-scale projects, as it gives developers the freedom to choose their tools and the simplicity of writing their applications efficiently.

Flask was chosen because it's a lightweight Python framework that can be used to create web APIs. Since the core of the application is the AI models that were implemented using Python it made it easier to integrate them into our project. For further developments, it is also suitable because it is ideal for projects on a bigger scale.

3.3 Facial Action Coding System - FACS

The Facial Action Coding System (FACS) is an anatomically based extensive system for describing facial movements. The system breaks down facial expressions into individual components of muscle movement called Action Units (AUs). Combining the different AUs allows us to express different facial emotions. The basics of the system were created by Carl-Herman Hjortsjö in 1970 when he taxonomized human facial movements [38] and defined 23 facial motion units. Later, two American psychologists, Paul Ekman, and Wallace Friesen, formalized the descriptions and developed the FACS manual, which describes the criteria for observing and coding each Action Unit (AU) [2]. They revised the manual in 2002 [3]. Each AU represents a specific movement or combination of movements of one or more facial muscles. For example, AU12 (*Zygomaticus Major*) represents raising the corners of the lips in a smile. Specific combinations of these facial muscle actions correspond to expressed emotions. For example, happiness (smile) is the combination of AU6 (cheek raiser) and AU12 (lip corner puller) (**Figure 3.1**).

FACS is widely used in various fields, including psychology, neuroscience, medicine, and computer science. Research uses it to study emotion, nonverbal communication, social interaction, and psychological disorders. Using the FACS system makes it possible to determine the participants' emotions. Traditionally, researchers have been constrained to manually decode video footage of subjects based on the Facial Action Coding System (FACS) action units. However, advancements in AI now enable us to analyze facial expressions automatically. This innovation drastically reduces time and cost burdens, as scoring no longer demands meticulous frame-by-frame scrutiny by a skilled analyst; instead, specialized software seamlessly handles the task [40].

3.4 Personality

Personality is commonly described as the expression of unique variations in thought, emotion, and action, which tend to endure over time. In the realm of personality com-



Figure 3.1: Happiness - Smile: Combination of AU6 and AU12 illustrated by one of our participants. She took part in the pilot testing and provided a written agreement to use her images captured by FaceGym for research purposes [39]

puting, it is typically identified by the fundamental Big Five traits (**Figure 3.2**), commonly abbreviated as OCEAN, derived from self-reported evaluations [41]. The acronym OCEAN stands for:

1. Openness
2. Extraversion
3. Conscientiousness
4. Agreeableness
5. Neuroticism

Each trait represents a broad dimension of personality, and individuals can fall at various points along each continuum. Unlike other trait theories that sort individuals into binary categories, the Big Five Model asserts that each primary personality trait is a spectrum, representing a range between two extremes. Extraversion is fundamentally a continuum between extreme extraversion and extreme introversion, but in the real world, most people fall somewhere in between. Despite a significant body of literature, researchers

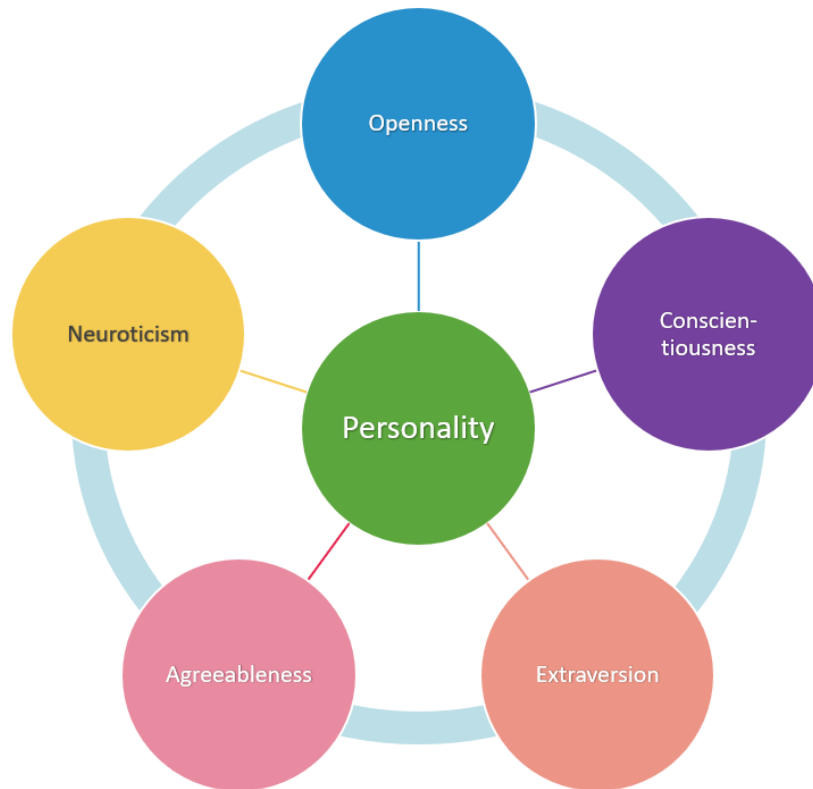


Figure 3.2: The Big Five Personality Traits. Reproduced from [Wikipedia](#) under Creative Commons Attribution-Share Alike 4.0 International license.

may not always reach a consensus on the precise labeling of each dimension, but these five traits are generally outlined as follows.

Openness reflects the degree to which a person is open-minded, imaginative, curious, and creative. Individuals high in openness tend to be adventurous, willing to try new things, enjoy exploring new ideas and experiences, and are happy to think about abstract concepts.

Conscientiousness refers to the degree of organization, responsibility, and reliability a person exhibits. Individuals high in conscientiousness are typically dependable, paying attention to the details, organized, and goal-oriented. They tend to be careful and diligent in their actions. Someone with lower scores in this personality trait can be less structured and organized, and sometimes they miss deadlines completely.

Extraversion represents the extent to which a person is outgoing, sociable, and energetic. Extraverts tend to seek out social interactions, enjoy being around others, and often display assertiveness and enthusiasm. People with low-level scores tend to be more reserved. They dislike making small talk and being in the center, making it difficult to start

conversations, but they can feel their emotions more deeply.

Agreeableness reflects the degree of kindness, cooperativeness, and compassion. Individuals with high scores tend to be friendly, empathetic, and considerate of others' feelings. They value harmony and cooperation in social interactions. People with low scores tend to be less competitive and sometimes manipulate.

Neuroticism refers to negative emotions such as anxiety, depression, and moodiness. Individuals high in neuroticism are more prone to experiencing stress and emotional instability. They may be more sensitive to perceived threats and are likely to react strongly to stressful situations. The evaluations often reverse the negative emotion scores.

3.4.1 Big Five Inventory 2 (BFI-2)

BFI - Big Five Inventory - is one of the most commonly used personality tests to measure the Big Five personality traits due to its brevity and effectiveness in capturing key aspects of personality. The original test was developed by Oliver P. John, Laura Naumann, and Christopher J. Soto in 1999 [30] and contains 44 different questions, characteristics which can measure personality traits based on the Five-Factor Model (FFM) of personality. Its popularity in research and practical applications is mainly due to the simple language, brevity, and free accessibility of the questionnaire [42]. The questionnaire is formed from simple, few-word statements and the responder has to assess on a scale how well they describe themselves: e.g. talkative, depressed, tense, lazy or emotionally stable. The rating is based on a five-point Likert scale, ranging from "Disagree Strongly" (1 point) to "Agree strongly" (5 points). The instrument has been translated into several languages (e.g. Hungarian, Danish, Dutch, German, Norwegian, Italian) and applied to a wide variety of study samples like adolescents, clinical patient samples, PTSD, elderly, physicians or academic students.

The creators of the test redesigned their measurement tool in 2007. They published an extended 60 characteristics version of the original test called **BFI-2**, which measures the Big Five personality domains (which we label Extraversion, Agreeableness, Conscientiousness, Negative Emotionality, and Open-Mindedness) and the 15 more specific facets traits [43]:

- Sociability
- Assertiveness
- Energy Level

- Compassion
- Respectfulness
- Trust
- Organization
- Productiveness
- Responsibility
- Depression
- Emotional Volatility
- Intellectual Curiosity
- Aesthetic Sensitivity
- Creative Imagination

The BFI-2 is also translated into several languages (e.g. Hungarian, Danish, Dutch, German, Norwegian, Italian), and its simple language and brevity allow it to be used even at the age of 12 [44]. All these advantages led us to choose and use the BFI-2 test in our experiment to make the participant's personality map. The test contains 60 characteristics that may or may not apply to the respondent. The respondent must rate his answer on a five-point Likert scale, ranging from "Disagree Strongly" (1 point) to "Agree strongly" (5 points).

3.4.2 HEXACO

HEXACO is an alternative personality test to the Big Five model and includes six major dimensions. Kibeom Lee and Michael C. Ashton developed it as an expansion of the traditional Big Five model to incorporate the dimension of Honesty-Humility[45]. The HEXACO model is very popular in personality research and assessment due to its comprehensive coverage of personality traits and its ability to capture nuances not fully addressed by the Big Five model. In our experiment, we used only the Hexaco Honesty-Humility dimension, which relates to sincerity, fairness, modesty, and avoidance of greed or manipulation. Extending the original BFI-2 with this dimension can provide a more comprehensive personality assessment. This could allow us to have a more nuanced understanding of an individual's personality profile beyond what is captured by the Big Five alone.

3.4.3 Energy Level

There are many different solutions to measure the energy level of the participants during an experiment. One of the most common tests is the Visual Analogue Scale to Evaluate Fatigue Severity (VAS-F). This is a specific type of visual analog scale designed to assess the severity of fatigue experienced by an individual [46]. It is commonly used in clinical and research settings to measure subjective experiences of fatigue. The scale consists of 18 items, which are simple instructions and can be completed with minimal time and effort. 13 instructions are related to fatigue, and 5 are related to energy level. The test includes a line or a scale from 0 to 10 where the participants have to choose a level of how tired they are or what their energy level is. We asked our participants to complete the test before and after the experiment. The results may show us who was more tired down by the series of tasks or whose energy level was higher.

3.5 Specification and Design

3.5.1 Exercises

Since we can detect action unit values and emotional expression, we decided to design exercises from scratch that would surround these values. We came up with four different types of exercises with multiple sub-tasks:

1. Action Unit maximization task
2. Face Coloring task
3. Drawing with emotions task
4. Imitation game

Action Unit maximisation task

This task mainly focuses on how an individual can control their action units. We call this a "face gymnastic" exercise where we indirectly teach the participants to recognize which facial muscle they should use to form an emotion. We selected some universal emotional expressions and more complex expressions for each sub-task. The FACS system defines which action units are actuated for a given expression [47].

On the user interface (**Figure 3.3-3.4**), we provided the user with a mirrored camera image (so it would look less strange for the user) on the left, and on the right we showed

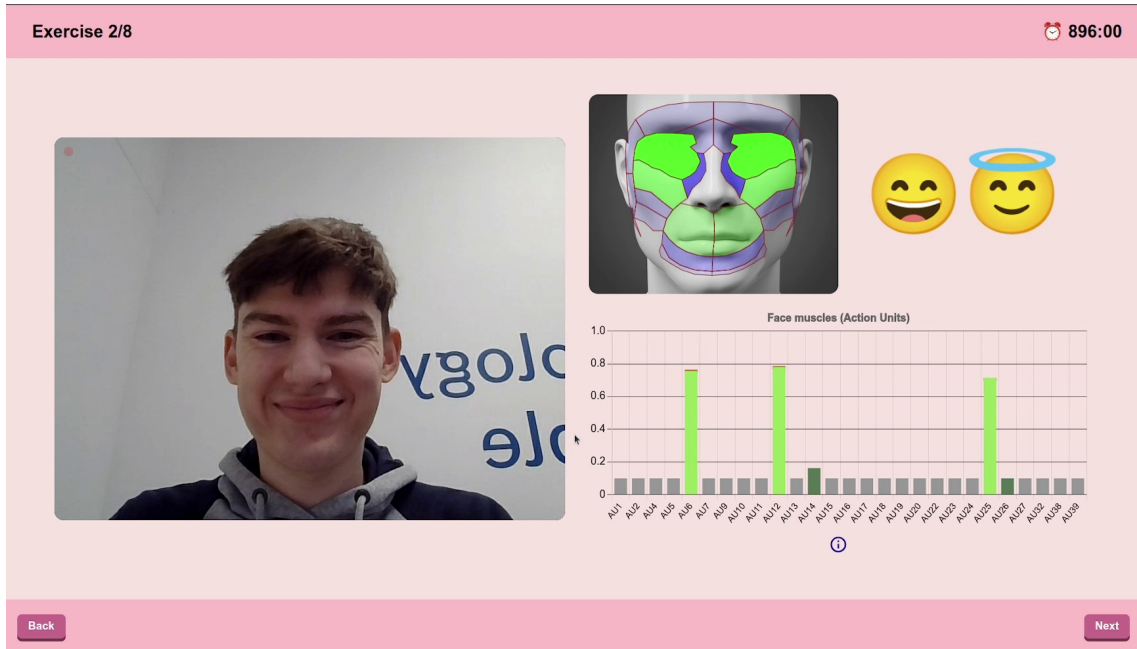


Figure 3.3: The action unit maximization task for happiness. The task here was to move the bars by moving some regions of the face. Each region is corresponding to an Action Unit and the bars show how much a given action unit is actuated.

a bar plot of the different Action Units. We also provided some graphics and a schematic figure of the face to help the user figure out what they needed to do. On the schematic face, the relevant regions for a given emotion were colored green, and the others were colored blue. The regions and coloring method are discussed in 3.5.1. We also gave a small description of the task before starting it (Figure 3.5).

Regarding the bar chart visualization, we split the relevant action units into two groups. The first is the primary group always activated for a given expression. The secondary group contains those action units that might get actuated based on how someone creates the given emotional expression; for example, some people smile with open mouths while others with closed ones. The bars corresponding to the primary group were colored light green while the ones in the secondary group were colored dark green. The bars that do not fall into either of the categories were set to 0.1, and their height did not change. This design choice was made specifically to reduce user distraction.

Overall, we included ten different sub-tasks for this game. Table 3.1 shows the sub-tasks with their corresponding name and defined action unit groups.

Expression	Primary	Secondary
Happiness	AU6, AU12, AU25	AU14, AU26

Anger (Rage)	AU4, AU6, AU9	AU1, AU2, AU5, AU7, AU12, AU25, AU26
Disgust	AU4, AU6, AU9, AU15, AU17	AU7, AU10, AU16, AU25, AU26
Sadness	AU1, AU4, AU15	AU7, AU11, AU17
Surprise	AU1, AU2, AU25, AU26	AU6, AU12
Fear	AU1, AU5, AU25, AU26	AU2, AU7, AU11, AU20
Contempt	AU12, AU14	AU6, AU11
Snarl, like a wolf	AU9, AU10	AU6, AU11
Give a small kiss	AU18, AU23	AU27, AU11, AU22, AU26
Stick out your tongue, like when you're mocking someone.	AU9, AU10, AU26	AU4, AU7

Table 3.1: The emotional expression and their corresponding primary and secondary action units. These primary and secondary action units were defined based on [39].

The exercise aimed to get the highest possible score on the AU values highlighted, i.e. to raise the AU bar as high as possible when the participant was asked to formulate a particular emotion. The task is unique to each person, as we have different skills in expressing our emotions using our facial muscles. Many people are more adept at moving certain muscle groups, but there is a wide variance in how each AU group is moved, varying from individual to individual. To give a simple example: if you ask someone to move their right eyebrow and left eyebrow, a very high percentage of people will find that they can only lift one eyebrow if the task is to lift only one and not parallel. We involved a FACS specialist in the designing process of this exercise. We aimed to select the most commonly used AUs, including those associated with basic emotions. Additionally, we devised unique tasks, such as 'wolf snarling' or 'giving a kiss' to explore AU groups that might not be typically used in shaping basic emotions.

Face Coloring task

This task focuses on showing a schematic representation of the face with some regions that can be colored. These regions are based on the facial landmarks and how each land-



Figure 3.4: The action unit maximization task for disgust. The task here was to move the bars by moving some regions of the face. Each region is corresponding to an Action Unit and the bars show how much a given action unit is actuated.

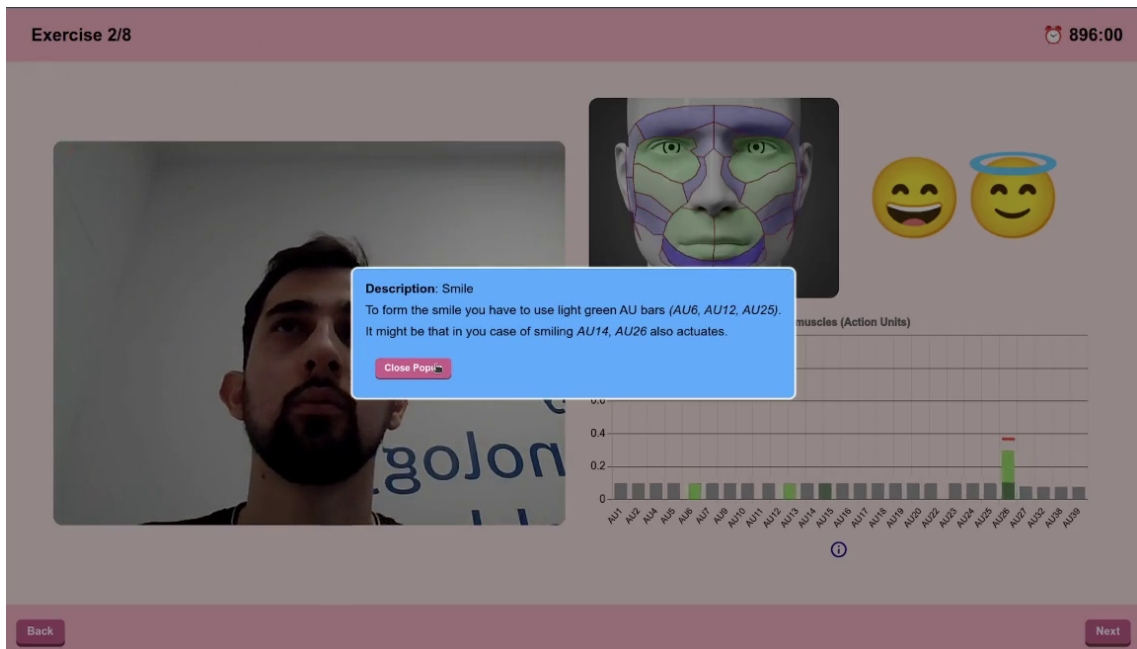


Figure 3.5: Sample description that was given for the user. It shows the name of the exercise such as smile or disgust, and names the highlighted action unit bars. This description is shown at the beginning of the task and it can be opened by clicking the info button.

mark corresponds to a facial muscle according to [48]. Kim et al. selected facial landmark points. In our case, we included the neighboring points of the selected ones thus defining regions. The regions are mapped with action units according to [40], and the regions are colored blue based on action unit intensity (the intensity defines the opacity of the color). The action unit region mapping is a many-to-one relation, so we took the average of the action unit intensities in case the case when two action units are mapped to the same region.

This can be a standalone exercise and a helper visualization for the previous task, as well, where our goal was to teach and be aware of the facial muscles for forming mimicry and emotions. This component has two modes: maximization and real-time. In the maximization mode, the task is to color the whole face so that all regions have the highest opacity. In the case of the real-time, the opacity changes in real-time. We can also define important action units here that would result in some regions being colored green instead of blue.

Action Unit	Landmark Index
Buccinator L	147, 123, 50, 205, 206, 216, 207, 187
Buccinator R	352, 376, 411, 427, 436, 426, 425, 280
Corrugator Supercilii L	151, 66, 107, 55, 221, 189, 193, 8, 9
Corrugator Supercilii R	151, 296, 336, 285, 441, 413, 417, 8, 9
Depressor Anguli Oris L	175, 152, 148, 176, 149, 150, 169, 210, 170, 140, 171
Depressor Anguli Oris R	175, 152, 377, 400, 378, 379, 394, 430, 395, 369, 396
Depressor Labii Inferioris L	200, 199, 175, 171, 140, 170, 210, 212, 202, 204, 194, 201
Depressor Labii Inferioris R	200, 199, 175, 396, 369, 395, 430, 432, 422, 424, 418, 421
Frontalis L	54, 103, 67, 109, 10, 151, 66, 105, 63, 68
Frontalis R	10, 338, 297, 332, 284, 298, 293, 334, 296, 151
Levator Labii Superioris L	193, 122, 196, 236, 198, 209, 129, 142, 100, 47, 114, 245, 189
Levator Labii Superioris R	417, 351, 419, 456, 420, 429, 358, 371, 329, 277, 343, 465, 413
Masseter L	172, 58, 132, 93, 234, 227, 116, 123, 147, 215
Continued on next page	

Action Unit	Landmark Index
Masseter R	454, 323, 361, 288, 397, 435, 376, 352, 345, 447
Mentalis L	200, 199, 175, 171, 140, 170, 210, 212, 202, 204, 194, 201
Mentalis R	200, 199, 175, 396, 369, 395, 430, 432, 422, 424, 418, 421
Orbicularis Oculi L	143, 156, 70, 63, 105, 66, 107, 55, 221, 189, 245, 114, 47, 100, 119, 229, 228, 31, 35
Orbicularis Oculi R	336, 296, 334, 293, 300, 383, 372, 265, 261, 448, 449, 348, 329, 277, 343, 465, 413, 441, 285
Orbicularis Oris Lower	164, 0, 11, 12, 13, 14, 15, 16, 17, 18, 200, 201, 194, 204, 202, 212, 216, 206, 203, 98, 97
Orbicularis Oris Upper	164, 0, 11, 12, 13, 14, 15, 16, 17, 18, 200, 421, 418, 424, 422, 432, 436, 426, 423, 327, 326
Risorius L	215, 147, 187, 207, 216, 212, 214, 192
Risorius R	376, 435, 416, 434, 432, 436, 427, 411
Temporalis L	234, 127, 162, 21, 54, 68, 63, 70, 156, 143, 116, 227
Temporalis R	284, 251, 389, 356, 454, 447, 345, 372, 383, 300, 293, 298
Zygomatic Major L	123, 116, 143, 35, 31, 228, 229, 119, 100, 142, 129, 203, 206, 205, 50
Zygomatic Major R	372, 345, 352, 280, 425, 426, 423, 358, 371, 329, 348, 449, 448, 261, 265

Table 3.2: The muscles and the corresponding region defined by the landmarks.

Drawing with emotions task

This task approaches the concept of facial expression control from another direction. Instead of relying on action unit values, it uses the detected emotional expression values to give user feedback and measure control capabilities. We split the screen again into two regions. The left side shows the mirrored camera image, and on the right side, we can see a graph where each axis is labeled with a basic expression. The task here is to move a cursor on a trajectory by making a facial expression. These trajectories are defined in such a way that the users are required to create each of the six emotion expressions to achieve better results.

The cursor's movement is controlled by the emotional expression value received from the model based on the face the user is making. When an emotional expression is re-

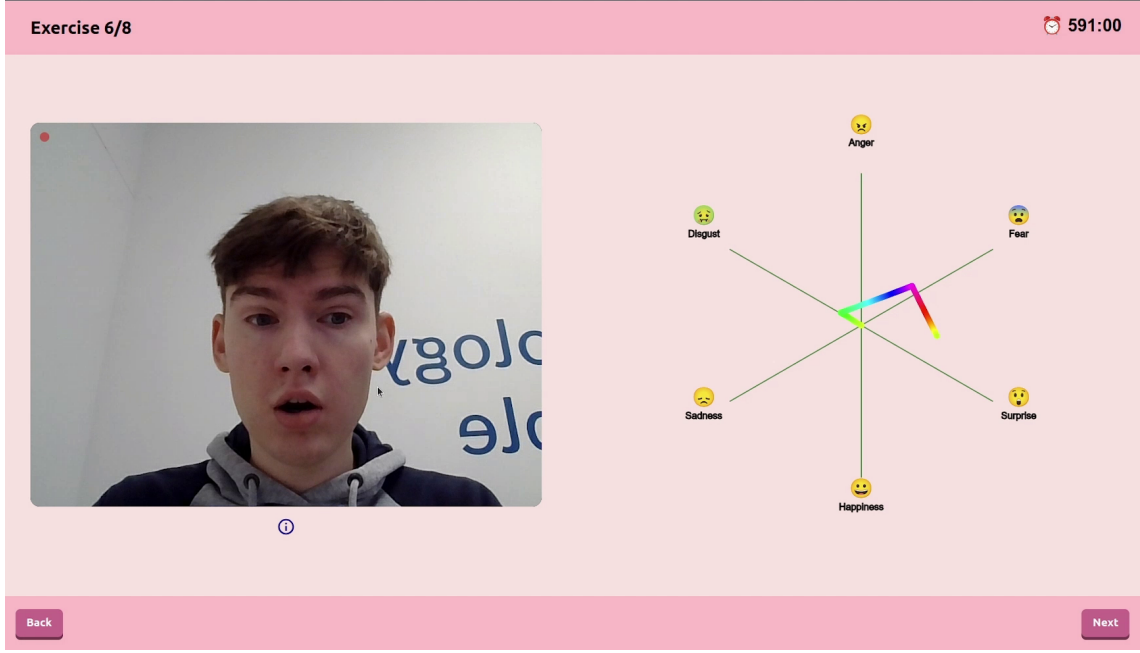


Figure 3.6: The "drawing with emotions" task. Here the user had to draw a trajectory by reaching pre-defined target emotional expression endpoints and moving the cursor between them. The cursor could be controlled by creating emotional expressions.

ceived, the cursor moves in the direction corresponding to that emotion on the screen. Each emotional expression value influences the direction and magnitude of the cursor's movement.

The position of the cursor in a given time t is calculated like this:

$$d_t = \begin{bmatrix} \tau \cdot \sum_i \varepsilon_{ti} \cdot \cos \theta_i \\ \tau \cdot \sum_i \varepsilon_{ti} \cdot \sin \theta_i \end{bmatrix} \quad (3.1)$$

$$p_t = (1 - v)p_{t-1} + v \cdot d_t \quad (3.2)$$

where

ε - The emotion expression values received from the model without the neutral value.

It contains the values for anger, fear, surprise, happiness, sadness, and disgust.

τ - The value that constrains how far we can go from the center of the graph. It is based on the size of the graph.

θ - The angles of each direction are defined for all of the emotion values (**Table 3.3**).

They are defined clockwise because the image that we draw on puts the (0, 0) coordinate on the top left corner of the image and the y-axis increases as we go down on the image. So the coordinate system is flipped on the x-axis.

Expression	Direction
Anger	$\frac{3\pi}{2}$
Fear	$\frac{11\pi}{6}$
Surprise	$\frac{\pi}{6}$
Happiness	$\frac{\pi}{2}$
Sadness	$\frac{5\pi}{6}$
Disgust	$\frac{7\pi}{6}$

Table 3.3: The emotional expression and their corresponding angles if we define the directions as polar coordinates. The angles are defined clockwise (surprise, happiness, sadness, disgust, anger, fear).

ν - Controls how much we consider the new direction and the last position into the new cursor position. It is based on the image's size so the cursor would move faster on a bigger image.

The d_t vector is the direction in which the cursor has to be moved. It contains the Cartesian coordinates of the direction vector calculated from the polar coordinates defined by the θ angles (**Table 3.3**) and ε as the radius of the circle. The τ value increases the size of this vector to match it with the size of the graph thus creating a target position at a given time.

The game is designed so the cursor moves to the target emotion expression endpoint calculated from the received values. The neutral position is in the middle. Since the output logits of the model are passed through the softmax function thus emotion values add up to 1. In this case, it can be considered that the neutral is the choice of "none of the above" if all the other emotion expressions are around zero, then the neutral would be much higher. Since we have defined ν then $(1 - \nu)$ is the "forget" rate, this results in that, if d_t happens to be a $[0\ 0]$ vector then we gradually move back to the center of the graph. In short, if none of the other emotions are present we move the cursor back to the neutral position.

During the two experiments, we slightly redesigned the game. The original goal of the participants was to draw freely with their emotions and draw a polygon by forming different emotions. In the second experiment, we slightly modified the task, such that we included arrows on the trajectory that they had to follow. They didn't need to follow it, it

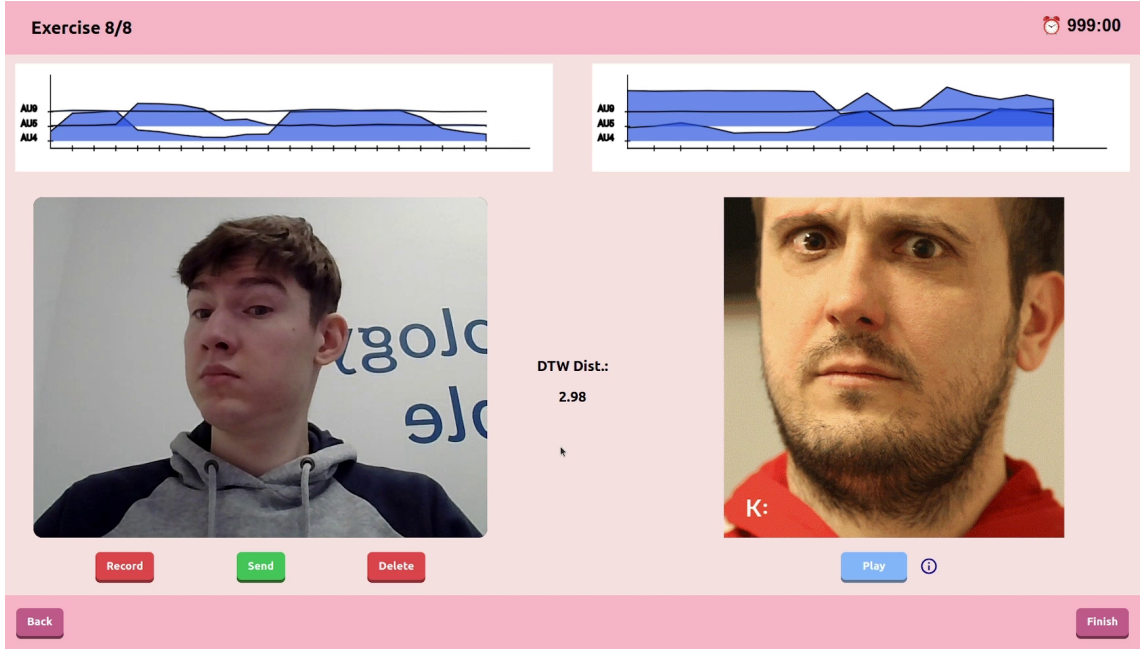


Figure 3.7: The imitation task. The figure shows FaceGym’s imitation exercise where the user had to recreate a small scene, compare the results with the reference video and reach a minimal DTW Dist.

acted as a guide on how to create the given polygon. Also, each of these exercises starts by placing the cursor in the neutral position (center of the graph).

One of the most common feedback from the pilot tests was that playing with the drawing game was ‘frustrating’. The cursor moved slowly, and the detection of the emotional expressions was noisy. To solve this problem, we adjusted v to consider the new direction even more and we replaced the originally used Py-feat [21] emotion detection model with a newer fine-tuned model, EmoNet, which was included in the Exordium [32] package.

Imitation task

The imitation game explores AU control using a different approach. We present a reference video to the player on the right, and the task is to imitate the scene. The task for the user is to record a video imitation of the reference video, and then submit it to the server. They can see the reference video and record their trial as many times as they want. As a convenience, we also play the two videos simultaneously after the recording so the user can compare them visually. As feedback, we show two ridge line graphs to the user showing the recordings and the reference three most significant AU changes over time.

To measure the similarity between the two series we utilize DTW for which we used **tslearn**’s [49] implementation. According to the publication, the problem can formalised

as follows:

$$DTW(x, y) = \min_{\pi} \sqrt{\sum_{(i, j) \in \pi} d(x_i, y_j)^2}, \quad (3.3)$$

where

$$d(X, Y) = \|X - Y\|, \quad (3.4)$$

which is the Euclidean distance, and $\pi = [\pi_0, \dots, \pi_K]$ is a path that satisfies the following properties:

1. $\pi_k = (i_k, j_k)$ with $0 \leq i_k < n$ and $0 \leq j_k < m$
2. $\pi_0 = (0, 0)$ and $\pi_K = (n - 1, m - 1)$
3. for all $k > 0$, $\pi_k = (i_k, j_k)$ is related to $\pi_{k-1} = (i_{k-1}, j_{k-1})$ as follows:
 - $i_{k-1} \leq i_k \leq i_{k-1} + 1$
 - $j_{k-1} \leq j_k \leq j_{k-1} + 1$

If we leave constrain (2) and allow the algorithm to choose its starting pair and ending pair then we get the DTW sub-sequence. **Figure 3.8** shows the difference between the two. The sub-sequence is less sensitive to the length difference between both time series and usually gives smaller distances. In most cases, the sub-sequence gives us better results, but when the other time series is very different from the reference, the sub-sequence will only contain one frame from the other time series, as shown in **Figure 3.9**. As a result, we have decided that we only show and consider the original DTW problem, and calculate DTW distance accordingly.

It was difficult to find the dataset of the imitation videos. Our original idea was to use the GEMEP [9] corpus, a benchmark dataset of emotions and facial expressions played by professional actors. The videos focus on full-body emotional expression and contain audio, which we did not need. After the first tests, we decided not to use GEMEP videos because their expressions weren't suitable for our experiment, as we only prioritized facial expressions, and some of the expressions were not suitable for the target users' age. In light of this, we decided to use the GIF Theatre dataset provided by Katona József Theatre for our research project instead. These GIFs feature short scenes performed by professional actors, focusing on facial expressions and head poses. As a side note, we only considered the facial action unit values in our evaluation.

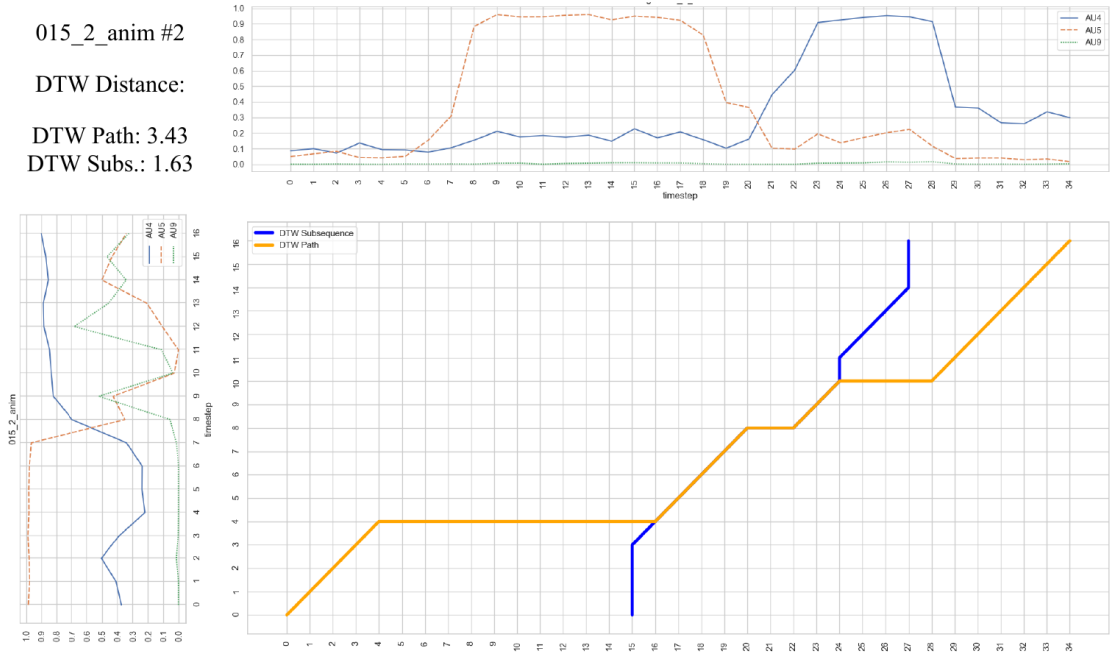


Figure 3.8: The results of the DTW for the original problem and the relaxation when the time series are similar. The orange line shows the original, while the blue line shows the sub-sequence frame matching.

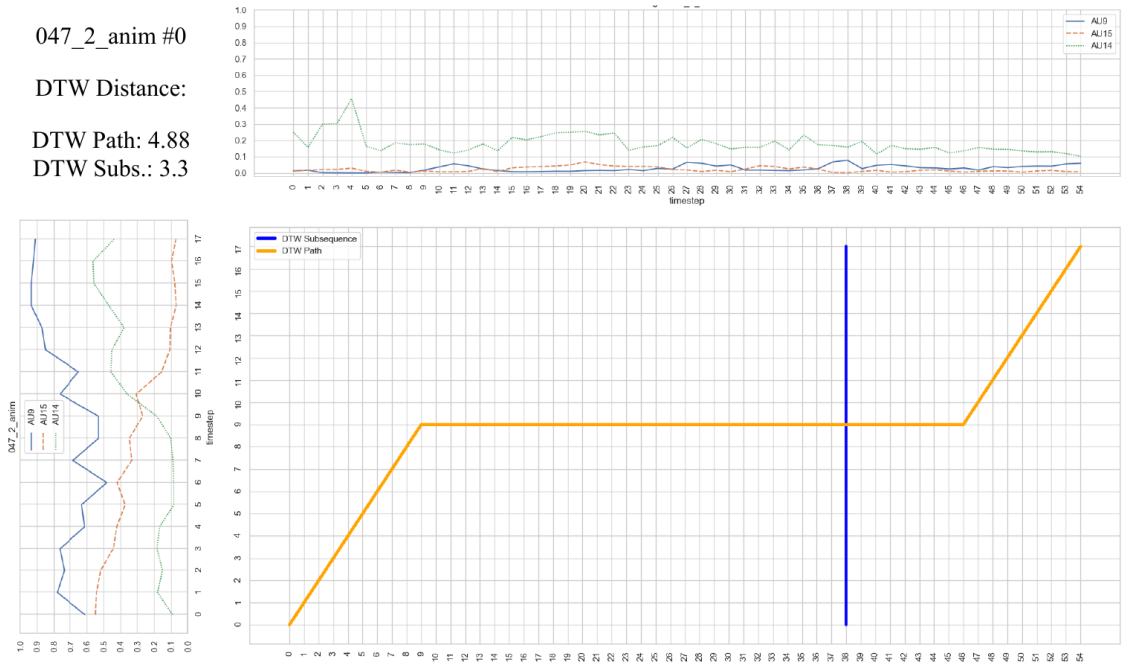


Figure 3.9: The results of the DTW for the original problem and the relaxation, when the time series are very different. The orange line shows the original, while the blue line shows the sub-sequence frame matching.

3.5.2 Architecture

For the first version of the software, the target platform was laptops because we wanted to conduct the experiment at the university on four identical laptops. The easier solution to this was to create a web application (**Figure 3.10**). This made it easier for the UI to run on a device with a lower hardware specification. The backend with GPUs could run on a server and provide an API. Then the UI can communicate with the server through this interface by sending the user data to the server, and the server sends back the data required for the feedback. In the meantime, the sent images can be saved and stored on the server.

We didn't store the measured data while the game ran because we didn't measure everything at the same moment. The initial plan was to measure everything and save everything, but it reduced the system's throughput. The servers only allowed us to use one model at a time (1 GPU was given out for 1-2 participants at a time). With this, we could guarantee near real-time feedback (we could process 2-4 frames per second).

For evaluation, we only required the following data:

1. The results from the questionnaires
2. Raw images
3. Timestamps when a given image was saved
4. The user to which the image belongs
5. The timestamps of starting and finishing a given task

While exploring various database management systems for our project, we considered options like MySQL [50], MongoDB [51], and InfluxDB [52]. MySQL is a widely used relational database management system (RDBMS), and MongoDB is known for its object-oriented database management capabilities. InfluxDB, optimized for time series data, also seemed like a promising choice due to its specific alignment with our data storage needs. However, after careful consideration, we decided against using these systems as they would introduce additional complexity and an extra step in data retrieval processes. Instead, we choose to store our data into CSV files, organizing them such that each individual's data is contained in their own 'periods.csv', simplifying both storage and access. To save the images recorded during gameplay we create a dedicated directory for each user, and we store the images there. Also, we add a new UUID and the timestamp of the save for each frame, thus avoiding collision.

For the long run, we would choose InfluxDB, because it would make sense since we only store time series data. We focus only on one user at a time and there are not a lot of relationships between the dataframes. It would also make it easier to process only sub-parts of time series, for example, collecting only the section that corresponds, for example, the AU maximization exercises. If the application would be released for commercial use, this database would be the target choice for this case, especially if, at the end of the game, we want to provide a comprehensive report on the findings.

Network latency

During our tests of the game with multiple users, we observed that the response time increased. After conducting further investigations, we found that using a higher image sampling rate - 15 FPS - was causing a significant reduction in throughput, even though we were only sending 2 FPS images for inference. This was mainly due to the storage we used to store the images we network mounted. Additionally, the network infrastructure was not capable of handling such a high volume of traffic at the time. We also attempted to use network cables to reduce the likelihood that the latency was due to using the wireless network.

As a result, we reduced the sampling rate to 2 FPS since it still captures the important moments of the game. Also, in the original architecture, we put a web API around each of the models because each model required a different Python environment. This also increased latency because the images have been put through the network multiple times. This made us rethink the architecture. Finally, we could unify the environment for all the used models, so we could include all the models under the same Python environment.

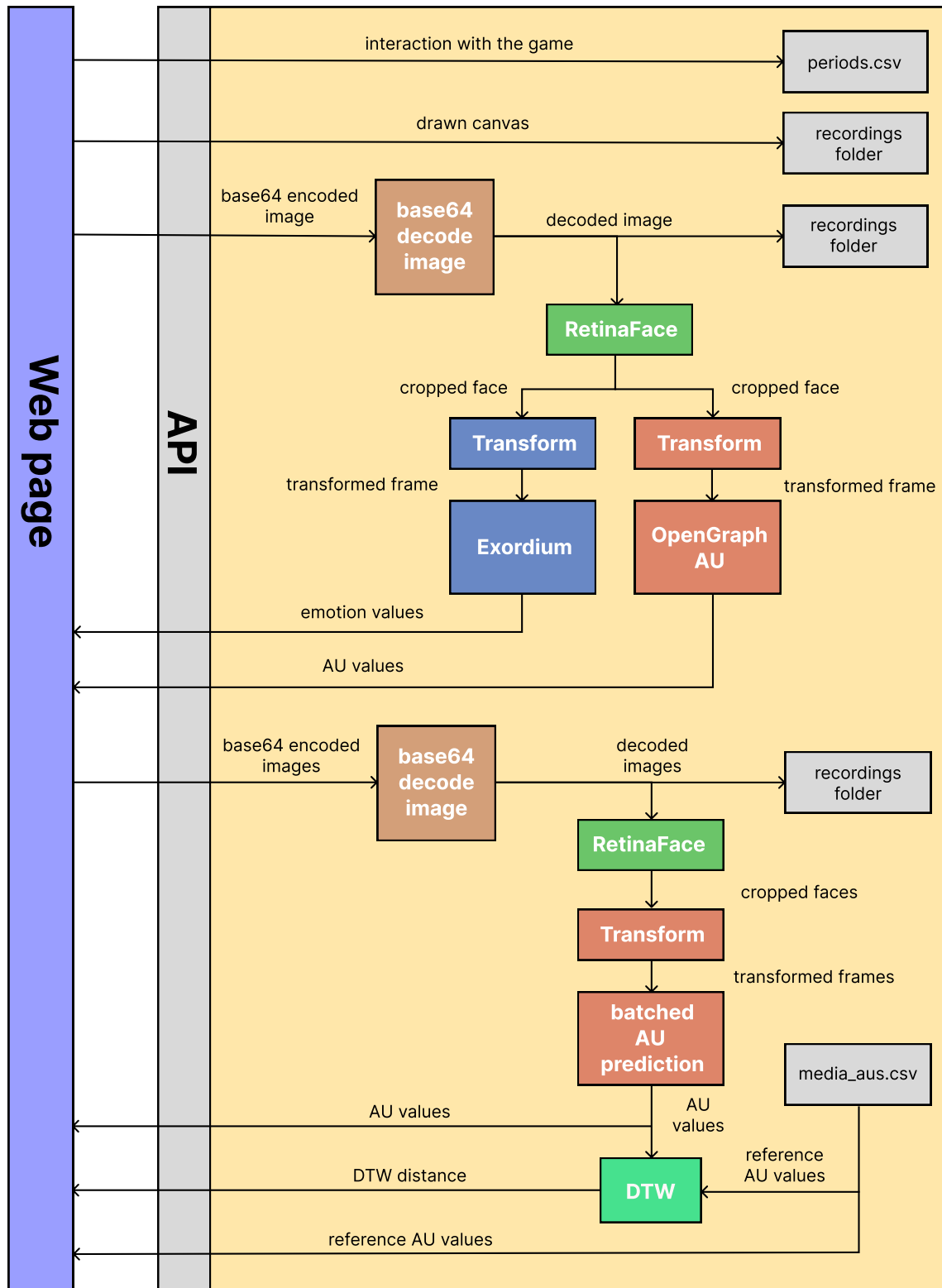


Figure 3.10: The architecture of the software. It depicts how we manage the incoming images, how each AI module is integrated, and what we are sending back to the user as feedback.



Figure 3.11: The diagram of the data. The periods table contains data related to what exercise the user was doing. The detections table contains the values received from the AU and emotional expression detection model for each recorded frame. The two tables share the **name** field which stores the users' name.

Chapter 4

Experiment

We conducted our pilot experiment in November on psychology students who happily volunteered. This occasion also acted as a test of our application on the whole scale. This was the easiest way to deploy and finalize the application for such occasions. The pilot gave us valuable insight into how the users would react to the user interface, whether those descriptions are understandable, and whether we need to include anything else to make the user experience as good as possible. Fortunately, we didn't find any issues related to data saving, however, we found mostly task-related issues.

We designed two experiments with a ninth-grade class, the first of which took place between 11-16 December 2023. The second experiment, a modified version of the first, was conducted between 8-12 April 2024. In this study, we focus on the results of the second experiment. We used the pilot experiment and the first experiment's outcomes to rigorously test our system. Subsequently, we made significant modifications to the game protocol and data handling methods, guided by the insights gained from the results.

4.1 Setup

- We set up a test room in the ELTE IK AI Department Library, where we provided four test stations separated by mobile walls (**Figure 4.1**).
- We had four Lenovo V330 laptops (14, Core I5, 8Gb, 256Gb SSD, Ubuntu) and a server with 2 NVidia Titan RTX.
- The backend was running on the servers, and there was a web server that was accessible through an SSH tunnel during the experiment.

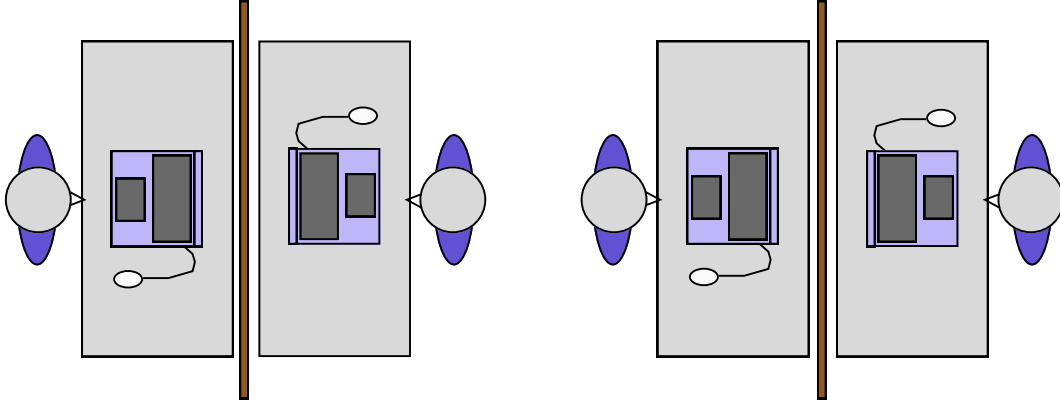


Figure 4.1: An illustration of our setup for the experiment. We tried to separate the participants as much as possible so they didn’t distract each other. Participants facing each other were separated by a movable wall.

4.2 Participants

We approached a biology-chemistry class of ninth-grade high school students to participate in our experiment, with a total of 29 participants. The class completed two different experiments, but for the purpose of this study, we only analyzed the results of the second experiment. All participants played the FaceGym game and filled out the accompanying questionnaires. The average age of the participants was 14.57 years old, with 17 females and 12 males. Each participant was tested individually in a personal testing station that we set up in the AI Department Library Room. To ensure privacy, we separated each test box with mobile walls (**Figure 4.1**).

Based on the different personalities, motivations, and gaming strategies, the experiment took between half to one hour. One task was set up with a time limit, but the other tasks were set up without any time limitations, and we let the participants play as many as they wanted. To keep the initial conditions for all participants, they were asked not to discuss the experiment with each other. Each explained and signed the GDPR form before the experimental phase began. Because the participants were minors, we also signed an Informed Consent of their legal representative parent, acting on behalf of an underaged person document. We have analyzed the recorded data anonymously. The database, which mainly consists of a collection of participant’s facial photos, is not publicly accessible for research purposes outside of the NIPG research group. We are not allowed to show any pictures of the participants in this study. However, to present exercises, problems, or results, we asked permission from an adolescent girl (not from the class) and her parents to

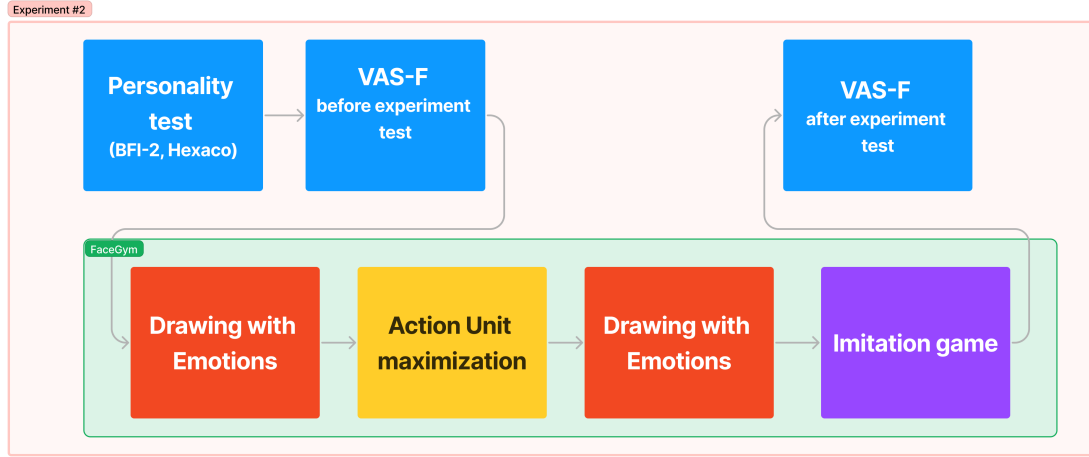


Figure 4.2: Second experiment setup. The blue boxes represent the questionnaires the participants filled out during the experiment. For the experiment, we designed the FaceGym for the Drawing with Emotions -AU maximization - Drawing with Emotions - Imitation game setup.

use the footage of her during the test for publication purposes.

4.3 Protocol of the second experiment

The FaceGym application contains several games, and we can design different experiments with slightly different game scripts. The schematic protocol for the second Experiment is illustrated in (Figure 4.2). We had two online Microsoft Form questionnaire blocks before and after the game. In the FaceGym section, we had four facial expression games.

4.3.1 Questionary part

In the first part of the questionnaire, the students completed the **personality tests** to calculate the OCEAN traits using **BFI-2** and the Honesty-Humility dimension using **Hexaco**. We utilized the validated Hungarian version of both tests. To assess their energy level, participants completed the **VAS-F**. Since we couldn't find a validated Hungarian version of the VAS-F test, we translated it. In the personality test and game, the students used specific nicknames, which allowed us to link the different datasets for a thorough analysis. In this TDK study, we have anonymized the nicknames and will use the notation system S1 to S29 to mark students. After participating in FaceGym, the students once again completed the VAS-F test to assess their energy levels post-experiment.

4.3.2 Game part

The gaming section involved four tasks that needed to be completed in sequence.

Drawing with Emotions

The first task was the Drawing with Emotions task, where the students had to draw polygons using emotions to control the cursor. We designed three different sets in the game without any time limitations so that the students could try each exercise as many times as they wanted:

- In the first action, we asked the students to visit all six emotions: anger, disgust, happiness, sadness, surprise, and fear. We designed this exercise for practice, to get familiar with the game and how they can control the cursor.
- In the second action we designed a triangle with anger, surprise, and sadness peaks. The student's task was to draw a triangle. The starting point was the neutral emotion, and the task was to go from neutral-anger-sadness-surprise-anger.
- In the third action, we rotated the triangle, whose peaks were at disgust-happiness-fear. The starting point was the neutral emotion, and the task was to go from neutral-disgust-happiness-fear-disgust.

Action Unit maximization task

The second game was the Action Unit maximization task, where they had to play on 10 different setups, and the goal was to maximize the dominant Action Units associated with the selected emotions or facial expressions. We set a time limitation; they only had 30 seconds to achieve the maximum AU score in each setup.

During the pilot experiment, we got a lot of questions regarding the descriptions. In the descriptions, we provided them with the name of the facial muscle that has to be moved for a given task. The following text shows a sample from the description:

Description: Happiness/Joy

Colored AU bars: 6 + 12

AU names:

- Cheek Raiser
- Lip Corner Puller

We replaced this description with a new UI where instead of naming the muscles we included the face coloring visualization with highlighted regions. We also separated the primary and secondary action units and gave a different description:

Description: Smile

To create a smile you have to raise the light green bars (**AU6, AU12, AU25**). It is possible that when you smile, you activate **AU14, AU26** as well.

Drawing with Emotions task

In the third game, the students had to play again with the Drawing with Emotions task, where we modified the setup. We excluded the trial game. They had to play with only the two triangle exercises with no time limitation. They could play as much as they wanted.

Imitation task

The last exercise with FaceGym was the Imitation game. We showed participants short videos of famous Hungarian actors (GIF theatre dataset of Katona József Theatre) expressing different emotions, and their task was to imitate them as accurately as possible. Participants could watch the imitation video as many times as they needed before recording their own video. After recording, they could send it to the server for analysis. The system calculated the DTW (Dynamic Time Warping) of the two videos and displayed the result to the player. The lower the DTW value, the better the sample video was reproduced. We showed participants nine different videos, and there was no time limit or limit on the number of videos they could record.

4.4 Data Pre-Processing

The core of the Imitation game was the GIFs that the users had to watch and recreate. In advance of the experiment these GIFs had to be split into frames, and then each frame had to be measured for Action Units so later we have the reference time series. We only

selected the top three Action Units that took part in the scene, thus preventing the less significant action unit values from being calculated in the distance measure.

4.4.1 Measurements and Post-Processing

After the experiment, we had to detect the Action Unit and Emotional Expressions on the frames. Originally, the models were only able to make inferences on one frame at a time, but this would have been slow, so we introduced batch inferencing for these models. This significantly reduced the measurement time. As a part of this, a custom **DataLoader** class had to be implemented that used the same image transformations as the inference.

In the case of **OpenGraphAU**, the results were quite noisy, so we had to introduce face detection, to exclude the background thus reducing the possibility of including noise in the images. Also, this model's input is a 224x224 image, so it is crucial to include only the face. To cut out the faces from the images we used RetinaFace. This extra pre-processing step is also included in the application.

Similarly, we had to implement batch processing for Exordium not because it didn't include one, but because its built-in video processor not only gives us the emotional expression values but also the action unit values and the face bounding boxes. We only need the emotional expression values. We used a similar approach as before, but we had to include different data pre-processing.

As a part of the data evaluation process, we had to filter out the relevant sections of the data. This meant selecting only those sections of the time series that were relevant to that specific task. The **periods.csv** contains the name of the task, the task description, the operation that can be either start or stop depending on whether the user start or leaves the task, and finally the timestamp of the recording of this event. The **detections.csv** contains the action unit values and emotion expression values for each image. It also contains the timestamp of the recording.

Based on these in further evaluations we can select the relevant images and detections for each of the task. During checking the results it was also helpful that by having these we could create a video where for example on the left we included the recorded frame and on the right we added the given stat of the drawing canvas in the drawing task. Thus we could check how the participants did the exercises.

4.4.2 Evaluation the AU maximization game

Since the task here was to move the bars corresponding to an Action Unit as high as possible we used the following approach. We grouped the detected action units based on the recorded periods. Each period has a start and a stop timestamp, so we have created the groups such that the detections fall between these two timestamps. Then, for each group we took maximum value for each Action Unit column.

4.4.3 Evaluation the Drawing game

To evaluate the drawing game, we took the same algorithm that was used to move the cursor, and using the Euclidean distance, we calculated what was the closest point to each of the emotion endpoints. It is worth mentioning that the two drawing games didn't have overlapping emotional end goals, so we only took those distances that were relevant to the given task. For selecting the relevant detection values, we used the same approach as [\[4.4.2\]](#).

Chapter 5

Results

5.1 Experiment

5.1.1 The distribution of personality traits

In each experiment, we asked the participants to fill out an online personality questionnaire; this was necessary because, during the period between the two experiments, we found validated Hungarian translations of the BFI-2 and the Hexaco test. When we made the first Experiment, this translation wasn't available, and we had to translate the original English test to Hungarian. As a result, we considered the second experiment data to be more credible. We conducted the personality analysis only based on the results of the second experiment.

To represent the participant's personality, we used the Big Five psychological model that describes five broad dimensions of personality:

- Openness,
- Conscientiousness,
- Extraversion,
- Agreeableness,
- Neuroticism

The original score of each personality trait in BFI-2 and Hexaco test is between 1 and 5. To use the data for further analysis, we normalized the values. Each trait, except Neuroticism, is positively associated with higher values and negatively associated with lower values. To have the same scale, we inverted Neuroticism's dimension, and we call

this trait Emotional Stability. To calculate the new values of Emotional Stability, we used the following formula:

$$\text{Emotional Stability} = 1 - \text{Neuroticism}$$

To keep the well-known OCEAN acronym, everywhere in further research, the N in OCEAN means Emotional Stability (**Table 5.1**).

Trait	Value range	mean $\pm$ STD	Difference
O	[0.355..0.980]	0.701 ± 0.162	0.625
C	[0.313..0.980]	0.598 ± 0.150	0.667
E	[0.355..0.918]	0.606 ± 0.141	0.563
A	[0.105..0.958]	0.588 ± 0.175	0.853
N	[0.125..0.958]	0.479 ± 0.216	0.833
P	[0.365..0.895]	0.654 ± 0.112	0.530
S	[0.375..0.805]	0.569 ± 0.102	0.430
H	[0.125..0.95]	0.594 ± 0.167	0.825

Table 5.1: Summary of the personality values. The table shows the statistics of the class's personality test (BFI-2 and Hexaco). We calculated the value range, mean, STD, and the difference between the OCEAN personality traits, Stability, Plasticity, and Honesty-Humility traits. We used these results to split the students into three classes on all personality traits.

For further analysis, we also calculated the Plasticity and Stability values. Stability represents the shared variance of Agreeableness, Conscientiousness, and Emotional Stability, while Plasticity represents the shared variance of Extraversion and Openness [42].

$$\text{Stability} = \text{Agreeableness} + \text{Conscientiousness} + \text{Emotional Stability}$$

$$\text{Plasticity} = \text{Extraversion} + \text{Openness}$$

We plot the personality trait results in a histogram for each trait, showing a Gaussian-like distribution (**Figure 5.1**). We have categorized each trait into three classes, with thresholds determined using trait-specific mean and standard deviation calculations (**Table 5.1**). We also calculated a sixth personality trait, the Honesty-humility factor, which measures honesty and genuine cooperativeness. This trait also shows a Gaussian-like distribution. We plotted (**Figure 5.2**) the class of all the personality values by using PCA where the traits are discretized as:

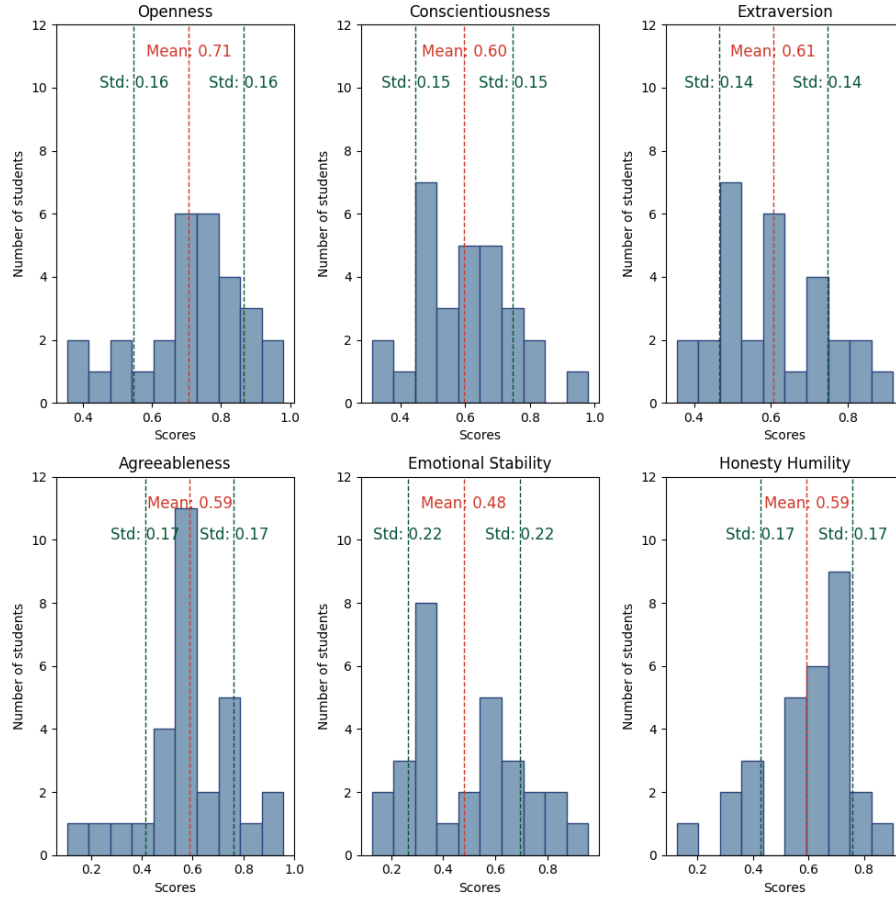


Figure 5.1: Personality trait class. The histograms display the distribution of students' personality traits, including the mean and standard deviation. We have categorized each trait into three classes, with thresholds determined using trait-specific mean and standard deviation calculations.

low (blue) : Trait value < mean - std

middle (grey) : mean - std ≤ Trait value ≤ mean + std

high (red) : Trait value > mean + std

It's difficult to identify a well-defined cluster in both of the three classes because our sample size is 29, and we have very little sample in the low and high category classes both in Plasticity and Stability traits. The Stability shows a very mixed picture. The Plasticity high and low traits, except for two students' data (S21 (high), S11(low)), are separated. These dimensions can be used to analyze the students' emotional expression results if we can see a correlation with the personality traits.

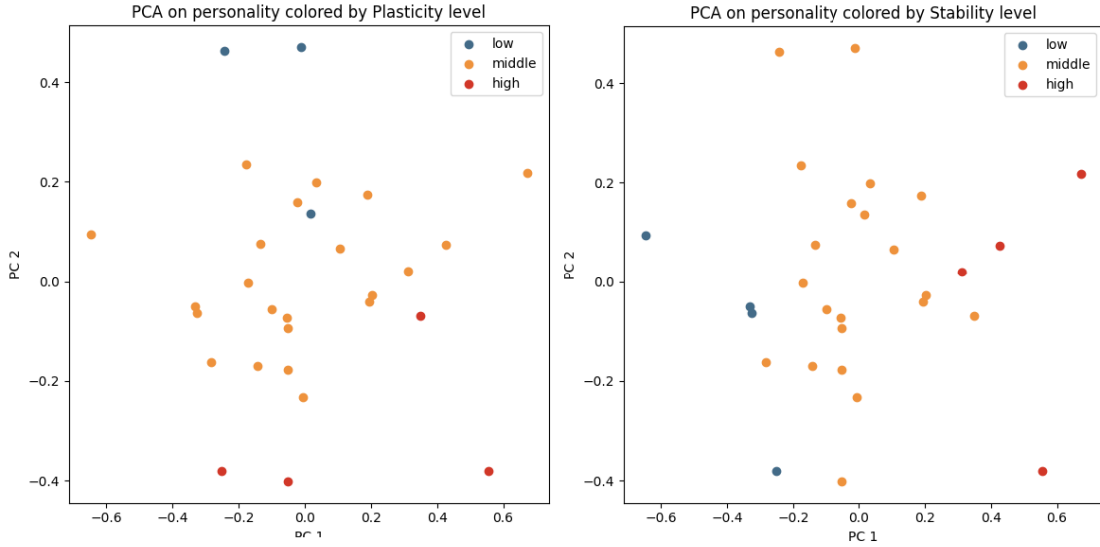


Figure 5.2: Visualization of 2-component PCA of the multidimensional Personality traits. In the left figure, we visualized the PCA using Plasticity values. In the right figure, we visualized the PCA using Stability values. Blue represents the low trait value class, red represents the high trait value class, and orange represents the middle trait value class, which includes the majority of the students. Our sample size is 29, and it is difficult to identify a well-defined cluster in the low and high trait classes.

5.1.2 Energy level

We analyzed the participants' energy levels during the experiment to observe the game's effect on the children. We utilized the VAS-F test to evaluate the data, measuring fatigue and energy levels.

Gender	Female	Male
Energy (↑)	9	7
Energy (↓)	8	5

Table 5.2: Summary of the energy values of the experiment. The table shows that 55% of the students felt refreshed, and 45% felt depleted after finishing the task.

The results (**Table 5.2**) indicate that 16 students experienced higher energy levels after the experiment, feeling somewhat refreshed, while 11 students experienced lower or unchanged energy levels, indicating that they felt a bit tired after the experiment. The students with the highest fatigue levels were S29, S21, S4 and S17, while the most energetic students were S3, S7, S8, S9, S12 and S26. We could not identify significant clusters based on the correlation between energy level and personality (**Figure 5.3**).

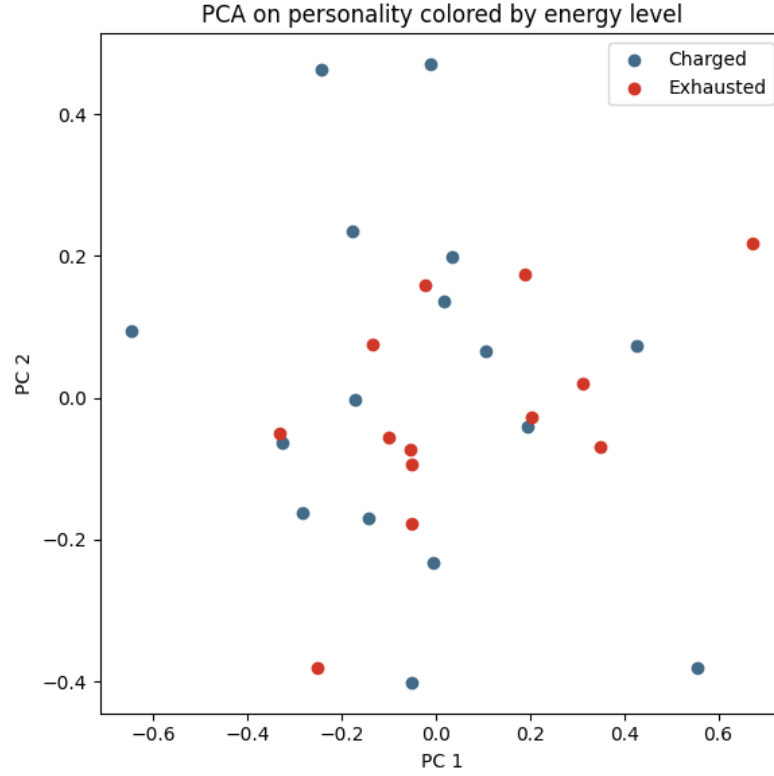


Figure 5.3: Visualization of 2-component PCA of the multidimensional personality traits colored by the energy level results. The figure shows that based on the personality traits and the energy level we cannot define any cluster.

5.1.3 Drawing with emotions task

To evaluate the task, we used two different methods:

- We selected the extent to which the player could approach the end of the triangle's vertices, i.e. how close they could move to the emotion and how steady they could hold the emotion. We calculated these results for both of the three different triangle exercises.
- We also analyzed the values of the EmoNet model on each player's image sequence, which was used to recognize emotions within the game. EmoNet gives us a score in each record frame for all six emotions. This statistic can show us which emotion could express the player.

5.1.4 Action Unit maximization task

To assess this exercise, we analyzed the highest recorded Action Units captured by OpenGraphAU across all players' time-series image data. We utilized the Principal Component Analysis (PCA) method to reduce the data's dimensionality and plotted the

outcomes for each player. By sorting the data based on the maximum values of Action Units, we compared the player's outcomes to their corresponding personality traits. This approach enabled us to form clusters of players who showed better control over their Action Units.

We have analyzed the data, like in the drawing task, to get the maximum values that EmoNet measured on basic facial emotions.

5.1.5 Imitation task

We also analyzed the last game measurement, where the task was to copy the emotional expression of a professional (a short, humorous video from an actor) as precisely as possible. To compare the original video and the players' recording, we calculated the AU values of the three most relevant AUs in each video sample - imitator and player - and compared this using the DTW method. We primarily used the original DTW distance results, but the DTW subsequences can also be used. Our aim was to evaluate which student performed better in these exercises. To provide a comprehensive analysis, it is yet to come to compare with the personality traits and the other exercises' results.

5.2 Evaluating and interpreting the results

5.2.1 Analysis of improvement as a result of AU practice

First, we looked at the drawing exercise results. In this game, the goal was to draw a trajectory by moving a cursor to the predefined target emotions as close as possible. To achieve this, the students had to create these emotional expressions for an extended period (1-3 seconds). As a side note, this task could be tried multiple times and some of the students used this opportunity to refine their drawn trajectory. **Figure 5.4** shows the individual results for these distances for the first and second drawing sessions. From this, it can be seen that **disgust**, **happiness**, and **surprise** were relatively easier, while **anger** and **sadness** were harder.

To investigate further, we looked at the output of EmoNet for these periods. We wanted to check why **anger** and **sadness** targets were harder to approach. We took the predicted emotion levels for the first and second drawing sessions. Then, we took overlapping 2-second-long slices of these series, calculated the maximum value in that slice, and checked if all of the other values in a given slice were within the 10% margin of that maximum

value. For those slices where this condition was satisfied, we kept the maximum value, the other slices were discarded. After that, we retrieved the maximum value from these slices for each period. **Figure 5.6** shows the retrieved maximums for the first and second sessions. This figure is related to **Figure 5.4** in such a way that those who reached smaller distances are close to 1.0.

We have also looked at the correlations between the reached distances as shown in **Figure 5.5**. We see strong correlations between surprise, disgust, happiness, and fear which indicates that for these tasks the participants produced similarly good results. As for anger and sadness, these were much harder, and we also see more variance in the before-after results in **Figure 5.4** compared to the previous emotions.

To check if the **Action Unit Maximization** game helped them to reach better results we have created a histogram showing how many students could achieve a certain distance for both the first and second try. **Figure 5.8** shows the distribution of distances achieved by the students. For **anger** we can see a slight improvement but overall the the results are the same for the sessions. Similarly, we checked these for the emotions (**Figure 5.9**). Here we couldn't see significant improvements. Looking at **Figure 5.7** we were expecting that by giving the participants hints about which action unit to move to create the given emotional expression most of them would produce better results. From this, we can conclude that they didn't make the assigned emotional expression for an extended time. This can be the result of that they may carried out the task by moving each selected action unit bar individually.

5.2.2 Who struggled with anger and sadness?

To investigate the reason why some of the students were struggling with these emotions we looked at their results in the two drawing sessions and the action unit maximization exercise (**Figure 5.6-5.7**). We are trying to find what is common with these students.

First, we had to define a threshold value that defines which results from this game we accept and reject. We have decided that we would accept every student's effort if they could reach the maximum value based on the model within a 2% margin in these three sessions. In this way, **Table 5.3** lists the selected student for further investigation.

To check our hypothesis on whether adolescent girls have problems expressing anger we looked at how many of these students are males or females. **Table 5.4** shows the number of males and females whose values were accepted. The table shows an intriguing

finding that for anger mostly girls and for sadness mostly boys struggled. This brings us momentum to further investigate this phenomenon in the near future.

Emotion	ID
Anger	S19, S8, S15, S11, S28, S9, S25, S19, S14, S21, S10, S23, S18, S27, S6, S20
Sadness	S29, S4, S2, S26, S25, S17, S9, S24, S1, S21, S22, S12, S23, S18, S14, S6, S13

Table 5.3: Selected students for the case study. The students are selected based on their extent of emotions. For those who could not reach 1.0 within a 2% margin, we considered that they did not create that emotional expression for a longer period.

Gender	F (17)		M (12)	
Could do	✓	✗	✓	✗
Anger	5	12	8	4
Sadness	9	8	3	9

Table 5.4: Genderwise results of the extent of anger and sadness. Based on **Figure 5.6** and **Figure 5.7** a summary on who could and couldn't create **anger** and **sadness** within 2% margin split by gender

5.2.3 Correlation between facial emotions and personality

Correlation analysis of emotional distances

We conducted a statistical analysis of the emotional data extracted from the Drawing with Emotions (DwE) game. We compared the students' maximum reached distance values from the two DwE exercises with their OCEAN personality traits. We performed an analysis using both Pearson (linear) and Spearman (monotonic) correlations (**Figure 5.10, 5.11**).

Both methods show the same result. We have not found any correlation between plasticity, stability, extraversion, open-mindedness, and the six emotions. The only small negative correlation we observed is between emotional stability and fear. We also found a small positive correlation between agreeableness and disgust. The highest, medium correlation we identified is between conscientiousness and anger, and we identified a small correlation with disgust and surprise.

Correlation based on the created emotion data (EmoNet)

We also conducted a statistical analysis of the emotional data extracted from the Drawing with Emotions (DwE) game. We compared the students' maximum reached values based on the evaluation mentioned in 5.2.1 from the EmoNet model with their OCEAN personality traits. We performed an analysis using both Pearson (linear) and Spearman (monotonic) correlations (**Figure 5.12-5.13**).

Both methods show more of the same result. Our findings in the Spearman correlation regarding fear exhibit the strongest medium correlation with personalities (emotional stability -0.5, stability -0.3). However, surprise and disgust do not show any notable correlation with the personalities. Happiness, on the other hand, expresses a small negative correlation with extraversion and plasticity. Sadness has a small negative correlation with plasticity, and anger also has a small negative correlation with conscientiousness.

Comparing the performance of emotional expressions with personality based on their highest scores

We calculated all the best emotional expression scores for anger, sadness, surprise, disgust, happiness, and fear based on the EmoNet model and the achieved distance score during the DwE game, both in the first and second exercises. We compared the results of the student's personality traits and showed the following Figures based on the EmoNet: **5.14, 5.15**.

We can see that students struggled more with expressing anger, sadness, and fear, while most of them achieved happiness, disgust, and surprise. Unfortunately, we cannot see any cluster relating to expressing an emotion better and having a higher or smaller personality trait score. Our sample is too small to make this analysis.

We achieved a similar result when we plotted the best-achieved distance of both DwE exercises for expressing the six emotions. The following Figures show the best distance result in the DwE exercise: **5.16, 5.17**.

Here again, the diagram illustrates how each emotion is expressed and the extent to which the students failed to express anger and sadness. Unfortunately, there is no visible cluster with the personality traits.

5.2.4 Personality and emotions considering the impact of learning

In our experiment, one of our goals was to find a correlation between personality traits and the production of emotional expressions. We calculated the best score in both Drawing with Emotions (DwE) tasks using the method described here: [4.4.3]. We calculated a score value for each participant, showing the performance improvement between the two tasks:

$$Score_i = DwE_{2i} - DwE_{1i}$$

where i is the index of the student. The student improves their facial expression if the $Score_i$ value is negative; if the $Score_i$ value is zero or positive, their performance is stable or poor. We consider this value to be the student's 'learning score'. We calculated this rate for each emotion, compared them with the personality, and plotted the results in the following (Figure 5.18)

In the context of the game, it was challenging to express certain facial expressions. The class generally did well in portraying happiness and disgust, with minimal differences in improvement and deterioration rates that make comparing them with personality traits insignificant. We found it interesting to compare the results of the two worst-performing emotions, anger, and sadness, with personality. On the plot, we can see the personality results with PCA, and the coloring is based on productivity; if the student could improve their performance, we colored with blue; if not, then red. In both diagrams, we can see some similar 'clustering'. Next, we recognized that 20 students had the same result in expressing the two emotions (Figure 5.19).

If they perform better in anger, they will also perform better in sadness (10 students: S2, S3, S5, S9, S12, S14, S18, S20, S22, S29). In the opposite direction, if they perform poorly in anger, they also have poor results in sadness (10 students: S1, S4, S6, S8, S11, S13, S16, S25, S26, S28). We do not receive this strong performance if we compare the fear and the surprise. Only 11 students had the same performance (Poor results in S1, S6, S11, S17, S22, S26, and S29; Improvement in both emotions: S3, S12, S13, S20, S27).

The intersecting set of four emotions is tiny: both of the four emotions, S1, S6, S11, and S25 students, had poor results, and S3, S12, and S20 had good results. S3, S12, and S20 are females with high Openmideness (Aesthetic Sensitivity) scores. S6 and S11 are females, and S1 and S26 are males; their Extraversion scores are similar (0.45 - 0.58, lower than the class mean (0.6)).

We further analyzed the data on these two groups that were identified in the anger-sadness relationship. We made a scatter plot with the identified group and compared it with the Personality trait values (**Figure 5.20-5.21**).

Comparing the results with the OpenMindness Personality trait, we can see that the improving group has high personality traits except for one student if we compare this with both anger and sadness. Regarding the other Personality traits, we couldn't define any findings.

As another approach, we have utilized one-way ANOVA to check if a personality trait makes someone more prone to improve their score. We split the participants into three groups similarly to (**Figure 5.19**). The fact that for Agreeableness the p-value is marginally close to the 0.05 threshold to reject the null hypothesis hints that it can be a great direction to check this on a larger scale in the future. As for the other personality traits, we have not found any impact on someone's capability for improvement in this sample.

	F	p
OpenMindedness	1.19	0.32
Conscientiousness	0.92	0.41
Extraversion	0.01	0.99
Agreeableness	3.00	0.07
NegativeEmotionality	0.99	0.39
Plasticity	0.58	0.57
Stability	1.87	0.17

Table 5.5: The results of ANOVA on the OCEAN + Plasticity + Stability personality traits where the groups are defined by the improvements (Improved in both, Improved in only one, and No improvement at all). From these results, only Angreeableness's p-value is close to the reject threshold.

5.2.5 AU Maximization task and Personality

We analyzed the data by correlating the students' results in the AU Maximization game with their personality scores. We gathered the top AU scores of each student, sorted the data, and used PCA to visualize the results (**Figure 5.22**).

Based on the Personality dimensions, we cannot define any clusters. We could only identify S25 as an outlier, who also had the worst performance in the Drawing with

Emotions (DwE) game. This means that on a larger scale, we may find more similar outliers.

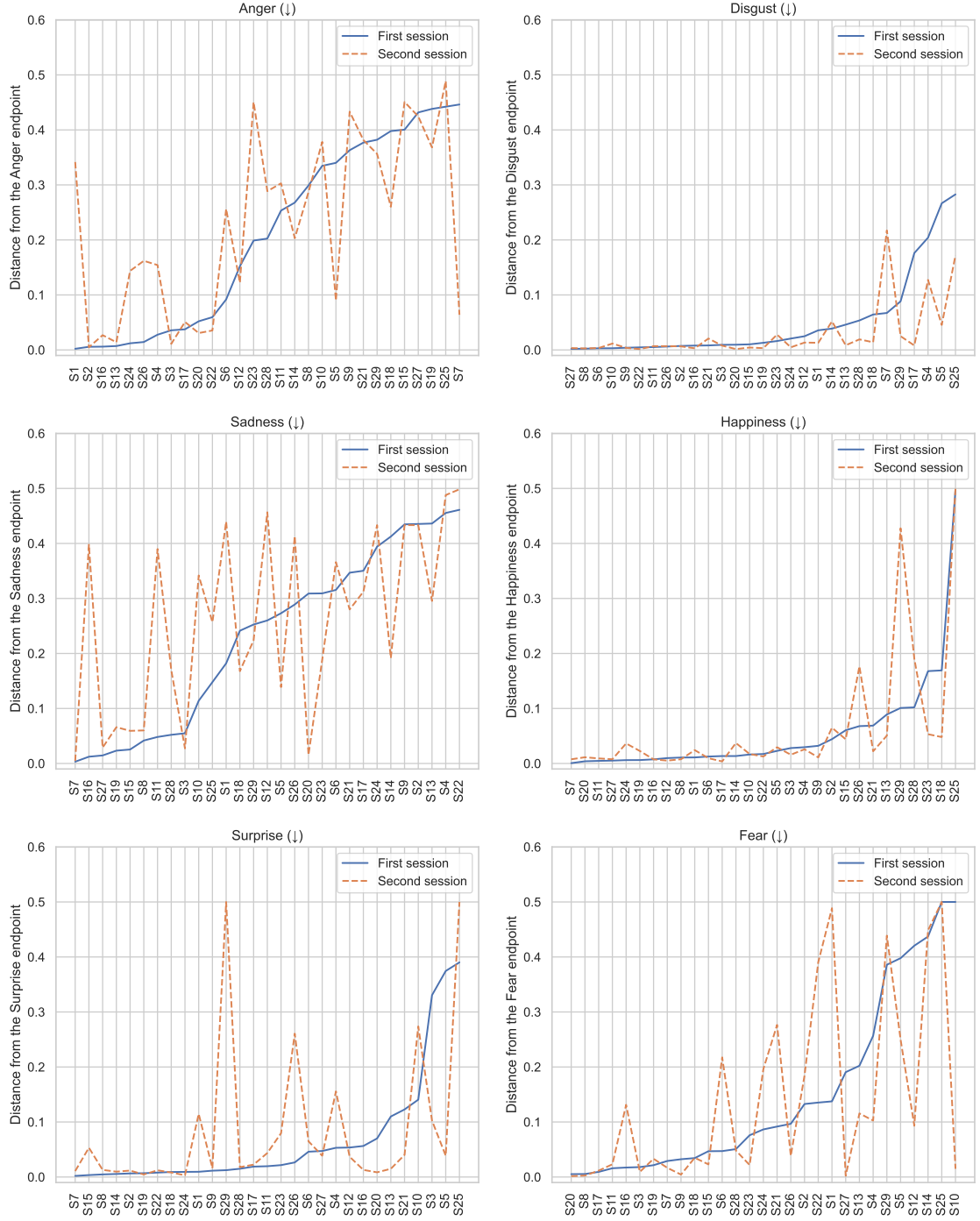


Figure 5.4: The reached distances for the first and second drawing game. The blue line shows the distances from the goal before the AU task and the dashed orange line shows the distances after the AU task. The y-axis shows the best minimal distance from the target emotion endpoint in the drawing exercise. The participants are sorted based on their results from the first session.

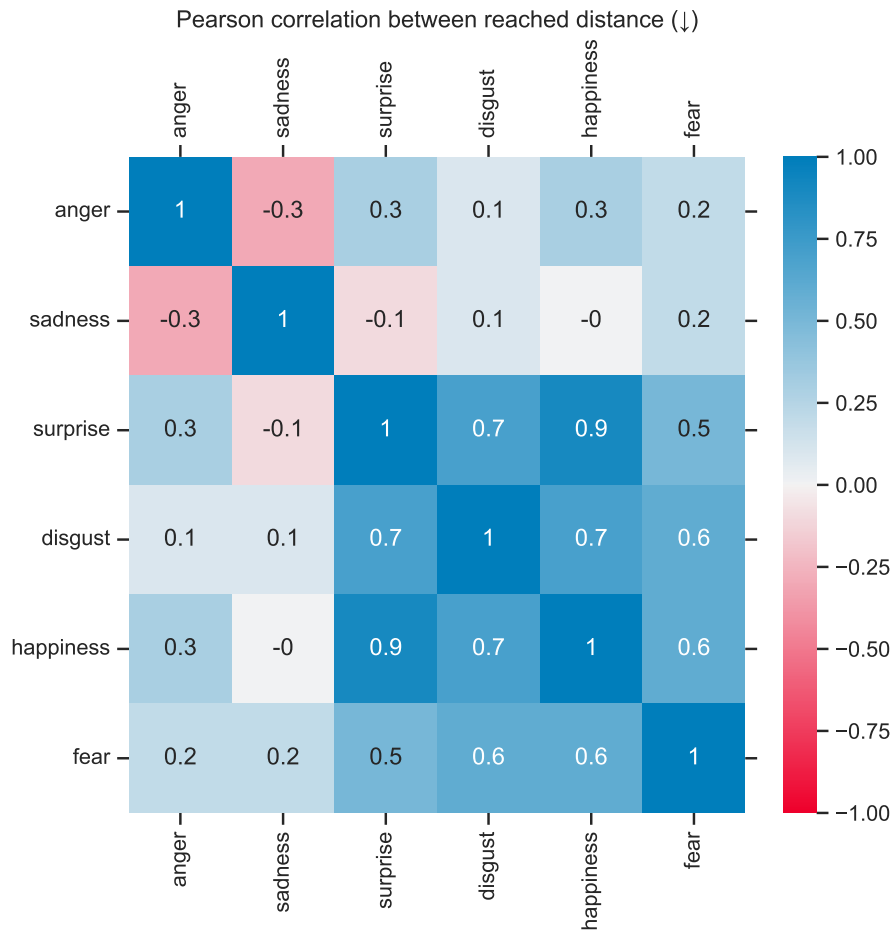


Figure 5.5: The correlation between the distances reached by the participants. We see that for surprise, disgust, happiness, and fear the results are positively correlated meaning that they produced similarly good results for these tasks. As for anger and sadness, there are significantly smaller correlations, because as we have shown these tasks were harder and there are larger variances in the results compared to the other tasks.

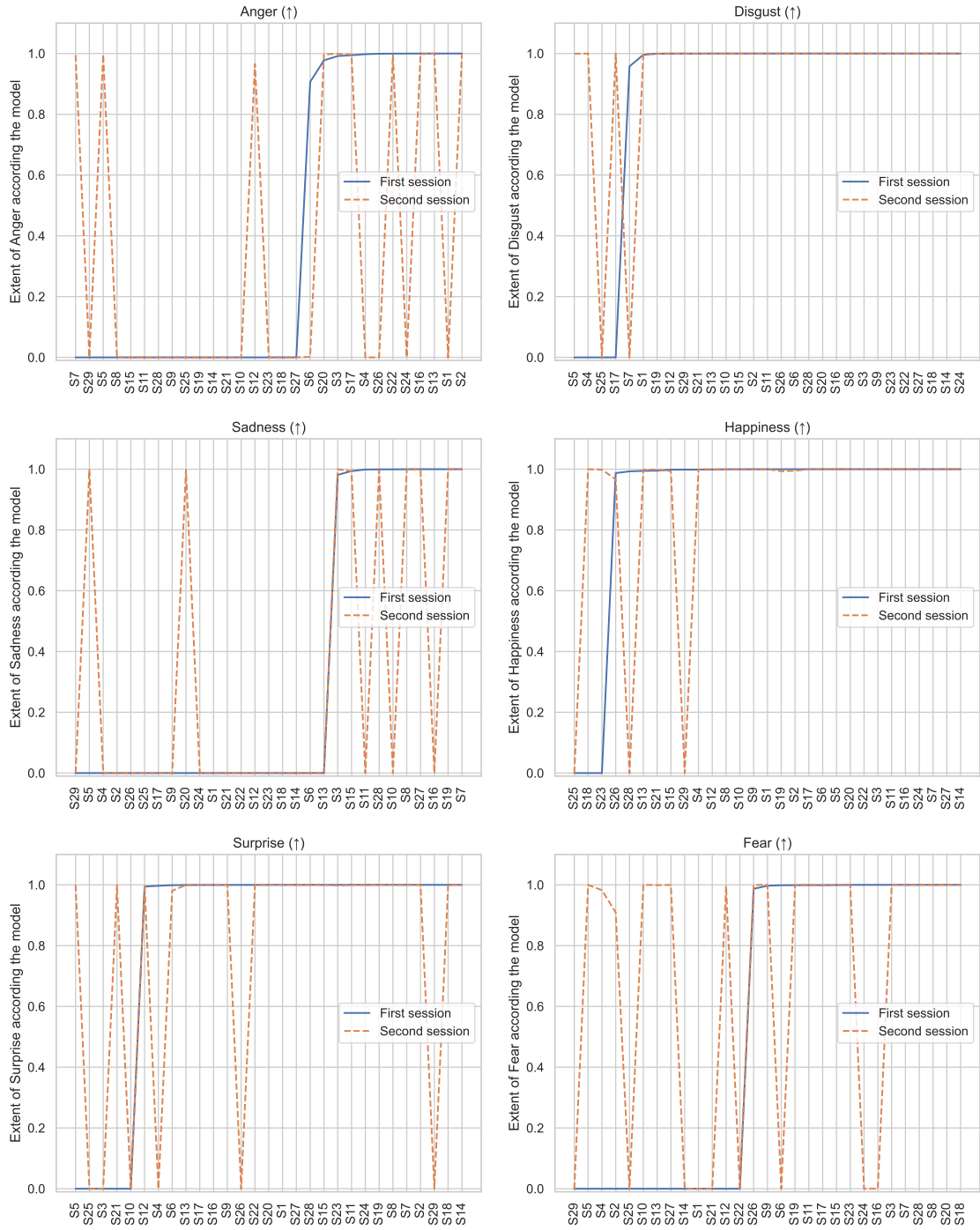


Figure 5.6: The extent of emotion for a longer period based on EmoNet. The blue and orange lines show the extent of the emotion according to the model before and after the AU task during the drawing exercise.

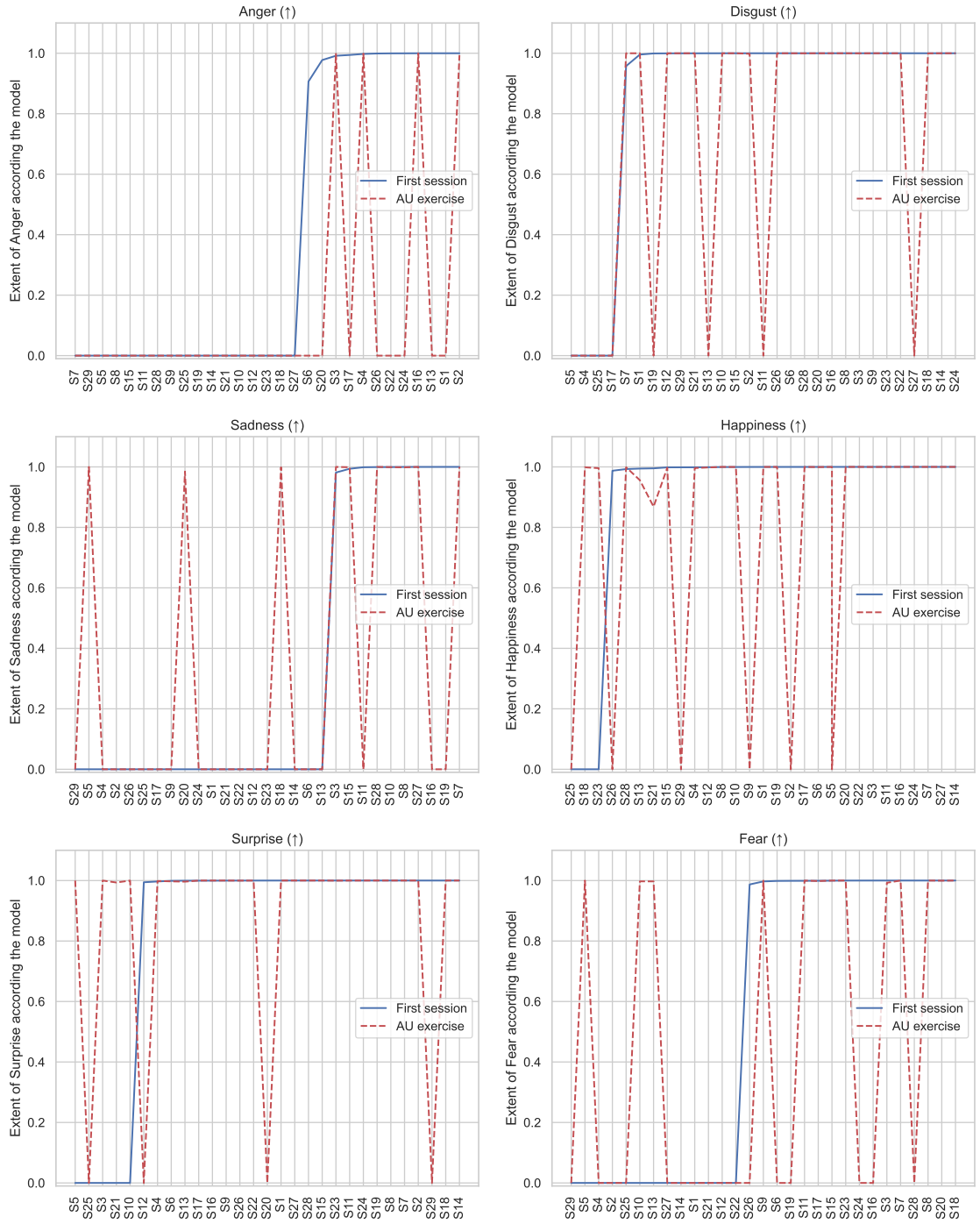


Figure 5.7: The extent of emotion for a longer period based on EmoNet. The blue and red lines show the extent of the emotion according to the model before and during the Action Unit bar maximization exercise.

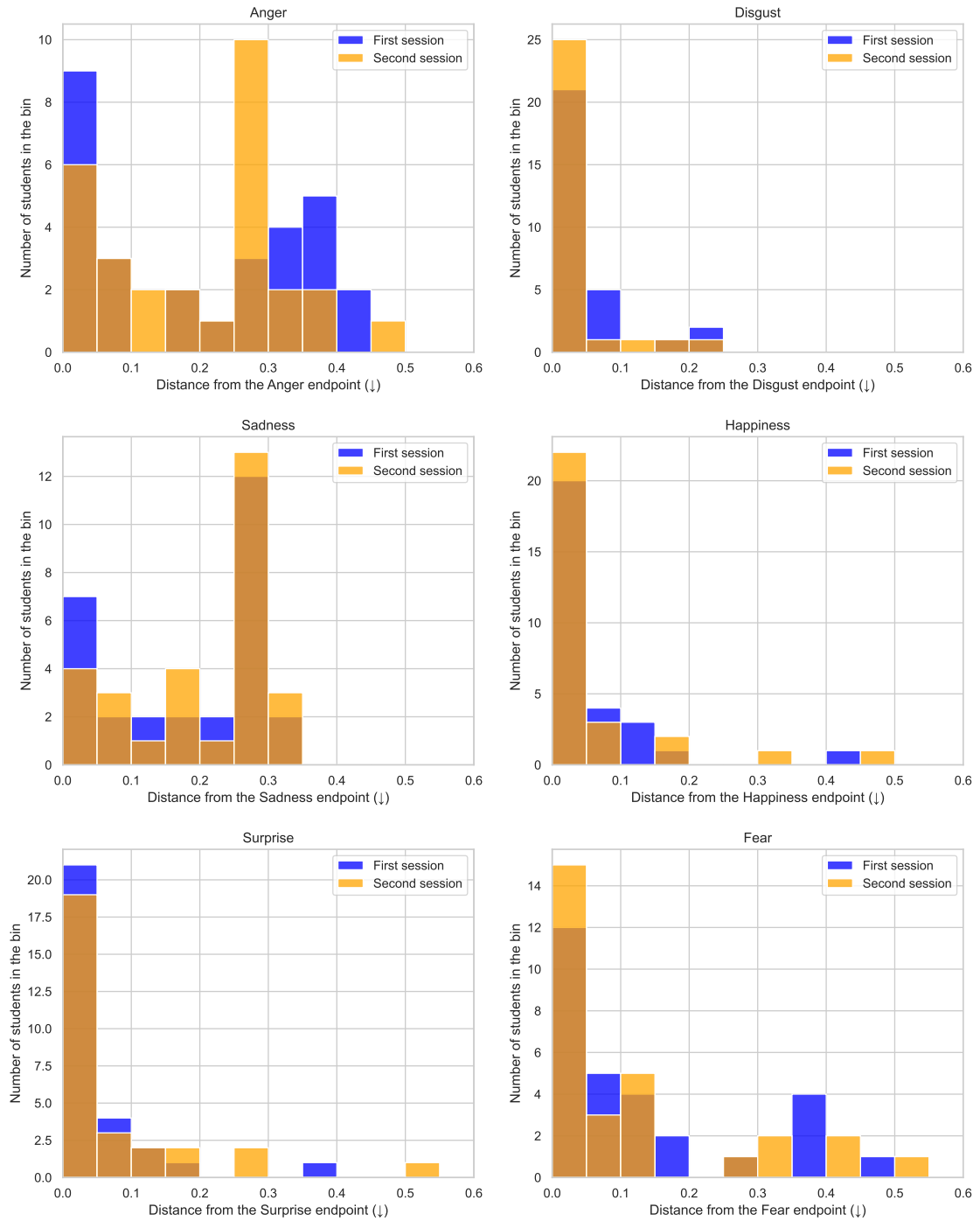


Figure 5.8: Reached distances for each emotion shown on a histogram. The blue and orange lines show the distances before and after the AU task during the drawing exercise.

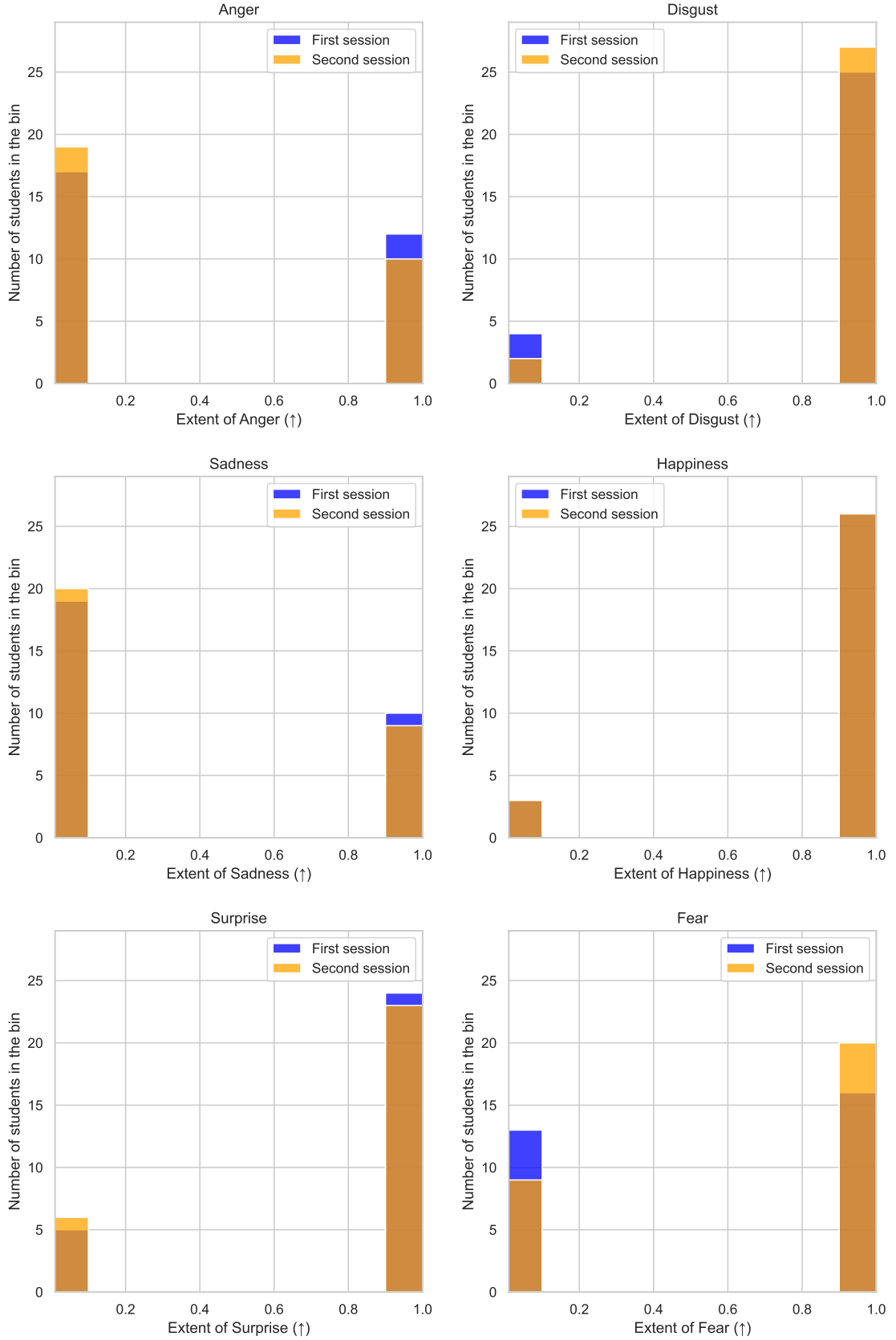


Figure 5.9: Reached extent for each emotion. The blue and orange lines show the extent of the emotion according to the model before and after the AU task during the drawing exercise.

Pearson correlation between reached distance ($\downarrow$) and OCEAN + Plasticity + Stability

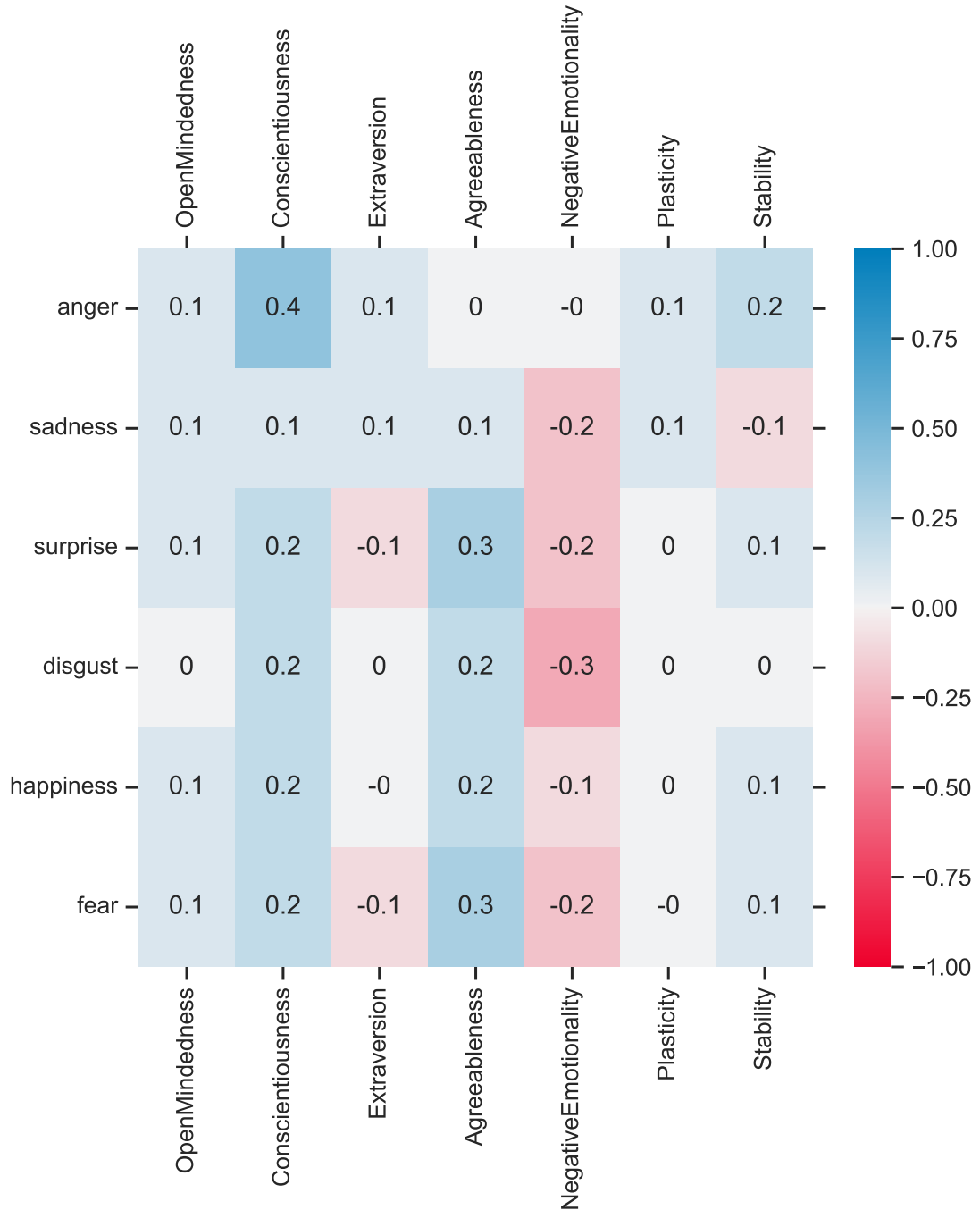


Figure 5.10: Pearson correlation between the reached distance from the emotion target and the OCEAN Personality traits in the DwE game. Blue shades indicate a positive correlation, red for negative. We observe a medium correlation between anger and Conscientiousness, a small correlation between fear and surprise with Agreeableness, and a small negative correlation between disgust and Emotional Stability.

Spearman correlation between reached distance ($\downarrow$) and OCEAN + Plasticity + Stability

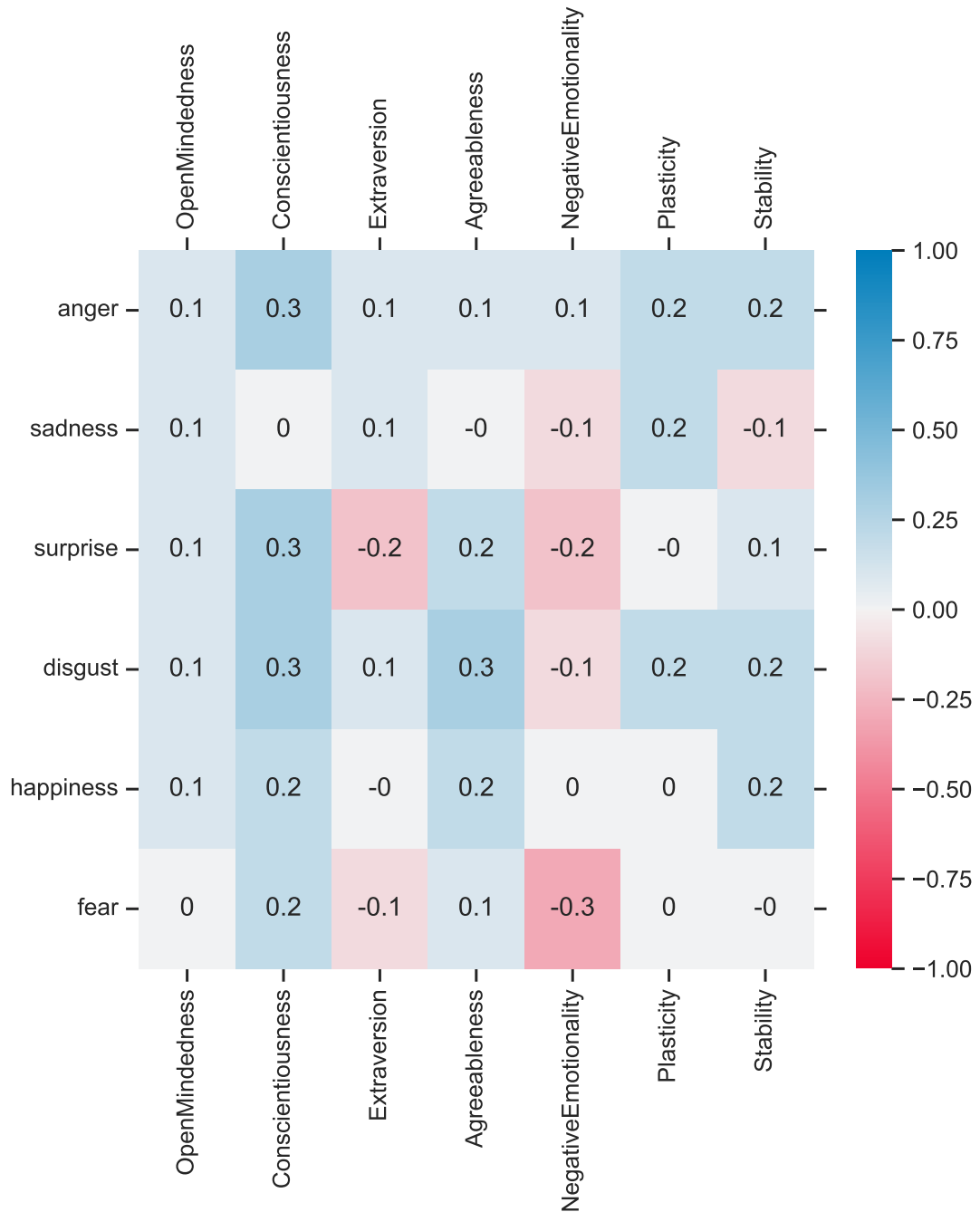


Figure 5.11: Spearman correlation between the reached distance from the emotion target and the OCEAN Personality traits in the DwE game. Blue shades indicate a positive correlation, red for negative. We observe a small correlation between anger, surprise, disgust and Conscientiousness, a small correlation between disgust and Agreeableness, and a small negative correlation between fear and Emotional Stability.

Pearson correlation between created emotion ($\uparrow$) and OCEAN + Plasticity + Stability

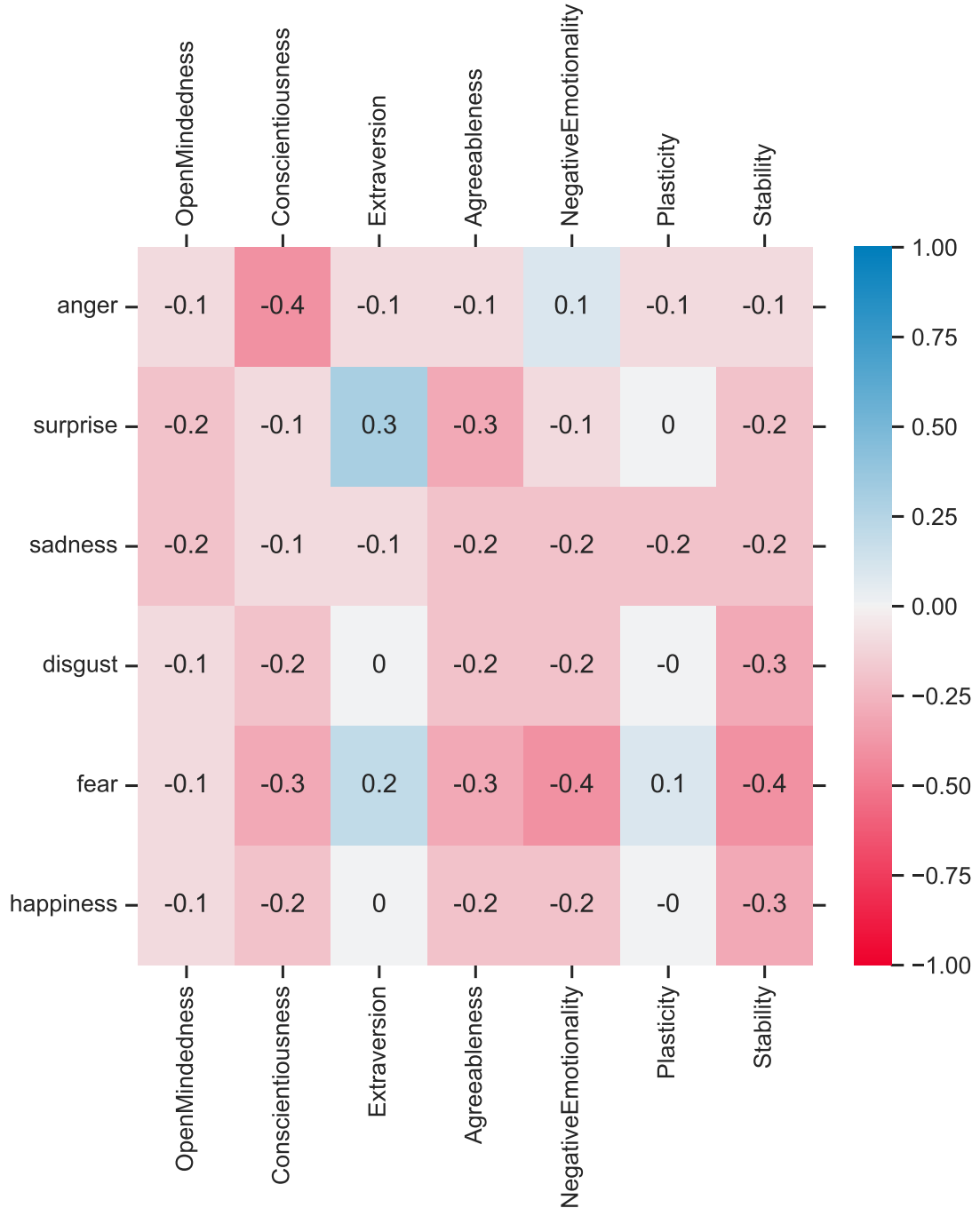


Figure 5.12: Pearson correlation between the emotions detected by EmoNet and the OCEAN Personality traits in the DwE game. Blue shades indicate positive correlation, red for negative. We observe a medium negative correlation between anger and Conscientiousness and a small negative correlation between fear and Conscientiousness. Surprise and fear have a small negative correlation with Agreeableness. Surprise also has a small positive correlation with Extraversion. Fear has a medium negative correlation with Emotional Stability and Stability. Disgust and happiness have also a small negative correlation with Stability.

Spearman correlation between created emotion ($\uparrow$) and OCEAN + Plasticity + Stability



Figure 5.13: Spearman correlation between the emotions detected by EmoNet and the OCEAN Personality traits in DWE game. Blue shades indicate positive correlation, red for negative. We observe a medium negative correlation between fear and Emotional stability. We have a small negative correlation between anger and Conscientiousness, a small negative correlation between happiness and Extraversion, and sadness and happiness with Plasticity. Fear also has a small negative correlation with Stability.

5. Results

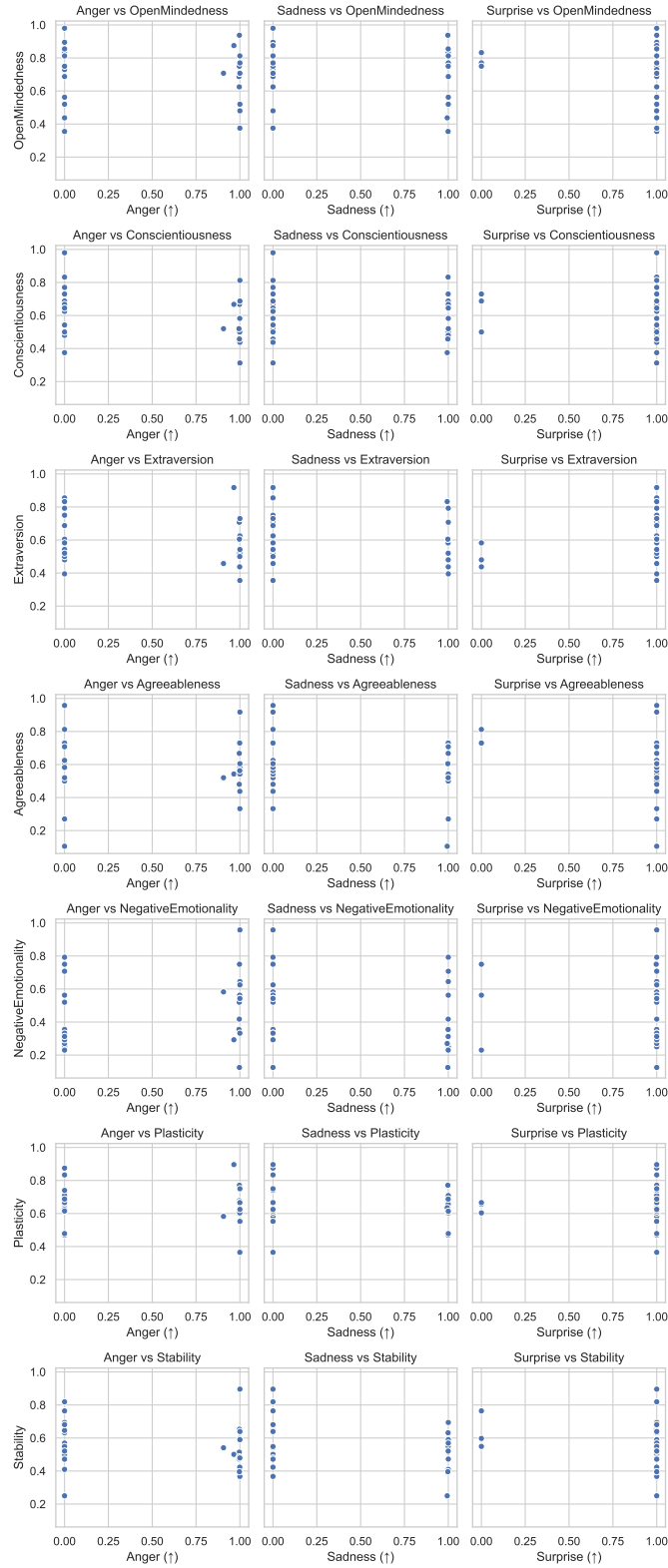


Figure 5.14: Scatter plot of the best-achieved emotion rate and OCEAN personality traits from EmoNet in the Anger-Sadness-Surprise DwE exercise. The x-axis represents the performance of the emotions, with 0 indicating an incorrect performance and 1 indicating a correct performance. No identifiable clusters are observed with the personality. The plot indicates that the students struggled to express anger and sadness effectively.

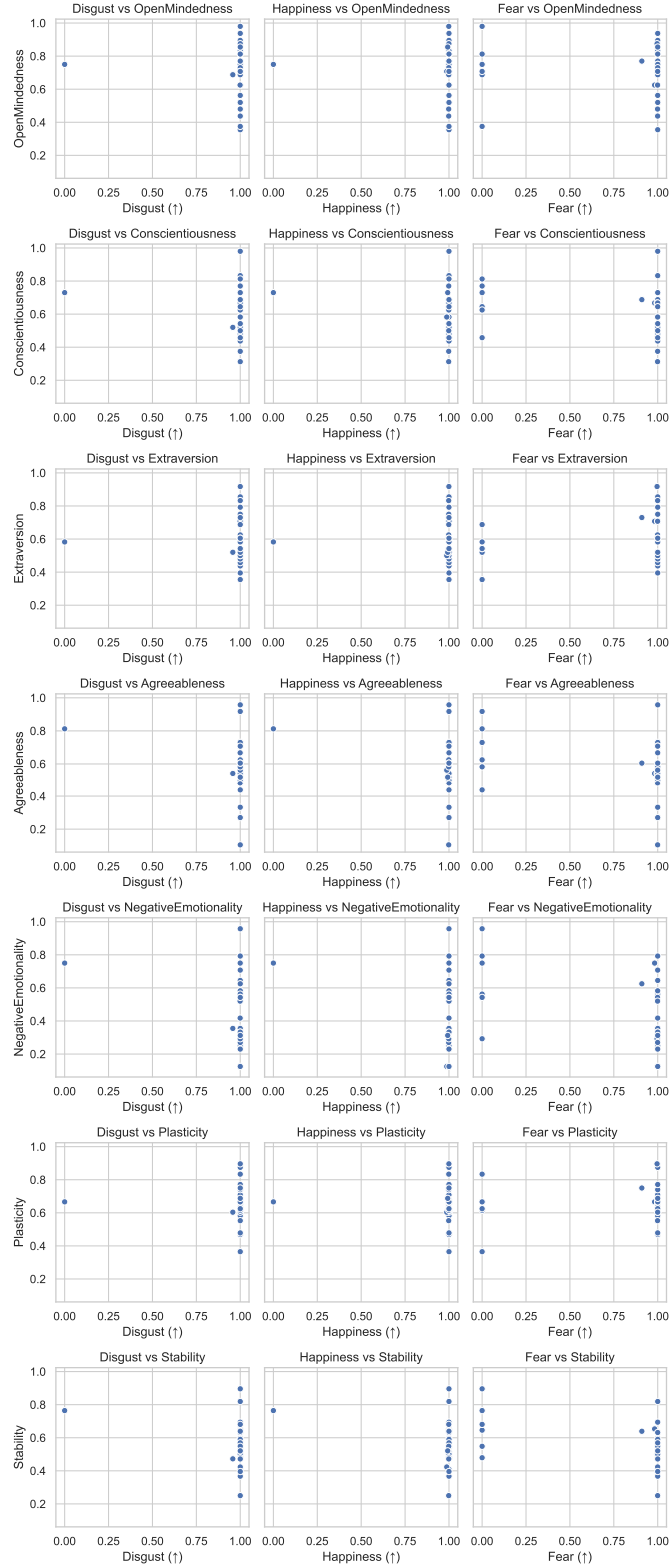


Figure 5.15: Scatter plot of the best-achieved emotion rate and OCEAN personality traits from EmoNet in the Disgust-Happiness-Fear DwE exercise. The x-axis represents the performance of the emotions, with 0 indicating an incorrect performance and 1 indicating a correct performance. No identifiable clusters are observed with the personality. The plot indicates that the students can effectively express both disgust and happiness.

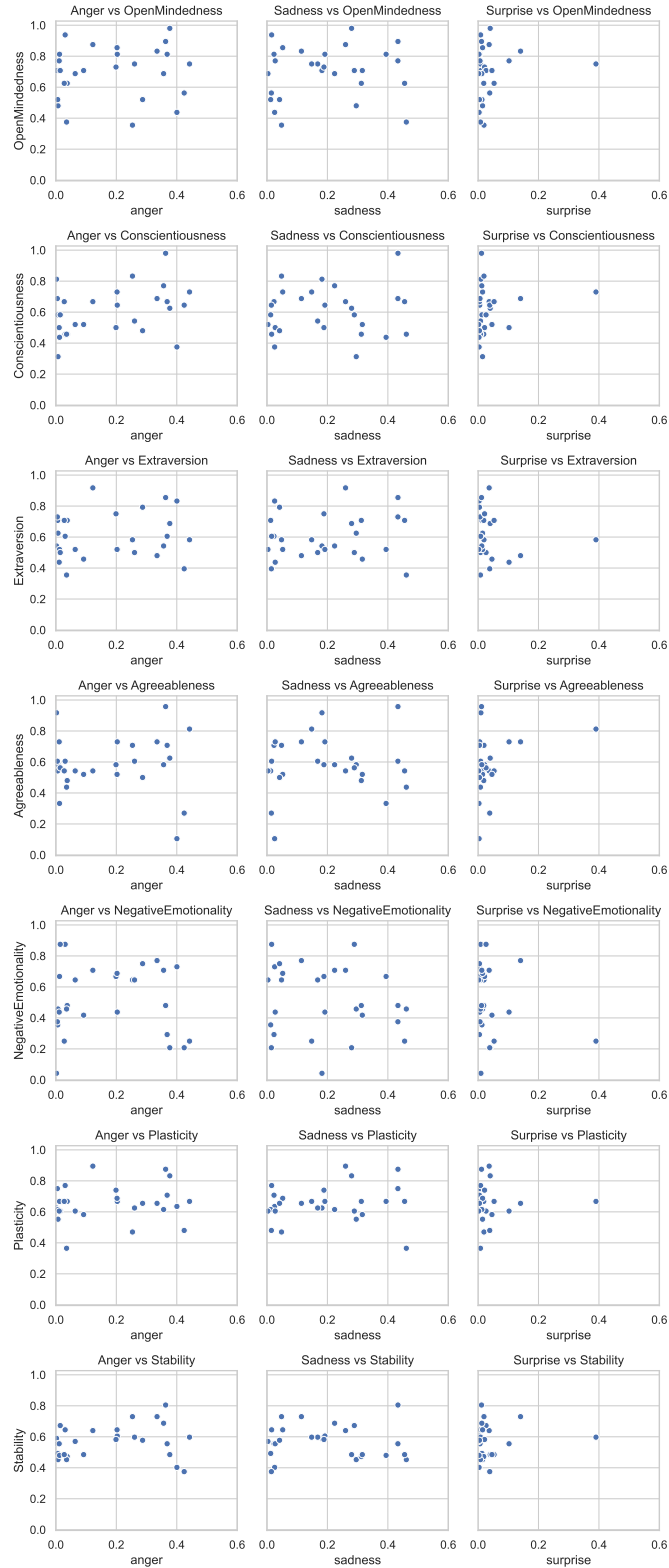


Figure 5.16: Scatter plot of the best-achieved emotion rate and OCEAN personality traits from the distances in the Anger-Sadness-Surprise DwE second exercise. The x-axis shows the distance from emotion on the triangle, where 0 indicates good performance and a value between 0.2 and 0.6 indicates poor performance. The figure shows the different performances of anger and sadness. We cannot identify clusters with the personality.

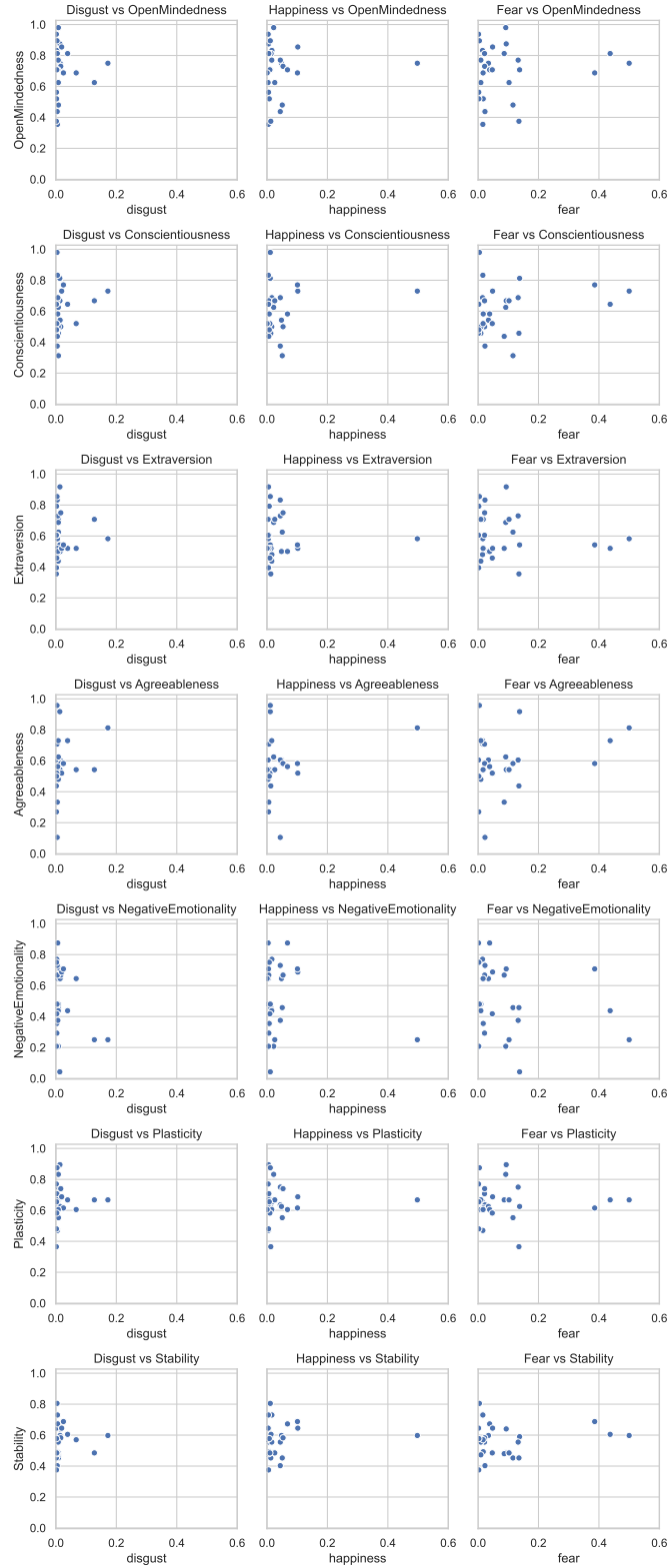


Figure 5.17: Scatter plot of the best-achieved emotion rate and OCEAN personality traits from the distances in the Disgust-Happiness-Fear Dwe second exercise. The x-axis shows the distance from emotion on the triangle, where 0 indicates good performance and a value between 0.2 and 0.6 indicates poor performance. The figure shows the different performances of fear. We cannot identify clusters with the personality.

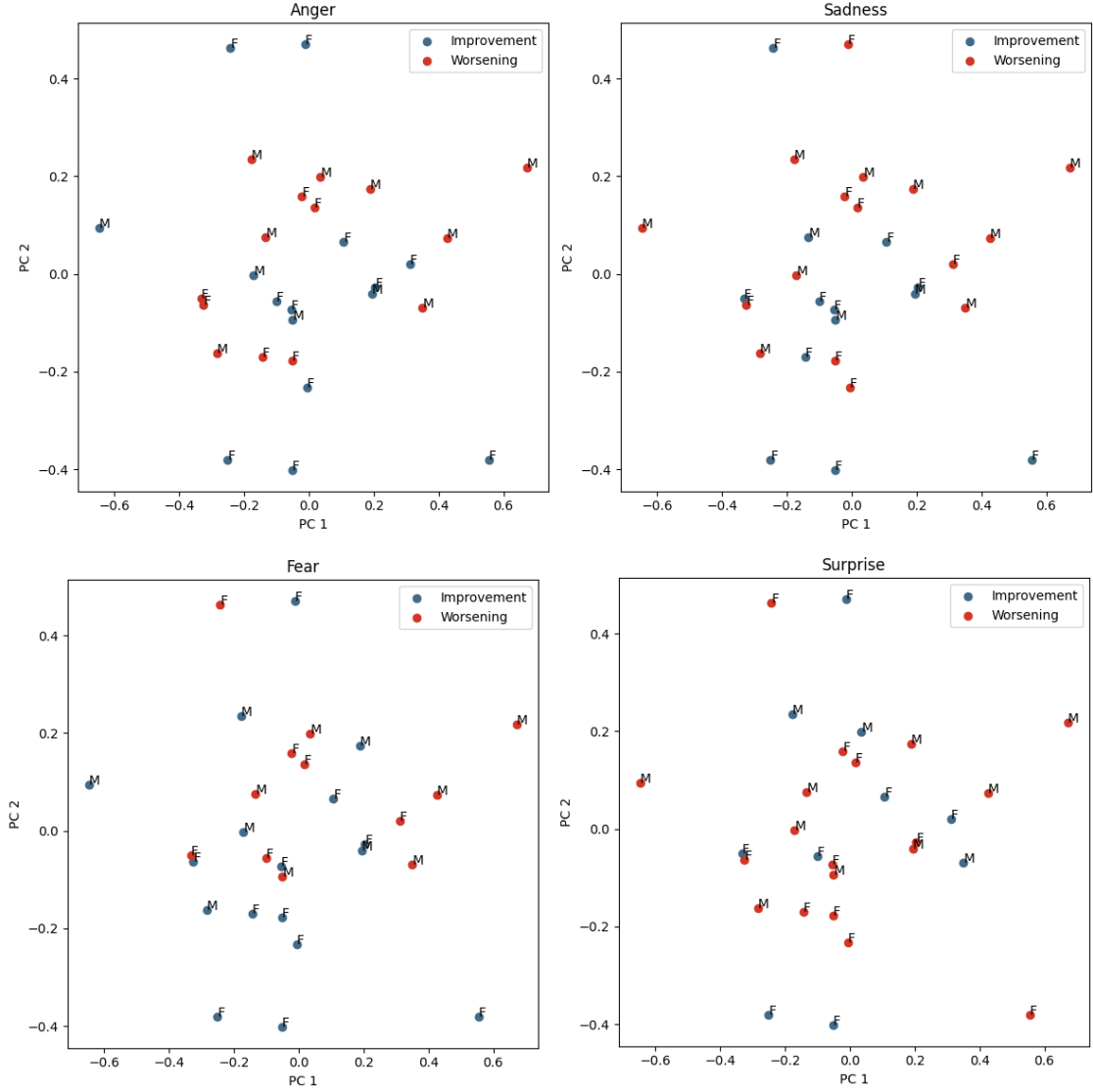


Figure 5.18: Visualization of 2-component PCA of the multidimensional personality traits colored by the 'learning score'. The blue and red dots show each person's 'learning score' based on improvement (blue) or decrease (red) in the OCEAN Personality Trait PCA, with a female/male (F/M) split. Clusters can be identified in the figure for anger and sadness. In anger, individuals whose Plasticity scores were in the low and high classes (the well-defined cloud) all improved their performance. The results of fear and surprise are mixed.

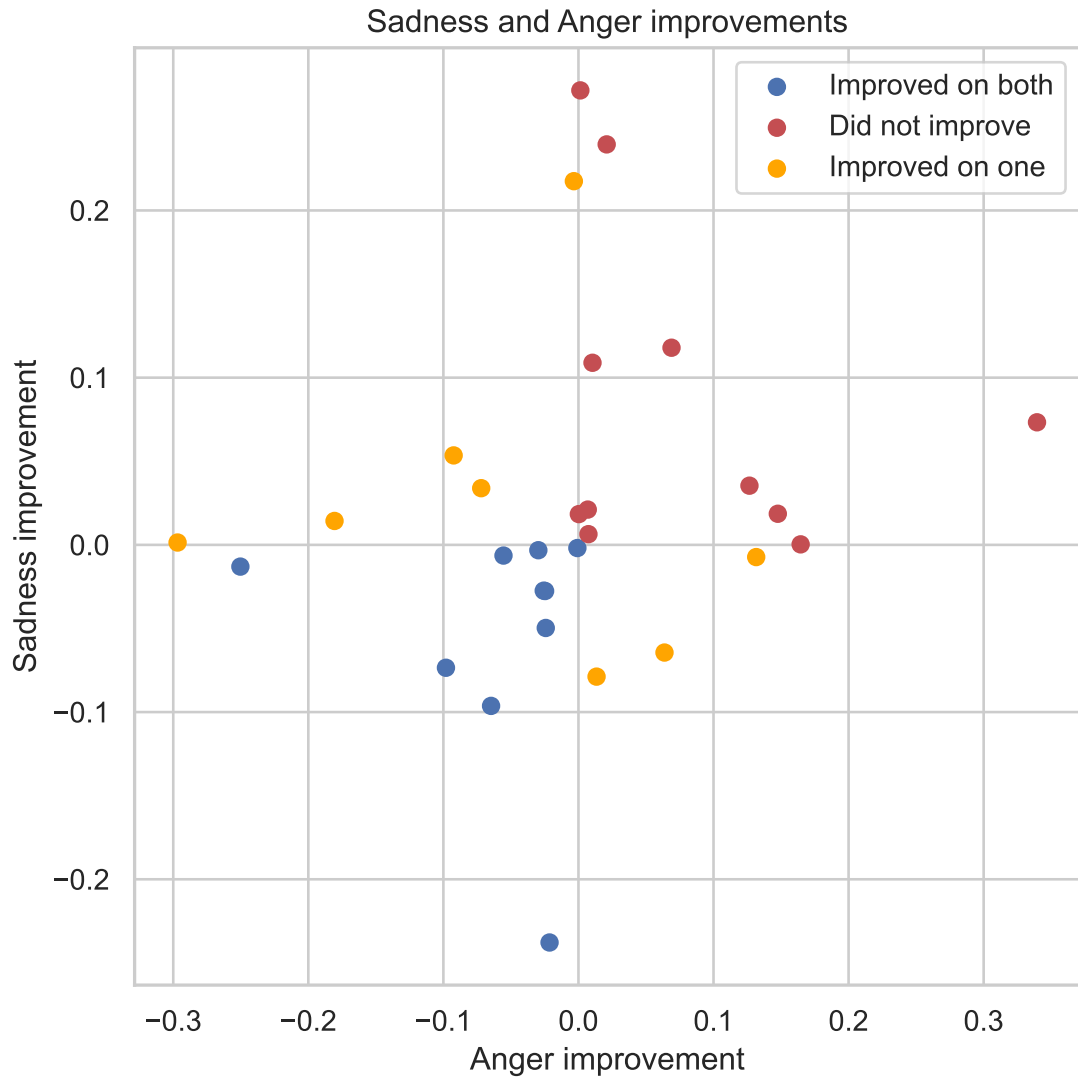


Figure 5.19: Improvement and deterioration of performance in expressing sadness and anger. The three groups are defined as follows: improved in both anger and sadness, improved in only one of them, or did not improve in either. The negative values mean they have improved because it means that their distance from the target emotion expression decreased for the second session.

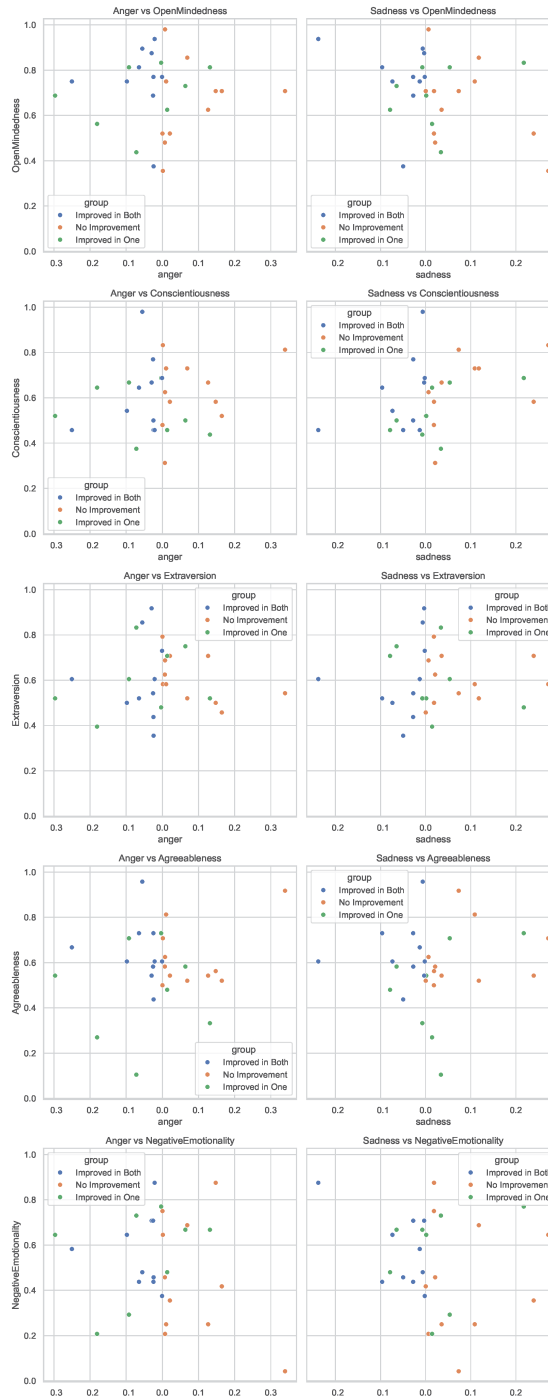


Figure 5.20: Scatter plot of 'learning score' and compared with OCEAN personality traits. The x-axis represents the 'learning score' value (negative values indicate improvement to the left of zero), while the y-axis represents the personality trait value. The color scheme distinguishes three groups: blue for those who improved in both, orange for those who didn't improve, and green for the mixed class. Based on the personality trait, we can detect only on Openmindedness that high scores are more likely related to improve.

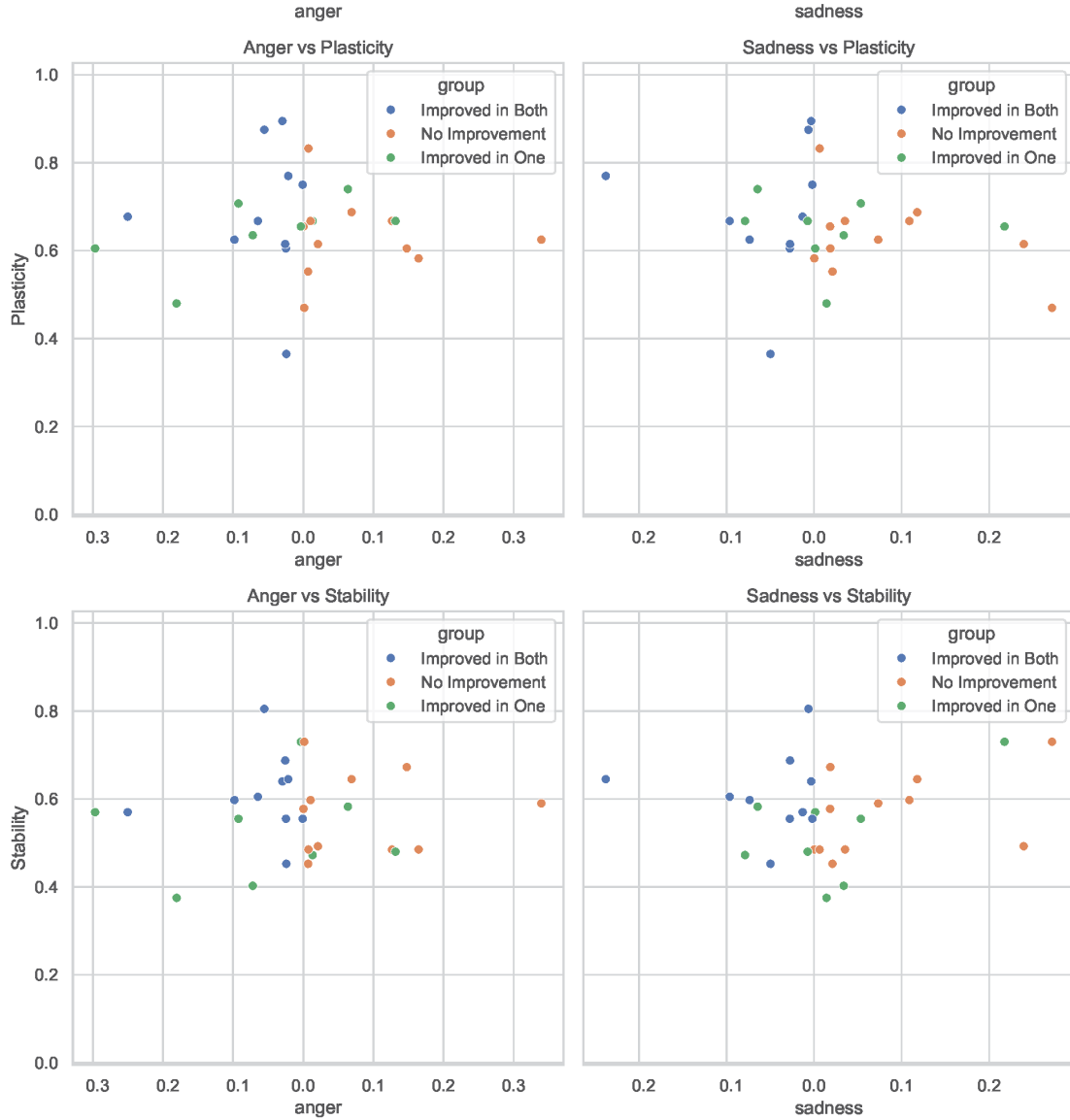


Figure 5.21: Scatter plot of 'learning score' compared with Stability and Plasticity traits. The x-axis represents the 'learning score' value (negative values indicate improvement to the left of zero), while the y-axis represents the personality trait value. The color scheme distinguishes three groups: blue for those who improved in both, orange for those who didn't improve, and green for the mixed class. Based on the personality trait, we are unable to identify any specific clusters.

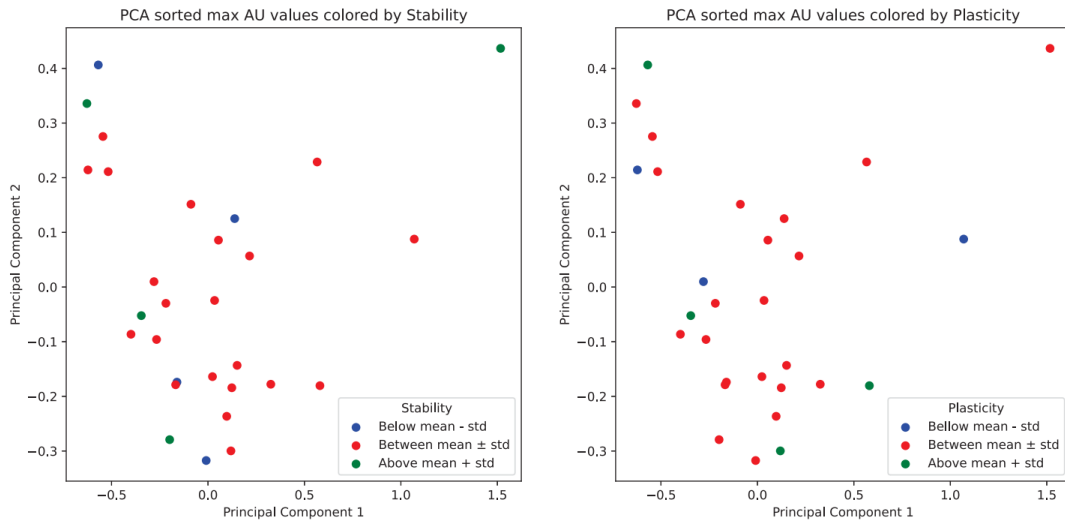


Figure 5.22: Visualization of 2-component PCA of the sorted max AU values. The two figures show the visualization of the results of the AU maximization tasks, where we colored the results by the identified Plasticity and Stability classes. Based on the personality trait, we cannot identify any specific clusters.

Chapter 6

Discussion

We have shown the 'FaceGym' capabilities and the data we can collect while a user plays with the application. With this data, we aimed to address our initial inquiries about the enhancements in users' emotional expression recreation and the correlations between the achieved results and the users' personality traits. We have reached the following conclusions using machine learning methods and different visual representations of the data.

Based on the results of the Drawing with Emotions game, we have found that many participants had difficulty expressing anger and sadness, which is consistent with our hypothesis. The gender-wise distribution (**Figure 5.4**) showed that **71%** (12 out of 17) of the girls weren't able to express anger, while **75%** of the boys (9 out of 12) and **47%** of the girls (8 out of 17) couldn't express sadness. These findings are consistent with the verbal claim by Francesca Happé, a professor of Cognitive Neuroscience at King's College London, that adolescent girls cannot express anger properly. Our goal is to bring up the opportunity to investigate this conclusion further with a larger group.

In analyzing the users' progress during the drawing exercise, we observed no significant improvement in their results. We did not anticipate a major impact from including the Action Unit maximization game, as the participants only completed the exercise once. To see a noticeable difference, these exercises should be repeated multiple times. The fact based on the **Figure 5.7**, the more problematic emotional expressions (anger and sadness) were not present for an extended period of time. However, we have found that the same 10 people struggled and the same 10 people improved on the anger and the sadness. To check if this would give us some direction how to redefine the game we used one-way ANOVA to check if this observation correlates with the OCEAN, Plasticity, and Stability. Based on the results (**Table 5.5**) we can rule out such a correlation for this sample. For

this, it still holds that for this small sample this hypothesis it can't be rejected confidently.

Most of the participants, even with the assistance that suggested which action units should be moved, indicated that we may need to redefine that part of the game's script, or we may need to use a different training method on how we can form these emotions. This may be because the users' goal was to maximize the bars, and some of them made these bar movements separately at a time. A solution could be to focus on designing a two-level training method. During the practice section of the drawing game, the system can measure the participant's level of expression of a selected emotion, like anger. In the AU maximization task, the system could give feedback to the user about the most problematic AU. The player can then select this and have a separate practice session. The system can measure the distance compared to the ratio of the ideal anger and the player's best performance. With this approach, we can train the participants to achieve an ideal (considering their individual abilities) AU curve of an emotion. Finally, the system can calculate the person's learning outcomes.

Regarding the correlation between personality traits and the results produced by the users for such a small sample (29 participants), it is hard to conclude exact results. Even though we had such a small sample size, strong correlations can still be expected, corresponding to the literature. For example, we detected a medium correlation between negative emotions and fear, a small correlation between conscientiousness, with anger and disgust, and agreeableness, with disgust.

Looking at the PCA utilized on the personality traits and the instances colored by the 'learning score' we were expecting to see that similar personalities are producing similar results. At the moment, from the available data, it is hard to say for sure that this pattern is present.

Overall, the sample size of 29 people was sufficient to test the capabilities of the 'FaceGym' system. The sample is unique because it represents the results of a ninth-grade class and their community. We have received positive feedback from the class, as they really enjoyed playing the game and participating in the experiment. In the GwE game experiment, the absence of a time limit and the repetition of tasks made it challenging to analyze the improvement between the two exercises more significantly. The presented results still show great opportunities for future work. It is worth mentioning that while designing this game, we had non-neurotypical children in mind, and the pilot testing was conducted on neurotypical adolescents.

Chapter 7

Conclusion and Future Work

In this TDK study, we introduced the pilot version of the FaceGym application. Leveraging the findings from this study, we are committed to a thorough evaluation of the second experiment protocol. This rigorous process will allow us to identify the current limitations of FaceGym confidently. Our goal is also to include head pose detection and gaze direction detection into the system and create a redesigned pipeline of the application, which would be suitable for larger traffic. Another crucial aspect of our research is integrating Gaussian Splatting methods, a promising new approach for creating the user's face 3D avatar. The model quality is photorealistic, and we can teach the model to generate and imitate different facial expressions. We aim to revise and expand the game protocol and design personalized exercises with the player's 3D avatar integration, where the avatar expresses the emotions or facial expressions that the player has to imitate. We aim to set up and run a third experiment by integrating this modification to find a sufficient sample size (number of participants in the experiment) to precisely define the system training capabilities.

By integrating this new technology into our framework, we envision a transformative potential for the FaceGym application. It could serve as a training tool for individuals grappling with physical disorders such as facial paralysis or ALS, offering a novel approach to rehabilitation. Additionally, it could aid in overcoming impairments stemming from mental health conditions like autism, providing a unique avenue for therapy and support.

Acknowledgements

We want to express our deepest gratitude to our supervisor, prof András Lőrincz, the leader of the NIPG research group, for his continuing guidance and mentorship during the one-year development.

We want to extend our sincere thanks to Kristian Fenech, research fellow of the NIPG research group, for his guidance on the statistical analysis of the data.

We are extremely grateful to Zsolt Csibi, an AI specialization master student, a founder and developer of the FaceGym application, for his indispensable work in development.

We are also thankful to the Katona József Theater for providing us with their GIF Theatre database.

We want to thank the 9.c biology-chemistry class of the Szinyei Merse Pál Gimnázium and their biology professor, Bence Együd, for making the pilot test of "FaceGym's application, for being part of the experiments and for the precious time we spent together.

We want to thank Anikó Mezősi from Argus Cognitive for her advice and insights on our project, and for providing us with resources for reference during the development.

The research was supported by the European Union project RRF-2.3.1-21-2022-00004 within the framework of the Artificial Intelligence National Laboratory Program. The authors thank Robert Bosch, Ltd. Budapest, Hungary, for their generous support to the Department of Artificial Intelligence.

Bibliography

- [1] A. J. Fridlund. ““The new ethology of human facial expressions, ” in *The Psychology of Facial Expression*”. In: eds J. A. Russell and J. Fernandez-Dols (Cambridge: Cambridge University Press), 103–129. doi: 10. (1997).
- [2] P. Ekman and W. V. Friesen. *Manual for the Facial Action Code*. Palo Alto, CA: Consulting Psychologist Press, 1978.
- [3] P. Ekman, Friesen WV, and Hager JC. *Facial Action Coding System: The Manual on CD ROM*. Salt Lake City: A Human Face, 2002.
- [4] Romina Palermo and Max Coltheart. “Photographs of facial expression: Accuracy, response times, and ratings of intensity”. In: *Behavior Research Methods, Instruments & Computers* 36.4 (2004), pp. 634–638.
- [5] Happé F. Sauter DA Panattoni C. “Children’s recognition of emotions from vocal cues”. In: 31 (2013), pp. 97–113.
- [6] Susanne A. Denham. “Emotional Competence During Childhood and Adolescence”. In: *Handbook of Emotional Development*. Ed. by Vanessa LoBue, Koralý Pérez-Edgar, and Kristin A. Buss. Cham: Springer International Publishing, 2019, pp. 493–541. ISBN: 978-3-030-17332-6. DOI: [10.1007/978-3-030-17332-6_20](https://doi.org/10.1007/978-3-030-17332-6_20).
- [7] C. Grossard et al. “Children Facial Expression Production: Influence of Age, Gender, Emotion Subtype, Elicitation Condition and Culture”. In: *Frontiers in psychology* 9.446. (2018).
- [8] C. R. Darwin. “The Expression of the Emotions in Man and Animals”. In: *1st Edn. London: John Murray*. doi: 10. (1872), pp. 1037/10001–000.
- [9] T. Bänziger, M. Mortillaro, and K. R. Scherer. “Introducing the Geneva multimodal expression corpus for experimental research on emotion perception”. In: *Emotion* 12, 1161–1179. doi: 10. (2012).

- [10] Benda MS and Scherf KS. “The Complex Emotion Expression Database: A validated stimulus set of trained actors”. In: *PLoS One*. 2020 Feb 3;15(2):e0228248. doi: 10. (2020).
- [11] X. Zhang et al. “(2014)”. In: *Bp4d-spontaneous: a high-resolution spontaneous 3d dynamic facial expression database*. *Image Vis. Comput* 32 ().
- [12] Rosalind W. Picard. *Affective Computing*. Cambridge, MA: MIT Press, 1997. ISBN: 978-0-262-16170-1.
- [13] Bernard Marr. *What Is Artificial Emotional Intelligence?*
- [14] Shao Liu and Sos S. Agaian. “3DEmo: For Portrait Emotion Recognition with New Dataset”. In: 17.2 (2024). ISSN: 1556-4673. DOI: [10.1145/3631133](https://doi.org/10.1145/3631133). URL: <https://doi.org/10.1145/3631133>.
- [15] Olufisayo Ekundayo and Serestina Viriri. “Facial Expression Recognition: A Review of Trends and Techniques”. In: *IEEE Access* PP (Sept. 2021), pp. 1–1. DOI: [10.1109/ACCESS.2021.3113464](https://doi.org/10.1109/ACCESS.2021.3113464).
- [16] Eva Krumbhuber et al. “The role of facial movements in emotion recognition”. In: *Nature Reviews Psychology* 2 (Apr. 2023). DOI: [10.1038/s44159-023-00172-1](https://doi.org/10.1038/s44159-023-00172-1).
- [17] R. Zhi, M. Liu, and D. Zhang. *A comprehensive survey on automatic facial action unit analysis*. 2020.
- [18] Tadas Baltrušaitis, Peter Robinson, and Louis-Philippe Morency. “OpenFace: An open source facial behavior analysis toolkit”. In: *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*. 2016, pp. 1–10. DOI: [10.1109/WACV.2016.7477553](https://doi.org/10.1109/WACV.2016.7477553).
- [19] Cheng Luo et al. “Learning Multi-dimensional Edge Feature-based AU Relation Graph for Facial Action Unit Recognition”. In: *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence*. IJCAI-2022. International Joint Conferences on Artificial Intelligence Organization, July 2022. DOI: [10.24963/ijcai.2022/173](https://doi.org/10.24963/ijcai.2022/173). URL: <http://dx.doi.org/10.24963/ijcai.2022/173>.
- [20] Tomas Gajarsky. *facetorch*. URL: <https://github.com/tomas-gajarsky/facetorch>.
- [21] *pyfeat*. URL: <https://py-feat.org/pages/intro.html>.

- [22] A. Toisoul, J. Kossaifi, A. Bulat, et al. *Estimation of continuous valence and arousal levels from faces in naturalistic conditions*. 2021.
- [23] M. Lüken et al. “(MEXCA - A simple and robust pipeline for capturing emotion expressions in faces, vocalization, and speech.” In: (2024).
- [24] McClure E. B. “A meta-analytic review of sex differences in facial expression processing and their development in infants, children, and adolescents”. In: *Psychological bulletin* 126.3 (2000).
- [25] Baron-Cohen S. “The extreme male brain theory of autism”. In: *Trends in cognitive sciences* 6.6 (2002).
- [26] Chaplin TM and Aldao A. “Gender differences in emotion expression in children: a meta-analytic review”. In: *Psychol Bull.* 2013 Jul;139(4):735-65. doi: 10. (2012).
- [27] M. Sonnby-Borgström, P. Jönsson, and O. Svensson. “Gender differences in facial imitation and verbally reported emotional contagion from spontaneous to emotionally regulated processing levels”. In: *Scandinavian journal of psychology* 49.2 (2008).
- [28] Hanna Drimalla, Isabelle Baskow, Behnoush Behnia, et al. “Imitation and recognition of facial emotions in autism: a computer vision approach”. In: *Molecular Autism* 12.1 (2021), p. 27. DOI: [10.1186/s13229-021-00430-0](https://doi.org/10.1186/s13229-021-00430-0). URL: <https://doi.org/10.1186/s13229-021-00430-0>.
- [29] Dacher Keltner. “Facial Expression, Personality, and Psychopathology”. In: *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)*. Ed. by Paul Ekman and Erika L. Rosenberg. Oxford University Press, Apr. 2005, pp. 548–550. ISBN: 9780195179644. DOI: [10.1093/acprof:oso/9780195179644.001.0001](https://doi.org/10.1093/acprof:oso/9780195179644.001.0001). URL: <https://doi.org/10.1093/acprof:oso/9780195179644.001.0001>.
- [30] O. P. John and S. Srivastava. “The Big-Five trait taxonomy: History, measurement, and theoretical perspectives”. In: *Handbook of personality: Theory and research* (Vol. Ed. by L. A. Pervin and O. P. John. Vol. 2. New York: Guilford Press, 1999, pp. 102–138.

- [31] Lucía Teijeiro-Mosquera et al. “What Your Face Vlogs About: Expressions of Emotion and Big-Five Traits Impressions in YouTube”. In: *IEEE Transactions on Affective Computing* 6.2 (2015), pp. 193–205. DOI: [10.1109/TAFFC.2014.2370044](https://doi.org/10.1109/TAFFC.2014.2370044).
- [32] Ádám Fodor. *exordium*. 2023.10.09. URL: <https://github.com/fodorad/exordium>.
- [33] A. "Toisoul et al. “Estimation of continuous valence and arousal levels from faces in naturalistic conditions””. In: *Nature Machine Intelligence* 3 (2021), pp. 42–45.
- [34] Jiankang Deng et al. *RetinaFace: Single-stage Dense Face Localisation in the Wild*. 2019. arXiv: [1905.00641](https://arxiv.org/abs/1905.00641) [cs.CV].
- [35] Adam Paszke et al. “PyTorch: An Imperative Style, High-Performance Deep Learning Library”. In: *Advances in Neural Information Processing Systems* 32. Curran Associates, Inc., 2019, pp. 8024–8035. URL: <http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf>.
- [36] *React*. URL: <https://react.dev/>.
- [37] Google. *mediapipe*. URL: <https://github.com/google/mediapipe>.
- [38] Carl-Herman Hjortsjö. *Man’s Face and Mimic Language*. Studentlitteratur, 1969, pp. 1–111.
- [39] Uldis Zarins. *Anatomy of Facial Expression*. Anatomy Next, Incorporated, 2019, p. 103.
- [40] Bryn Farnsworth. *Facial Action Coding System (FACS) – A Visual Guidebook*. URL: <https://imotions.com/blog/learning/research-fundamentals/facial-action-coding-system/>.
- [41] Robert R. McCrae and Oliver P. John. “An introduction to the five-factor model and its applications.” In: *Journal of Personality* 60.2 ().
- [42] S. Srivastava et al. “Development of personality in early and middle adulthood: Set like plaster or persistent change?” In: *Journal of Personality and Social Psychology* 84 (2003), pp. 1041–105.

- [43] C. J. Soto and O. P. John. “The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power”. In: *Journal of Personality and Social Psychology* 113 (2017), pp. 117–143.
- [44] P. Rózsa, Zs. Tárnok, and P. Nagy. *A gyermekpszichiátriában alkalmazott kérdőívek, interjúk és tünetbecslő skálák*. Budapest: Állami Egészségügyi Ellátó Központ, 2020, pp. 56–60.
- [45] K. Lee and M. C. Ashton. “Psychometric properties of the HEXACO-100”. In: *Assessment* 25 (2018), pp. 543–556.
- [46] K. A. Lee, G. Hicks, and G. Nino-Murcia. “Validity and reliability of a scale to assess fatigue”. In: *Psychiatry Research* 36 (1991).
- [47] U. Zarins. *Anatomy of Facial Expression*. Anatomy for sculptors. Anatomy Next, Incorporated, 2019. ISBN: 9781735039046. URL: <https://books.google.hu/books?id=8UV5zQEACAAJ>.
- [48] Junsik Kim et al. “Numerical Approach to Facial Palsy Using a Novel Registration Method with 3D Facial Landmark”. In: *Sensors* 22 (Sept. 2022), p. 6636. DOI: [10.3390/s22176636](https://doi.org/10.3390/s22176636).
- [49] Romain Tavenard et al. “Tslearn, A Machine Learning Toolkit for Time Series Data”. In: *Journal of Machine Learning Research* 21.118 (2020), pp. 1–6. URL: <http://jmlr.org/papers/v21/20-091.html>.
- [50] *MySQL*. URL: <https://www.mysql.com/>.
- [51] *MongoDB*. URL: <https://www.mongodb.com/>.
- [52] *InfluxDB*. URL: <https://www.influxdata.com/>.

List of Figures

3.1	Happines - Smile: Combination of AU6 and AU12 illustrated by one of our participants. She took part in the pilot testing and provided a written agreement to use her images captured by FaceGym for research purposes [39]	15
3.2	The Big Five Personality Traits. Reproduced from Wikipedia under Creative Commons Attribution-Share Alike 4.0 International license. . . .	16
3.3	The action unit maximization task for happiness. The task here was to move the bars by moving some regions of the face. Each region is corresponding to an Action Unit and the bars show how much a given action unit is actuated.	20
3.4	The action unit maximization task for disgust. The task here was to move the bars by moving some regions of the face. Each region is corresponding to an Action Unit and the bars show how much a given action unit is actuated.	22
3.5	Sample description that was given for the user. It shows the name of the exercise such as smile or disgust, and names the highlighted action unit bars. This description is shown at the beginning of the task and it can be opened by clicking the info button.	22
3.6	The "drawing with emotions" task. Here the user had to draw a trajectory by reaching pre-defined target emotional expression endpoints and moving the cursor between them. The cursor could be controlled by creating emotional expressions.	25
3.7	The imitation task. The figure shows FaceGym's imitation exercise where the user had to recreate a small scene, compare the results with the reference video and reach a minimal DTW Dist.	27
3.8	The results of the DTW for the original problem and the relaxation when the time series are similar. The orange line shows the original, while the blue line shows the sub-sequence frame matching.	29

3.9	The results of the DTW for the original problem and the relaxation, when the time series are very different. The orange line shows the original, while the blue line shows the sub-sequence frame matching.	29
3.10	The architecture of the software. It depicts how we manage the incoming images, how each AI module is integrated, and what we are sending back to the user as feedback.	32
3.11	The diagram of the data. The periods table contains data related to what exercise the user was doing. The detections table contains the values received from the AU and emotional expression detection model for each recorded frame. The two tables share the name field which stores the users' name.	33
4.1	An illustration of our setup for the experiment. We tried to separate the participants as much as possible so they didn't distract each other. Participants facing each other were separated by a movable wall.	35
4.2	Second experiment setup. The blue boxes represent the questionnaires the participants filled out during the experiment. For the experiment, we designed the FaceGym for the Drawing with Emotions -AU maximization - Drawing with Emotions - Imitation game setup.	36
5.1	Personality trait class. The histograms display the distribution of students' personality traits, including the mean and standard deviation. We have categorized each trait into three classes, with thresholds determined using trait-specific mean and standard deviation calculations.	43
5.2	Visualization of 2-component PCA of the multidimensional Personality traits. In the left figure, we visualized the PCA using Plasticity values. In the right figure, we visualized the PCA using Stability values. Blue represents the low trait value class, red represents the high trait value class, and orange represents the middle trait value class, which includes the majority of the students. Our sample size is 29, and it is difficult to identify a well-defined cluster in the low and high trait classes.	44
5.3	Visualization of 2-component PCA of the multidimensional personality traits colored by the energy level results. The figure shows that based on the personality traits and the energy level we cannot define any cluster. . .	45

5.4	The reached distances for the first and second drawing game. The blue line shows the distances from the goal before the AU task and the dashed orange line shows the distances after the AU task. The y-axis shows the best minimal distance from the target emotion endpoint in the drawing exercise. The participants are sorted based on their results from the first session.	53
5.5	The correlation between the distances reached by the participants. We see that for surprise, disgust, happiness, and fear the results are positively correlated meaning that they produced similarly good results for these tasks. As for anger and sadness, there are significantly smaller correlations, because as we have shown these tasks were harder and there are larger variances in the results compared to the other tasks.	54
5.6	The extent of emotion for a longer period based on EmoNet. The blue and orange lines show the extent of the emotion according to the model before and after the AU task during the drawing exercise.	55
5.7	The extent of emotion for a longer period based on EmoNet. The blue and red lines show the extent of the emotion according to the model before and during the Action Unit bar maximization exercise.	56
5.8	Reached distances for each emotion shown on a histogram. The blue and orange lines show the distances before and after the AU task during the drawing exercise.	57
5.9	Reached extent for each emotion. The blue and orange lines show the extent of the emotion according to the model before and after the AU task during the drawing exercise.	58
5.10	Pearson correlation between the reached distance from the emotion target and the OCEAN Personality traits in the DwE game. Blue shades indicate a positive correlation, red for negative. We observe a medium correlation between anger and Conscientiousness, a small correlation between fear and surprise with Agreeableness, and a small negative correlation between disgust and Emotional Stability.	59

5.11	Spearman correlation between the reached distance from the emotion target and the OCEAN Personality traits in the DwE game. Blue shades indicate a positive correlation, red for negative. We observe a small correlation between anger, surprise, disgust and Conscientiousness, a small correlation between disgust and Agreeableness, and a small negative correlation between fear and Emotional Stability.	60
5.12	Pearson correlation between the emotions detected by EmoNet and the OCEAN Personality traits in the DwE game. Blue shades indicate positive correlation, red for negative. We observe a medium negative correlation between anger and Conscientiousness and a small negative correlation between fear and Conscientiousness. Surprise and fear have a small negative correlation with Agreeableness. Surprise also has a small positive correlation with Extraversion. Fear has a medium negative correlation with Emotional Stability and Stability. Disgust and happiness have also a small negative correlation with Stability.	61
5.13	Spearman correlation between the emotions detected by EmoNet and the OCEAN Personality traits in DwE game. Blue shades indicate positive correlation, red for negative. We observe a medium negative correlation between fear and Emotional stability. We have a small negative correlation between anger and Conscientiousness, a small negative correlation between happiness and Extraversion, and sadness and happiness with Plasticity. Fear also has a small negative correlation with Stability.	62
5.14	Scatter plot of the best-achieved emotion rate and OCEAN personality traits from EmoNet in the Anger-Sadness-Surprise DwE exercise. The x-axis represents the performance of the emotions, with 0 indicating an incorrect performance and 1 indicating a correct performance. No identifiable clusters are observed with the personality. The plot indicates that the students struggled to express anger and sadness effectively.	63
5.15	Scatter plot of the best-achieved emotion rate and OCEAN personality traits from EmoNet in the Disgust-Happiness-Fear DwE exercise. The x-axis represents the performance of the emotions, with 0 indicating an incorrect performance and 1 indicating a correct performance. No identifiable clusters are observed with the personality. The plot indicates that the students can effectively express both disgust and happiness.	64

5.16	Scatter plot of the best-achieved emotion rate and OCEAN personality traits from the distances in the Anger-Sadness-Surprise DwE second exercise. The x-axis shows the distance from emotion on the triangle, where 0 indicates good performance and a value between 0.2 and 0.6 indicates poor performance. The figure shows the different performances of anger and sadness. We cannot identify clusters with the personality.	65
5.17	Scatter plot of the best-achieved emotion rate and OCEAN personality traits from the distances in the Disgust-Happiness-Fear DwE second exercise. The x-axis shows the distance from emotion on the triangle, where 0 indicates good performance and a value between 0.2 and 0.6 indicates poor performance. The figure shows the different performances of fear. We cannot identify clusters with the personality.	66
5.18	Visualization of 2-component PCA of the multidimensional personality traits colored by the 'learning score'. The blue and red dots show each person's 'learning score' based on improvement (blue) or decrease (red) in the OCEAN Personality Trait PCA, with a female/male (F/M) split. Clusters can be identified in the figure for anger and sadness. In anger, individuals whose Plasticity scores were in the low and high classes (the well-defined cloud) all improved their performance. The results of fear and surprise are mixed.	67
5.19	Improvement and deterioration of performance in expressing sadness and anger. The three groups are defined as follows: improved in both anger and sadness, improved in only one of them, or did not improve in either. The negative values mean they have improved because it means that their distance from the target emotion expression decreased for the second session.	68
5.20	Scatter plot of 'learning score' and compared with OCEAN personality traits. The x-axis represents the 'learning score' value (negative values indicate improvement to the left of zero), while the y-axis represents the personality trait value. The color scheme distinguishes three groups: blue for those who improved in both, orange for those who didn't improve, and green for the mixed class. Based on the personality trait, we can detect only on Openmindeness that high scores are more likely related to improve.	69

5.21 Scatter plot of 'learning score' compared with Stability and Plasticity traits. The x-axis represents the 'learning score' value (negative values indicate improvement to the left of zero), while the y-axis represents the personality trait value. The color scheme distinguishes three groups: blue for those who improved in both, orange for those who didn't improve, and green for the mixed class. Based on the personality trait, we are unable to identify any specific clusters.	70
5.22 Visualization of 2-component PCA of the sorted max AU values. The two figures show the visualization of the results of the AU maximization tasks, where we colored the results by the identified Plasticity and Stability classes. Based on the personality trait, we cannot identify any specific clusters.	71

List of Tables

3.1	The emotional expression and their corresponding primary and secondary action units. These primary and secondary action units were defined based on [39].	21
3.2	The muscles and the corresponding region defined by the landmarks. . . .	24
3.3	The emotional expression and their corresponding angles if we define the directions as polar coordinates. The angles are defined clockwise (surprise, happiness, sadness, disgust, anger, fear).	26
5.1	Summary of the personality values. The table shows the statistics of the class's personality test (BFI-2 and Hexaco). We calculated the value range, mean, STD, and the difference between the OCEAN personality traits, Stability, Plasticity, and Honesty-Humility traits. We used these results to split the students into three classes on all personality traits.	42
5.2	Summary of the energy values of the experiment. The table shows that 55% of the students felt refreshed, and 45% felt depleted after finishing the task.	44
5.3	Selected students for the case study. The students are selected based on their extent of emotions. For those who could not reach 1.0 within a 2% margin, we considered that they did not create that emotional expression for a longer period.	48
5.4	Genderwise results of the extent of anger and sadness. Based on Figure 5.6 and Figure 5.7 a summary on who could and couldn't create anger and sadness within 2% margin split by gender	48
5.5	The results of ANOVA on the OCEAN + Plasticity + Stability personality traits where the groups are defined by the improvements (Improved in both, Improved in only one, and No improvement at all). From these results, only Angreeableness's p-value is close to the reject threshold. . . .	51