Development of a Wearable Vision Substitution Prototype and Determination of a Suitable Sensory Feedback Method

Anna Kushnir
Socio-Informatics and Societal Aspects of Digitalization
Faculty of Computer Science and Business Information Systems
University of Applied Sciences Würzburg-Schweinfurt
Würzburg, Germany
e-mail: info@anna-kushnir.de

Nicholas H. Müller
Socio-Informatics and Societal Aspects of Digitalization
Faculty of Computer Science and Business Information Systems
University of Applied Sciences Würzburg-Schweinfurt
Würzburg, Germany
e-mail: nicholas.mueller@fhws.de

Abstract—**This paper introduces a Sensory Substitution Device (SSD) prototype, which aims to support conversation situations. The purpose of the SSD is to convert facial expressions into emotions based on the seven basic emotions according to Paul Ekman and the Facial Action Coding System. The paper describes a study by which vibration and temperature stimuli have been compared and a suitable feedback method for the SSD prototype was selected. As a result of the study, vibration stimuli were selected for the feedback of the SSD, as these were recognized more reliably, more quickly and were perceived as less distracting on average by most of the subjects. The SSD prototype was further developed on the basis of the results. The work in progress of the SSD prototype design is described in detail.**

Keywords-Non-verbal communication; Basic Emotions; Vibrotactile Interface; Visually impaired; Vision Substitution; Sensory Substitution Device; FACS.

I. INTRODUCTION

The following paper is based on the work in progress paper about the development of a wearable vision substitution prototype, presented at ACHI 2020 [1]. Everyday face-to-face communication situations use a variety of communication channels. In addition to the verbalized information, several non-verbal cues are communicated, which have to be interpreted by the communication partners. These include, e.g., facial expressions, intonations, and gestures. Sighted people can use all these communication channels, which allows them, e.g., to interpret facial expressions. According to studies, in this way, they are able to read emotions to understand their interlocutor better. Since our emotions are involuntarily reflected in our facial expressions, they reveal valuable information [2].

Blind and visually impaired people are limited in the interpretation of a communication situation as they are not able to process the visual non-verbal information. For this reason, the communication situation can only be analyzed incompletely. Although an emotional valence can be determined through the interlocutor's intonation, this is only possible when the other person is speaking. While a visual impaired person speaks, he or she is not able to determine the emotional valence, since the interlocutor is listening and, therefore, only non-verbal cues are transmitted.

A survey carried out with focus groups of blind people and disability experts proves that there are several key needs of non-verbal information that blind people may need to access during social encounters [3]. These include, but are not limited to, the facial expressions of a person standing in front of the user. Based on this demand, the purpose of this work is to design an interface prototype that assists people with visual disabilities in everyday conversations. The proposed system is based on vision substitution, and thus, it can be classified under Sensory Substitution Devices. The goal of the SSD is to communicate the facial expressions of the conversation partner to the blind user in a conversation situation by converting the visual information into tactile stimuli.

The paper is structured as follows. In Section II, the theoretical basics for sensory substitution and related work are presented. Section III describes the relationship between emotions and the Facial Action Coding System (FACS), which is used for the SSD prototype emotion recognition. Section IV presents the procedure of choosing a suitable feedback method for the SSD. In Section V, the prototype of the SSD and important design decisions are described. Finally, Section VI summarizes the paper and describes the next steps.

II. SENSORY SUBSTITUTION AND RELATED WORK

The term sensory substitution was introduced at the end of the sixties and describes the transformation of sensory stimuli from one sensory modality into another, in order to substitute a sense that is missing or impaired by illness or disability [4]. At that time, Paul Bach-y-Rita and his colleagues developed the first SSD, which was dedicated to the scientific investigation of the plasticity of the brain in congenitally blind people. This SSD converts image information into pressure stimuli on the skin and enables blind people to find their way around the room. For this purpose, the image from the video camera is displayed on the skin using a vibrotactile belt that is worn on the stomach.

Nowadays, sensory substitution devices for vision are largely divided into tactile and auditory substitution systems.

These consist of the sensor, in this case, a video camera, and a human-machine interface (HMI), which consists of a coupling system and a stimulator [5]. The coupling system receives the stimuli recorded by the sensor, interprets them, and forwards them to the stimulator. In this way, the recorded information is analyzed, converted, and sent to the user through either the tactile sense or the hearing.

In the following, examples of vision substitution are presented, and the research needs are derived.

Lykawka et al., for instance, presented a tactile interface that allows users to navigate in environments including obstacles and to detect the movements of people and objects. The system converts the visual information into tactile feedback and conveys it with the help of a vibrotactile belt [6].

Bernieri et al. dealt with visually impaired people's mobility. The authors describe a prototype of a smart glove that complements the classic cane. The proposed glove provides vibrotactile feedback on the position of the next obstacle in the range [7].

In [8], a text reading system called FingerEye is proposed, which translates text into audio information or braille.

Bhat et al. also presented a system that aids reading texts. Additionally, it assists in recognizing objects. Both stimuli are translated into audio output [9].

The interaction assistant ICare, described in [10], deals with choosing an appropriate face recognition algorithm to build an assistant for social interactions. It also describes the prototype, of which the feedback method is also aural.

While a lot of research has been done to meet a wide range of needs of people with visual disabilities, less attention has been given to the development of assistive devices that satisfy the need for access to non-verbal communication in social interactions. However, there are a few systems that deal with social interaction, among other functions, and are available for purchase.

Orcam MyEye 2.0 [11] is a wearable device, which is worn on the temple stem of eyeglasses and it combines several features. With the help of a camera on the front and a loudspeaker on the back, the device can read texts, recognize the time, identify goods by their barcodes and recognize people, by storing and voicing their name out loud. All recognitions are translated into audio information and conveyed through a loudspeaker.

Microsoft SeeingAI [12] is an application for the Smartphone, which shares a lot of features with the Orcam MyEye. Moreover, it offers a feature, which recognizes and describes scenes, people, and people's emotions. All types of recognition are translated and represented by audio feedback. To enable the recognition and translation, a photo of the object to be analyzed has to be taken.

An important shortcoming of SeeingAI and Orcam MyEye is that the solutions provide only aural outputs. People with visual impairment rely on their hearing to perceive their environment. Therefore, aural signals that are played by an assistance system during social interactions, such as face-to-face communication, could be perceived as disturbing as they may interfere with the hearing of the speech of the communication partner or the own speech. Moreover, SeeingAI is able to recognize people and emotions, but not to communicate these in real-time. Instead of this, the user has to take a photo first. Orcam and SeeingAI thus are not sufficient solutions to support face-to-face communication.

What is needed is a system that communicates non-verbal cues in real-time and whose output is based on a different sense than the hearing, on which verbal communication is perceived.

A common alternative to audio-vision substitution is to use vibrotactile feedback, which was already used for a haptic belt in the described work about navigation [6]. McDaniel et al. also presented a haptic belt to assist in communication situations [3]. The focus of the work is on communicating non-verbal cues, like the number of people in the visual field, the relative direction and distance of the individuals with respect to the user. The output of the belt is created and delivered to the user continuously and in real-time through the haptic belt with vibrotactile feedback. Experiments have shown that non-verbal communication can be successfully conveyed through vibrotactile cues.

In addition to the vibrotactile feedback, it is also conceivable to use other tactile feedback methods for the SSD to be designed. Temperature feedback would be one such possibility. However, this has not been used in SSDs yet, but for direct heating or cooling of the human body [13], [14].

Since temperature feedback has not yet been used in SSDs, it should be examined which of the tactile feedback methods described is best suited to convey emotions during a conversation. Building on this, the SSD prototype presented in [1] can then be adapted and expanded.

III. Emotions and the Facial Action Coding System

Every emotion sends signals, which are most visible through our voice and facial traits. According to Paul Ekman, there are so-called basic emotions that are understood by all cultures in the world, since they can be recognized through universal facial expressions. The seven basic emotions include the emotion groups anger, happiness, sadness, disgust, contempt, fear, and surprise. Emotion groups mean that there can be different forms and intensity levels of emotions in a group [2], [15].

Although the basic emotions can be separated well in terms of their nature and their characteristic facial expression, mixed emotions occur more often than pure emotions [16]. Some examples of mixed emotions are listed below:

- Rejection of a loved one can evoke both sadness and anger.
- When we feel threatened, we are afraid, but often we are also angry. For example, we could be angry with ourselves for feeling fear if we found out afterwards that it was completely unfounded.
- Anger can come with contempt. If we find that we have reacted angrily, we can despise ourselves for it.

- It often happens that the emotion of contempt alternates with the emotion disgust. For example, if a person does something disgusting, we may find it disgusting and, therefore, despise the person.

Moreover, emotions should not be confused with moods, because unlike emotions, they can last up to two days [2]. However, emotions come and go within minutes or seconds. For example, the emotion surprise lasts at most a few seconds. After that, it ends in fear, happiness, anger, or other emotions, depending on the quality and nature of what surprises us [2]. For the development of the prototype, this means that it is important, among other things, that the emotions can be recognized quickly by the user as these can change quickly.

The basis for the analysis of facial expressions and the emotion recognition in this project is the FACS, an anatomically based coding system for all visually perceptible facial muscle movements and also a common standard for the recognition of basic emotions and their intensity. The system assigns Action Units (AUs) to almost every visible movement of facial muscles. AUs can be divided into two categories. One part of them refers to the upper face and one part to the lower face. An Action Unit can combine single or multiple muscle movements. A combination of certain AUs can be assigned to emotions. Besides, the intensity of the muscle movements is differentiated based on a five-tier ranking. In order to determine the emotions of a facial expression using FACS, it has to be deconstructed in AUs [17], [18]. For example, if a face has the combination of the AUs "Inner brow raiser", "Outer brow raiser", "Upper Lid Raiser" and "Jaw drop", this facial expression would indicate the emotion of surprise [2], [18].

IV. Choosing a feedback Method

As previously reported in [19] this section describes the process of deciding on a suitable feedback method for the stimulator of the proposed SSD. The goal of this project is to create an SSD that conveys the emotions of the interlocutor in real-time, meaning that the stimulator feedback should be conveyed during a conversation. The study described below focuses on tactile feedback methods since these are received through a different sense than the hearing and, therefore, would not interfere with verbal communication.

A. Approach

In order to investigate which tactile stimuli are best suited to convey emotions during a communication situation, a quantitative study was carried out in which the tactile stimuli heat, cold, and vibration were compared. Data were collected through a laboratory experiment in which test subjects received various tactile stimuli. During the experiment, detection rates of the stimuli and the reaction times to them were measured. The experiment's within-subject design ensured that each test subject could test all three stimuli types. After the experiment demographic data, information on the perceived distraction by the stimuli as well as other dependent factors was collected using a questionnaire.

B. Experimental Setup

The test subject’s task in the experiment was to read a text aloud and to respond to heat, cold, and vibration stimuli while reading. Depending on the stimulus type sent, the reading volume should be adjusted in the following way: The vibration stimulus calls to read on in a normal volume, the heat stimulus is the signal to continue reading aloud and the cold stimulus signals to read on in a whisper.

The stimuli were always sent at the same text passages and appeared both at the beginning of a paragraph and in the running text and lasted about 2 seconds. After the signal stopped, the volume had to be maintained until the next stimulus would be sent. In order to not influence the reaction times, the test subjects sat with their backs to the experimenter and, therefore, could not see when the stimuli were sent.

To generate the tactile signals a Groove vibration motor and two Peltier elements of the size 8 x 8 x 2,6 mm were used. The temperatures chosen are 40 °C for the heat stimulus and 10 °C for the cold stimulus. When choosing the temperatures, the orientation was based on the selected temperatures in related works [13], [14], a pretest carried out, and the 45 °C limit, which is the temperature that could lead to first degree burns [20].

C. Evaluation

In the study, a total of 55 test subjects were tested of which 36.21% were female, and 64.79% were male. The average age of the subjects was 24.34, while the youngest person being 17 and the oldest 40 years old.

Each test person received 12 stimuli during the reading process and thus 4 of each stimulus type. However, not all signals were always recognized correctly. Sometimes certain stimuli were not noticed and, therefore, there was no reaction to measure.

Before statements could be made about reaction rates and reaction times, the data first had to be cleared of outliers with the help of the SPSS software. Except for one outlier, all data were used for the calculation. This was the reaction time of a test person to the heat signal. The person recognized only one of four transmitted heat signals and reacted to it only after more than 9 seconds. Since this is a far above average reaction time, it was assumed for the subsequent calculations that the person did not recognize any of the four heat signals.

After cleansing the data, the reaction rates and response times could be calculated. Table I shows the frequencies of tactile stimuli recognition. When analyzing the frequency of reactions to the stimuli, it was found that it was often the case that people did not recognize a particular stimulus type in all four interventions. However, this does not apply to the recognition of the vibration signal, as it was recognized by all 55 test subjects at least once.

TABLE I. STIMULI RECOGNITION AS PREVIOUSLY REPORTED IN [19].

Stimuli	Recognition rate	Number of recognized stimuli / Number of interventions	Number of test subjects, who recognized stimuli
Vibration	98.63%	217 / 220	55
Cold	91.36%	201 / 220	52
Heat	80.45%	177 / 220	49

Furthermore, the vibration signal was most frequently recognized. Of 220 stimuli sent in total, 217 were correctly recognized, resulting in a recognition rate of 98.63%.

The cold stimulus did a little worse with a recognition rate of 91.36%. Moreover, only three subjects out of 55 did not react to any of the cold signals sent.

The heat stimulus has been recognized least frequently, resulting in a recognition rate of 80.45%. Out of 55 people, 49 people responded to the heat stimulus at least once. The remaining six people did not adapt their volume when reading.

A reason for the differences in the recognition rates of the vibration and the temperature feedback could be the different perceived intensity of the sent stimuli (see Fig. 1). In the questionnaire, 58.63% of the test persons classified the intensity of the vibration stimulus as strong or too strong and 0% classified it as weak or too weak. In contrast, the other two stimuli were more often classified as weak or too weak. The heat stimulus was classified by 58,62% and the cold stimulus by 39% as weak or too weak. The weaker perceived intensity of the temperature stimuli could also be related to the temperature of the test subject's hands, on which the actuators were worn. The temperature of the hands was not measured, however.

The reaction times to the stimuli were compared using the dependent t-Test. Since normally distributed samples are a prerequisite for the application of a t-test, the Kolmogorov-Smirnov test was carried out to test the whether the samples for reaction times were normally distributed [21]. The SPSS software was used for data analysis of the reaction times to the different stimuli. The test showed in the SPSS examination that the samples of reaction times for the vibration, heat and cold stimuli are normally distributed. For this reason, the t-test to dependent samples was applied.

For the calculation, the signal types were compared in pairs (see Table II). Looking at the number of comparison pairs for the groups, it is noticeable that they vary from group to group. The reason for this is that a comparison can only be made if a person has reacted to both signal types. This is not the case with all of them, since not all test persons have recognized every stimulus type. For example, the vibration signal was reliably detected at least once by each

TABLE II. RESULTS OF THE DEPENDENT T-TEST AS PREVIOUSLY REPORTED IN [19].

Groups for t-test	Cold / Vibration	Warm / Cold	Warm / Vibration
Number of pairs	52	46	49
Reaction time mean values	2.097 / 1.376	3.059 / 2.059	3.091 / 1.1372
t-statistic	9.386	9.356	17.173
p-value	.000	.000	.000
Effect size r	.80	.81	.93

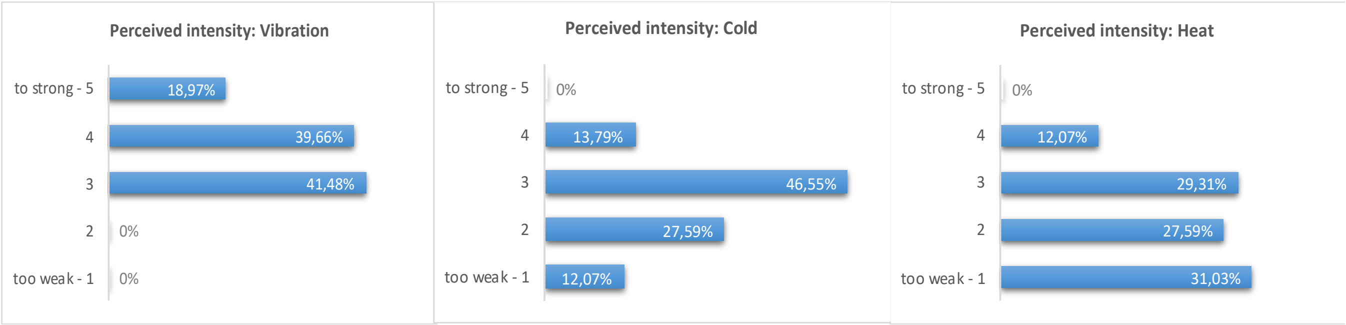

Figure 1. The perception of the stimuli types in terms of intensity.

test person, but only 49 of 55 test persons detected the heat signal. As a result, only 49 comparison pairs could be formed for the vibration/cold group. In the following, the results of the t-test are presented in pairs.

On average, the response time to the vibration stimulus was significantly shorter than to the cold (t=9.386, p=.000, n=52). The effect size r=.80 corresponds to a large effect.

A similar result is obtained when comparing heat and cold. The reaction time to the cold stimulus is significantly shorter than to the heat stimulus (t=9.356, p=.000, n=46). The effect size is r=.81 and also corresponds to a large effect.

The mean values of the reaction times to heat and vibration stimuli differ the most. The response time to the vibration stimulus is significantly shorter compared to the heat stimulus (t=17.173, p=.000, n=49). Therefore, there is also a larger effect size with r=.93.

A similar outcome to the t-test can be seen when the response times are correlated with the intervention type (see Table III). This results in strong, highly significant positive correlations between the intervention type warm and the reaction time for all 12 interventions. In contrast, there is a highly significant negative correlation between the intervention type vibration and the corresponding reaction time. There are no significant correlation values between the cold intervention type and the reaction time (p<0.05).

The reason why the temperature stimuli also performed worse in terms of reaction times could be that the Peltier elements also need some time to adapt to the set heat and cold temperatures. Besides, the differently perceived intensity of the stimuli may also play a role here (see Fig. 1). For example, a person who is not very sensitive to heat would probably only notice the heat stimulus at a maximum temperature of 40 °C. In contrast, a person who is more sensitive to heat would determine the increase in temperature more quickly. It should be noted that the heat temperature has already been chosen very close to the limit of a potential burn trauma [20].

In the questionnaire the test subjects were asked, among other things, to sort the signals according to the degree of distraction from speaking. In this way, the vibration stimulus was ranked 1.6 as the least distracting on average. The heat signal followed with the rank 2.1 and the cold stimulus with the rank 2.7 on average distracted most from reading.

This ranking is also reflected in the evaluation of the perceived comfort of the signals (see Fig. 2). The vibration signal was rated as rather pleasant or pleasant by 65.51% and as rather unpleasant or unpleasant by only 12%. In contrast, the two temperature signals were perceived as pleasant less often. The heat signal was even classified by 20.69% and the cold signal by 39.66% as rather unpleasant or unpleasant.

The perceived intensity also seems to influence the degree of distraction. Some test subjects reported that they were distracted by a signal that was too weak, as they were too focused on not missing it. This is also reflected in Fig. 1. The vibration was not rated too weak by any, whereas the heat of 58.62% and cold of 39.66% were rated as weak or too weak. A stimulus that is perceived too strongly does not seem to have an influence on the degree of distraction. Even if 58.63% perceived the vibration stimulus to be strong or too strong, it still distracted the least on average.

TABLE III. PEARSON CORRELATIONS OF THE VARIABLES REACTION TIME AND INTERVENTION TYPE AS PREVIOUSLY REPORTED IN [19].

Interventions	**Intervention type warm**	**Intervention type vibration**
1. Intervention (n=54)	r = .474**; p = .000	r = –.570**; p = .000
2. Intervention (n = 53)	r = .656**; p = .000	r = –.500**; p = .000
3. Intervention (n = 45)	r = .555**; p = .000	r = –.585**; p = .000
…	…	…
12. Intervention (n = 45)	r = .471**; p = .000	r = –.613**; p = .000

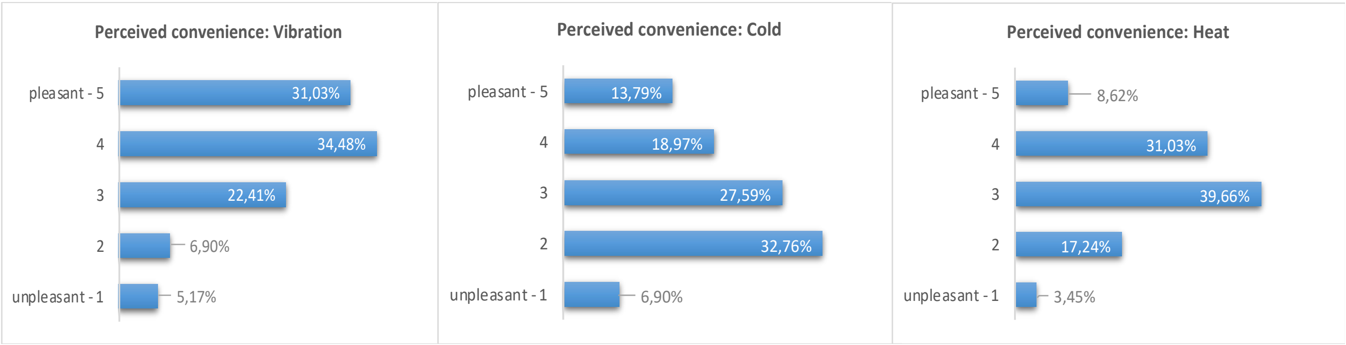

Figure 2. The perception of the stimuli in terms of convenience.

D. *Feedback method selection*

Based on the evaluation of all stimuli tested, the vibration stimulus is best suited to convey emotions and is, therefore, selected as a feedback method for the tactile interface. The vibration stimulus impresses with higher detection rates and shorter reaction times, which have the advantage that even emotions that only last a few seconds could be recognized quickly.

Moreover, since the aim of the project is to support communication rather than divert from it, the degree of distraction is also a key criterion. The vibration feedback, which was classified as the least distracting on average, is convincing at this point, too.

The temperature feedback did worse and is hence, not suitable for conveying emotions. Nevertheless, it can prove useful to aid communication.

One possible application would be to convey other non-verbal information, such as moods, as these have higher latency. The detection rates, however, are significantly lower than those of the vibration stimulus. The temperature stimuli were also more often perceived as unpleasant, which makes the use of Peltier elements unattractive for an interface.

For this reason, it should first be investigated how the perception is and whether the temperature stimuli are better recognized if larger Peltier platelets are used or if the stimuli are sent for longer than two seconds.

V. PROTOTYPE CONSTRUCTION AND DESIGN

This section discusses the architecture of the SSD prototype and design decisions made in this project. The prototype’s architecture is formed by four main components:

- Camera unit
- Smartphone
- Notebook / FaceReader
- Haptic device

The camera unit records the interlocutors face during the communication and is, therefore, the sensor of the SSD. It is attached to glasses so that the face of the communication partner is always in focus. The camera uses a smartphone for the power supply, which sends the captured photos continuously and in real-time to the FaceReader software. The smartphone thus represents a supplement to the main sensor.

A notebook is used to run the FaceReader software [27], which analyzes and categorizes the photos by utilizing the FACS. Both together form the coupling system of the substitution system. After categorization, the results are sent to the haptic device.

The haptic device is the stimulator of the SSD. It consists of a controlling unit and a vibrotactile glove. The controlling unit receives the signals from the FaceReader Software via a Bluetooth module and controls the vibrotactile glove. Depending on the classified emotions, the associated vibration motors vibrate. In the following, the technical procedure and components of the system will be described in greater detail.

A. *Camera-Unit and Smartphone*

As previously reported in [1] the predecessor prototype was working with a Logitech Brio 4K webcam, whereby the device was not mobile. For this reason, the webcam is now replaced by the HD Mini camera from Spyschool [22], which is attached to the temple of eyeglasses, similar to the proposed face recognition device in [10]. For the power supply of the HD Mini Camera, an android smartphone is used, which can be carried in the pocket. The smartphone also runs the application CameraFI for the HD Mini Camera und sends the image of the camera to the notebook in real-time. Fig. 3 shows the Camera-Unit connected to a Smartphone. In order not to strain the battery of the smartphone, a special power bank can be used for the power supply of the camera.

In the previous paper, the use of the smartphone camera was described as an alternative. In this case, the smartphone would be carried in the breast pocket. Although this could be very convenient and cheap, this turned out to be impractical for this application, as the camera would not be pointed at the face of the conversation partner and would, therefore, lead to restrictions in emotion recognition. By attaching the HD mini camera to glasses, the conversation partner's face is always in view.

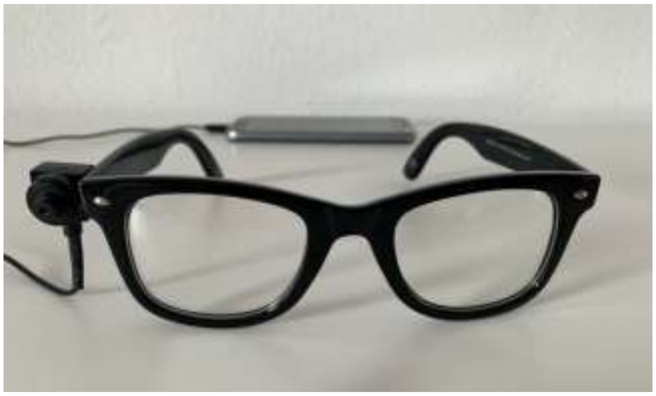

Figure 3. Camera-Unit connected to a smartphone.

B. *Notebook / FaceReader*

The notebook is used to run the FaceReader software [27]. In order to recognize the emotions of the interlocutor, it is not necessary to develop software that will recognize faces and analyze facial expressions. This task can be undertaken by the FaceReader, which is an automatic analysis tool for facial expressions. It utilizes the FACS and is, therefore, able to recognize the seven universal emotions and their intensity, which are described in Section III as well as neutral facial expressions. The functionality of the software in relation to emotion recognition is described below.

When the software has detected the presence of a face with the Viola-Jones algorithm [23], a 3D model of the face is created. The model is generated using an algorithm based on the Active Appearance Method of Cootes and Taylor [24]. With the help of the model, points are placed around and in the face, and around those parts of the face that are usually easy to recognize such as eyebrows, lips, nose, and eyes. Furthermore, the texture of the face is also determined.

In this way, not only the position and shape of the face but also the shape of the eyebrows, wrinkles, and the like are described.

The classification of facial expressions is based on the training of an artificial neural network [25]. Training material from over 10,000 images was used for this purpose. The FaceReader also uses the Deep Face classification method [26], which allows the face classification from image pixels and the recognition of patterns with a neural network. Hence, the face does not necessarily have to be completely visible. It is sufficient if the position of the eye area can be determined. The Deep Face classification method is used when modeling with the Active Appearance method is not possible. More detailed information about the functionality of the system can be found in the FaceReader white paper [27].

After the classification in FACS, the recognized emotions are forwarded to the haptic device. If the FaceReader cannot recognize a face because the communication partner has moved or the image quality is poor, no analysis is possible. This information is also forwarded to the haptic device. Due to the privacy aspects of having a camera recording during a conversation, we constructed the prototype as a closed-loop-system. This means the recordings are interpreted by the FACS software instantaneously and no video recording remains on the server.

C. *Controlling unit*

The controlling unit's main component is an ESP32-WROOM-32U (ESP32). It controls the vibrotactile glove, which will be described in section D and is controlled by a built-in Bluetooth module. In this way, the filtered real-time data from the analysis of the FaceReader are sent to the ESP32 controller via Bluetooth. The data sent include on the one hand the emotions of the conversation partner in real-time. On the other hand, the user is also notified if no analysis is possible because the conversation partner has moved, or the lighting conditions are insufficient. A notification is also sent as soon as the analysis can be continued. The vibrotactile glove is controlled on the basis of the received data. A detailed description of the hardware used for the controlling unit and the vibrotactile glove can be found in the wiring diagram in Fig. 4. The ESP32 is in this graphic the U2-component. The close-up view of the controlling unit and vibrotactile glove hardware is presented in Fig. 5.

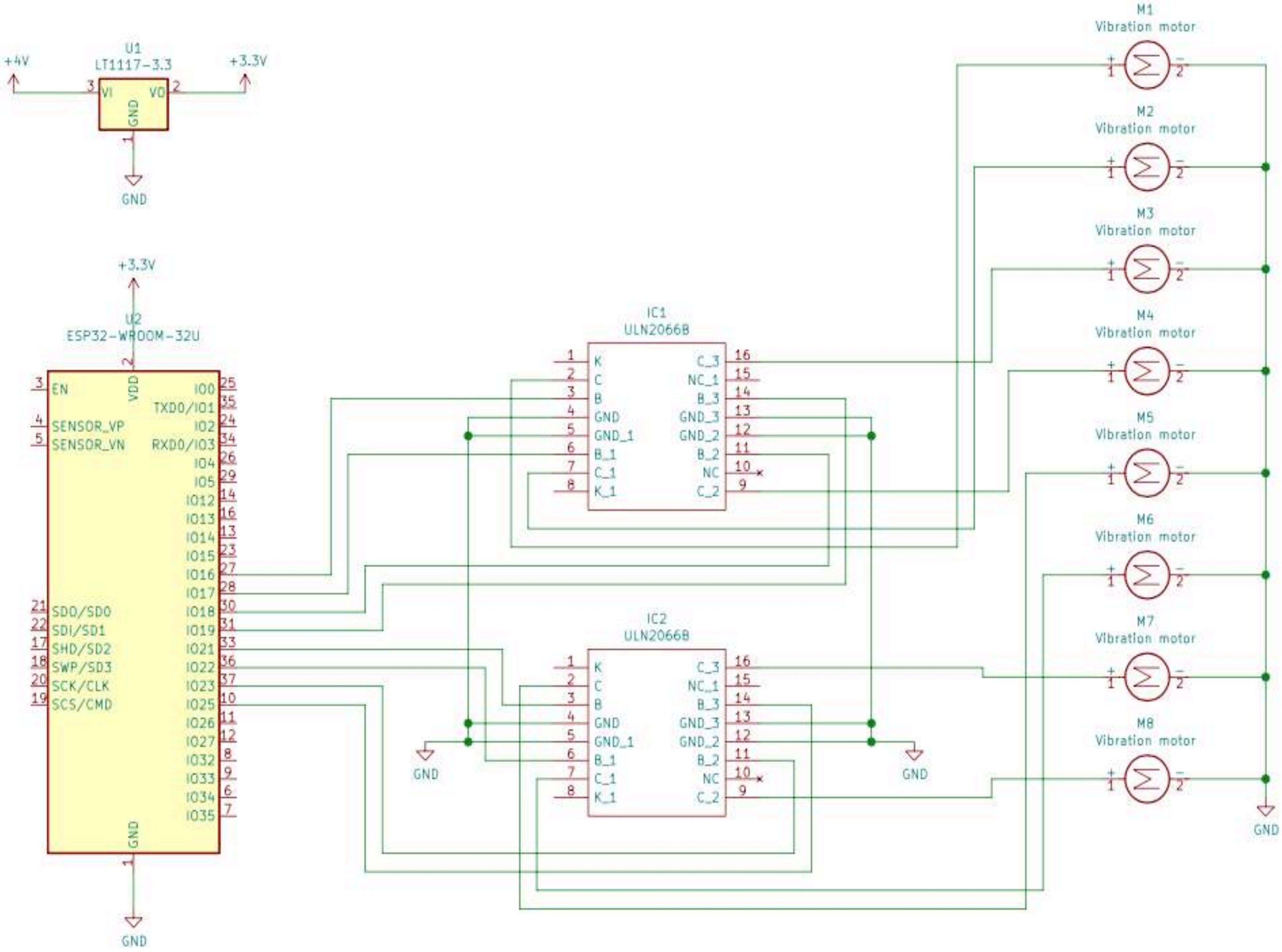

Figure 4. Wiring diagram of the haptic device.

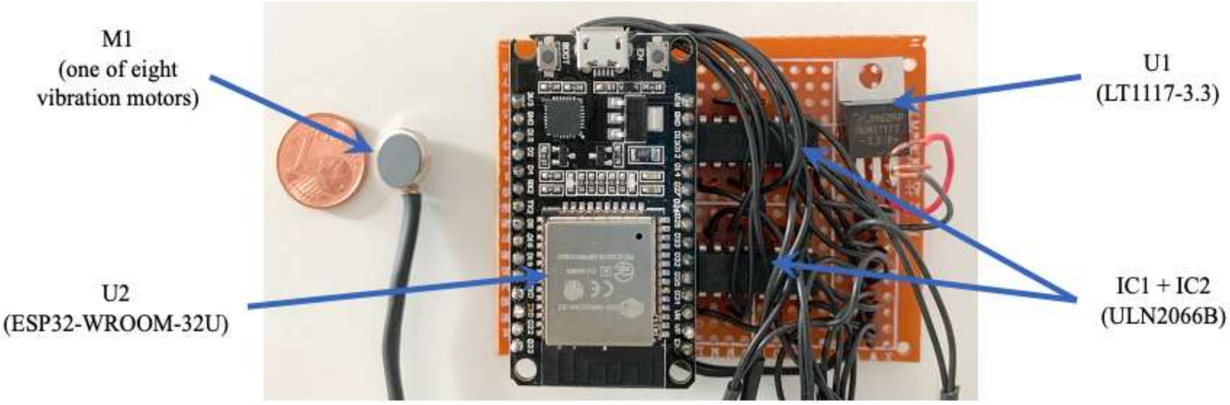

Figure 5. Close-up view of controlling-unit and vibrotactile glove hardware.

Important upgrades to the last prototype are the use of the ESP32 and the light and flat BL-5C Nokia battery instead of a power bank. In contrast to most Arduino controllers, the ESP32 has a Bluetooth module and also offers WiFi connection to the chip. For future developments, the categorization of the data recorded by the camera could possibly be shifted to the chip. It would therefore no longer be necessary to use a notebook, which runs the FaceReader software for the analysis.

Thanks to the built-in Bluetooth module and the use of the BL-5C battery for the power supply, it is also possible to make the controlling unit very light and compact. The controlling unit including the battery is therefore no longer worn on the neck, but on the wrist. The length of the cables that connect the controlling chip with the vibrotactile glove is thus reduced to a minimum and ensures more freedom of movement.

D. Vibrotactile glove

As previously reported in [1], it was planned to expand the predecessor prototype with Ekman's seven basic emotions. The task of the predecessor prototype was to convey the emotional valence of the interlocutor to the user. The haptic interface thus consisted of two vibrating rings, one of which represented the positive and the negative valence. In the following, the improvements of this haptic component of the prototype are described.

One of the most important decisions that were made concerning the haptic device was the selection of the feedback method. As described in Section II, vibrotactile cues for vision substitution have proven to be a good alternative to aural cues in terms of navigation and social interactions [1][2]. Moreover, in the study described in Section IV, in which various haptic signals were compared, vibration signals also proved to be particularly suitable for conveying emotions. For this reason, it was decided in this project to use vibration signals for the haptic interface.

Another design decision that was made was related to the placement of the vibration actuators. Some described vision substitution systems in Section II have successfully used and tested vibrotactile cues in haptic belts or gloves [1][3]. Since SSDs for navigation use both vibrotactile cues in haptic belts and gloves, it follows that this also applies to communication for which only haptic belts were previously presented.

In order to keep the haptic device small and easy to put on for future experiments, it was decided to develop a vibrotactile glove that can be worn on the non-dominant hand. The haptic glove is based on a set of vibration motors attached to an elastic bicycle glove.

Initially, it was planned to use fewer vibration motors to save costs, so that an emotion could be signaled by one or a combination of different vibration motors, but this idea was discarded. Although the basic emotions can be separated well in terms of their nature and the associated facial expression, it is necessary to take into account that mixed emotions occur in addition to pure emotions [2]. If an emotion were addressed via a combination of several vibration motors, it would not be possible to convey mixed emotions. For this reason, the prototype was developed so that a vibration motor always addresses exactly one emotion. Figures 6 and 7 show the haptic device when it is worn.

Each of the seven vibration motors, which are attached to the fingers the back of the hand, signal a basic emotion. The eighth vibration motor is attached to the palm and vibrates when the software cannot recognize a face. This happens on the one hand due to poor lighting conditions or if the communication partner has moved and, therefore, cannot be recognized by the camera.

As more vibration motors are necessary, it was decided to use smaller vibration motors so that they can be attached to all fingers without the motors touching each other and distorting the signal.

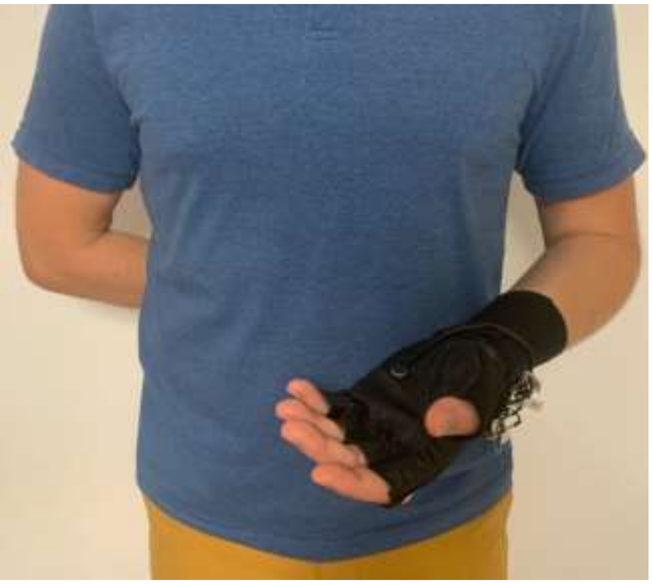

Figure 6. Front of the haptic device.

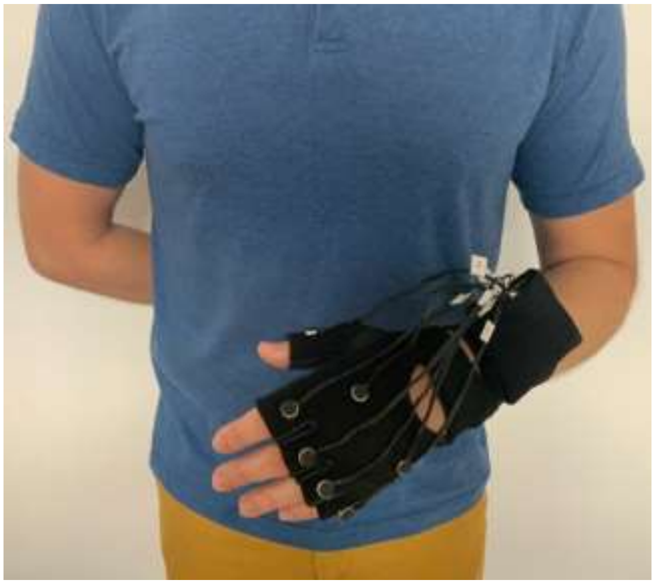

Figure 7. Back of the haptic device.

VI. Conclusion and Outlook

This paper has introduced the prototype of an SSD designed to assist people with visual impairment in daily face-to-face communication situations.

An important sub-goal was to select a suitable feedback method for the SSD stimulator. This was done by means of a within-subject design study in which test subjects tested and compared vibration and temperature stimuli. Unlike many substitution solutions that send aural feedback signals, the focus here was on tactile stimuli, since these are received on a different channel than verbal communication.

Since the SSD was developed to aid communication, great importance was also attached to creating a test environment in which the test subjects should react to the stimuli while they are speaking.

The vibration stimulus emerged as the clear winner in this study as it convinced with higher detection rates and shorter reaction times than the ones of the temperature feedback. Furthermore, the vibration stimulus was also classified as less distracting during speech and perceived more rarely as unpleasant. For this reason, the temperature stimuli were classified as unsuitable for conveying emotions in communication.

Nevertheless, it is necessary to investigate how the perception of the temperature stimuli changes when the signals are sent for longer periods or larger Peltier platelets are used. With unchanged reaction times and improved detection rates, the use of temperature stimuli to support everyday conversational situations would also be possible, e.g., to convey moods, as these have a higher latency.

However, since this project is focused on conveying emotions that can last only a few seconds, it was decided to choose the vibrational stimulus as a feedback method for the SSD.

In the next step the previously presented prototype of [1], which conveys the emotional valence of the conversation partner, was expanded. The further development of the prototype was aimed at making the prototype more mobile and providing the user with more detailed information about the emotional state by extending the prototype with the basic emotions according to Ekman [2]. This is made possible by recording the interlocutor during communication with a portable camera and determining the emotions in real-time, using the FaceReader software. Subsequently, the recognized emotions are translated into tactile information and transmitted to the user via a vibrotactile glove, which can be worn on the non-dominant hand.

Thus, the camera-based recording of facial expressions, the recognition of emotions, and finally, the conversion into mute vibration movements on a hand, an unobtrusive signal transmission can be ensured. Additionally, the device is discreet, and the user is free to gesticulate with the dominant hand.

The next milestone is the implementation of a qualitative study using a guided interview. The cognitive interest of the work is to discuss which non-verbal cues people perceive in face-to-face conversation situations, which of them are important for a smooth conversation, e.g., to avoid misinterpretations or to better understand their interlocutor. On the other hand, the SSD prototype will be experimentally validated as part of a master's thesis. It is planned to survey to what extent the developed substitution system helps to recognize emotions in a conversation situation and how this could be expanded or improved in order to optimize the support of a communication.

It is important to note that there is still room for improvement in terms of mobility. Here, for example, the

FACS analysis could be shifted to the haptic device or the smartphone. However, the degree of mobility of the SSD solution is sufficient for the planned experiments.

Future developments could include passing on individual AUs to the user. Since these can occur without indicating an emotion, their feedback could also provide valuable information for communication.

Furthermore, there might be even more communication situations, which might benefit from the SSD and are currently not in focus of our research. Besides substituting physical limitations of being able to visually perceive the interlocutor, there might be a useful application for people who simply are not capable of interpreting facial actions correctly. Therefore, training scenarios for various communication settings could benefit from a direct and discreet form of an independent non-verbal-cue interpretation device. In order to get a better insight into the advanced applications for the SSD, further studies together with the applied social sciences department are being prepared. Especially during the training of nursing professions and social workers, the nonverbal feedback of their counterparts might provide useful insights.

References

[1] A. Kushnir and N. H. Müller, "Development of a Wearable Vision Substitution Prototype for Blind and Visually Impaired That Assists in Everyday Conversations", In: Jaime Lloret Mauri, Diana Saplacan, Klaudia Çarçani, Prima Oky Dicky Ardiansyah and Simona Vasilache (eds). Proceedings of ACHI 2020, The Thirteenth International Conference on Advances in Computer-Human Interactions, Nov. 2020, pp. 189-192, ISBN: 978-1-61208-761-0.

[2] P. Ekman, S. Kuhlmann-Krieg, M. Reiss, Gefühle Lesen [In English: Emotions Revealed], 2nd. edition. Heidelberg: Spektrum Akademischer Verlag, 2010.

[3] T. McDaniel, S. Krishna, V. Balasubramanian, D. Colbry, S. Panchanathan, "Using a haptic belt to convey non-verbal communication cues during social interactions to individuals who are blind", in 2008 IEEE International Workshop on Haptic Audio visual Environments and Games, Oct. 2008, pp. 13–18, doi: 10.1109/HAVE.2008.4685291.

[4] P. Bach-Y-Rita, C. C. Collins, F. A. Saunders, B. White, L. Scadden, "Vision Substitution by Tactile Image Projection", Nature, vol. 221, no. 5184, Art. no. 5184, Mar. 1969, doi: 10.1038/221963a0.

[5] P. Bach-y-Rita and S. W. Kercel, "Sensory substitution and the human–machine interface", Trends in Cognitive Sciences, vol. 7, no. 12, pp. 541–546, Dec. 2003, doi: 10.1016/j.tics.2003.10.013.

[6] C. Lykawka, B. K. Stahl, M. d B. Campos, J. Sanchez, M. S. Pinho, "Tactile Interface Design for Helping Mobility of People with Visual Disabilities", in 2017 IEEE 41st Annual Computer Software and Applications Conference (COMPSAC), Jul. 2017, vol. 1, pp. 851–860, doi: 10.1109/COMPSAC.2017.227.

[7] G. Bernieri, L. Faramondi, F. Pascucci, "A low cost smart glove for visually impaired people mobility", in 2015 23rd Mediterranean Conference on Control and Automation (MED), Jun. 2015, pp. 130–135, doi: 10.1109/MED.2015.7158740.

[8] Z. Liu, Y. Luo, J. Cordero, N. Zhao, Y. Shen, "Finger-eye: A wearable text reading assistive system for the blind and visually impaired", in 2016 IEEE International Conference on Real-time Computing and Robotics (RCAR), Jun. 2016, pp. 123–128, doi: 10.1109/RCAR.2016.7784012.

[9] P. G. Bhat, D. K. Rout, B. N. Subudhi, T. Veerakumar, "Vision sensory substitution to aid the blind in reading and object recognition", in 2017 Fourth International Conference on Image Information Processing (ICIIP), Dec. 2017, pp. 1–6, doi: 10.1109/ICIIP.2017.8313754.

[10] S. Krishna, G. Little, J. Black, S. Panchanathan, "A Wearable Face Recognition System for Individuals with Visual Impairments", in Proceedings of the 7th International ACM SIGACCESS Conference on Computers and Accessibility, New York, NY, USA, 2005, pp. 106–113, doi: 10.1145/1090785.1090806.

[11] OrCam, OrCam MyEye 2. [Online]. Available from: https://www.orcam.com/de/myeye2/ [Accessed: 2020.11.25].

[12] Microsoft, Seeing AI. [Online]. Available from: https://www.microsoft.com/en-us/ai/seeing-ai [Accessed: 2020.11.25].

[13] G. Lopez, K. Takahashi, K. Nkurikiyeyezu, A. Yokokubo, "Development of a Wearable Thermo-Conditioning Device Controlled by Human Factors Based Thermal Comfort Estimation", in 2018 12th France-Japan and 10th Europe-Asia Congress on Mechatronics, Sep. 2018, pp. 255–259, doi: 10.1109/MECATRONICS.2018.8495727.

[14] G. Lopez, T. Tokuda, N. Isoyama, H. Hosaka, K. Itao, "Development of a wrist-band type device for low-energy consumption and personalized thermal comfort", in 2016 11th France-Japan 9th Europe-Asia Congress on Mechatronics (MECATRONICS) /17th International Conference on Research and Education in Mechatronics (REM), Jun. 2016, pp. 209–212, doi: 10.1109/MECATRONICS.2016.7547143.

[15] P. Ekman and D. Cordaro, "What is Meant by Calling Emotions Basic", Emotion Review, Sep. 2011, doi: 10.1177/1754073911410740.

[16] P. Ekman, Ich weiß, dass du lügst: Was Gesichter verraten [In English: Telling Lies], 6th. edition. Reinbek bei Hamburg: Rowohlt Taschenbuch Verlag, 2011.

[17] P. Ekman, Emotion in the Human Face, 2nd. edition. New York: Cambridge University Press, 1982.

[18] P. Ekman, W. V. Friesen, J. C. Hager, Facial Action Coding System. The Manual on CD ROM. Salt Lake: Network Information Research Corporation, 2002.

[19] A. Kushnir and N. H. Müller, "Haptic Feedback in Everyday Conversation Situations", in HCI International 2020 - Posters, Cham, 2020, pp. 239–244, doi: 10.1007/978-3-030-50726-8_31.

[20] T. Trupkovic and G. Giessler, „Das Verbrennungstrauma – Teil 1 [In English: The burn trauma - Part 1], Anaesthesist, vol. 57, no. 9, p. 898, Aug. 2008, doi: 10.1007/s00101-008-1428-5.

[21] A. Field, Discovering Statistics Using SPSS. SAGE Publications, 2007.

[22] Spyschool, Spycam. [Online]. Available from: http://www.spyschool.de/usb-full-hd-spionkamera-neu/ [Accessed: 2020.11.25].

[23] P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features”, in Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, Dec. 2001, vol. 1, p. I–I, doi: 10.1109/CVPR.2001.990517.

[24] T. F. Cootes and C. J. Taylor, Statistical Models of Appearance for Computer Vision. 2000.

[25] C. M. Bishop, Neural Networks for Pattern Recognition. USA: Oxford University Press, Inc., 1995.

[26] A. Gudi, H. E. Tasli, T. M. den Uyl, A. Maroulis, “Deep learning based FACS Action Unit occurrence and intensity estimation”, in 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), Ljubljana, May 2015, pp. 1–5, doi: 10.1109/FG.2015.7284873.

[27] Noldus, White paper on FaceReader methodology. [Online]. Available from: https://info.noldus.com/free-white-paper-on-facereader-methodology [Accessed: 2020.11.25].