Development of a Wearable Vision Substitution Prototype for Blind and Visually Impaired That Assists in Everyday Conversations

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 an idea of a Sensory Substitution Device (SSD), which supports everyday conversations by conveying to the user the emotional valence of its interlocutor. It describes the work in progress of a SSD prototype design that aims to remap visual stimuli into tactile information, by utilizing the Facial Action Coding System (FACS).**

Keywords-Non-verbal communication; Emotional valence; Visually impaired; Vision Substitution; Sensory Substitution Device.

I. INTRODUCTION

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 facial expressions, intonations or gestures. Sighted people can use all these communication channels, which allows them, for example, to interpret the facial expressions in order to understand their interlocutor better.

Blind and visually impaired people are limited in interpretation as they are not able to process the visual non-verbal information. This makes it hard to determine the emotional valence of the communication partner, which indicates whether an emotional status is positive or negative. Although the emotional valence can be determined through the interlocutor's intonation, it is only possible while the 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 [1]. 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, therefore, it can be classified under Sensory Substitution Devices (SSDs).

The aim of the proposed SSD is to supply blind people with visual information by converting it into tactile representation in order to convey emotional valence of the user's interlocutor.

The paper is structured as follows. In Section II, related work is presented, and the research needs are derived. Section III describes the relationship between Emotions, Facial Action Coding System (FACS) and emotional valence. In Section IV, the prototype of the SSD and important design decisions are presented. Finally, Section V summarizes the paper and describes the next steps.

II. BACKGROUND AND RELATED WORKS

Most of the related works about vision substitution for blind and visually impaired people focus on navigation, reading texts, object recognition and face recognition. 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 [2].

Bernieri et al. dealt in [3] with visually impaired people's mobility. The authors describe a prototype of a smart glove that complements the classic cane. The described glove provides vibrotactile feedback on the position of the next obstacle in the range.

In [4], a text reading system called FingerEye is proposed, which translates text into audio 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 [5].

The interaction assistant ICare, described in [6], deals with choosing an appropriate face recognition algorithm to build an assistant for social interactions. It also describes the prototype, of which the output is also in audio.

While a lot of research has been done to meet a wide range of needs of people with visual disabilities, not enough 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 [7] 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 is able to read texts, recognize barcodes and time, identify goods by their barcodes and recognize people, by saving their name. All recognitions are translated into audio information and conveyed through a loudspeaker.

Mircrosoft SeeingAI [8] is an application for the mobile phone, which shares a lot of features with the Orcam MyEye. Moreover, it offers a feature, which recognizes and describes scenes, people and their emotions. All types of recognition are translated and represented by audio output. 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 audio outputs. People with a visual impairment rely on their hearing to perceive their environment. Audio 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 one's own speech or the one of the communication partners. 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 are, therefore, not sufficient solutions to support face-to-face communication.

What is needed is a system that communicates non-verbal cues in real-time to a blind person 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 [2]. McDaniel et al. also presented a haptic belt to assist in communication situations [1]. 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.

III. Emotions and the Facial Action Coding System

Every emotion sends signals, which are most noticeable through our voice and facial traits. For this reason, the FACS is used as the basis for the emotional valence recognition in this project. FACS is an anatomically based system for describing all visually perceptible facial muscle movements. The system assigns Action Units (AU) to almost every visible movement of facial muscles [9]. A combination of certain AUs can be assigned to emotions. The emotions of anger, happiness, sadness, disgust, contempt, fear and surprise are considered universal and cross-cultural, according to Paul Ekman. These can be recognized through facial expressions using FACS.

To get back to the emotional valence, every emotion has a value that categorizes emotions into positive and negative ones. Thus, it is possible to deduce the emotional valence from the emotion. Happiness is seen as a positive emotion. In contrast, the rest of the universal emotions, except surprise, belong to the negative emotions. The emotion surprise is a special case as it lasts at most a few seconds. After that, it ends in fear, pleasure, anger or other emotions, depending on the quality and nature of what surprises us. Therefore, a surprise can lead to positive valency, as well as negative valency [10].

IV. Prototype Construction and Design

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

- Camera unit
- Laptop (for emotional valence recognition)
- Wearable microcontroller
- Wearable haptic device

Figure 1 shows a simplified representation of the interaction of these components.

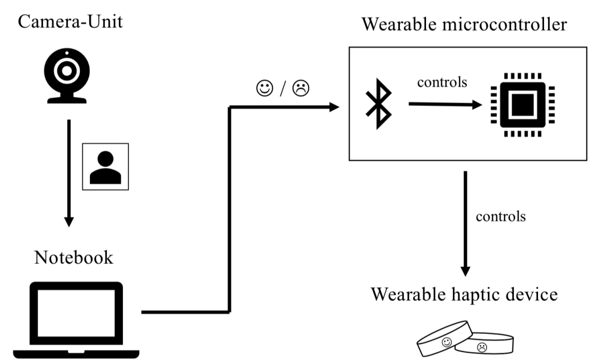

Figure 1. Prototype components interaction

The camera unit is recording the interlocutors face during the communication and sends the captured photos continuously and in real-time to the FaceReader Web API. A laptop is used to run the FaceReader Software, which analyzes and categorizes the photos with regard to the emotional valence. After categorisation, the FaceReader Web API sends the results to an Arduino nano board by using a Bluetooth module for the transmission. This microcontroller controls the haptic device consisting of a set of two vibrating rings. Depending on the emotional valence, either the ring which stands for a positive emotional valence or the one for the negative emotional valence will vibrate. In the following, the technical procedure and components of the system will be described in greater detail.

A. Camera-Unit

Currently the prototype is working with a Logitech Brio 4K webcam, which is able to make high quality pictures. Similar to the proposed face recognition device in [6], it is planned for the future work to use eyeglasses with an included camera or a portable camera, which can be attached on the temple of eyeglasses instead, to make the device more mobile. As an alternative, it is also conceivable to use the smartphone camera while the smartphone is in the breast pocket. However, the webcam is sufficient for the planned experiments with the system during the second quarter of 2020.

B. Notebook

The notebook is used to run the FaceReader software from Noldus [11]. In order to recognize the emotional valence of the interlocuter, it is not necessary to develop a software which 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 universal emotions and their intensity, which are described in Section III. The emotional valence is automatically calculated by the FaceReader during an analysis. It results from the difference between the intensity of the positive emotion and the intensity of the most pronounced negative emotion. Happy is the only positive emotion, while sad, angry, scared and disgusted are considered to be negative emotions in the calculation. A special case is the emotion surprised, which can be either positive or negative [12]. 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. Wearable microcontroller

The wearable microcontroller is an Arduino Nano V3.3 board which can be controlled via a Bluetooth module, the HC-05-6. The Arduino board can be placed around the neck and controls two vibration motors, which are part of the haptic device. Figure 2 shows a wiring diagram for these three components and Figure 3 shows how the device can be worn. For the power supply of the microcontroller, a power bank can be used. The Arduino Nano board was chosen because of its extensibility. During the second quarter of 2020, it is planned to expand the device with the seven universal emotions.

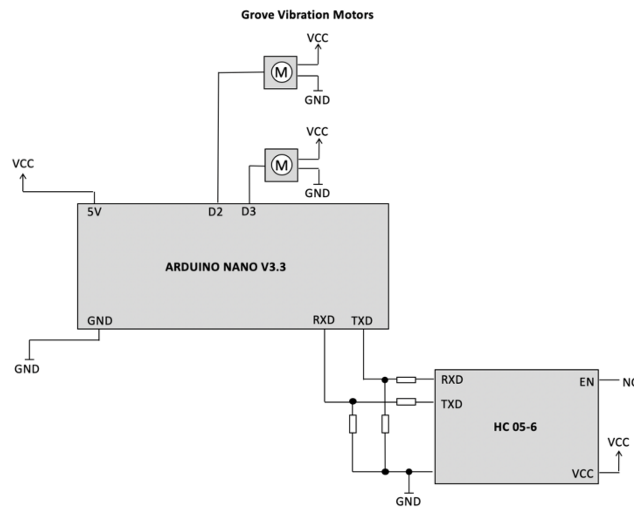

Figure 2. Wiring diagram of the wearables

D. Wearable Haptic Device

The goal is to create an interface that conveys the emotional valence in real-time, meaning that the signals should be conveyed during a conversation. As a result, these signals are also sent while the user is talking or listening to his communication partner. 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]. However, the crucial point why vibrotactile cues were chosen as the transmission method for this project is that they are received through a different sense than the hearing, to not interfere with the verbal

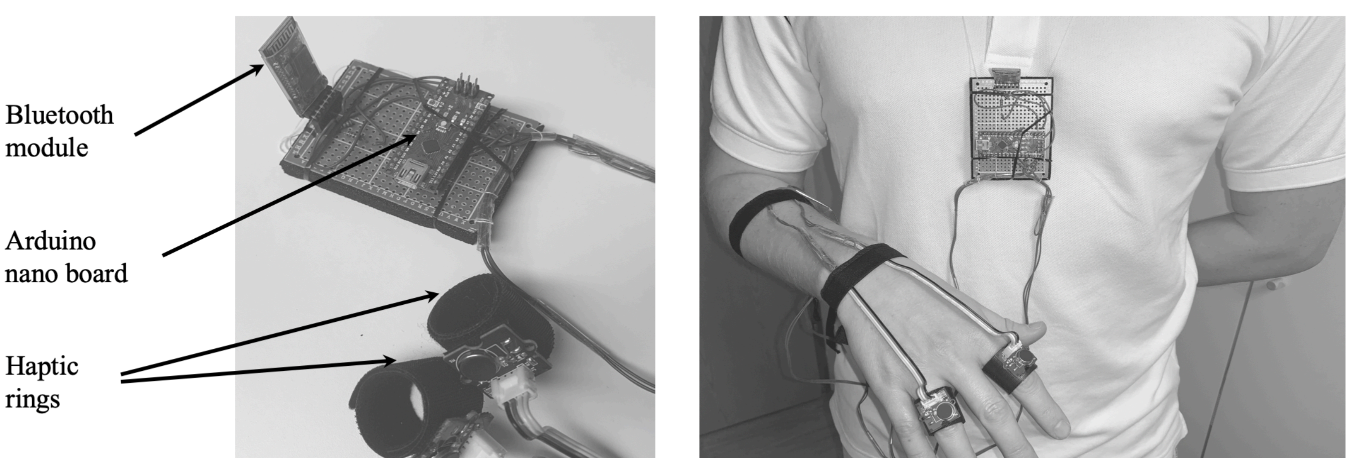

Figure 3. Haptic rings with Grove vibration motors connected to Arduino board with cables

communication. Some described 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. To keep the device small, the system is based on a set of two Grove vibration motors attached to resizable rings which can be worn on the non-dominant hand. Thus, the camera-based recording of facial expressions, the conversion into emotional valence 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 other hand. Figure 3 shows the rings, each with a fixed motor connected to the Arduino nano board with cables. The cables are attached to the arm with two elastic bands so that the cables do not interfere with gesturing. In order to make the device more comfortable to wear, it is planned to create a wireless version in the future.

V. CONCLUSION AND OUTLOOK

This paper has introduced the idea and prototype of an SSD designed to assist people with visual impairment in daily face-to-face communication situations. This is made possible by recording the interlocutor during communication and determining the emotional valence in real-time, using the FaceReader software from Noldus. Subsequently, the recognized emotional valence is translated into tactile information and transmitted to the user via vibrating rings, which can be worn on the non-dominant hand.

Designing the prototype was the first step towards communicating the emotional state of the conversation partner to assist in everyday communications. The next steps will be to evaluate which tactile interfaces could also be used for the design of the prototype. For example, the vibration could be compared with the tactile stimuli heat and cold. In addition, it is planned to expand the prototype with the seven universal emotions, so that the user will be able to access a more detailed emotional state of the interlocutor. In the second quarter, the functionality of the overall system is to be experimentally validated as part of a master's thesis. It is planned to carry out the experiment in cooperation with visually impaired as well as blindfolded test subjects.

REFERENCES

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

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

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

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

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

[6] S. Krishna, G. Little, J. Black, and 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.

[7] "OrCam MyEye 2", *OrCam*. [Online]. Available from: https://www.orcam.com/de/myeye2/ [Accessed: 2020.02.01].

[8] "Seeing AI | Talking camera app for those with a visual impairment". [Online]. Available from https://www.microsoft.com/en-us/ai/seeing-ai [Accessed: 2020.02.01].

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

[10] P. Ekman, *Gefühle lesen [In English: Emotions Revealed]*, 2nd. edition. Heidelberg: Spektrum Akad. Verl., 2010.

[11] Noldus Information Technology, "Free white paper on FaceReader methodology". [Online]. Available from: https://info.noldus.co
m/free-white-paper-on-facereader-methodology [Accessed: 2020.02.01].

[12] "FaceReader Webshop". [Online]. Available from: https://ww w.noldus.com/facereader-webshop [Accessed: 2020.02.01].