Making Facial Expressions of Emotions Accessible for Visually Impaired Persons

Hendrik Buimer
Radboud University
Nijmegen, The Netherlands
h.buimer@donders.ru.nl

Thea van der Geest
HAN University of Applied Sciences
Arnhem, The Netherlands

Abdellatif Nemri
University of Twente
Enschede, The Netherlands

Renske Schellens
Radboud University
Nijmegen, The Netherlands

Richard van Wezel
Radboud University
Nijmegen, The Netherlands

Yan Zhao
University of Twente
Enschede, The Netherlands

Keywords
Vision-impaired persons; facial expressions; emotion recognition; vibrotactile output; wearables

1. INTRODUCTION
Our R&D project is set up in a Human-Centered Design and Engineering process; the prospective user group is involved in various stages of the process. We are designing, developing, testing, and evaluating a wearable system that helps people with visual impairments to perceive facial expressions of emotions. In earlier studies concerning the needs and problems that visually impaired persons encounter in daily life, including interviews we have conducted amongst visually impaired persons, respondents identified their inability to perceive non-verbal cues of conversation partners as a critical issue in their daily life contacts. They ranked this issue as one of the most pressing problems that should be high on the R&D agenda. [1, 4, 6]. We took their expressed need as a starting point for a R&D project on facial expression recognition.

To convey the emotions expressed by the conversation partner, we used a vibrotactile belt, because the haptic modality does not interfere with the other modalities used in most social interactions, such as sound and vision.

The system (see Figure 1) consists of:

1. An eye-level camera which captures the face of the conversation partner in real time, which can be clipped on to any pair of spectacles;
2. The emotion recognition software FaceReader which detects facial expressions and classifies those into one of six basic emotions (i.e. anger, disgust, enjoyment, fear, sadness, surprise) [3, 7, 12];
3. A belt worn on the waist, connected through USB or Bluetooth, with six vibrotactile units each assigned to one of the emotions [11]. The principle is that when the system detects an emotion, the user feels a vibration on the location associated to the detected emotion.

ASSETS '17, October 29-November 1, 2017, Baltimore, MD, USA

ACM ISBN 978-1-4503-4926-0/17/10.
https://doi.org/10.1145/3132525.3134823

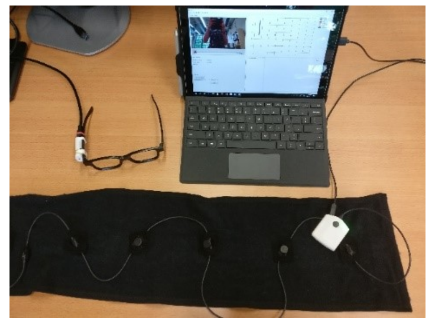

Figure 1 Picture of the current version of the system, existing of a camera clipped-on to spectacles, emotion recognition software running on a tablet, and a vibrotactile belt.

In an earlier study ([2], full paper currently under revision), we have presented the results of a controlled experimental study where people with visual impairments (VIPs) and sighted persons were asked to determine what emotions were presented to them. The images displayed were photos, videos, and additional auditory cues (affect bursts) which are used as validated stimuli in emotion research [5, 8–10]. In this controlled experimental setting, VIPs proved to be able to perceive and recognize emotions from facial expressions at almost the level of sighted persons, and significantly more than without the support of the system. For example, the correct identification of emotional facial expressions from photos without sound cues increased from 14.2% to 82.5%. More importantly, VIPs ventured to imagine how such a system could be applied in situations in daily life, such as work related meetings. To see whether this promise of usefulness for daily life situations like work place conversations could be fulfilled, we wanted to answer an important question: how well would the system perform in more realistic usage contexts, such as an actual one-on-one conversation? In such situations, new challenges arise, such as changing lightning conditions and moving conversations partners who are not always watching directly at the wearer of the system.

The goal of the study we are presenting was to determine whether the earlier developed wearable assistive technology would benefit visually impaired users in realistic usage contexts.

2. METHOD
2.1 Participants
Eight people with visual impairments were participating in the study. The participants, four women and four men, had an average age of 46(±15.8) years old. Five of whom were fully blind, the others were partially sighted. Four of the participants had congenital visual impairments, while the others became visually

impaired at a later stage in life and had been able to see facial expressions in the past.

2.2 Procedure

Participants were invited to participate in two sessions: a training session and an experimental session. The training session, which took place at a location preferred by the participants – most often at their home – lasted about two hours. The participants were presented with pictures and videos of facial expressions of emotions, with and without auditory cues [8, 9]. Over the course of the training, the number of complementary auditory cues decreased, meaning that ultimately the participants had to rely on the vibrotactile cues to determine the emotion expressed in the images.

Two weeks later, the participants were invited to the lab for the experimental session. First, the participants went through another (short) training session. Then the VIPs were asked to engage in a mock job interview with an actor, who played the role of the director of a fictive company. The actor was instructed to keep the conversation as realistic as possible, without the use of any exaggerated facial expressions. The mock job interview consisted of two 15-minute conversations, during one of these the system was worn by the participant.

After the job interview conversations, we conducted semi-structured exit interviews with the participants. We asked questions regarding the users' experiences, perceived usefulness, possible contradictions between their perceptions and the emotions signaled by the system, the economic and financial value of the system, possible improvements of the system, and social acceptance.

2.3 Data collection and preliminary results

During the earlier controlled experimental study, participants showed that the vibrotactile cues were easy to learn and to interpret. In the training sessions, these findings were confirmed, as participants achieved percentage-correct scores of over 95% in the stimuli without complementary auditory cues.

During the observations, various data collection methods were used. Firstly, video-captures were made of the facial expressions of both the actor and the participant of the study, which were combined with a graphical representation of the belt output. This gives us the opportunity to analyze the performance of the system, and checking whether the facial expressions of the actor are recognized by the system and mirrored by the visually impaired conversation partner. The actor annotated his performance in several captured sessions, so we can check whether the system does recognize the emotion expressions that the actor intended to show. The videotapes will be analyzed for emotion expressions by two independent coders.

Additionally, the emotion recognition software logged its performance frame by frame. This so-called emotion log will be used to determine at what times the software surpassed the threshold that was set for each emotion, enabling us to determine the exact times when the system conveyed emotions to the user. We can combine those timestamped data with the video recordings.

From the analysis of the video footage, combined with the emotion log, we expect to draw conclusions about the accuracy of emotion recognition software in realistic settings, as well as insights in the ability of system users to interpret and use cues conveyed by the system while being engaged in a conversation. Preliminary analysis showed that the wearable system is not always conveying correct information to its users, which is problematic and requires improvements before it can be applied as assistive technology. Possible explanations for this difference compared to the earlier controlled experimental setting, are changing lightning conditions and a moving conversation partner, resulting in different viewing distances and angles, making it more challenging to aim the camera at the target.

Finally, we interviewed the participants about their experiences with using the system in a realistic conversation. Besides the findings from the analysis of the video footage and the emotion log, participants also perceived a mismatch between the emotion conveyed by the system and the emotion they perceived from auditory cues. Nevertheless, during moments of silence, the system was perceived helpful to have some idea of what was going on. Even though the recognition of emotions was not always correct, participants were positive about location and the easiness of the vibrotactile cues conveyed by the system. While three participants found it difficult to interpret the cues during the conversation, the others stated they could stay engaged in the conversation while being able to interpret the vibrotactile cues at the same time.

3. ACKNOWLEDGMENTS

There are no conflicts of interest to declare. The research was funded by INTERREG and was approved by the Research Ethical Committee of the University of Twente.

4. REFERENCES

[1] Broerse, J. et al. 2015. *Een onderzoeksagenda vanuit cliëntenperspectief.*

[2] Buimer, H.P. et al. 2016. *Enhancing emotion recognition in VIPs with haptic feedback.*

[3] Ekman, P. 1992. An argument for basic emotions. *Cognition & Emotion*.

[4] Van der Geest, T.M. and Buimer, H.P. 2015. User-centered priority setting for accessible devices and applications. *Mensch und Computer 2015 – Workshopband* (Stuttgart, 2015), 383–389.

[5] Hawk, S.T. et al. 2009. "Worth a thousand words": absolute and relative decoding of nonlinguistic affect vocalizations. *Emotion*. 9, 3 (2009), 293–305. DOI:https://doi.org/10.1037/a0015178.

[6] Krishna, S. et al. 2008. A Systematic Requirements Analysis and Development of an Assistive Device to Enhance the Social Interaction of People Who are Blind or Visually Impaired. *Workshop on Computer Vision Applications for the Visually Impaired*. (2008).

[7] Kuilenburg, H. Van et al. 2008. Facereader 6.1 – How do they do it? (2008), 10000.

[8] Lima, C.F. et al. 2013. When voices get emotional: a corpus of nonverbal vocalizations for research on emotion processing. *Behavior research methods*. 45, 4 (2013), 1234–45. DOI:https://doi.org/10.3758/s13428-013-0324-3.

[9] Olszanowski, M. et al. 2015. Warsaw set of emotional facial expression pictures: A validation study of facial display photographs. *Frontiers in Psychology*. 6, JAN (2015), 1–8. DOI:https://doi.org/10.3389/fpsyg.2015.00516.

[10] Van der Schalk, J. et al. 2011. Moving faces, looking places: Validation of the Amsterdam Dynamic Facial Expression Set (ADFES). *Emotion (Washington, D.C.)*. 11, 4 (2011), 907–920. DOI:https://doi.org/10.1037/a0023853.

[11] Science Suit: 2014. *http://www.elitac.nl/science-suit/.*

[12] Den Uyl, M. et al. 2005. FaceReader: an online facial expression recognition system. *Measuring Behavior 2005 5th International Conference on Methods and Techniques in Behavioral Research* (2005), 589–590.