Tactile Facial Expressions and Associated Emotions toward Accessible Social Interactions for Individuals Who Are Blind

Troy McDaniel, Diep Tran, Samjhana Devkota, Kaitlyn DiLorenzo,
Bijan Fakhri, Sethuraman Panchanathan
Center for Cognitive Ubiquitous Computing (CUbiC), Arizona State University, Tempe, AZ

ABSTRACT

For individuals who are blind, much of social interactions are inaccessible: The majority of information exchanged is non-verbal, e.g., facial expressions and body language. Little work has been done toward building social assistive aids for individuals who are blind. This work presents a mapping between facial action units and vibrotactile representations that may be presented through haptic displays. We present a study exploring how well individuals who are blind can learn to recognize universal emotions of happy, sad, surprise, anger, fear and disgust from vibrotactile facial action units. Results show promising recognition accuracy and subjective feedback, demonstrating that individuals who are blind can learn to understand the emotional content of facial movements presented through vibrations.

CCS CONCEPTS

• CCS → Human-centered computing → Accessibility → Accessibility technologies

KEYWORDS

Vibrotactile displays, Haptics, Assistive technology, Blindness

ACM Reference format:

T. McDaniel, D. Tran, S. Devkota, K. DiLorenzo, B. Fakhri, S. Panchanathan. 2018. Tactile Facial Expressions and Associated Emotions toward Accessible Social Interactions for Individuals Who Are Blind. In 2018 Workshop on Multimedia for Accessible Human Computer Interface (MAHCI'18), October 22, 2018, Seoul, Republic of Korea, ACM, NY, NY, USA. 8 pages. https://doi.org/10.1145/3264856.3264860

MAHCI'18, October 22, 2018, Seoul, Republic of Korea

ACM ISBN 978-1-4503-5980-1/18/10…$15.00
https://doi.org/10.1145/3264856.3264860

1 Introduction

Social interactions are essential to our livelihood, personal development, and overall health and well-being. It is well known that social interactions involve speech and verbal components, but lesser known that most information exchange during interactions is non-verbal: approximately 65% of information exchanged is non-verbal [1] including facial expressions, eye gaze, body language, interpersonal distance, appearance, and contextual cues. Most non-verbal cues are communicated visually; therefore, individuals who are blind are not able to access much of the information exchanged during social settings.

When we consider the implications of social barriers created from limited access to non-verbal communicative cues, the significance of solving this problem becomes apparent: incomplete exchanges can lead to miscommunications, which in turn may create embarrassing circumstances, which may result in social avoidance, isolation, and eventually, psychological problems, e.g., social anxiety and depressions [2]. Consider the following: facial expressions convey basic emotions to complex emotional states, such as sarcasm; body language conveys rich non-verbal communication including interest and agreement; and eye-gaze conveys attentional focus. A lack of any or all of the aforementioned social cues is an incomplete representation of the original communicative message. The work presented here focuses on emotions perceived from facial expressions toward building useful social assistive aids for individuals who are blind.

To achieve this goal, we propose a mapping from the basic universal emotions of *happy*, *sad*, *surprise*, *fear*, *anger*, and *disgust*, as conveyed through facial expressions, to vibrotactile representations for discreet and unobtrusive communication through haptic displays used as part of social assistive aids. Software is now available for reliable recognition of emotions and detailed facial movements [3-5] for integration within body-worn or table-top social assistive aids. Our proposed vibrotactile representations are inspired by *facial action units*, which are the building blocks of facial expressions. Any facial expression can be broken down into its fundamental facial action units using the Facial Action Unit Coding System, or FACS [6]. For example, *happy* is often expressed as lip corners pulled up and cheeks raised; *sad* is often expressed as lip corners pulled down and eyes squinted. Significant work has been carried out to identify which facial action units occur most frequently with the basic, universal emotions [7]. It is this information that guides our selection of facial action units to ensure sufficient coverage of the basic emotions. A simpler approach would be to present the emotions directly, e.g., *"… your interaction partner is surprised …"* or *"… Susan expressed happiness …"*, but we argue that it is more useful to leave the decision-making up to the user (i.e., human-in-the-loop), by providing rich information such as facial action units to aid interpretation. This work makes two significant contributions: (i) A novel set of vibrotactile facial action units for emotion understanding in social assistive aids for individuals who are blind; and (ii) A study exploring how individuals who are blind learn and recognize emotions from these facial movements.

2 Related Work

Very little research has been conducted on the topic of social assistive aids for individuals who are blind or visually impaired. Previously, we explored vibrotactile communication of interpersonal distance [8] and direction [9] between an individual who is blind and his or her interaction partner. In the case of distance, the analogy of a heartbeat was used as inspiration for the design of a set of vibrotactile rhythms, which were mapped to different distances associated with personal space. For direction, a vibrotactile belt, worn around the waist, stimulated different body sites to convey the direction of the interaction partner. The same device was used to convey interpersonal distance. The haptic device was driven by a vision-based module that performed simple face detection from a wearable camera. In [10], Qiu et al. used a vibrotactile band around the head to communicate eye gaze information of interaction partners to individuals who are blind. A visual glance was indicated through a short vibrotactile burst, whereas fixation was conveyed through a repeating vibrotactile pattern.

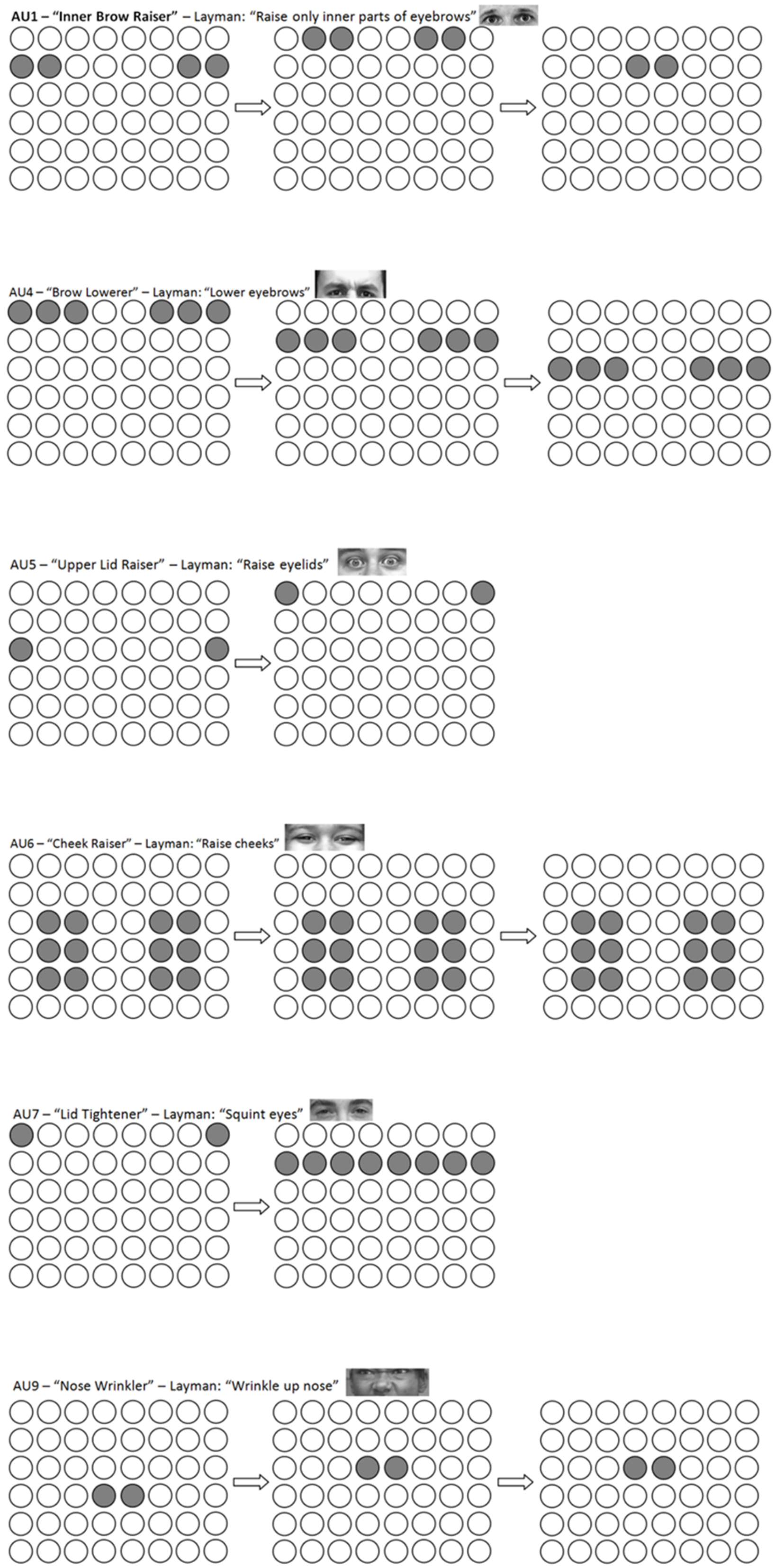
In the context of social assistive aids, presenting facial expressions and emotions to the skin have received the most attention. Réhman et al. [11] embedded a three-axis vibrotactile display in the back of a chair to map each axis to a different emotion—*happy*, *sad*, or *surprise*—of which progression along an axis was correlated with emotional intensity. Using a waistband embedded with vibrotactile actuators, Buimer et al. [12] mapped each of the six basic emotions to a unique body site around the waist. The VibroGlove [13], developed by Krishna et al., presented the six basic emotions to wearers through spatiotemporal vibrotactile patterns inspired by emoticons. For example, *sad* was presented as a vibration in the shape of a frown on the back of the wearer's hand. An example of auditory output, rather than haptic output, is work by Rahman et al. [14] in which speech output was used to communicate behavioral expressions such as yawn, sleepy, or looking away, as well as dimensions of affect such as valence and arousal.

Our previous work [15] represents the first exploration of vibrotactile facial action units. A mapping was proposed between facial action units and spatiotemporal vibrotactile patterns for use in social settings by individuals who are blind. While an important contribution, our work aims to address several shortcomings. First, the selected set of action units in [15] does not provide sufficient coverage of the most commonly occurring units frequently expressed in the six universal emotions. This work aims to select a more comprehensive set for further exploration. Second, much redundancy was present in the previous set, and therefore, could be simplified to ease learning. Our current work removes this redundancy. Lastly, the proposed mapping of [15] was not evaluated on individuals who are blind. To ensure usability by the target population, our current work includes an evaluation with individuals who are blind.

3 Proposed Approach

Our proposed set of vibrotactile facial action units is depicted in Fig. 1. Action units were chosen such that the six universal emotions are well represented. The proposed vibrotactile patterns, which vary both spatially and temporally, are intended for use with two-dimensional vibrotactile displays. The patterns shown in Fig. 1 were finalized after extensive pilot testing and refinement with individuals who are blind. Some patterns are inspired by our previous work of [15], particularly AU4, AU15, and AU26; the remaining patterns were redesigned following pilot testing to enhance distinctness and intuitiveness. During actuation of a single facial action unit, pulse widths are 250 ms, and the gap between individual actuations within a sequence are 50 ms. These timing parameters were identified as desirable for fast communication, yet highly perceptible by participants based on the pilot testing we conducted where a large set of durations were explored.

In Fig. 2, a screenshot of the graphical user interface used to run the user study (Section 4) is depicted. This screenshot shows examples of how combinations of facial action units can communicate specific emotions. In this work, we chose 12 combinations of facial action units, two combinations per each basic emotion for good distribution. Each combination may be one action unit or more in length. To ensure the user study was of reasonable length for participants, combinations of no more than two facial action units in length were chosen. From Fig. 2, *happy* is represented by either AU12 or AU6+AU12; *anger* is represented by either AU4+AU5 or AU4+AU7; *sad* is represented by either AU1 or AU6+AU15; *fear* is represented by either AU5 or AU5+AU20; *surprise* is represented by either AU5+AU26 or AU5+AU27; and *disgust* is represented by either AU9 or AU10. When an emotion is presented, its action units are presented sequentially, rather than concurrently, to avoid overlapping and occluding patterns. There is gap between action units of 1 s. Reduction of the length of this pause will be explored as part of future work to speed up communication and understand perceptual limitations of sequential action units presentation toward practical use.


AU1 – "Inner Brow Raiser" – Layman: "Raise only inner parts of eyebrows"
AU4 – "Brow Lowerer" – Layman: "Lower eyebrows"
AU5 – "Upper Lid Raiser" – Layman: "Raise eyelids"
AU6 – "Cheek Raiser" – Layman: "Raise cheeks"
AU7 – "Lid Tightener" – Layman: "Squint eyes"
AU9 – "Nose Wrinkler" – Layman: "Wrinkle up nose"

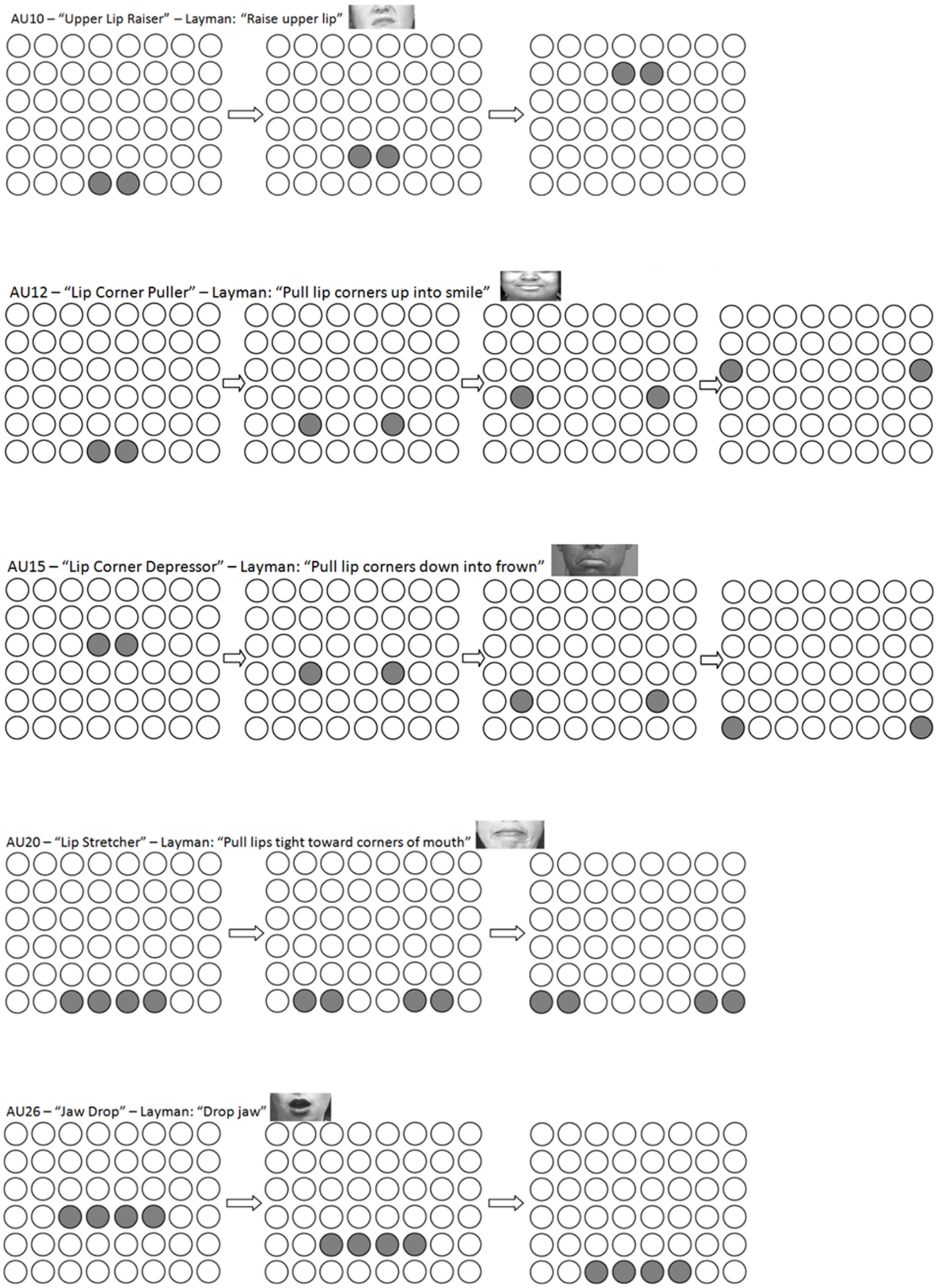


Figure 1. Visual depiction of proposed mapping from facial action units to spatiotemporal vibrotactile representations. Arrows convey sequence of spatial pattern presentation, and filled in circles indicate actuated vibration motors. While designed for a 6 by 8 two-dimensional array of vibration motors, the proposed patterns could be adapted for smaller or larger displays. For each facial action pattern above, included is its number (e.g., AU15), technical name (e.g., "Lip Corner Depressor"), and layman name ("Pull lip corners down into a frown"). Referencing action units in their layman terminology simplifies procedures for user study participants (see Section 4).

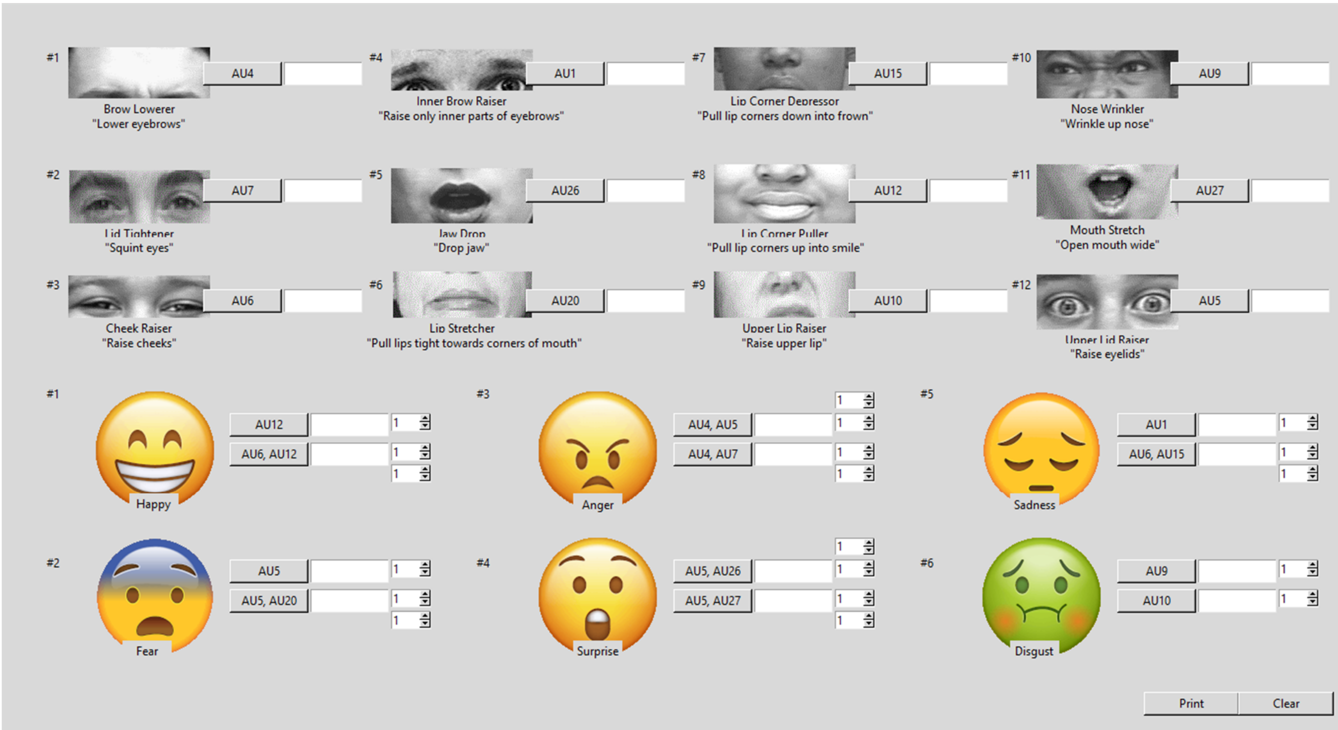


Figure 2. Screenshot of graphical user interface for experiment presented in Section 4. The top portion of the user interface was used to run Part 1 of the experiment, and the bottom portion was used to run Part 2.

4 Experiment

4.1 Aim

The purpose of this experiment is two-fold: (i) Evaluate the proposed set of vibrotactile facial action units in terms of distinctness and intuitiveness; and (ii) Understand how well individuals who are blind can learn and recognize emotions from tactile representations of facial movements. We employ an absolute identification (AI) paradigm (i.e., patterns are randomly presented and recognized in isolation) to better gauge learnability, recognition, and naturalness of each individual pattern design. Future work will improve ecological validity by introducing contextual factors commonly found in social interactions including verbal cues (speech content and nonverbal sounds).

4.2 Participants

A total of 8 participants who self-identify as either blind or visually impaired were recruited for an Institutional Review Board (IRB) approved study. Of the 8, six participants passed training, and therefore, data presented below are for the following 6 participants: 4 females, 2 males; 5 identified as blind, 1 as visually impaired; 1 was born blind, 5 became blind later in life; with an age range of 24 to 63 (*M*: 42.5, *SD*: 14.5). All but one subject participated in previous studies by the research team, using a similar apparatus as described in Section 4.3. To reduce unwanted learning effects, a few months have passed between studies.

4.3 Apparatus

A two-dimensional array of eccentric rotating mass (ERM) vibration motors, placed on the back of an ergonomic mesh chair, was used as the display for this experiment. The display consists of 6 rows and 8 columns of actuators spaced 2 cm horizontally (center-to-center) and 5 cm vertically (center-to-center). To ensure sufficient localization accuracy, spacing was guided by research findings of van Erp [16] who explored vibrotactile spatial acuity limitations on the back and torso. The display was embedded on the back of a chair to simulate how such an assistive technology might be used in real settings. Hardware was custom built using the Arduino open source development platform. All software was implemented in Python.

4.4 Procedure

Participants first read and signed an informed consent form to provide their voluntary consent to participate. They then were given $25 compensation for their time. Participants then completed a subject information form to collect basic demographics. During the entire duration of the study, facial action units were referred to in their layman terminology (Fig. 1).
Participants then underwent a two-part study. Part 1 began with a familiarization phase, followed by a training phase. During the

familiarization phase, participants were introduced to each of the 12 vibrotactile facial action units. Two passes through all 12 patterns were conducted, and participants were given an opportunity to request repeats of specific patterns. Next, during the training phase, participants were asked to recognize each of the 12 vibrotactile facial action units. Correct guesses were confirmed, and incorrect guesses were clarified. Participants were allowed to request a pattern be repeated as many times as needed before they made a guess. Patterns were randomly presented.

Part 2 consisted of three phases: Familiarization, training, and testing. During the familiarization phase, participants were introduced to 12 different vibrotactile facial action unit combinations, each representing one of the six basic emotions. Similar to Part 1, two passes through all 12 patterns were conducted, and participants were given an opportunity to request repeats. During the training phase, participants were asked to recognize the vibrotactile facial action units contained within each combination pattern as well as the associated emotion. Participants could guess in any order they preferred. Correct guesses were confirmed, and incorrect guesses were corrected. As many repeats as desired were allowed before participants made a guess. Patterns were randomly presented. To move to Part 2 testing, participants had to score 80% or better (at least 10 out of 12 patterns correctly guessed); otherwise, the Part 2 training phase was repeated. A maximum of four training phases were allowed. Part 2 testing involved the presentation of 4 pattern sets, where each set consisted of all 12 vibrotactile facial action unit combinations, randomly ordered. Participants were asked to recognize the facial action units and associated emotion, as before, but without confirmation from the experimenter. Participants were allowed to ask for repeats before guessing. Finally, participants completed a brief post-experiment questionnaire.

4.5 Results

Due to consistent malfunctions during the presentation of the AU6 Cheek Raiser pattern, participant responses to facial action unit combinations with this action unit were thrown out; these included AU6+AU12 and AU6+AU15. Three individual actuators ceased operation before the first participant, and therefore, were inoperable for all participants. All three motors were on the last row (row #8, columns #2, #6, and #7), and affected only vibrotactile facial action units AU20 and AU26. Pilot tests revealed these changes to have little impact on the perception of the proposed vibrotactile facial action units given the redundancy encoded in the spatiotemporal patterns.

In this paper, we define *recognition accuracy* as the number of presentations guessed correctly divided by the total number of presentations during the testing phase. Average Part 2 training phase attempts was *M*: 2.7, *SD*: 0.8. Average facial action unit combination recognition accuracy was *M*: 92.5, *SD*: 19.1. Average emotion recognition accuracy was *M*: 86.2, *SD*: 24.5. Average recognition accuracy for both emotions and facial action unit combinations was *M*: 84.1, *SD*: 26.4. Recognition accuracies for individual facial action unit combinations are depicted in Fig. 3. Recognition accuracies for individual emotional expressions are depicted in Fig. 4.

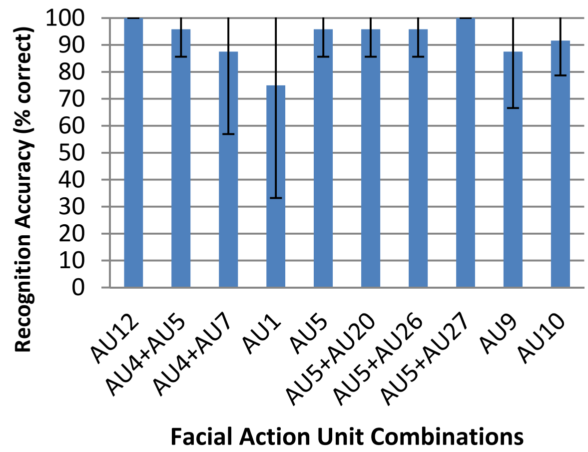


Figure 3. Mean recognition accuracy per facial action unit combination. Error bars are standard deviations.

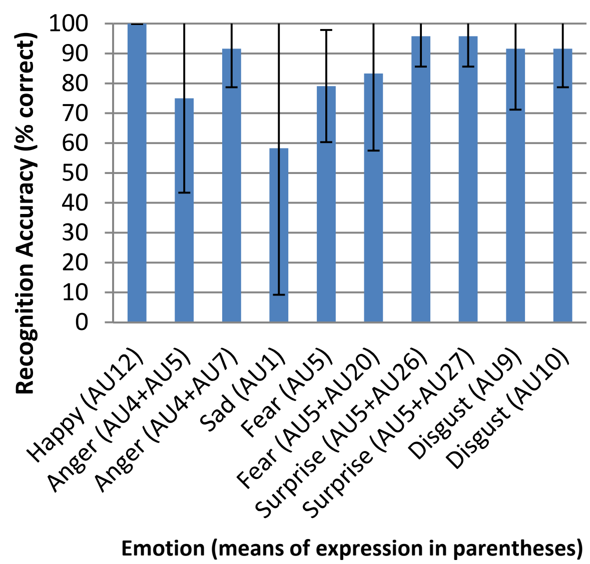


Figure 4. Mean recognition accuracy per emotion. Happy and Sad were expressed through one method each; all other emotions had at least two expressions. Error bars are standard deviations.

Table 1 and 2 summarize participants' subjective responses for the following questions, respectively: "How easy was it to recognize emotions based on vibration patterns represented by the following facial action units?" and "How natural (intuitive) was the mapping between emotion and vibration patterns represented by the following facial action units?"

Table 1. Mean participant responses to the question "How easy was it to recognize emotions based on vibration patterns represented by the following facial action units?" using a Likert scale from 1 (very hard) to 5 (very easy).

Questions	M	SD
Happy "Pull lip corners up into smile"	4.6	0.5
Anger "Lower eyebrows" + "Raise eyelids"	4.1	0.7
Anger "Lower eyebrows" + "Squint eyes"	4.3	0.8
Sad "Raise only inner parts of eyebrows"	3.5	1.5
Surprise "Raise eyelids" + "Drop jaw"	3.8	1.1
Surprise "Raise eyelids" + "Open mouth wide"	4.3	0.8
Fear "Pull lips tight toward corners of mouth"	3.5	1.4
Fear "Raise eyelids" + "Pull lips tight…"	3.8	0.8
Disgust "Wrinkle up nose"	4.5	0.8
Disgust "Raise upper lip"	4.1	0.9

Table 2. Mean participant responses to the question "How natural (intuitive) was the mapping between emotion and vibration patterns represented by the following facial action units?" using a Likert scale 1 (very hard) to 5 (very easy).

Questions	M	SD
Happy "Pull lip corners up into smile"	4.8	0.4
Anger "Lower eyebrows" + "Raise eyelids"	3.3	0.5
Anger "Lower eyebrows" + "Squint eyes"	3.6	0.8
Sad "Raise only inner parts of eyebrows"	2.6	1.6
Surprise "Raise eyelids" + "Drop jaw"	4.3	1.2
Surprise "Raise eyelids" + "Open mouth wide"	4.5	1.2
Fear "Pull lips tight toward corners of mouth"	2.8	1.1
Fear "Raise eyelids" + "Pull lips tight…"	3.0	1.0
Disgust "Wrinkle up nose"	4.3	1.0
Disgust "Raise upper lip"	4.0	1.2

4.6 Discussion

The majority of participants passed the training phase of Part 2 (6 out of 8 participants) with an average of 2.7 passes through training needed. This is impressive considering the limited familiarization and training participants underwent before they had to associate and learn emotional meanings of facial action unit combinations. Two one-way repeated measures (RM) Analysis of Variance (ANOVA) were performed to explore the effect of pattern type on emotion recognition and facial action unit recognition, respectively. All data assumptions for repeated measures ANOVA were met. An alpha value of 0.01 was selected and divided by the number of dependent variables (k=2) to account for the multiple significance tests. The final alpha value used was 0.005. No statistically significant difference was found between recognition accuracies for emotion, $F(9,45) = 2.178$, $p = 0.042$, nor facial action unit combinations, $F(9,45) = 0.939$, $p = 0.501$. No significant differences indicate that participants, overall, found no particular pattern more difficult to recognize in terms of its emotional content nor facial action units compared to other patterns.

While some participants clearly struggled with recognizing *Sad* (AU1)—see Fig. 3 and 4 as well as subjective ratings for both ease of recognition and intuitiveness in Table 1 and 2—many participants did not have difficulty, resulting in large variability in terms of recognition accuracy for this pattern (see error bars in aforementioned figures). Other difficulties worth noting are that some participants had challenges recognizing emotions from facial action unit combinations representing *Anger* and *Fear* (see Fig. 4). Indeed, we do see low subjective marks for ease of recognition and naturalness of the mappings in Table 1 and 2. Even though the majority of participants performed well in terms of both emotion and facial action unit recognition, struggle with respect to the aforementioned emotions among some participants warrants attention.

Facial action unit combinations for *Fear* were incorrectly classified as *Surprise* 8 times. Both expressions of *Fear* as well as *Surprise* involve AU5 (raising eyelids) together with variations in mouth movements (stretching the lips, opening the mouth, and dropping the jaw). We hypothesize that confusion between these two emotions was a result of the significant overlap, forcing participants to have to rely on recognizing subtle differences in mouth movements. Similarly, facial action unit combinations for *Anger* were incorrectly classified as *Fear* 4 times, most likely due to the overlap between these patterns in terms of AU5 (raise eyelids). Again, such overlap forces users to rely on other cues, in this case, subtle eyebrow or eye movements. We hypothesize that with further training, participants will be able to better recognize subtle differences between expressions to become more proficient at interpreting emotions. Since *Sad* was expressed through AU1 (raise only inner parts of eyebrows), difficulties with recognizing this facial action unit impacted the association of this pattern with the correct emotion. The vibrotactile facial action unit for AU1 will be redesigned to develop an easier to recognize and more natural pattern for this facial movement.

5. Conclusion

This work proposed a novel vibrotactile facial action unit set, and is the first work to explore how well individuals who are blind can perceive emotions from vibrotactile representations of facial movements. The results of this work may be useful in building social assistive aids for those with visual impairments toward enhancing access to the non-verbal information present in social interactions. The findings of the proposed experiment show promise in terms of participants' ability to learn a mapping between emotions and facial action units, especially given such short training periods.

Future work will explore: (i) The role of visual experience between participants born blind and participants who acquired blindness later in life toward emotion understanding from facial movements. We hypothesize that visual experience of facial movements may ease the training process of associating emotions to facial action units. Given that the current study had only one participant who was born blind, this question will be explored as part of future work when more subjects from this target group are recruited; (ii) Many parameters of the current study were held constant, and therefore need to be investigated further. For example, a gap of 1 s was introduced between facial action units; we hypothesize that with further training this gap could be reduced to speed up communication. While the presentation of facial action unit combinations were randomized, the ordering of individual facial action units within each combination were not. For example, AU5+AU26 was always presented as AU5 followed by AU26. As part of future work, this ordering will be randomized to understand its effect; (iii) Contextual factors will be investigated toward achieving better ecological validity including multimodal presentations (e.g., facial action units with speech cues); and (iv) A longitudinal study will be conducted to understand how competence changes with continual training and use.

REFERENCES

[1] M.L. Knapp. 1996. *Nonverbal communication in human interaction.* Harcourt College.

[2] C. Segrin, J. Flora. 2000. Poor Social Skills Are a Vulnerability Factor in the Development of Psychosocial Problems. Human Communication Research, 26(3), 489–514.

[3] Seeing Machines, https://www.seeingmachines.com, last accessed 2017/11/24.

[4] IMOTIONS, https://imotions.com, last accessed 2017/11/24.

[5] F. De la Torre et al. 2015. IntraFace. In *Proceedings of 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition, 8 pages.*

[6] P. Ekman, W.V. Friesen, J.C. Hager. 2002. The Facial Action Coding System: A Technique for the Measurement of Facial Movements. Consulting Psychologists.

[7] M.F. Valstar, M. Pantic. 2006. Biologically vs. Logic Inspired Encoding of Facial Actions and Emotions in Video. In *Proceedings of IEEE International Conference on Multimedia & Expo, 4 pages.*

[8] T. McDaniel, D. Villanueva, S. Krishna, D. Colbry, S. Panchanathan. 2010. Heartbeats: A Methodology to Convey Interpersonal Distance through Touch. In *Proceedings of ACM Conference on Human Factors in Computing Systems, 6 pages.*

[9] T. McDaniel, S. Krishna, V. Balasubramanian, D. Colbry, S. Panchanathan. 2008. Using a Haptic Belt to Convey Non-Verbal Communication Cues During Social Interactions to Individuals Who Are Blind. In *Proceedings of IEEE Haptics Audio-Visual Environments and Games Conference, 6 pages.*

[10] S. Qiu, M. Rauterberg, J. Hu. 2016. Designing and Evaluating a Wearable Device for Accessing Gaze Signals from the Sighted. In M. Antona, C. Stephanidis (eds.) *Universal Access in Human-Computer Interaction, LNCS, vol. 9737.* Springer, Cham, pp. 454–464.

[11] S.U. Réhman, L. Liu. 2008. Vibrotactile Rendering of Human Emotions on the Manifold of Facial Expressions. Journal of Multimedia, 3(3), 18–25.

[12] H.P. Buimer, M. Bittner, T. Kostelijk, T.M. van der Geest, R.J.A. van Wezel, Y. Zhao. 2016. Enhancing Emotion Recognition in VIPs with Haptic Feedback. In C. Stephanidis (ed.) *HCI International 2016 – Posters' Extended Abstracts, vol. 618.* Springer, Cham.

[13] S. Krishna, S. Bala, T. McDaniel, S. McGuire, S. Panchanathan. 2010. VibroGlove: An Assistive Technology Aid for Conveying Facial Expressions. In *Proceedings of ACM Conference on Human Factors in Computing Systems, 6 pages.*

[14] A. Rahman, A.I. Anam, M. Yeasin. 2017. EmoAssist: Emotion Enabled Assistive Tool to Enhance Dyadic Conversation for the Blind. Multimedia Tools and Applications, 76(6), 7699–7730.

[15] S. Bala, T. McDaniel, S. Panchanathan. 2014. Visual-to-Tactile Mapping of Facial Movements for Enriched Social Interactions. In *Proceedings of IEEE International Symposium on Haptic, Audio and Visual Environments and Games, 6 pages.*

[16] J.B. van Erp. 2005. Vibrotactile Spatial Acuity on the Torso: Effects of Location and Timing Parameters. In *Proceedings of First Joint Eurohaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems, 6 pages.*