VibroGlove: An Assistive Technology Aid for Conveying Facial Expressions

Sreekar Krishna, Shantanu Bala, Troy McDaniel,
Stephen McGuire and Sethuraman Panchanathan

Center for Cognitive Ubiquitous Computing (CUbiC)
School of Computing, Informatics and Decision Systems Engineering
Ira A. Fulton School of Engineering
Arizona State University
Tempe AZ 85281, USA
Ph: (480) 727 3612
Fax: (480) 965 1885
http://cubic.asu.edu

Sreekar.Krishna@asu.edu,
Shantanu.Bala@asu.edu,
Troy.McDaniel@asu.edu,
Panch@asu.edu

Abstract

In this paper, a novel interface is described for enhancing human-human interpersonal interactions. Specifically, the device is targeted as an assistive aid to deliver the facial expressions of an interaction partner to people who are blind or visually impaired. Vibro-tactors, mounted on the back of a glove, provide a means for conveying haptic emoticons that represent the six basic human emotions and the neutral expression of the user's interaction partner. The detailed design of the haptic interface and haptic icons of expressions are presented, along with a user study involving a subject who is blind, as well as sighted, blind-folded participants. Results reveal the potential for enriching social communication for people with visual disabilities.

Keywords

Vibrotactile glove, Basic facial expressions, Bilateral interpersonal interaction, Haptic interface

ACM Classification Keywords

H.5.2. Information interfaces and presentation (e.g., HCI): User Interfaces, *Haptic I/O*. K.4.2. Computing Milieux: Computers and Society, *Assistive technologies for persons with disabilities*.
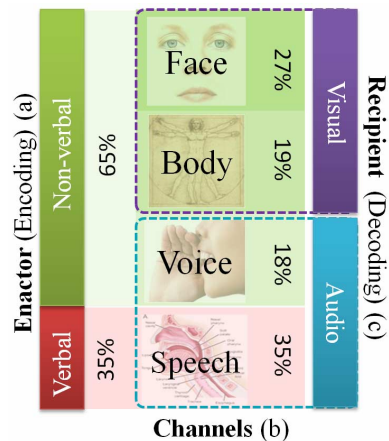
Figure 1: Relative importance of a) verbal vs non-verbal cues, b) four channels of non-verbal cues, and c) visual vs. audio encoding & decoding of bilateral human interpersonal communicative cues.
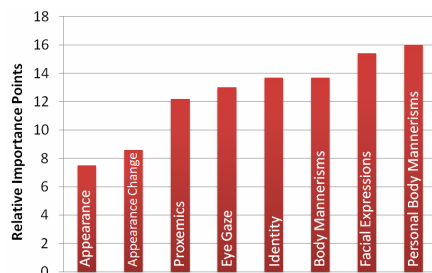

Figure 2: Self report importance (scaled over 100 points) of visual non-verbal cues obtained through an online survey of target population and specialists [4].

General Terms
Design, Human Factors, Experimentation, Verification

Introduction
Nearly 65% of all human interpersonal communications happen through non-verbal communication cues [3]. In a bilateral interpersonal interaction, while speech encodes all the information, non-verbal cues facilitate an elegant means for delivery, interpretation and exchange of this verbal information. For example, eye gaze, iconic body or hand gestures, and prosody enable effective and seamless role play in social interpersonal interactions. People communicate so effortlessly through both verbal and non-verbal cues in their everyday social interactions that they do not realize the complex interplay of their voice, face and body in establishing a smooth communication channel. Nearly 72% of non-verbal communication [1] takes place through visual cues encoded on the face and body of the interaction partners (see Figure 1). Unfortunately, people who are blind or visually impaired cannot access this huge portion of interpersonal information independently.

While most persons who are blind or visually impaired eventually make accommodations for the lack of visual information, and lead a healthy personal and professional life, the path towards learning effective accommodations could be positively effected through the use of assistive aids. Specifically, children with visual disabilities find it very difficult to learn social skills while growing amongst sighted peers, leading to social isolation and psychological problems [2]. Social disconnect due to visual disability has also been observed at the college level [7] where students start to learn professional skills and independent living skills.

Any assistive technology aid that can enrich interpersonal social interactions could prove beneficial for persons who are visual disabled.

Motivation
In order to understand the importance of visual social cues, we conducted a web based survey where the participants (16 persons who are blind, 9 with low vision and 2 sighted specialists in the area of visual impairment) rated the importance of 8 social needs that were identified from two open-ended focus groups [4]. The participants responded on a 5 point Likert scale; 5, implying strong agreement, to 1, implying strong disagreement. Figure 2 shows a non-parametric rank average analysis of the participants' responses. The rank-ordered social needs list shows that participants' most important need corresponds to feedback on their own body mannerism and how it was affecting their social interactions. Following this was their need to access facial expressions, body mannerisms, identity, eye gaze, proxemics (location) and appearance of their social interaction partners, in the presented order. Recently, in [5], we proposed and demonstrated a methodology to detect stereotypic body mannerism (body rocking) towards providing social rehabilitation for people who are blind or visually impaired. Focusing on the next important social need of accessing facial mannerisms, in this paper, we propose a methodology to deliver facial expressions of the social interaction partner to a person who is visually disabled.

Design Considerations
The human face is very dynamic when it comes to generating important non-verbal communicative cues. Subtle movements in the facial features can convey great amounts of information. For example, slight
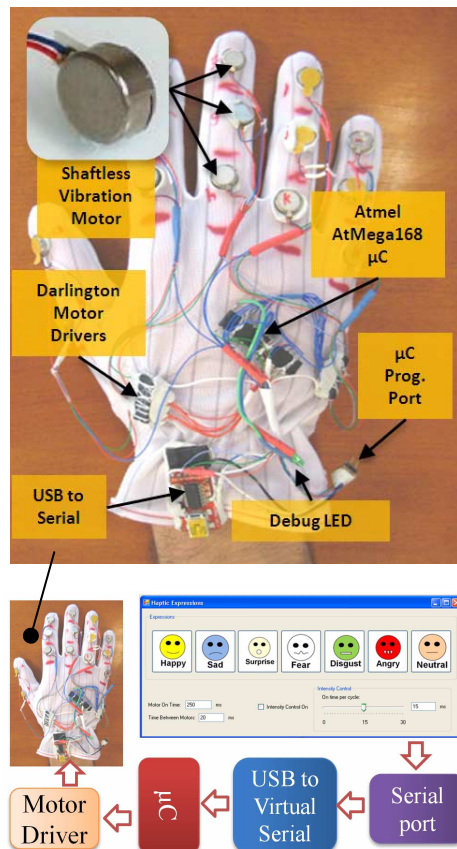
Figure 3: System level architecture of the VibroGlove. Software running on a PC controls three dimensions of the vibration patterns, including magnitude, location and duration. Temporal concatenations of vibrations results in the haptic pattern representing the facial expressions.

opening of the eyelids conveys confusion or interest, whereas a slight closing of the eye lids conveys anger or doubt. Thus, the human face can be considered to be a very high bandwidth information stream, where careful design considerations need to be taken into account if this data has to be encoded optimally and effectively through other modalities.

In the past, most researchers and technologists have resorted to auditory cueing when information has to be delivered to persons with visual disabilities; but there is a strong growing discomfort in the target population when it comes to overloading their hearing. People with visual disabilities have a natural tendency to accommodate for the lack of a primary sensory channel by relying on hearing. For example, with the aid of ambient noise in a room, they can gauge approximately how big a room is. Thus, when designing assistive devices aimed at social aid, we need to carefully consider how to deliver high bandwidth data streams to users relating to the facial movements of interaction partners. Touch or haptic based delivery is a growing area of research which is relatively underutilized, except for Braille. To this end, we explore the use of vibrotactile cueing on the back of the human palm (the human hand has a very large representation in the somatosensory cortex of the brain) to be both versatile and unobtrusive.
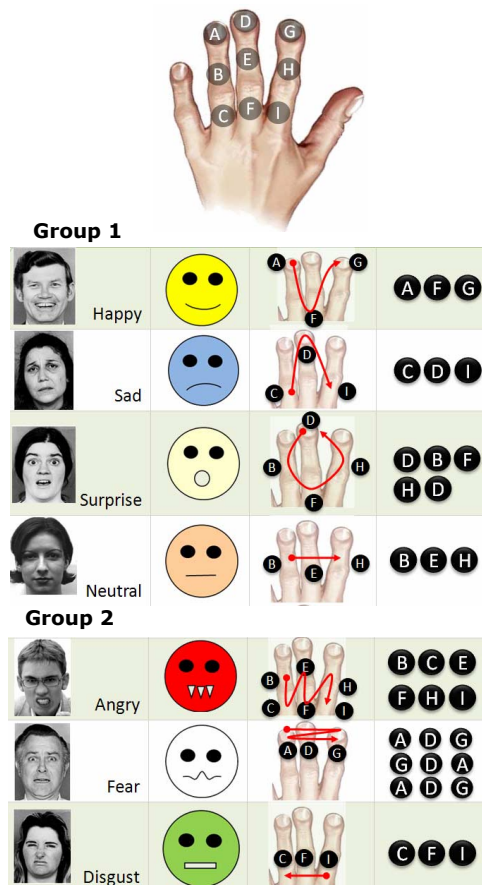
Related Work
Very few researchers have addressed the development of social assistive devices for persons with visual disabilities. Only recently, social assistance has started to emerge as an important assistive technology problem [2] [7]. To the best of the authors' knowledge,

only one related work focuses on the issue of delivering facial information to people who are blind: [6] developed a *haptic chair* for presenting facial expression information to people who are blind. A chair was equipped with vibrotactile actuators on its back rest forming an inverted Y. A camera mounted elsewhere tracks the mouth of an interaction partner, and actuators vibrate along any one of the three axes of the Y, based on whether the interaction partner was neutral, happy, sad or surprised. No formal experiments were conducted with the target population, except for a brief pilot study with sighted students. Further, this solution had the obvious limitation that users need to be sitting in the chair to use the system. In this paper, we discuss a more versatile solution that is portable, and has very high potential for future extensions.

The VibroGlove
Construction:
The VibroGlove consists of 14 tactors (vibration motors) mounted on the back of the fingers, one per phalange. The 14 motors correspond to the 14 phalanges (3 each on the index, middle, ring and little finger, and 2 on the thumb) of the human hand. Figure 3 shows the vibrotactile glove with the details of its implementation. Pancake shaftless vibrators are used as the primary vibrotactile actuator, which provides a net lateral vibration across the skin, above the phalanges of the fingers. The control software (implemented on a Windows OS based PC) interacts with the glove enabling or disabling motors according to pre-programmed spatio-temporal vibration patterns. The glove functions independently via USB interface without the need for external power.

Figure 4: Mapping of Group 1 and Group 2 haptic expression icons to the central three fingers (9 Phalanges) of the vibrotactile glove. Columns 1 to 3 represent the expression. Column 4 shows the spatial mapping of vibrations. Column 5 shows the temporal mapping of the vibrations.

Haptic Expression Icons:

While the versatility of the VibroGlove allows it to be used for various applications, here we discuss the specific application of delivering the six basic facial expressions, along with the neutral face, of an interaction partner to a user who is visually disabled. Humans rely heavily on the shape of the mouth and the eye area to decipher facial expressions. Motivated from this, we focused only on the mouth area to design spatio-temporal haptic alternates for facial expressions. We used only the three central fingers on the glove: 9 vibrators, as shown in Figure 4. In order to represent the seven facial expressions, we designed haptic expression icons that were motivated by two important factors: 1) Icons similar to the *visual emoticon* that are already in popular use, like Happy, Sad, Surprise and Neutral, where the mouth shapes prominently represent the expression, and 2) Icons like Anger, Fear and Disgust where the mouth area alone does not convey the expression, thereby forcing us to create haptic icons that could evoke a sense of the expression in question. Figure 4 provides details of the haptic expression icons. All 7 patterns were designed to be 750ms long with each motor vibrating for at least 50ms. These numbers were determined based on pilot studies where we found that participants could not isolate vibrations if the duration was less than 50ms long. Further, patterns longer than 800ms were considered to be too long by the participants, while patterns shorter than 600 ms were confusing, and training phase accuracies were unacceptable.

GROUP 1 – THE VISUAL EMOTICON MOTIVATED HAPTIC ICONS:
The Group 1 haptic expression icons primarily represent popular emoticons that are in wide use within the Instant Messaging community. These icons mostly

model the shape of the mouth. 1) *Happy* is represented by a U shaped pattern, 2) *Sad* by an inverted U, 3) *Surprise* by a circle, and 4) *Neutral* by a straight line.

GROUP 2 – THE AUXILIARY HAPTIC ICONS:
Anger, Fear and Disgust cannot be conveyed through the mouth appearance alone. To this end, we resorted to defining haptic patterns that were unique from what was already defined for Group 1, while keeping in mind a need to represent the underlying expression in question. 1) *Anger* is represented by successive vibrations on six lower phalanges representing an open mouth showing its teeth during an expression of anger; 2) *Fear* is represented by very brief vibrations to the top phalanges (tips of the central fingers) in three quick successive vibration sequences representing a fast emotional response that people show towards fear, and 3) *Disgust* is represented through a vibration pattern going from right to left on the bottom phalanges of the central fingers corresponding to a slightly opened mouth during the display of disgust.

Experiment

The primary goal of the experiment was to determine how well participants were able to recognize the seven haptic patterns. Along with the accuracy of recognizing the spatio-temporal vibrotactile cues, we were also interested in knowing how quickly the participants were able to recognize the expressions. The duration for recognition is very important in social interactions as the human face changes drastically over short time. Experiments have shown that expressions vary anywhere from 1 to 5 seconds ([3], Page 322). Conforming to these time scales, it is important that any device developed towards enriching social
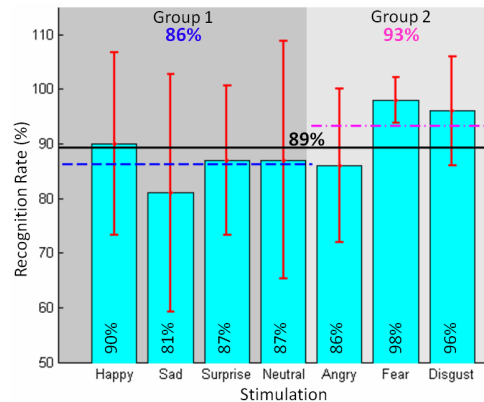
Figure 5: Recognition rate across all 12 participants. Individual expression, Group 1, Group 2, and the overall recognition rate are presented in the graph. SD of the recognition rates across subjects are shown in red.
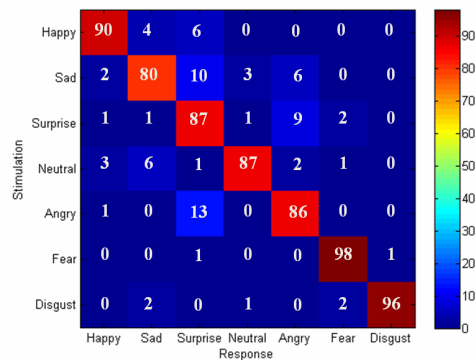

Figure 6: Confusion Matrix across the 12 participants. The rows are the stimulation and the columns are the responses of the participants. Each row adds to 100 % (rounding error of 1%).

experience should react in real-social-time towards facilitating smooth interpersonal interaction.

Participants: The experiment was conducted with one individual who is blind and 11 other participants who are sighted, but were blindfolded during the experiment. It is important to note that the individual who is blind had lost his sight after 25 years of having vision. To a large extent, this individual could correlate Group 1 haptic expression icons to his visual experiences from the past.

Procedure: Once the subjects wore the glove, they were seated in a chair with a blindfold and asked to keep their hand on their lap in the most comfortable position. Subjects were first familiarized with all 7 vibration patterns by presenting them in order, during which time the expression corresponding to the pattern was spoken aloud by the experimenter. The familiarization was continued until the subjects were comfortable in remembering all the seven expressions. This was followed by the training phase in which all seven patterns were presented in random order, in multiple sets, and subjects were asked to identify the expressions by depressing an appropriate key on a keyboard. The experimenter confirmed any correct response, and corrected incorrect responses. Subjects had to demonstrate 100% recognition on one set of all 7 expressions before moving to the testing phase. A 15 minute time limit was placed on the training irrespective of the training accuracy. The testing phase was similar to the training phase except the experimenter did not provide feedback to subjects, and each expression pattern was randomly presented 10 times making a total of 7 expressions x 10 = 70 trials. The subjects were given 5 seconds per trial to respond.

Results and Discussions

Recognition Accuracies:
Figure 5 shows the average recognition rate across all 12 participants for the seven haptic patterns. The overall recognition rate was 89%, with a one-way ANOVA [$F(6,77)=1.71$, $p=0.129$] supporting our first hypothesis that the responses across the seven expressions did not differ significantly. Our null hypothesis regarding the two groups was that there would be no significant difference in performance, and if the null hypothesis is rejected, Group 1 would perform better as the expressions were motivated by popular visual emoticons. A one-way ANOVA between groups rejected the null hypothesis [$F(1,82)=4.24$, $p=0.042$)] showing a difference between group performance. A Tukey test on the two group means $M_1=86.28$ & $M_2=93.46$, gave a standard error of $T_s=4.3$, which is less than the first mean difference ($M_2-M_1=7.17$). Thus, Group 2 performance was much higher than Group 1 rejecting the extension to the null hypothesis. Studies are underway to determine the nature of the haptic cues in Group 2 that make them significantly better than Group 1.

Figure 6 shows the confusion matrix for all the seven expressions. The diagonals correspond to the bar graph shown in Figure 5. The off-diagonal elements represent the confusion between expressions. These off-diagonal elements provide insight into the parameters that control effective and responsive haptic patterns. While subjects confused Sad and Neutral expressions with various others (mostly in Group 1), Anger and Surprise show exchangeability, where there is strong confusion between each other. Fear and Disgust are strongly isolated from the rest of the expressions as they were very well recognized by the subjects.
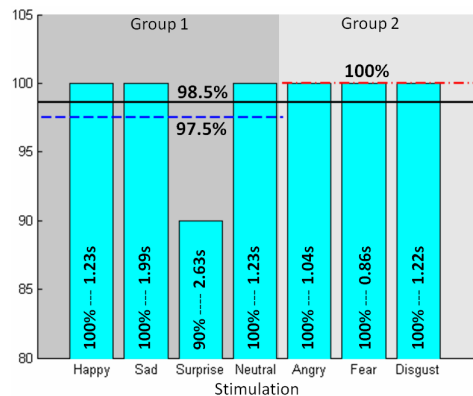
Figure 7: Average recognition rate and response time for the subject who is blind, for over 70 trails.
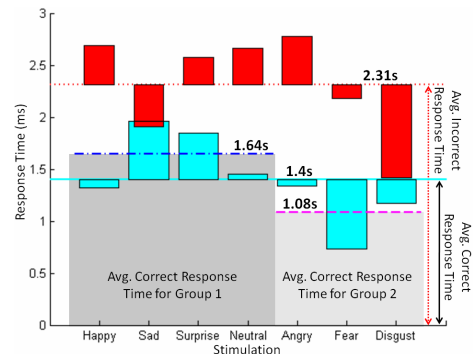

Figure 8: Average response time for all 12 participants. Four important results are shown above, 1) Avg. correct response time per expression (Cyan), 2) Avg. incorrect response time per expression (Red), 3) Avg. correct response time for Group 1 (Blue), and 4) Avg. correct response time for Group 2 (Magenta).

Figure 7 shows the average recognition performance and the average time of response for the subject who is blind. The individual was able to recognize most of the expressions at 100%, over the 70 trails.

Time for Recognition:
Figure 8 shows the average time taken by the subjects per expression when they recognized the haptic patterns correctly (cyan), and when they misclassified them (red). The bar graph shows excess or shortage of response time around the mean value. It can be seen that correct identification happened in just over a second (1.4s). When the subjects were not sure of the haptic pattern, they took more time to respond. This can be seen from the inverse correlation of the response time and recognition rates in Figure 5. The pattern for Sad had the worst performance of 81% and the corresponding response time was the highest (2s). Pattern for Fear had the best performance (98%) and least response time (765ms). This analysis can be extended to the Group level where Group 1 has a higher recognition time when compared to Group 2. Whenever the subjects responded wrong, they seem to take more time, as seen by the average incorrect response time of 2.31s (red), almost a second more than the response time for correct responses. We could not find any significant relevance between the response time for incorrect answers and the recognition rate graph. We conclude that subjects were responding with random answers once they crossed a self imagined time limit less than the 5 seconds that was provided.

Conclusion and Future Work
In this paper we demonstrated a novel interface, a vibrotactile glove, used as an assistive device for delivering seven facial expressions to persons who are

visually disabled. Results are convincing that it is possible to convey basic facial expressions through haptic interfaces. Work is in progress to make the system more dynamic for delivering all facial movements, thereby allowing the user to make the judgment of what facial expressions someone is displaying. This would allow independent access to all interpersonal communicative cues, and not just the basic expressions. Efforts are underway to recruit more individuals who are blind and visually impaired towards testing the efficacy of the system.

References
[1] N. Ambady and R. Rosenthal, "Thin Slices of Expressive behavior as Predictors of Interpersonal Consequences : a Meta-Analysis," *Psychological Bulletin*, vol. 111, 1992, pp. 274, 256.
[2] D. Jindal-Snape, "Generalization and Maintenance of Social Skills of Children with Visual Impairments: Self-Evaluation and the Role of Feedback," *J. of Visual Impairment and Blindness*, vol. 98, 2004, pp. 470-483.
[3] M.L. Knapp, *Nonverbal Communication in Human Interaction*, Harcourt College Pub, 1996.
[4] S. Krishna, D. Colbry, J. Black, V. Balasubramanian, and S. Panchanathan, "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 (CVAVI 08), ECCV 2008*.
[5] S. Krishna, N.C. Krishnan, and S. Panchanathan, "Detecting Stereotype Body Rocking Behavior through Embodied Motion Sensors," *Annual RESNA Conference*, 2009.
[6] S. Rehman, L. Liu, and H. Li, "Vibrotactile Rendering of Human Emotions on the Manifold of Facial Expressions," *J. of Multimedia*, vol. 3, 2008, pp. 18-25.
[7] K. Shinohara and J. Tenenberg, "A blind person's interactions with technology," *Commun. ACM*, vol. 52, 2009, pp. 58-66.