Exploring the Dorsal Surface of the Fingers for Visuo-Haptic Sensory Substitution

Sreekar Krishna, Shantanu Bala, Sethuraman Panchanathan
Center for Cognitive Ubiquitous Computing (CUbiC)
Arizona State Univeristy, Tempe AZ 85281
Sreekar.Krishna@asu.edu, Shantanu.Bala@gmail.com, Panch@asu.edu

Abstract— **In this paper, a novel haptic interface is presented for exploring the dorsal surface of the finger phalanges as visuo-haptic sensory substitution modality. The constructional details of the proposed haptic glove are detailed along with two experiments that explore vibrotactile localization and spatio-temporal cue identification among users who are blind and sighted, but blind-folded. The use of the dorsal surface of the fingers ensures that there is no functional inconvenience to the fingers or the hand. The results of the experiments are promising and offer an effective means for delivering high intensity (bandwidth) visual data in auxiliary sensory channels when vision is deprived or overloaded.**

Keywords: Vibrotactile Interfaces, Sensory Substitution, Spatio-temporal Cueing, Affect Interfaces.

I. Introduction

Over the past few decades, the environments in which humans live and operate have become increasingly information rich. Audio and visual modalities have been occupied by more than one source at the same time. Increasing audio-visual stimulations in the human's surroundings have increased the need for divided attention, which competes with the need for selective and sustained attention to complete a task at hand [1]. While audio and video have evolved as a medium for immersing humans in rich sensory experience, touch [2], taste [3] and smell [4] have only recently being considered for sensory augmentation and substitutions – note the subtlety between augmentation and substitution. When augmenting the already utilized sensory channels, the newer medium is not in demand to reproduce all of the information, but only enrich the already delivered experience. On the other hand, substitutions have to deliver information that was once being provided by a certain sensory channel on a newer medium, while maintaining similar or lesser cognitive load.

Vision is the primary sense organ for most mammals and for a few primates, including humans, trichromatic vision is so highly evolved [5] that a major portion of the neuronal pathway in the brain is dedicated to sensing, perceiving and cognizing visual stimuli. This allows the human vision to process high intensities of data that stimulates the eyes – Koch et. al. estimated that human eyes, with 10^6 ganglion cells, could transmit up to 10 Mbits/s [6] to the brain. Hence substituting vision with any other sensory channel is a challenging task that requires appropriate design taking into account the high intensity of data generated by visual stimuli.

In this paper, we present a visuo-haptic sensory substitution device that intends to replace visual channel with somatosensory vibrotactile stimulations for specific applications. The high intensity visual stimuli are mapped to a matrix of haptic actuators placed in contact with the dorsal surface of the fingers, which allows both spatial and temporal mapping of vibration patterns. The fingers have the largest tactile representation in the brain after the tongue. Together, the fingers have the largest projection on the cortical surface when compared to any other body part [7]. The concentrated neuronal mapping of the fingers allow for a very high sensitivity (both spatial and temporal resolution) making them an ideal candidate for sensory substitution. Further, to allow functional operation of the user's hands, the vibratos are placed on the dorsal surface of the fingers.

In order to test the effectiveness of the visuo-haptic sensory substitution of the proposed vibrotactile glove, we explore one specific application area of delivering facial expressions via the glove. We targeted an assistive technology to translate facial expressions of social interaction partners into haptic spatio-temporal cues which could then be interpreted by people who are blind. 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 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 are needed if this data has to be encoded optimally and effectively through haptic modalities.

II. Related Work

A prominent visuo-haptic sensory substitution system is the TVSS (Tactile-Visual Sensory Substitution) [8] that substitutes visual data into a 400 point tactile array worn on the back of the user. A similar effort by Rahman et. al. [9] focused on delivering facial emotions of interaction partners using vibrators installed on a chair such that the user's back is in direct contact with the vibrators. Both of these technologies have the obvious disadvantage that the user is restricted to a seated position with immobile augmentations. Recently we

978-1-4244-6509-5/10/$26.00 ©2010 IEEE

have explored the use of vibrotactile technologies on a belt like form factor for delivering direction and distance information [10]. Unfortunately, the waist (combined with a belt like form factor) did not prove suitable for delivering high intensity data like facial expressions. Cappelletti et. al. [11] developed a vibrotactile glove to deliver color information to the tips of the fingers as vibrations. But the glove required that the users wear the vibrators on the inside which restricts functional use of one's hand. From a sensory augmentation perspective, vibrotactile gloves have been explored for applications ranging from assistive technologies to telesurgery. Uchiyama et. al. [12] have developed a glove that provides directional information to users of a semi autonomous wheelchairs, while Brell et. al. [13] used haptic gloves for providing feedback during telesurgery. These technologies have proved the viability for glove based vibrotactile sensory augmentation. Adding to these findings, the experiments described in this paper addresses the vibrotactile sensory abilities of the dorsal surface of the fingers, especially for sensory substitution.

III. THE VIBROTACTILE GLOVE

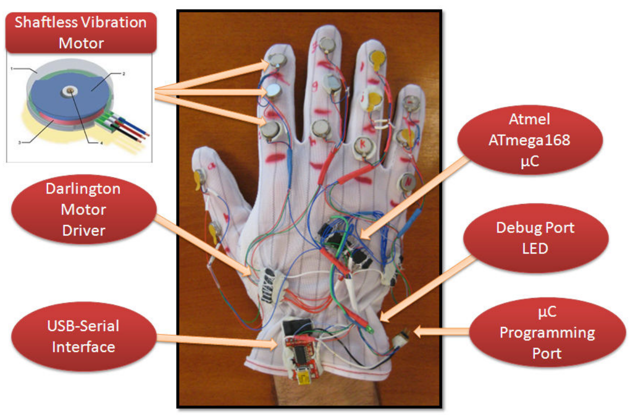


Figure 1: The Vibrotactile Glove

As shown in Figure 1, 14 shaftless vibration motors are installed on the dorsal surface of the phalanges of a stretchable anti-static glove – corresponding to the 14 phalanges of the human hand. Each vibrator has an effective displacement of 1.5mm @ 55Hz with an effective acceleration along X, Y and Z of $X_g = 0.38g$, $Y_g = 0.29g$ and $Z_g = 1.08g$, respectively, with Z axis perpendicular to the skin. The vibrators are individually controlled through a dedicated output port of a microcontroller that acts as the Communication and Control Bridge between a computer and the glove. The commands are sent to the microcontroller via a serial port that is translated to USB for interfacing with any generic computing element. The USB port also provides the necessary power for the operation of the glove. Through these commands, the microcontroller allows precise simultaneous control of three dimensions of vibrations, namely, the intensity of vibration, the location of vibration and the duration of vibration. The location of vibration is controlled by choosing the appropriate output port of the microcontroller; the duration of vibration is controlled by the onboard timer; the intensity of the vibration is controlled via simulated Pulse Width Modulation (PWM) on the output ports. In order to isolate and protect the controller from the vibration motor induced back-EMF (electromotive force) two 7-array Darlington transistors with opto-isolation are used between the controller and the motors.

The software to control the vibrations on the glove is shown in Figure 2. Two independent programs where developed to explore the localization capabilities of users, and for testing the ability of users to identify spatio-temporal cues. The serial port interface for the glove is designed similar to the popular Hayes AT command set. Activation commands are passed as ASCII strings that are interpreted by the microcontroller to activate the appropriate motor for the requested duration and intensity of vibration. The details of the cueing patterns are presented in the next section.

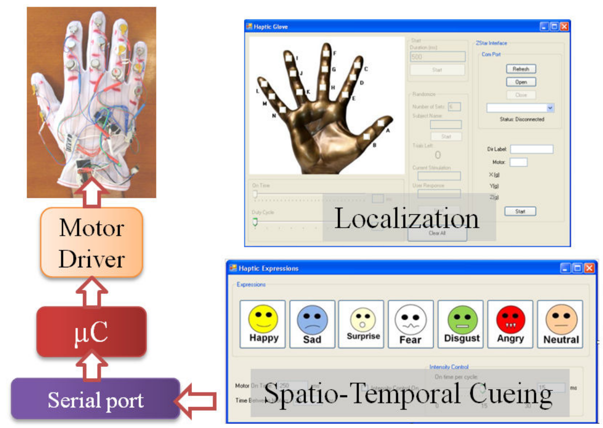


Figure 2: Localization and spatio-temporal cueing software used for the vibrotactile glove.

IV. HAPTIC CUEING

A. Localization

To determine how well users were able to perceive the vibratory patterns on the phalanges, vibrators were excited at randomly selected locations. Each excitation was applied at 100% intensity and duration of 5 seconds. The localization experiments were focused on studying the vibrotactile detection capabilities of the individual phalanges, fingers as a whole, and groupings based on the distance of the phalanges from the palm (distant, intermediate and proximal phalange) as shown in Figure 3.

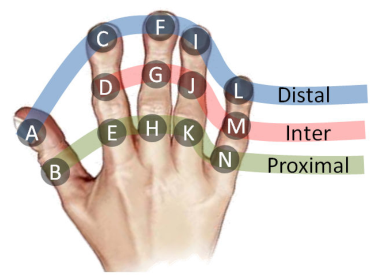


Figure 3: Phalange naming convention and grouping bases on the anatomical distances.

B. Spatio-Temporal Cueing

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.

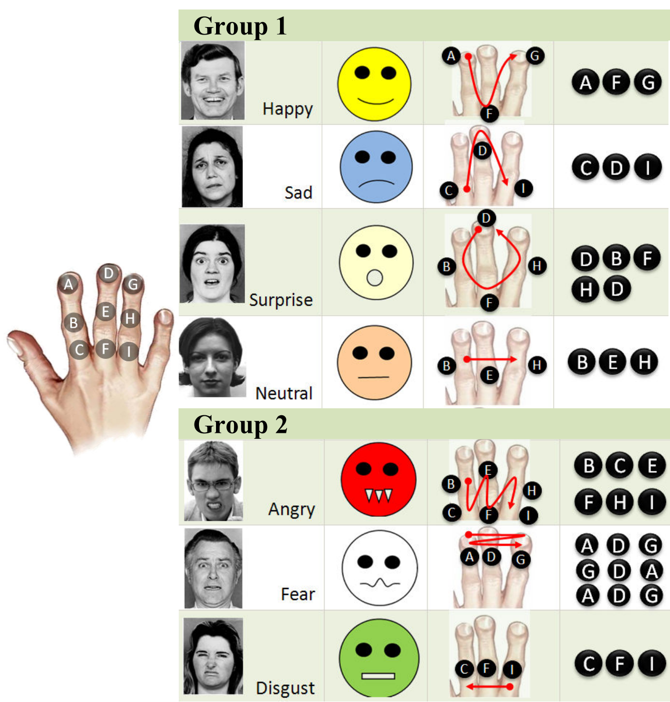


Figure 4: Mapping of Group 1 and Group 2 haptic expression icons to the central three fingers (9 Phalanges) of the vibrotactile glove. In the expression mapping chart, Columns 1 to 3 represent the expression. Column 4 shows the spatial mapping of vibrations. Column 5 shows the temporal mapping of the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

1) 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.

2) Group 2 – The auxiliary haptic icons:
Anger, Fear and Disgust cannot be conveyed through the appearance of mouth 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 teeth during an expression of anger; 2) *Fear* is represented by very brief vibrations on the dorsal phalanges of the central 3 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.

V. RESEARCH HYPOTHESES

A. Localization

While testing the localization capabilities of the haptic glove, three distinct and correlated hypotheses were tested. These hypotheses are related to the individual phalange localization, localization per finger, and localization on the phalange groups based on their distance from the palm.

1) Hypothesis 1: a) The recognition rates per phalange will be above chance (50%); b) The mean recognition rate per phalange will not be significantly different between any twophalanges.

2) Hypothesis 2: a) The recognition rates per finger will be above chance (50%); b) The mean recognition rate per finger will not be significantly different between two fingers.

3) Hypothesis 3: a) The recognition rates per phalange group (distal, intermediate, proximal) will be above chance (50%); b) The mean recognition rate per phalange group will not be significantly different between two phalange groups.

B. Spatio-Temporal Cueing

Similar to the localization experiments, the hypotheses relating to the spatio-temporal cueing relates to the ability of the users to recognize the individual expression and also the two groups of expressions as identified in Section IV B.

1) Hypothesis 4: a) The recognition rates for the spatio-temporal expression patterns will be above chance (50%); b) The mean recognition rate per expression will not be significantly different between any two expressions.

2) Hypothesis 5: a) The recognition rates per expression group (Group 1 and 2) will be above chance (50%); b) The mean recognition rates between the two groups will not be significatnly different.

VI. EXPERIMENTS & ANALYSIS METHODOLOGY

Two independent and consecutive experiments were conducted to test the localization and spatio-temporal cue identification capabilities of uses. Participants were engaged for the entire time of the two experiments and the localization experiments preceded the spatio-temporal cueing experiments.

A. *Participants:*

The experiments were conducted with one individual who was blind and 11 other participants who were sighted but blindfolded during the experiment. It is important to note that the individual who was blind had lost his sight after 25 years of having vision. To a large extent, this individual could correlate with the Group 1 haptic expressions, of the spatio-temporal cueing experiment, to his visual experiences from the past. None of the participants had any obvious medical conditions that prevented them from perceiving the vibrotactile stimulations on their right hand.

B. *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. Both the localization and spatio-temporal cueing experiments were conducted in three successive phases, namely, *Familiarization phase*, *Training phase* and *Testing phase*. Subjects were first familiarized with the various vibration patterns by presenting them in order – each phalange for the localization experiment and each facial expression for the spatio-temporal experiments. During this phase, the corresponding location or the facial expression was spoken aloud by the experimenter. The familiarization was continued until the subjects were comfortable in remembering all the locations and expressions. This was followed by the training phase in which all the fourteen vibration locations and seven facial expression patterns were presented in random order, in multiple sets, and subjects were asked to identify them by speaking them out. The experimenter confirmed any correct response, and corrected incorrect responses. Subjects had to demonstrate 100% recognition on at least one set of all fourteen locations and seven 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 location and expression pattern was randomly presented 10 times making a total of 14 locations x 10 trials = 140 localization results, 7 expressions x 10 trials = 70 expression results. The subjects were given 5 seconds per trial to respond.

C. *Analysis:*

In order to test the hypotheses presented in Section V (relating to the localization and spatio-temporal cueing experiments), three related analyses were carried out, namely, a) location or expression recognition rate, b) One-way analysis of variance (ANOVA) on the recognition accuracies, and c) Tuckey Honestly Significant Difference (HSD) test to determine the mutual performance of location and expression results. The details of the three techniques are discussed below.

1) *Recognition Accuracies*

As the name suggests, the recognition accuracies measure the average true positive rate of recognition on the localization and expression recognition experiments. Along with the mean recognition rate, the deviation in the recognition rates across the 11 participants is also shown.

2) *ANOVA*

One-way analysis of variance is a statistical tool used for comparing two or more sample groups to test null hypothesis that the samples were drawn from different populations using an F-distribution. The F statistic is derived from the sample means and the group means. Following the central limit theorem, if the samples are drawn from the same population, the variance between group means have to be smaller than the sample variance. A higher ratio of the variances justifies the null hypothesis, else it's rejected. The results of ANOVA are reported as p-value scores from the F-statistic with the dimensions *(k-1)* and *(n-1)*, where the k is the number of groups and n is the number of samples. Lower the p-value higher is the chance of accepting the null hypothesis and vice versa.

3) *Tuckey HSD*

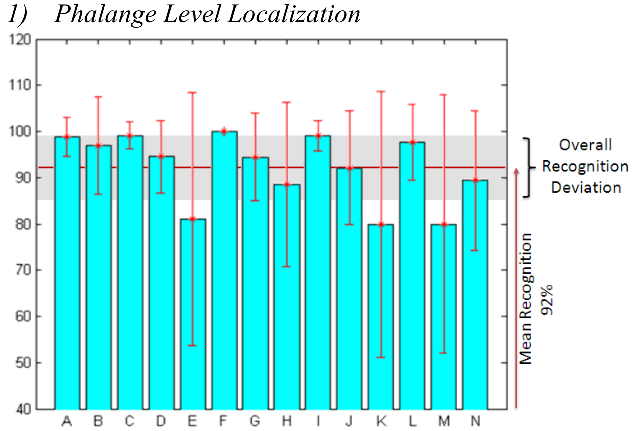
While ANOVA tests for a chance that the samples could have been derived from different populations, the Tuckey Honestly Significant Difference (HSD) test relies only on the group means to determine if there is a significant difference between groups of samples. A significant difference calls for reasoning to suspect/explain performance differences within groups derived from a single sample set. Unlike ANONA, where all the groups and samples are combined into the dimensions of comparison, HSD allows individual group-wise comparisons, providing for an opportunity to identify which groups are performing differently from others. Mostly reported as a ratio of the group mean difference and the standard group mean error, it is possible to quickly identify the significant group differences. In the results section below, the group means and the standard errors are plotted as circles and whiskers, respectively. Significant difference is established when one group's standard error stretch is beyond the scope of any of the other groups.

VII. RESULTS

In this section, five sets of results (Figure 5 – 9) are presented. Each set presents the three analyses that were described in Section VI C. Three of the five sets correspond to the localization experiments, while the other two correspond to the spatio-temporal experiments.

A. *Localization Experiments*

1) *Phalange Level Localization*



	SSE	DoF	Mean Sq.	F	p-value
Between Groups	8488.7	13	652.97	2.68	0.002
Within Groups	37529.6	154	243.69		
Total	46018.2	167			

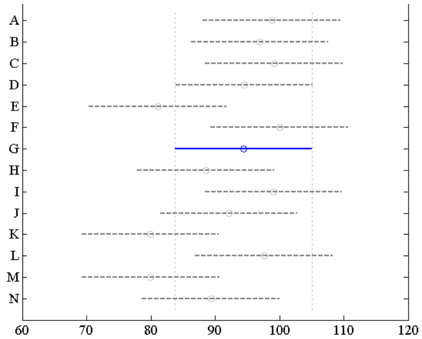
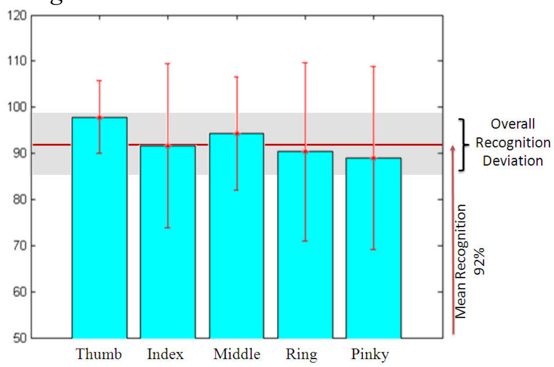


Figure 5: (a) Recognition Accuracies; (b) ANOVA; (c) HSD

2) *Finger Level Localization*



	SSE	DoF	Mean Sq.	F	p-value
Between Groups	1440.9	4	360.234	1.32	0.2657
Within Groups	44577.3	163	273.48		
Total	46018.2	167			

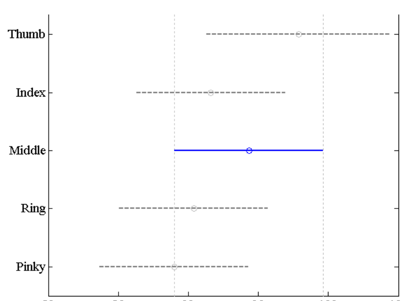
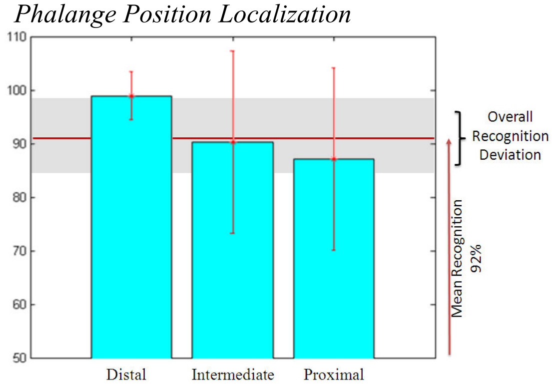


Figure 6: (a) Recognition Accuracies; (b) ANOVA; (c) HSD

3) *Phalange Position Localization*



	SSE	DoF	Mean Sq.	F	p-value
Between Groups	4408.7	2	2204.33	8.74	0.0002
Within Groups	41609.6	165	252.18		
Total	46018.2	167			

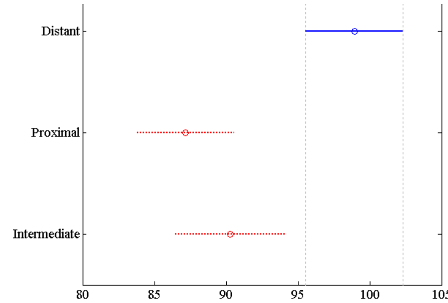
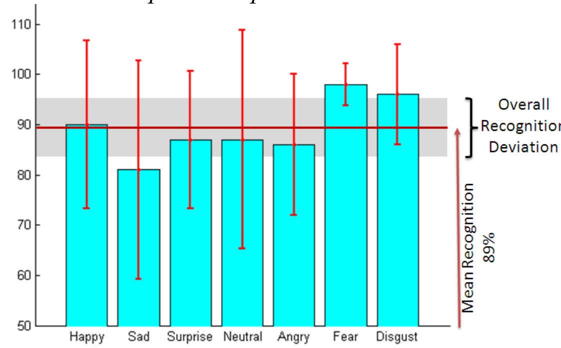


Figure 7: (a) Recognition Accuracies; (b) ANOVA; (c) HSD

B. *Spatio-Temporal Experiments*

1) *Individual Spatio-temporal Patterns*



	SSE	DoF	Mean Sq.	F	p-value
Between Groups	2530.4	6	421.72	1.71	0.1299
Within Groups	18990.9	77	246.63		
Total	21521.3	83			

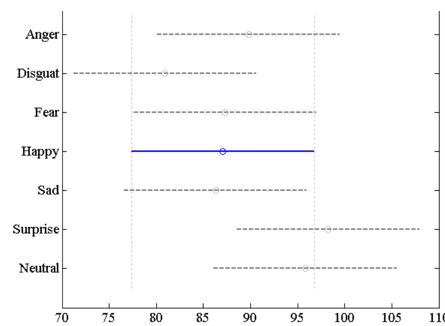
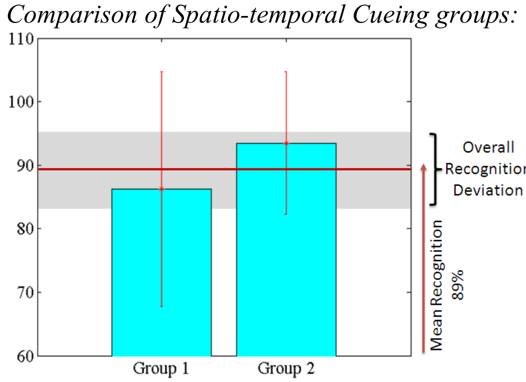


Figure 8: (a) Recognition Accuracies; (b) ANOVA; (c) HSD

2) *Comparison of Spatio-temporal Cueing groups:*



	SSE	DoF	Mean Sq.	F	p-value
Between Groups	1058.2	1	1058.16	4.24	0.0426
Within Groups	20463.1	82	249.55		
Total	21521.3	83			

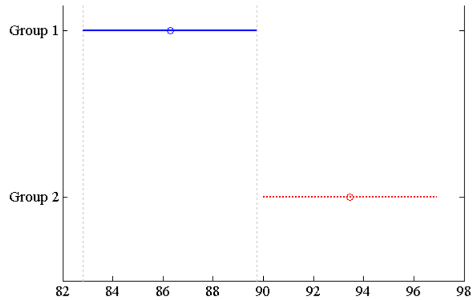


Figure 9: (a) Recognition Accuracies; (b) ANOVA; (c) HSD

VIII. DISCUSSION OF RESULTS

From Figure 5 through 9, it can be seen that the part (a) of the five research hypotheses can be answered immediately. Users of the device found it convenient to localize vibration patterns and identify expressions easily. The localization recognition accuracy was at 92% (SD: 7.5%), while the expression recognition rate was measured at 89% (SD: 5.9%). This validates the null hypothesis that the users were able to localize and identify vibrotactile patterns well above average.

Investigating part (b) of the five hypotheses reveals interesting insights into the user's abilities to detect and localize vibrotactile stimulations. From Figure 5(b) and 5(c) it can be concluded that Hypothesis 1(b) is accepted; both ANOVA HSD tests reveal no significant difference between phalange performances. Similarly from Figure 6(b) and 6(c), it can be concluded that there is no significant difference in the performance between fingers. Figure 8(b) and 8(c) accepts the Hypothesis 4(b) and we see no significant difference in the mean performance of the seven spatio-temporal cueing patterns of facial expressions.

In contrast, from Figure 7(b) and 7(c), we see that the Hypothesis 3(b) is rejected as user performance diminished at the proximal phalanges. This could be attributed to the fact that the vibration motors are very closely placed next to one another at the proximal phalanges which may cause inter-motor vibrations. From Figure 9(b) and 9(c), we see that Group 2 performance was much higher than Group 1 rejecting the Hypothesis 5(b). Studies are underway to determine the nature of the haptic cues in Group 2 that make them significantly better than Group 1. This could have been due to the fact that Group 2 cues were designed based on extensive user feedback when compared to Group 1 expressions which were designed based on popular visual emoticons.

IX. CONCLUSION AND FUTURE WORK

In this paper, we have explored the dorsal surface of the finger phalanges to determine the localization and spatio-temporal haptic cue detection capabilities. From the results presented above, we conclude that the hand can be used as an effective location of the body to achieve visuo-haptic sensory substitution without causing any functional inhibition to the fingers or the hand. Further studies are needed to quantify the bandwidth of information that can be conveyed effectively without inducing cognitive load on the users. Future work will focus on experiments where the users are involved in cognitive tasks while interpreting the haptic cues.

X. REFERENCES

[1] M.M. Sohlberg and C.A. Mateer, *Introduction to Cognitive Rehabilitation: Theory and Practice*, The Guilford Press, 1989.

[2] M. Grunwald, *Human Haptic Perception: Basics and Applications*, Birkhäuser Basel, 2008.

[3] D. Maynes-aminzade, "Edible Bits: Seamless Interfaces between People, Data and Food," *Proceedings of the 2005 ACM Conference on Human Factors in Computing Systems (CHI'2005) - Invited Talk*, 2005.

[4] Y. Chen, "Olfactory Display: Development and Application in Virtual Reality Therapy," *International Conference on Artificial Reality and Telexistence*, Los Alamitos, CA, USA: IEEE Computer Society, 2006, pp. 580-584.

[5] N.J. Dominy and P.W. Lucas, "Ecological importance of trichromatic vision to primates," *Nature*, vol. 410, Mar. 2001, pp. 363-366.

[6] K. Koch, J. McLean, R. Segev, M.A. Freed, M.J. Berry II, V. Balasubramanian, and P. Sterling, "How Much the Eye Tells the Brain," *Current Biology*, vol. 16, 2006, pp. 1428-1434.

[7] A. Nakamura, T. Yamada, A. Goto, T. Kato, K. Ito, Y. Abe, T. Kachi, and R. Kakigi, "Somatosensory homunculus as drawn by MEG," *NeuroImage*, vol. 7, May. 1998, pp. 377-386.

[8] P. Bach-y-rita, C.C. Collins, F.A. Saunders, B. White, and L. Scaden, "Vision Substitution by Tactile Image Projection," *Nature*, vol. 221, Mar. 1969, pp. 963-964.

[9] S.U. Réhman and L. Liu, "Vibrotactile Rendering of Human Emotions on the Manifold of Facial Expressions," *Journal of Multimedia*, vol. 3, 2008.

[10] 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," *Haptic Audio visual Environments and Games, 2008. HAVE 2008. IEEE International Workshop on*, 2008, pp. 13-18.

[11] L. Cappelletti, M. Feeri, and G. Nicoletti, "Vibrotactile colour rendering for the visually impaired within the VIDET project," *Telemanipulator and Telepresence Technologies V*, 1998, pp. 92- 96.

[12] H. Uchiyama, M.A. Covington, and W.D. Potter, "Vibrotactile Glove guidance for semi-autonomous wheelchair operations," *Proceedings of the 46th Annual Southeast Regional Conference on XX*, Auburn, Alabama: ACM, 2008, pp. 336-339.

[13] M. Brell, D. Roßkamp, and A. Hein, "Fusion of Vibrotactile Signals Used in a Tactile Display in Computer Aided Surgery," *Proceedings of the 6th international conference on Haptics: Perception, Devices and Scenarios*, Madrid, Spain: Springer-Verlag, 2008, pp. 383-388.