2014 11th International Conference on Wearable and Implantable Body Sensor Networks

A Usability User Study Concerning Free-Hand Microgesture and Wrist-Worn Sensors

David Way, Joseph Paradiso
Responsive Environments Group, MIT Media Laboratory
Cambridge, MA -USA-
e-mail: dway@mit.edu, joep@media.mit.edu

***Abstract*—Wrist-worn sensors (microphones, time-of-flight cameras, etc) have gained the attention of Human Computer Interaction (HCI) and body sensor researchers for their potential ability to aid interaction with wearable devices. In this paper, we use wrist-worn sensor modalities to evaluate free-hand microgesture usability. Our goal was to determine which microgestures should be included in a potential universal microgesture language and identify any underlying microgestural usability principles. Through a brief pilot user study recording microgesture task time and user accuracy, we were able to explore trends in common usability aspects. Results of the user study showed that free-hand microgestures, even at small physical and temporal scale, have significant effects on task time and user accuracy. Further analysis through multiple comparisons identified which microgestures produce relatively more accurate and efficient interaction. Physical commonalities between such microgestures prompted theories concerning why certain microgestures produce more efficient results. Based on findings and proposed theories, we give suggestions concerning a universal microgestural language and microgestural application development.**

Keywords- HCI; wearable computing; usability; computer vision; gesture recognition

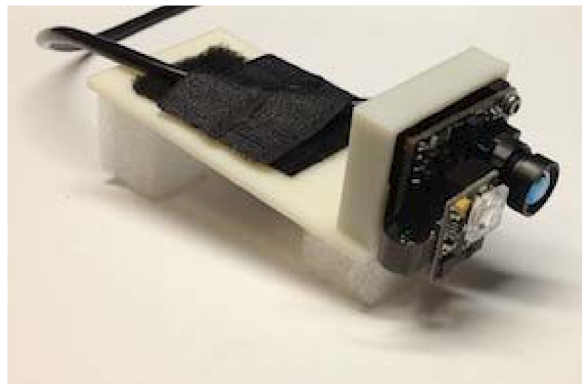

Fig. 1. Time-of-flight camera used to build EMGRIE.

I. Introduction

Many current ubiquitous gesture applications track hands in front of the user, requiring large fatiguing movement of the shoulder and arm (commonly called 'gorilla arm' [12]). The space in front of the user is socially awkward and requires considerable volume; we believe that hand-at-side finger/wrist based gesturing will dominate for instant on-the-go use cases - such as everyday wearable interfaces. To progress towards such interfaces, wrist-worn microgesture (small finger/thumb based movements) capturing devices have begun to emerge [7], [14], [1]. Such devices greatly diminish the problem of fatigue and have considerable power for application, yet microgesture evaluation and microgesture interaction design is still an underexplored territory. In this particular study, we hoped to extract various trends that relate microgesture to task time and user accuracy.

We built our own wrist-worn microgesture recognition system to robustly detect free-hand microgesture performances. Our system, which we deem EMGRIE (Ergonomical MicroGesture Recognition and Interaction Evaluation), extracts multiple concepts from past wrist-worn microgesture recognition systems to produce similar recognition results. We use EMGRIE as a mode to examine microgesture application usability through task time and user accuracy. Given user data, we applied common HCI statistical methods (ANOVA, multiple comparisons) to extract evidence that microgesture has significant effect on task time and user error rate, and further investigate which microgestures produce such effects.

Post-hoc tests showed that microgesture is correlated to input efficiency, which we theorize is due to physical and attentional differences between hand poses. Based on such physical differences, we believe certain microgestures may hold a physiological and psychological contrast bias [9] over other microgestures and attract more attention during initial learning. We believe such biases should be acknowledged in order to acheive good usability during application design. Upon similar statistical examination of user accuracy rates for each microgesture, it is clear that the fastest microgesture is not the most accurate across users. The lift microgestures (in particular the index lift) are significantly less user error-prone. While there was no apparent tradeoff between speed and accuracy, we believe microgestures that involve less muscle groups are less error-prone but not necessarily more efficient. Hence, microgestural application developers should heed specific application requirements concerning usability and universal microgestural language. Based on sources of data variation, we also believe user preferences regarding microgestures are mandatory. We encourage developers to include some of the microgestures studied here in a universal microgestural language, while keeping in mind our findings in user accuracy and efficiency.

II. Related Work

To construct a wrist-worn microgesture recognition system within a minimal time frame, we draw upon many previously built systems. In order to capture the potential robustness

978-1-4799-4959-5/14 $31.00 © 2014 IEEE
DOI 10.1109/BSN.2014.32

Fig. 2. The camera and microphone on the user's wrist, user performing the index pinch microgesture.

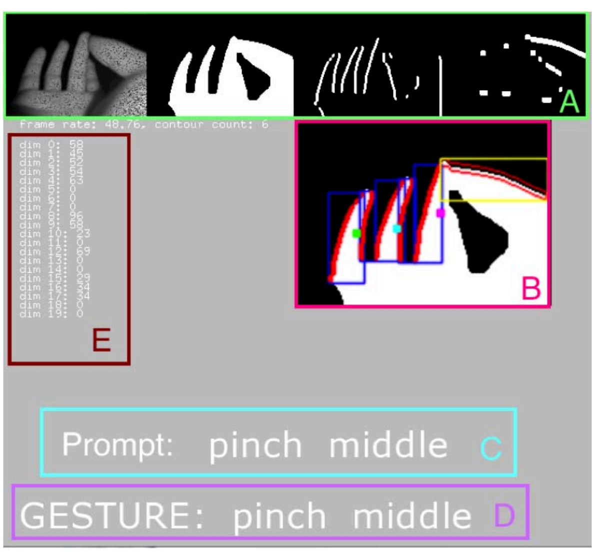

Fig. 3. The screen the user viewed during a user test session. We displayed the current gesture classification (D), goal gesture (C), depth camera perspective (B), computer vision pipeline (A), and bag-of-words data vector (E).

of Microsoft Digits [7], we use a similar form factor and feature extraction methodology. However, we simplify the Digits system by building a customized feature vector rather than estimating all joint angles found in the hand. By taking advantage of this tradeoff between simplicity and full hand pose extraction, we can quickly build a system that can robustly detect only the microgestures we are interested in.

Similar systems to Digits and EMGRIE have been researched as well, and we would like to recongize them here. The WristCam [14], DataGlove [11], RFID rings [2] are good examples of a microgesture recognition system. While these studies produced good results in recognition, we continue this research and present theories concerning microgesture usability.

Stern [13] and Nielsen [10] produced applicable methodologies for usability study concering gestural interaction. Even though these studies did not specifically target free-hand microgesture, we would like to emphasize the importance of following such methodologies (and attempted to follow them ourselves over the course of the EMGRIE study). Our study mainly concerns itself with user accuracy and task time and how it relates to usability, but there are more aspects mentioned in these works that should be considered.

III. Sensing Modalities

To detect microgestures, we utilize a time-of-flight (TOF) camera PMD CamBoard Nano (www.pmdtec.com/products_-services/reference_design.php) and a small MEMS microphone. Both sensors are wrist-worn, where the camera and microphone are situated on the inside of the wrist. Figure 1 shows the hardware system components. While the size of the device is currently too large and power inefficient for everyday use, we believe that such devices are currently improving to warrant this research. To process data taken from the sensors, we utilize open source software OpenCV and Gesture Recognition Toolkit (GRT). Using OpenCV and the TOF camera, we extract finger-level features produced by various microgestures similarly to [7]. Such features allow individual finger and thumb tracking. However, rather than applying tracking and finger-level signals to estimate joint angles and build a virtual hand model, we formulate a bag-of-words support vector machine model (SVM) through GRT. We use the microphone to detect vibrations made by the thumb and finger during a pinch gesture, and utilize the vibration for accurate recognition, similar to [1]. Constructing EMGRIE in this manner did not require extensive engineering resources and produced viable gesture recognition error rates for usability study. While such algorithmic methods have been researched extensively, we have novelly applied them in a simplistic manner to recognize microgestures. Since the benefits of this methodology are unclear, we used EMGRIE to primarily study user interaction. EMGRIE system design is available in more detail: [15].

IV. Microgesture Choice and Experiment Design

There are many potential microgestures that could deemed useful in applicaton. However, we found that users had difficulty learning more than seven gestures in one test session. Hence, we decided to only test seven microgestures from two very different yet basic free-hand microgesture categories: the finger lifts and the finger pinches. [15] gives an extensive description of free-hand microgesture categories. We describe how to perform example microgestures in each group: figure 2 shows EMGRIE on a user's wrist while performing the index pinch. To perform a finger lift, the user simply bends the corresponding finger toward the palm at the first finger joint (MP joint). Our motivation for choosing these microgesture groups stems from the ability of the wrist-worn camera to recognize finger lifts with high accuracy [14] and the hands natural ability to perform finger to thumb opposition [8]. Through examining these basic free-hand gestures, we can extract statistically significant differences between them (if differences exist).

Specifically, we decided to use 3 finger lifts, the index lift (IL), middle lift (ML), and ring lift (RL). We study 4 different fingertip pinch gestures: index pinch (IP), middle pinch (MP), ring pinch (RP), and pinky pinch (PP). 11 participants (8 male 3 female ages 19-30) were asked to train the EMGRIE system to accommodate their specific interpretation of each gesture. To train the EMGRIE SVM model, users held the corresponding hand pose for 5-10 seconds while EMGRIE collected data (about 500 data points). Once each gesture was trained and users felt mastery of gesture performance, the user completed 10 sequences of 7 random gestures. Gestures were displayed one at a time via text description commands (i.e. ”lift index”) while time to classification and user accuracy was recorded. Each sequence covered each gesture and was displayed in random order to mitigate learning and fatigue effects. The user rests both hands by their side or on their lap while orchestrating a series of such specified microgestures displayed on-screen about two feet in front of him/her. The user was asked to not view their own hand during performances and were encouraged to maintain eyesight on visual feedback (wrist-worn camera image and gesture instruction) on-screen. Errors were flagged as either classification errors or user errors and were rescheduled in the sequence. To simulate an eyes-free situation, participants were asked to maintain their eyesight on the screen detailing the current microgesture to perform. The user viewed the screen shown in figure 3. There was a user-determined length break between each sequence and 3 seconds break between each gesture performance. Due to frequent breaks, each user test took on average 50 minutes to complete. Users did not express discomfort concerning the length of the test nor wearing the device.

Our experiment design examines within-subject microgesture and iteration factors and their relation to the continuous response factor task time. The experiment follows a randomized two-way repeated measures design: each user performs each microgesture 10 times ordered randomly. The results of this experiment upon two-way repeated measure ANOVA should show any effect of short term training (a within-subject iteration factor) or microgesture differences. Since task times are skewed in the positive direction, we removed outliers over a 95% confidence interval. In order to handle missing data, we used a Mixed Factor model. After finding that the interaction term between iteration and gesture was not significant, we converted to a additive repeated measures model to only examine simple main effects and pairwise significance. Since task time data was recorded as a response factor to microgesture, task times corresponding to user and classification errors were discarded.

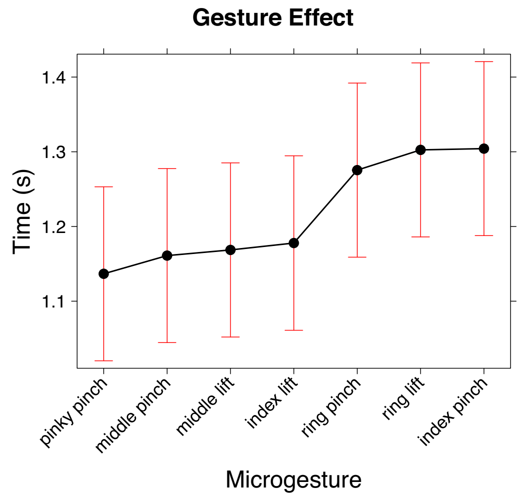

Fig. 4. Box plot showing the task time average estimtion for microgestures over all users and all performance iterations. The pinky pinch gesture appears to be the most efficient.

V. Results

Two-way repeated measures ANOVA showed significant effect in the gesture factor ($p < .001$). We did not find a significant effect in the iteration factor. Through Tukeys pairwise mutliple comparison test, we found significant variations between pinky pinch and index pinch, and between the ring lift and pinky pinch microgestures ($p < .02$). Figure 4 shows the mixed model plot (average task time estimation and corresponding error thresholds for each microgesture). Estimation error thresholds were about the same for each microgesture, as well as were standard errors for each coefficient in the resulting regression model. Risidual standard deviation (variation caused by microgesture) was about .22 seconds and random effect standard deviation (variation caused by the participant) was about .24 seconds. We note that total variance is large over each microgesture, but also see that most variation was caused by participants and not by microgestures.

We used similar statistical tools to examine trends between microgesture and user error (summing over all users and iterations) assuming independent measures (task time and user error). Figure 5 shows the user error rate of each microgesture. Upon adding the microgesture factor to a repeated measures logistic regression model, model improvement showed microgesture having a significant effect on user error rate ($p < .0001$). Based on multiple comparisons (Tukey test) over microgestures, we found the index lift to have significant differences between the middle and ring pinches ($p < .02$). EMGRIE yields low system error rates out of 809 total gestures, 39 were classified as system errors (about 4.91% misclassifications over all gestures and users, similar to mini-QWERTY typos [4]). Since the focus of this research was on user interaction, EMGRIE classification errors were not examined further.

VI. Discussion

Before experimentation, it was not clear that gesture at this scale (simple free-hand pose) will show statistical differences

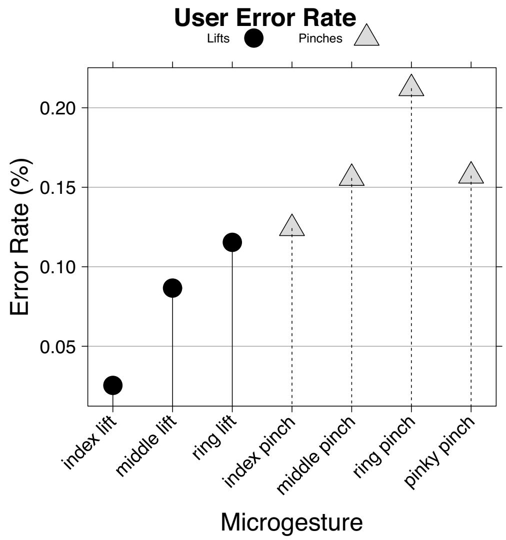

Fig. 5. A table showing the user error rates (%) for each microgesture summed over each user. Clearly, the lift gesture group has lower error rates when compared with the pinch group.

in performance times. One would expect that iteration of gesture performances should have an effect on task time, but it seems that 10 iterations over a 20 minute test session is either a too little or too large sample size to show trend. Hence, we hypothesize that learning microgesture is similar to learning to keyboard - the required detailed control over ones hand requires large amounts of training to produce effects on task time. We believe that the estimated average task time for each microgesture shown in 4 is seemingly high - but attribute this to novice user's cognitive processing time caused by prompt randomization. Since variation is high, we note that some users performed gestures in under .5 seconds. This shows promise in potential learnability effects: users may be able to perform microgestures consistently at a much higher rate after training.

In terms of the effect microgesture has on task time, we note that the pinky pinch and index pinch are physically different, and the ring lift and pinky pinch are physically different. It follows that we formulate a hypothesis as to why the pinky pinch microgesture was the faster microgesture: we believe that the pinky pinch holds a bias over the other microgestures during learning. The other microgestures may hold a continuous psychological representation or gradient of stimulus between themselves, where the pinky pinch (being the arguably most unnatural gesture concerning pinches [8]) may show a disconnect. For instance, users may view the pinky pinch to be more unusual compared to the other microgestures and hence give the pinky pinch special attention due to the contrast bias [9] (as known in psychology). Further experimentation is needed to explore this resulting hypothesis. If true, then we can infer an interesting generalization: certain unnatural microgestures are more efficient than natural (everyday) microgestures, due to less distinct natural microgestures warranting more cognitive processing to remember and perform.

We believe that resulting differences in efficiency between microgestures are only substantial over longer periods of application use. Based on figure 4, the greatest difference in task time between microgestures is about .15 seconds. While we hypothesize that the index pinch did not gain as much attention during learning, we believe that the resulting decrease in efficiency (about 13%) will have a large effect on usability during repeated gesture performances in applications (such as text entry). Since performance errors were discarded, we can modify and apply the Keystroke-Level Model (KLM) [3] to further explore this effect. To continue with KLM analysis, we believe a more in-depth user study involving expert users and consecutive input sequences is required.

Based on error plot in Figure 5, we find smaller user error rates in the lift microgesture group compared to the pinch microgesture group. The lift microgesture group requires less muscle groups to perform and are generally more simple than the pinch group. While the lift microgestures are not the most efficient microgesture group on average, we think they are relatively simple for users to understand and are easier to perform correctly. We also find an interesting trend between fingers concerning user error: error rate increases nearly linearly from index to ring in both groups. This may be an effect of ring and middle fingers not having as much natural use as the fore finger. The pinky pinch user error rate follows our theory concerning attention and contrast bias: it does not follow the trend of increasing error rate and we see a marked improvement from the ring pinch. We do not see a similar trend between the other fingers in task time that we see in user accuracy. More exploration is needed to examine the relationship between user error and task time in the microgestural application space.

Based on these findings, we believe that since the middle and ring pinches are not significant concerning user accuracy and efficiency, they may not be the best choice to be included in a universal microgestural language. However, we would like to emphasize, from our own empirical observation (and large data variation found between users), that users showed preferences for various microgestures. We suggest that microgestural application designers take into account the simplicity of the microgesture, the contrast bias, and the ability for the user to customize microgestures when building an application.

VII. Applications

There are many potential microgesture applications for use with wearable interfaces. Since wearable interfaces are limited in screen space, we believe microgesture will be a powerful mode for text input. Text entry interaction design should follow proper fit-to-function: less efficient and more error prone microgestures should be used for less common characters, such as 'z' or 'x'. Through applying KLM and user test results given in this study, interaction designers can study various microgesture mappings to build a generalized and efficient text entry application.

We also believe that short-lived interactions involving data visualization, multiple choices, and shortcuts possess just as much potential. Since such applications require shorter interactions, user accuracy results of this study are more applicable than the efficiency results. Hence, application designers should be more concerned over choosing microgestures that are easy and less error-prone for the general population. We conducted exploratory experiments concering such applications [15] and leave associated discussion for a future publication.

VIII. Summary and Future Work

Based on task time results, we would like to formulate a theory concerning microgesture usability: due to the contrast effect, groups of natural hand pose (defined as microgestures commonly used in everyday activity) often require more cognitive processing time to perform than less natural microgesture groups require. More natural microgesture groups may have psycho-physiologically closer representations and not require much attention during learning, while less natural microgesture groups require more attention. Hence, added attention yields more efficient performance time. Based on user accuracy results, we predict simplistic microgestures that involve less muscle groups will yield better usability. However, if a user has trouble performing a certain microgesture, then applications should have the ability for the user to customize microgesture mappings for ease of use.

To test these theories more extensively, task time experimentation regarding microgestural applications are required. Upon large scale experimentation exploring a universal microgesture language and action mappings we will know more towards developing highly usable microgestural applications.

Acknowledgment

The authors would like to thank Matt Aldrich for aid in ANOVA application, Nicholas Gillian for the construction of GRT and its application, and the many users who helped in extracting usability data. We thank the rest of our colleagues in the Responsive Enviornments Group, especially Nan Zhao, Jie Qi, Nan-Wei Gong, Gershon Dublon, and Pragun Goyal.

References

[1] Brian Amento, Will Hill, and Loren Terveen. The sound of one hand: A wrist-mounted bio-acoustic finger gesture interface. *CHI*, 2002.

[2] Rachel Bainbridge and Joseph Paradiso. Wireless hand gesture capture through wearable passive tag sensing. In *International conference on Body Sensor Networks*, pages 200-204, 2011.

[3] Stuart K. Card, and Thomas P. Moran. The Keystroke-Level Model for User Performance Time with Interactive Systems. In *Communications of the ACM* 23(7):396-410. July, 1980.

[4] Clawson, J., K. Lyons, E. Clarkson, T. Starner. Moble Text Entry: An Empirical Study and Analysis of mini- QWERTY Keyboards. (2005) Presented at ISWC 05.

[5] Nicholas Gillian. Gesture recognition toolkit. http://www.nickgillian.com/software/grt/, June 2013.

[6] Bruce Howard and Susie Howard. Lightglove: Wrist-worn virtual typing and pointing. pages 172-173, 2011.

[7] David Kim, Otmar Hilliges, Shahram Izadi, Alex D. Butler, Jiawen Chen, Iason Oikonomidis, and Patrick Oliver. Digits: Freehand 3D interactions anywhere using a wrist-worn gloveless sensor. In *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology*, UIST '12, pages 167-176, New York, NY, USA, 2012. ACM.

[8] Christine L. MacKenzie; Thea Iberall. *The Grasping Hand*. North-Holland, 1994.

[9] Gordon B. Moskowitz. *Social cognition: understanding self and others*. Guilford Press, 2005.

[10] Michael Nielsen, Moritz Strring, Thomas B. Boeslund, and Erik Granum. A procedure for developing intuitive and ergonomic gesture interfaces for HCI. *International Gesture Workshop 2003*, pages 409-420, 2004.

[11] Vitor Pamplona, Leandro Fernandes, Joo Prauchner, Luciana Nedel, and Manuel Oliveira. The image-based data glvoe. *Revista de Informtica Terica e Aplicada*, 15(3):75-94, 2008.

[12] D. Saffer, *Gorrilla Arm*, Available at: http://www.kickerstudio.com/blog/2009/02/gorilla-arm/ visited 11th September 2013.

[13] Helman I. Stern, Juan P. Wachs, and Yael Edan. Designing hand gesture vocabularies for natural interaction by combining psycho-physiological and recognition factors. *International Journal of Semantic Computing*, 02(01):137-160, 2008.

[14] A. Vardy, J. Robinson, and Li-Te Cheng. The wristcam as input device. In *Wearable Computers, 1999. Digest of Papers. The Thrid International Symposium on,* pages 199-202, 1999.

[15] David Way and Joseph Paradiso. EMGRIE: Ergonomic Microgesture Recognition and Interaction Evaluation, A Case Study. M.Eng. Thesis, MIT EECS Department, February 2014.