Human Factors for Design of Hand Gesture Human - Machine Interaction

Helman I Stern* *Member, IEEE*, **Juan P. Wachs** *Student Member*, **Yael Edan** *Member, IEEE*
Department of Industrial Engineering and Management
Ben-Gurion University of the Negev, Be'er-Sheva 84105, Israel.
{helman, yael, juan}@bgu.ac.il

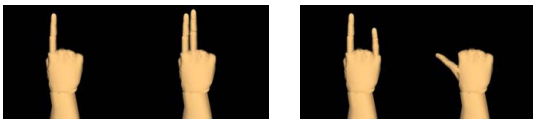
Abstract - *A global approach to hand gesture vocabulary (GV) design is proposed which includes human as well as technical design factors. The method of selecting gestures for preconceived command vocabularies has not been addressed in a systematic manner. Present methods are ad hoc, although sometimes attention is given through a set of rules. In an analytical approach technological factors of gesture recognition accuracy are easily obtained and well studied. Conversely, it is difficult to obtain measures of human centered desires (intuitiveness, comfort). These factors, being subjective, are costly and time consuming to obtain, and hence we have developed automated methods for acquisition of these data through specially designed applications. Results of the intuitiveness experiments showed when commands are presented as stimuli the gestural responses vary widely over a population of subjects. This result refutes the hypothesis that there exist universal common gestures to express user intentions or commands.*

Keywords: Hand gesture, optimal vocabulary, human factors, man-machine interaction, intuitive interfaces

* Corresponding author

1 Introduction

Gesture interfaces are needed to fill the need for more natural intuitive communication with non-human devices such as computers and robots. The design of gesture vocabularies (GV) to support these interfaces is a virgin area of research. In a global approach to hand gesture vocabulary (GV) design, human as well as technical factors must be considered. The method of selecting gestures for preconceived command vocabularies has not been addressed in a systematic manner. Present methods are ad hoc, although sometimes attention is given through a set of rules. The basic machine aspect of a *GV*, is that of gesture recognition accuracy, and this is easily obtained and well studied. For a few examples, see [1], [2], [3], and [4]. In all of these papers the aim is to attain high recognition accuracy. However, gestures that are easily discriminated with good recognition accuracy may not be intuitive and easy to perform from a human centered standpoint and vice a versa. Fig (a) and (b) show a pair of gestures with "bad accuracy, good intuitiveness" and "good accuracy, bad intuitiveness", respectively.



(a) (b)

Figure 1. Gestures for "One" and "Two" signs.

Few researchers have considered the human and technical factors jointly. One of them [5], where human factors are considered, gives limited attention to technical aspects. An analytical method for designing an optimal gesture vocabulary, which considers both human and machine based factors, can be found in [6]; and will not be

elaborated on here. This method is based on finding a subset of gesture poses to be matched with a set of commands of the same size. The subset is selected from a large master set of gestures, and evaluated using human factor and technological indices. Finding the recognition accuracy of the subset is no problem, as there are numerous classification algorithms for this purpose. Measuring intuitiveness and stress (or its inverse comfort) is another matter, as these factors are subjective and must be obtained by empirical methods. This seems to be the bottle neck in the design of optimal or near optimal gesture vocabularies for human-machine interfaces.

This paper focuses on the development and collection of numerical indices for the representation of human factors of intuitiveness and comfort (or its inverse effort). Intuitiveness is the cognitive association between a command or intent and its physical gestural expression. With respect to comfort, care must be taken to insure suitability of gestures to avoid muscle strain and fatigue during long term use of gesture interfaces. These factors being subjective are costly and time consuming to obtain, and hence we have developed automated methods for acquisition of this data through specially designed applications.

In section 2 to follow the basic gesture design research problem is defined. Section 3 discusses current approaches for measuring hand gesture intuitiveness and effort. The automated methods for obtaining intuitiveness and effort data appear in Section 4. Results and analysis of both experiments appear in 5 and 6 respectively. Section 7 provides conclusions.

2 The Gesture Design Problem

The basic research problem here is to provide an analytical method to find an optimal hand *GV*. Three performance measures are considered; intuitiveness, comfort, and recognition accuracy designated as *Z1* (*GV*), *Z2* (*GV*) and *Z3* (*GV*), respectively, each of which are to be maximized. The first two are human valued, while the third is machine valued. Intuitiveness of a *GV* is the sum total of the intuitiveness of each gesture-command pair in the vocabulary. Comfort is inversely related to strength needed to perform a gesture. Total comfort is equal to the comfort values of gestures and transitions between them, weighted by the frequencies and durationsof use. Accuracy, obtained from a gesture recognition algorithm, measures the percent of gestures successfully recognized. Weights can be used to reflect the relative importance of the criteria, but are difficult to determine, and hence subjective weights are usually obtained from the decision maker. The resulting multiobjective optimization problem can be found in [6]. The optimal hand GV architecture (Fig. 2) consists of: (a) determination of the human psycho-physiological input factors, (b) Stage 1- a gesture subset search subject to machine gesture recognition accuracy, and (c) Stage 2 - a command - gesture matching.

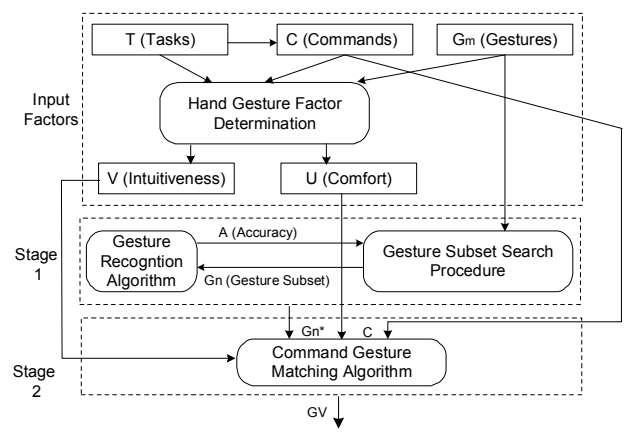


Figure 2. Architecture of hand GV methodology

The task set(s) T, command set for each task C, and master gesture set G_m are exogenous inputs. In this paper we focus only on the hand gesture factor determination stage to determine the indices required to compose the U (intuitiveness)and V(comfort) matrices

3 Current Approaches to Measure

Intuitiveness and Effort (Stress)

Intuitiveness is the cognitive association between a command or intent and its physical gestural expression. In [5] two approaches are mentioned for obtaining intuitiveness measures; (a) *bottom-up* - takes functions (commands) and finds matching gestures, and (b) *top-down* - presents gestures and finds which functions are logically matched.

Kolsch ,et.al. [7] provide a very interesting discussion on postural comfort based on a "comfort dimension" along which the human feelings are placed in states of comfort, discomfort, fatigue, and pain. Approaches to the measurement of stress, comfort, etc. can be divided into mathematical model based, physical measurement, and subjective methods. Brook, et al [8], construct a dynamic model representing the biomechanics of the index finger's flexion-extension and abduction-adduction motion. Yasumuro, et al [9], construct a biomechanical model of the entire hand comprised of tendons, muscles and bones, where physical stress is simulated through the natural constraints of the hand. An et al [10] develop a three-dimensional normative model of the hand. The authors state that the model can be used to perform force and motion analyses, but do not extend it to estimate stress. Harling and Edwards [11] use a rod string model to estimate finger tension although no comparison is made with perceived tension of users. The use of EMG measurement is popular, but the main problem is it usually only measures the activity of part of the muscles involved in structuring a pose, Shrawan and Anil [12]. The model and measurement approaches are prone to errors and have not, for the most part, been satisfactorily validated by user studies. We therefore initiate a series of experiments to obtain subjective measures by elucidating responses from human subjects. These can be used in the future to validate some of the other approaches**,**

For this study, we consider only static hand poses, although the methods proposed here can be easily adapted to consider dynamic gestures as well. To collect intuitive data we use the "bottom- up approach". The actual acquisition of gesture responses is not trivial, and we considered the following three methods; (a) *Direct Video Capture* - the subject physically forms the gesture and a camera image is taken. Here there may be errors in recognizing the gesture, (b) Use of a *Gesture Database* of candidate gesture images. Browsing a large database is time consuming, and difficult for the subject to remember and make comparative judgments, and (c) *Coded Gesture Entry***,** the subject physically generates the gesture, and enters configuration information. The coded gesture entry method was selected as one combining reasonable time demands, and accuracy of gesture labeling. Each pose configuration is represented by a coded string representing 11 degrees of freedom of basic hand elements. For example, finger flexion-extension, palm up or down, adjacent finger spread or not, etc.

To collect gesture effort, the subject holds a pose, and records the level of effort perceived. In the next section two experiments are described; (a) intuitive selections of gesture poses to represent commands, and (b) effort to configure and hold a static pose.

4 Methods

4.1 Subjects

Subjects were undergraduate students (approximately equal numbers of male and female) in an Ergonomics course. Most were 20 to 35 to years of age, and none had experience with the gestures interfaces. Thirty five and nineteen subjects participated in the intuitiveness and effort experiments, respectively.

4.2 Equipment

Subjects sit by a monitor, with a camera capture system on the side to allow them to visualize what portion of the

hand is viewed. Two applications automate the collection of subject responses; one for intuitiveness, and one for effort. The intuitive screen for a car is shown in Fig 3.

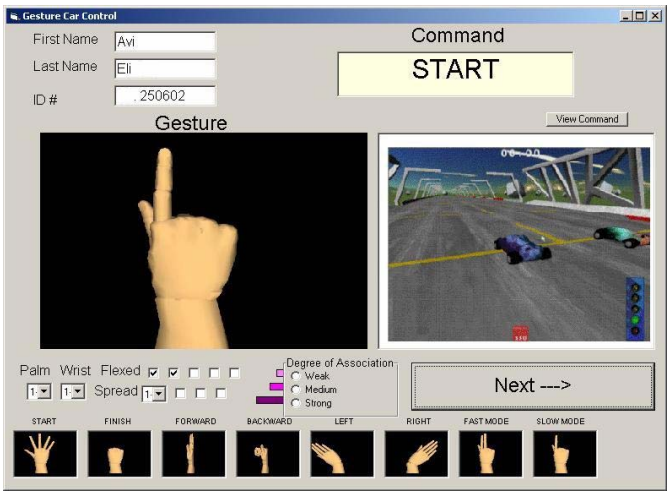


Figure 3. Interface for intuitive gesture- command data

In the application for effort pre-generated images are presented to each subject; and by a button press, the desired level of stress**,** based the [0-10] Borg scale of perceived exertion [26], is recorded. Radio buttons also allowed for the entry of level of belief of the effort selection. The bottom of the interface shows the set of gestures used in the experiment.

4.3 Procedure

A hand pose is represented by a configuration string of 11 elements of the hand. The "*basic elements*" are the moveable parts of the hand; finger flex, adjacent finger spread, wrist position, and palm orientation. The intuitiveness application uses commands for a car task, which are presented to each subject in random order. Upon the presentation of a command the user is asked to physically compose the gesture most associated with the command, and then replicate the pose on the screen by selecting buttons which control an animated 3D virtual hand model. Code representing the hand configuration is placed in a database. In addition, the subject selects a level of belief, reflecting the strength of his assertion that the gesture intuitively represents the command. During the experiment only poses within a selected set of configurations are accepted.

For collection of effort data, each trial requires a 25 second static hand pose, followed by a 15 second rest.. The subject then indicates the level of effort. This was repeated 3 times and an average recorded. The gesture poses were presented in random order. Users were asked not only to consider physical stress, but also mental effort in composing and holding the gesture.

5 RESULTS

5.1 Intuitiveness Experiment

The commands are C = [Start, Finish, Forward, Backward, Left, Right, Fast Mode, Slow Mode], which in Table 1 are numbered according to the list. As there were 35 respondents and 8 commands, a total of 280 gestural responses were made. Of these possible hand configurations, only 59 were selected by the respondents.

	Car Commands									
G	1	2	3	4	5	6	7	8	q
1	2	9	6		1		4	2	24
2	3	2	11	2			2	1	21
3	4	4	1	2	1			3	15
4	1	5		3			3	2	14
5	5	4					5		14
6	2				9				11
7						10		1	11
8				2		9			11
9					10				10
10				6				4	10
11			1		8		1		10
12	2	1	1	3		1	1		9
13				1		8			9
..									
..									
57					1				1
58						1			1
59						1			1
	35	35	35	35	35	35	35	35	280

Table 1. Intuitive Matrix - gesture-command pairs

If subject k associated gesture i with command j then

$a^k(i,j)=1$, and 0 otherwise The entry a(i,j) in Table 1(only a portion is shown) represents the number of respondents selecting the gesture in row i to represent the command in column j, i.e.;

$$a(i,j) = \sum_{k=1,..,K} a^k(i,j)\ , \quad q(i) = \sum_{j=1,..,8} a^k(i,j) \qquad (1)$$

Values of q are located in the right hand column, and represent the number of respondents (popularity) using a gesture configuration in the corresponding row. The value Nq represents the number of distinctively different gestures that had popularity value q. There were (Nq=22) least popular gestures; those selected by only one respondent (q = 1). The distribution of these responses according to popularity can be seen in Fig 4.

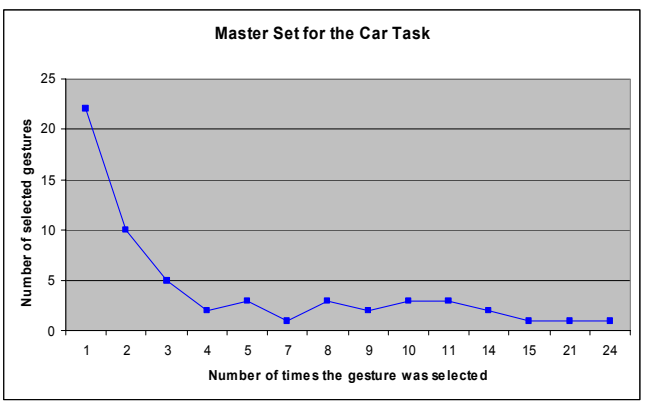


Figure 4. Popularity Graph for Selected Car Gestures

Of the 59 gestures selected for the car experiment, 32 of them were selected by only 1 or 2 respondents. Removing unpopular gestures (q = 1, 2, and 3) reduced the number of gestures to 22 as shown in Fig 5. The reduced set, represented gestures where there was some level of agreement (consensus).

Although there is useful information in the above table, it should be enhanced by considering the subjects degree of belief of gesture-command selections.

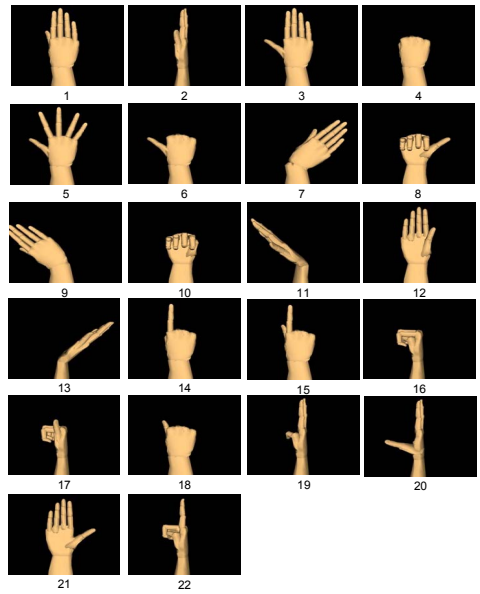


Figure 5. Reduced gesture set, car task

Each subject weighed their response by the values 1= weak, 2= medium and 3= strong to represent this belief. Consider subjects indexed as k = 1,…, K. Denote w^k (i,j) as the strength of belief of the k^{ith} respondent of associating gesture i with command j. A weighted intuitive matrix can be constructed where all the selections in Table 1 are multiplied by the strength of the association. For each command-gesture pair, a composite intuitive measure w(i,j) is obtained.

$$w(i,j) = \sum_{k=1,..,K} w^k(i,j) \qquad (2)$$

5.2 Effort Experiment

Effort responses were obtained from 29 subjects for each of 27 gesture poses. Several hard poses (with effort in the range of 8, based on the Borg scale of 0 to 1) were added to the set to widen the range of samples... For pose effort the mean and standard deviation ranges were found to be [2.0 - 4.3] and [1.1 - 1.8] (2- weak, 3- moderate, 5- strong, 10-extremely strong), respectively. Fig 6 shows examples of hard and easy poses, and there respective effort values.

Figure 6. Example of easy (2.21), and hard pose (8.14)

6 Analysis

The 3 most popular gestures were selected by 24, 21 and 15 respondents. The next 2 most popular were tied with 14 respondent selections each. Images of these five most popular gestures are shown in Fig 7.

(a) 24 (b) 21 (c) 15 (d) 14 (e) 14

Figure 7. Most popular gestures for the car set (number of users selecting gesture)

As expected these gestures are very simple to compose, one being a closed fist and another possessing an open palm with all fingers extended.

From rows 7, 8, and 13 of Table 1 it can be seen that there are strong associations between these gestures and the "right command" (col 6). As can be seen in Fig 8 these gestures are very intuitive for this command as they all tilt or point to the right.

(a) 10 /11 (b) 9/11 (c) 8/9

Figure 8. Multiple Gestures with Strong associations selected to represent the RIGHT command (times selected for Right Command/total times selected)

An agreement measure is used for determining the proportions of overall and specific gesture agreements on representing commands. S(i) is the specific agreement ratio [0, 1], i.e.; the fraction of agreement over all possible agreements, given that gesture i is selected. $S_{poss}(i)$ is the maximal possible number of agreements for gesture i. Let P_r (i) = the probability g_i was selected, then unconditioning on g_i we obtain the mean or overall agreement ratio, Φ.

$$\sum P(i) = 1 \qquad (3)$$

$$\Phi = \sum S(i) * P_r(i) = 0.38, \Phi$$

Consider gesture 6, in Table 1. This gesture was associated with command 1 and 5 with the frequency $n_{61} = 2$, $n_{65} = 9$.

$$P(j) = \frac{S(j)}{S_{poss}(j)} = \frac{2(2-1)+9(9-1)}{(2+9)(10)} = \frac{74}{110} = 0.67 \qquad (4)$$

This means there was only 67% of the maximum possible pair wise subject agreements given the selection of gesture 6 to represent commands (in this case 2 of them). Percent agreements for given gestures ranged from 11% to 100% with a mean ratio of 0.38.

In addition, there was strong evidence of the pairing of complementary gestures to complementary commands. A complementary command pair are two commands with an opposite connotations... Two gestures may be considered complimentary if they possess opposing configuration elements. Fig 9 provides three examples obtained from the experiments.

(a) (b) (c)

Figure 9. Compl. command – gesture pairings (a) Start-Finish, (b)-(c) Left-Right

With regard to the effort results the lower values obtained with means of 2 to 4.3 (2- weak, 3- moderate, 5- strong, 10-extremely strong), substantiates the notion that subjects when selecting gestures in intuitive experiment inadvertently filtered out difficult gestures.

7 Conclusion

GV design research is presently an ad-hoc procedure, and few researchers have considered the human and technical factors jointly. Finding the recognition accuracy

of a GV is no problem, as there exist many exelent classification algorithms. Measuring intuitiveness and stress (or its inverse comfort) is another matter, as these factors are subjective and must be obtained by empirical methods. This seems to be the bottle neck in the design of optimal or near optimal gesture vocabularies.

Intuitiveness results showed when commands are presented as stimuli, the gestural responses vary widely over a population of subjects. An agreement measure is developed to quantify the amount of agreement. Percent agreements of the maximal possible for given selected gestures, ranged from 11% to 100% with a mean ratio of 38 %. This result refutes the hypothesis that there exist universal common gestures to express user intentions or commands. In addition, most popular gestures possess the property that they have complimentary counterparts, which where collectively matched with complimentary command pairs. With regard to the effort results, the lower values obtained with means of 2.0 to 4.3 (2- weak, 3- moderate, 5-strong, 10-extremely strong), substantiates the notion that subjects when selecting gestures in the intuitive experiment inadvertently filter out difficult gestures.

The resultant ant methodology is useful to obtain human factors indices which are surely needed to complement technical factors in the design of optimal gesture command vocabularies. Although, our methodology does require effort to obtain human ergonomic and cognitive indices, it provides a structure for future replacement and expansion. More accurate effort (or its inverse comfort) or intuitiveness indices can easily replace old data by updating the gesture knowledge database. This work will not be lost as it can provide a database for subsequent studies. Future work includes a study of complimentary gestures and gestures transitions.

Acknowledgment

This project was partially supported by the Paul Ivanier Center for Robotics Research & Production Management, Ben-Gurion University of the Negev.

References

[1] J. P. Wachs, H. Stern and Y. Edan, “Cluster labeling and parameter estimation for automated set up of a hand gesture recognition system”, *IEEE Trans. SMC, Part A*, vol. 35, no. 6, Nov., 2005, pp. 932-944.

[2] T. Agrawal and S. Chaudhuri, “Gesture recognition using position and appearance features,” in *Proc. IEEE Int. Conf. Image Processing*, Barcelona, Spain, 2003.

[3] K. Abe, H. Saito, and S. Ozawa, “Virtual 3-D interface system via hand motion recognition from two cameras,” *IEEE Trans. SMC, Part A*, vol. 32, no. 4, pp. 536-540, Jul. 2002.

[4] C. W. Ng and S. Ranganath, "Real-time gesture recognition system and application," *Image and Vision Computing*, vol. 20, no. 13-14, pp. 993-1007, Dec. 2002.

[5] M. Nielsen, M. Storring, T. B. Moeslund, and E. Granum, “A procedure for developing intuitive and ergonomic gesture Interfaces for man-machine interaction,” Technical Report CVMT 03-01, CVMT, Aalborg University, March, 2003.

[6] H. I Stern, J. P. Wachs, and Y. Edan, “Hand gesture vocabulary design: a multicriteria optimization”, *IEEE SMC 2004, International Conference on Systems, Man and Cybernetics*, The Hague, Netherlands, Oct 10-13 2004.

[7] M. Kölsch, A Beall, and M. Turk, "An objective measure for postural comfort", *Proc. HFES 47th Annual Meeting,* 2003.

[8] N. Brook, J. Mizrahi, M. Shoham, and J. Dayan, "A biomechanical model of index finger dynamics," *Medical Engineering & Physics, 17*(1), 1995, pp. 54-63.

[9] Y. Yasumuro, Q. Chen, and K. Chihira, “Three dimensional modeling of the human hand with motion constraints,” *Image and Vision Computing*, Vol. 17, 1999, pp. 149-156.

[10] K. N. An, E. Y. Chao, W. P. Cooney, III, and R. L. Linsheid, “Normative model of human hand for biomechanical analysis,”

[11] P.A. Harling, and A. D. N. Edwards,”Hand tension as a gesture segmentation cue,” *Proceedings of Gesture Workshop on Progress in Gestural Interaction*, 1996, pp. 75 – 88.

[12] K. Shrawan and M. Anal, Electromyogaphy in Ergonomics, Taylor and Francis, 1996.

[13] G. Borg. "Psychophysical bases of perceived exertion," *Med. Sci. Sports Ex.* 14(5), 1982, pp.377-381.