Optimal Consensus Intuitive Hand Gesture Vocabulary Design

Helman I Stern[1], Juan P. Wachs[2], and Yael Edan[1]

[1]Depart. of Industrial Engineering and Management, Ben-Gurion University of the Negev, Beersheva, Israel, {helman, yael}@bgu.ac.il
[2]Depart. of Computer Science, Naval Postgraduate School, Monterey, CA, jpwachs@nps.edu

Abstract

Gesture interfaces are needed for natural intuitive communication with machine devices. Hand gesture intuitiveness is the cognitive association between a command or intent, and its physical gestural expression. Using an automated tool we quantified intuitive indices for static gesture commands for a car navigation task. A small number of gestures were selected to express most of the commands with 1/3 used only by single individuals. This followed a power function analogous to Zipf's Law for languages. We found gesture preferences to be highly individualized, providing evidence to refute the hypothesis of the universality of gestures. A mathematical program was formulated to obtain a consensus gesture vocabulary for a car navigation system with the objective of maximizing total intuitiveness. We also introduced the notion of complex consensus gesture vocabularies in which multi-gestures are associated with single commands and multi-commands are associated with single gestures. We recommend hybrid gesture vocabularies, decided by consensus with several gestures selected individually.

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 hand gesture vocabularies (GV) to support these interfaces is a relatively uncharted area of research. In a global approach to GV design, human as well as technical factors must be considered. The basic machine aspect of a GV is gesture recognition accuracy, which is well studied [1]. 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. Few researchers have considered human and technical factors jointly. In

[2] human factors are considered, but little attention is given to technical aspects. An analytical method for designing an optimal GV which does consider human and machine factors can be found in [3] and will not be detailed here. This method is based on finding a subset of gestures selected from a large master set to be matched with a given set of commands. The subset is evaluated using both human and technological factors. Finding the recognition accuracy of the subset is no problem, as there are numerous classification algorithms for this purpose. Measuring human factors in terms of intuitiveness and comfort (or its inverse - effort) are subjective and must be obtained by empirical methods. This seems to be the bottleneck in the design of optimal or near optimal GVs for human-machine interfaces, and this motivates the research reported here, although we limit our study to the intuitiveness aspect.

Although there is much reference to the intuitiveness of a gesture interface, little has been done to actually quantify this notion. This paper focuses on the development, collection and analysis of numerical indices that quantify hand gesture intuitiveness. We define intuitiveness as the cognitive association between a command or intent, and its physical gestural expression. Intuitive factors are costly and time consuming to obtain; and, hence, we develop automated methods for their acquisition through specially designed software. The association (matching) of each command with a gestural expression is defined here as a gesture vocabulary (GV). It must be emphasized, since we are interested in gesture-command vocabularies for machine interaction, the vocabularies obtained will be task dependent. The application domain considered is that of a car navigating through a 2D maze. We do not actually execute these tasks, but represent them in terms of a required set of commands.

People express themselves using gestures, such as body movements and facial expression. However, current interface technology rarely adopts such

methods when designing human-machine interfaces and, as a consequence, the expressive element is lacking [4]. Most human-machine interfaces are based on joysticks, keyboards, and keypads with few using gestures. Machine interactions based on hand gestures are commonly gloved or vision-based. A glove based system requires the user to be tethered to the computer, thus reducing user comfort and constraining the user's work space. Such devices are expensive and hard to calibrate. Because of such difficulties, we consider unencumbered, vision-based gestures in this research. Vision based gestures can be static postures (or poses), or dynamic movements. In this study we use static hand postures only. Although dynamic gestures allow more freedom, this decision was made in order to reduce the computational load of this study.

2. Hand gesture vocabulary design

The theory of universality of gestures states that some gestures have standard cross-cultural meanings [5]. This is also true for the most part within societies, but some gestures may have different meanings to different individuals [6]. Device control gestures, however, can be freely chosen to have specific meanings related to the particular device [7]. For example, there are no universally known gestures for commanding a robotic arm to "go to the home position." There has been virtually no research concerned with the issue of how to design an optimal gesture-based control vocabulary. The first step is to decide on a task dependent set of commands to be included in the vocabulary, such as "move left" or "increase speed." The second step is to decide what gestural forms are the candidates. The third step is to find the association (matching) of each command with a gestural expression.

2.1. Gesture designers

GVs can be overtly or inadvertently designed. The thumbs up and down signs come to us from Roman times, whereas the "OK" sign is more recent. Both can be considered as inadvertently designed or naturally evolved (emblems is the current term). More complete sign vocabularies have appeared in this manner without overt determination of the vocabulary by a designer. For straightforward designed vocabularies, the most common practice is for a single individual (usually a system developer) to decide which GV should be used for all users. This can be considered as the *Centrist or Authoritarian Approach* [8]. Alternatively, we can define a *Democratic or Consensus Approach*, where a group of users, either implicitly or explicitly, decide on a common vocabulary to express a given set of commands [9]. At the lowest level is the *Individual* or *Customized Approach*, where each individual defines his/her very own vocabulary [10]. One may hypothesize that the consensus and customized approaches will be more comfortable, easier to remember, and more natural to execute. The disadvantage is that the users will not consider other design factors such as the speed and accuracy of a gesture recognition system. In the centrist approach the system developer may select a GV that has high accuracy, but is unnatural for users to execute. In this paper we primarily consider the consensus approach to find gestures designed for multiple users. We use the term consensus loosely, in an implicit sense, as our subjects are unaware of the choices of others. The goal here is to focus on the problem of finding an intuitive consensus GV.

2.2. Gesture design solution methods

One of the works that explores the process of gesture design may be found in [11], where the application is that of a pen-based user interface. In a more recent work [2] a procedure and a benchmark for identifying gestures based on nine usability heuristics was presented. Below we classify solution methods into three types.

2.2.1 The specificity method. The specificity method is the prime method of determining a GV, whereby the designer of the GV pre specifies the GV to be used by all, usually with no mention of the method used or the rationale for the choices made. Many examples prevail in the literature; see for example [12].

2.2.2 Rule based approach. An example of the use of design rules can be found in [13]. They recommend such guidelines as "favor ease of learning," "use hand tension at the start of a dynamic gesture," and "use relaxed position of the hand at the end." No mention is made on how these guidelines are implemented to generate the actual vocabulary. The application allows a user to give a lecture by navigating through a set of slides with data glove-based gestural commands. In a vision-based computer interaction setting [14] a set of permissible motions users can make for control actions are defined. Stating that "the choice of such control movements is more art than science," they proceed to consider what good control actions for different types of tasks are.

2.2.3 Mathematical. Mathematical methods are scientifically based, involving the quantitative use of

human factors and other aspects such as; ergonomics, hand biomechanics, cognitive science, experimental statistics and machine recognition. Other than [15] for an ergonomic keyboard, we have not found any comprehensive work using analytical methods for the design of a GV. In a comprehensive approach there are two basic factors, technological and human. The technological factor refers to the automatic recognition of a gesture in a command interface. This depends on a good gesture recognition algorithm. With respect to human factors of intuitiveness and comfort, care must be taken to select gestures that avoid muscle strain and fatigue and are intuitive [16].

In this work we use a mathematical approach for finding an optimal intuitive GV. This will provide important insights for its use in more comprehensive future approaches.

3. Measuring hand gesture intuitiveness

In [2] two approaches are given 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. An example of the top-down approach is found in [17] whereby gestures are presented by the user and learned by an entertainment system through reinforcement learning. Nielsen exemplifies the top-down approach in a benchmark designed to test the user's chosen GV. In this paper to collect intuitive data we use the "bottom-up approach". The actual acquisition of gesture responses is not trivial. 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 similar gestures, (b) *Gesture Image Database* - 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 in gesture labeling.

An intuitive experiment containing commands for the command of a virtual car is described below. Subsequently we utilize the intuitive measures in a mathematical program to find an optimal consensus GV.

3.1. Methodology

The thirty five subjects participating in the intuitive experiment were university undergraduate students with equal numbers of male and females between 20 to 35 years of age. The subjects had no experience using GV interfaces. It is very important to establish the nature of the gesture capture system at the outset, as this will constrain the use and type of permissible gestures. We used a camera placed overhead and to the side of a monitor to allow the subjects to visualize the portion of the hand viewed when placed on the table. A software program developed to automate the collection of subject gestural responses to command stimuli is shown in Fig. 1.

Figure 1. Interface for collection of intuitive gesture - command data

A static hand pose is represented by a string of 11 elements of the hand. The "*basic elements*" are the moveable parts of the hand; finger flexion-extension, palm orientation, adjacent finger spread and wrist position. Eight commands were selected for controlling the maneuvers of a car through a maze, i.e.; C = [Start, Finish, Forward, Backward, Left, Right, Fast, Slow].

The intuitiveness application presents command words in random order. Only single hand poses were allowed. The subject then physically composes the gesture under the camera that he/she most associates with the command. The subject then replicates the pose on the screen using buttons to control an animated 3D virtual hand. A coded string representing the hand configuration is then placed in a database. During the experiment only poses within a prescribed set of configurations are accepted. Rejected poses are those physically impossibly to compose (a closed fist with all fingers spread apart) or those with features indiscernible by the camera because of occlusion (for example, a vertical palm with fingers closed or spread appear similar as viewed from above). As the gestures are selected there icons appear at the bottom of the screen, so that subject's subsequent selections may take them into account

4. Results

4.1. Intuitiveness indices

The results are shown as aggregate intuitive indices in Table 1 where the columns represent the commands for a car task, and ordered according to the command vector *C*, described previously. For 35 subjects and 8 commands, 280 gestural responses were made, 59 of them unique. The rows of Table 1 correspond to these gestures.

Table 1. Aggregate intuitive indices

Gesture i	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					10
11			1		8		1		10
12	2	1	1	3		1	1		9
13				1	8				9
14	3		3				1	1	8
15	2		2				2	2	8
16		2		2			1	3	8
17	1	1		3			1	1	7
18		2		1			1	1	5
19	1		1	2					5
20		1					4		5
21		1				1	2		4
22	4								4
23			1				2		3
24	1			1			1		3
25			1	1				1	3
..									...
..									...
..									...
57					1				1
58						1			1
59					1				1
	35	35	35	35	35	35	35	35	280

j	Command
1	START
2	FINISH
3	FORWARD
4	BACKWARD
5	LEFT
6	RIGHT
7	FAST MODE
8	SLOW MODE

Denote *(k = 1,…, K), (j = 1,..,C)* and *(i=1,….,G)* as the subject, command and gesture indices, respectively. If subject *k* is found to associate gesture *i* with command *j* then $a^k_{ij}=1$, and 0 otherwise. The entry a_{ij} in Table 1 represents the number of respondents selecting the gesture in row *i* to represent the command in column *j*, which is referred to as an *intuitive index* or measure. Values of q_i, found in the right hand column, represent the number of respondents (popularity) that selected the gesture in row i.
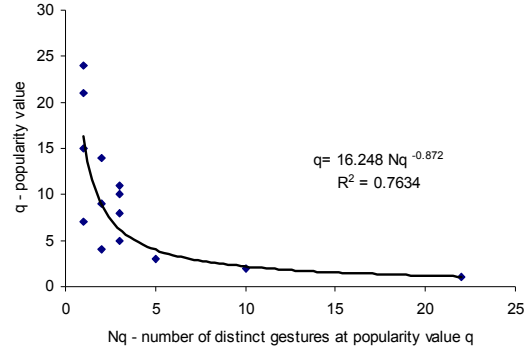
$$ a_{ij} = \sum_{k=1,..,K} a^k_{ij} \qquad q_i = \sum_{j=1,..,C} a_{ij} \qquad (1) $$

4.2 Popularity graph

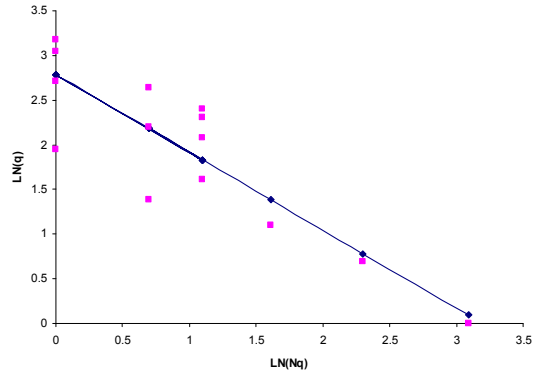Let V_q and N_q be the set and number of distinctively different gestures with popularity value $q = 1,…,Q$, respectively.

$$ V_q = \{i \ni q_i = q\}, \qquad N_q = |V_q| ; q = 1,...,Q \qquad (2) $$

The number of responses *q* vs. the number of uniquely different gestures N_q is plotted in Fig. 2(a).

q= 16.248 Nq $^{-0.872}$
$R^2 = 0.7634$

Nq - number of distinct gestures at popularity value q

(a) Gesture popularity graph

(b) Log log plot of popularity graph

Figure 2. Gesture popularity graphs

Of the 59 gestures selected for the car experiment, there were N_1=22 least popular gestures i.e.; those selected by only one respondent ($q = 1$), 37 of them were selected by only 1, 2 and 3. The most popular value was $q = 24$ with $N_{24} = 1$.

4.3 Gesture reduction

A procedure was used to find a vocabulary of manageable size. Static hand configurations were based on: 4 finger flexion-extension positions, 3 thumb positions (in, out, perpendicular to palm), 3 palm orientations (up, down, side of palm), 4 different adjacent finger spreads and wrist ulnar deviation (straight, left, right). Eleven bits in a coded string were

used to represent a posture. The total number of postures available using the coding is $2^8*3^3 = 6912$. This was reduced to 648 feasible candidates by eliminating those possessing ambiguous camera views and those that violated inter/intra finger constraints [18]. One example of an inter/intra finger constraint violation rule is: Two adjacent fingers, one open and one bent are not allowed to be spread.

Thirty five subjects and 8 command stimuli resulted in 280 responses. Of all the responses there were only 59 unique gestures types selected. The criteria for dropping gestures was to examine the gestures starting with the least popular and increasing the popularity until a gesture was found with an intuitive index greater than 2. This reduced the number of gestures to 22 (those above the solid line in Table 1). The reduced gesture set appears in Fig. 3 arranged in decreasing order of popularity. This set represents gestures where there is some level of agreement (consensus). Of these 37 of them were selected by only 1, 2 or 3 respondents.
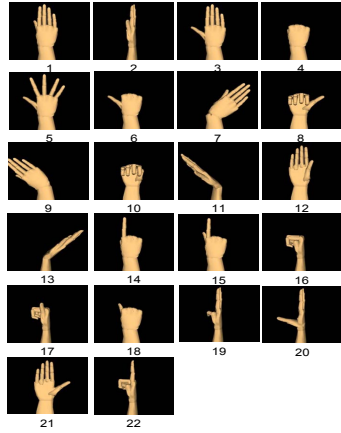

Figure 3. Reduced gesture set by popularity

4.4 Gesture popularity

The most popular gestures were selected by 24, 21 and 15 respondents. The next two were tied with 14 respondents each. Images of these gestures, labeled 1 to 5, appear in Fig. 3. As expected these gestures are very simple to compose. Gestures 7, 8, and 13 are strongly associated with the "right command" (see column 6 of Table 1), and are very intuitive as they all tilt or point to the right.

Zipf's Law, [19] found that a very small number of words make up most of what we say. Zipf's law states that the probability of occurrence of words or other items starts high and tapers off according to the power-law function $P_i \sim a/i^b$ with the exponent b close to unity. As words verbalize an intent, desire, command,

etc., hand gestures express them physically. When gestures are applied to command intentions we found an analogous result to Zipf's Law, in that a small number of gestures out of the number of possibilities are used to express command vocabularies. The function obtained is $q = 16.2/(Nq)^{.9}$ along with its log-log plot in Fig 2(a) and (b), respectively. In our experiments the maximum possible number of different gesture responses is 280. In fact, because different subjects often selected the same gesture types, the number of uniquely different gestures was 59 representing 21% of all possibilities. This reduced set of gestures represents the *basic vocabulary*. Examining the frequency of use of each gesture we found a 70:30 rule, where 70% of respondents selected 30% of gesture types.

Of the 59 different gestures in the basic set 37% (22) were selected once. It is quite surprising to find that of the gestures selected one third of them were each selected by only one subject. The commands these gestures were matched to appear in the lower rows (those with $q = 1$) of Table 1.

Formal agreement measures are developed for determining the proportions of specific S_i and overall Φ gesture agreements on representing commands (see eq. (3) and (5)).

$$S_i = \frac{\beta_i}{\beta_i^{poss}}$$
$$= \sum_{j=1,..,C} a_{ij}(a_{ij}-1) \left/ \left(\sum_{j=1,..,C} a_{ij} \right) \left(\sum_{j=1,..,C} a_{ij} -1 \right) \right. \tag{3}$$

S_i is the fraction of actual agreements β_i over all possible agreements β_i^{poss}, given that gesture i is selected.

For example, consider gesture 6 in Table 1. This gesture is associated with commands 1 and 5 with frequencies $n_{61} = 2$ and $n_{65} = 9$, respectively. Note, that the 9 respondents that associate gesture 6 with command 5 constitute 72 separate agreements; each respondent agreeing with 8 other respondents. The 2 that selected commanded 1 constitute 2 individual agreements.

$$S(6) = \frac{\beta_6}{\beta_6^{poss}} = \frac{2(2-1)+9(9-1)}{(2+9)(10)} = \frac{74}{110} = 0.67 \tag{4}$$

Thus, out of the total number of possible agreements only 67% of those selecting gesture 6 agreed on their command association selections. Agreements for given gestures vary from 11 to 100 % with a 0.38 mean rate. Let P_i be the probability gesture i is selected, Unconditioning on gesture i the overall agreement rate as shown in (5). This gives an

indication of the measure of agreement of the group studied.

$$\Phi = \sum_{i=1,...,G} S_i * P_i = 0.38 \qquad \sum_{i=1,..,G} P_i = 1 \qquad (5)$$

5. A Consensus Gesture Vocabulary

Table 1 provides a set of intuitiveness measures, but because of the variability between subjects there is no agreement on a single gesture to represent a given command. To overcome this problem one may try the following scheme. For each command, mark the largest value in its column. The rows of the marked cells represent the gestures matched to the commands. The solution contains conflicts for gestures 5 and 10 as both are assigned to different commands. In order to resolve this, a linear programming problem is formulated, where the objective is to find a one to one matching of gestures to commands which maximizes the total intuitiveness measure (6).

$$\max Z = \sum_i^G \sum_j^C a_{ij} x_{ij} \qquad (6)$$

Subject to:

$$\sum_{i=1}^G x_{ij} = 1, \qquad j = 1,...,C \qquad (7)$$

$$\sum_{j=1}^C x_{ij} \leq 1, \qquad i = 1,..,G \qquad (8)$$

$$x_{ij} \in \{0,1\}; \qquad i = 1,...,G, \qquad j = 1,...,C \qquad (9)$$

Here a binary assignment variable x_{ij} represents an assignment of gesture i to command j when set to 1, and 0 otherwise. The index a_{ij} represents the intuitiveness of representing command j by its matched gesture i. The value a_{ij} enters the sum on the right hand side of (6) only if its associated decision value x_{ij} takes on the value of 1. The values of x_{ij} are constrained by (7) and (8) which insure that each command j is assigned a unique gesture i, and each gesture i is assigned to no more than one command j, respectively. According to a well known result from linear programming theory the integer restriction on the variables in (9) may be dropped, since integer solutions are assured by the unimodularity of the constraint coefficient matrix. Solving the above results in several tied *consensus gesture vocabularies*. One is shown in Fig. 4 and Table 1 by the squared cells. One notes the presence of *complementary gesture pairs* for left-right and start-finish (thumb out-in). The recognition accuracy of the set of gestures was determined as .92

by a fuzzy c means classifier. If this accuracy is not acceptable, one may consider ranking the linear programming solutions until one is found with an acceptable recognition accuracy.
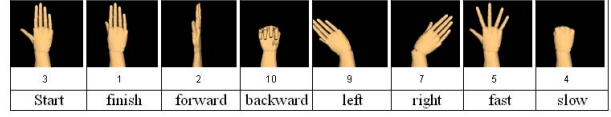

Figure 4. Consensus gesture vocabulary

6. Discussion

6.1 Static vs. other gesture types

The total number of hand configurations based on our hand primitives was in the order of 7000. This was paired down because of ambiguity due to occlusion. For example, since the camera was overhead a fist with the thumb up or thumb down could not be discerned. Such gestures, to be recognized, would require a more sophisticated capture device such as a stereo camera or multiple cameras for non ambiguous 3D object capture. This indicates the importance of the relationship between the camera capture environment and the types of permissible gestures. Although, for this study, we considered only static hand postures, the methodology could be applied to dynamic gestures. This would, however, require a video capture device. Such responses are unstable due to trajectory variations and the deformable nature of the hand. Also, a robust classification algorithm is needed for clustering similar gestures into gesture types is needed.

6.2. Complex gesture vocabularies

One notes that there appear several high intuitive indices within some of the rows and columns of Table 1. This raises the possibility of constructing *complex consensus vocabularies*. The high indices suggest that gesture 5 could be assigned to commands 1 and 7, and command 5 could be represented by gestures 6 and 9. The present application limits subjects to a single gestural response to a command stimulus. The possibility of allowing users to select multi-gestures to represent a given command can easily be obtained by restructuring the application to allow a subject to select their first, second, etc. gestural choices (this is useful for constructing individual complex GVs). Moreover, if it is desired to allow different commands to be evoked by a single gesture this is also possible. In such vocabularies multiple command matchings to single gestures can be problematic. However, in an

application, contextual awareness may be used to infer the correct intention. For example, consider the case of an up hand gesture intended to represent the two actions; "climb up" and "drive up" to command a movable object within a video game. The intent of the gesture may become clear by the presence of a human avatar or a car in the scene.

The mathematical program allows the flexibility to construct these complex consensus vocabularies of various sorts. For example, placing a 2 in the right hand side of constraint (7) for a given command allows 2 gestures for this command. For other commands, (for example commands with many low intuitive values associated with it) it may be desired to retain the 1 on the right hand side. Placement of a 2 on the right hand side of constraint (8), for a given gesture, allows it be assigned to no more than two commands.

Furthermore, optimal consensus gesture vocabularies exhibit tied solutions. This is mostly due to the fact that there are commands that are associated with a wide spread of gestures, all with low intuitive indices (little agreement amongst subjects). In such cases, the members of the cohort could not decide collectively on a common gesture to associate with that particular command. It may then be advisable to allow users to specify there own individual gestures for these commands. For example, see column 8 in Table 1 for the slow command. A *hybrid GV* is thus suggested, whereby several gestures are selected individually while the rest are determined by the consensus results.

6.3. Consensus vs. individual vocabularies

Often there exist tied optimal consensus gesture vocabularies, and others that are near optimal. This fact can be exploited through trade offs with other factors such as gesture recognition and comfort in a more comprehensive gesture design system. Although the design of a GV presented here was based on the establishment of intuitive indices which are aggregate measures of a cohort of subjects, individual gesture vocabularies can also be obtained using our methodology. One need only run the application once to obtain an individual (custom) set of gesture-command pairs.

7. Conclusions and future work

Human GV design is an emerging area of research. Few researchers have dealt with this subject in a comprehensive manner considering human (intuitiveness) and technical factors (recognition) jointly. Finding the recognition accuracy of a GV is not a problem, as many excellent algorithms exist.

Measuring intuitiveness is another matter, as this factor is subjective and must be obtained empirically. This seems to be the bottleneck in the design of an optimal GV. An Application was developed to automate the collection of subject gestural responses to command stimuli for a car navigation task. This allowed us to quantify the notion of intuitiveness by defining intuitive indices for gesture – command associations.

Our intuitiveness results show when commands are presented as stimuli, the gestural responses vary widely over a population of subjects. However, there is some agreement, and of the maximal possible number of agreements among subjects agreement percentages ranged from 11 to 100 with a mean of 38 for a car navigation experiment. We found a power function analogous to Zipf's Law, where frequency of gesture selection was high for only a few unique gesture types (following a 70:30 rule: 70% of respondents selecting 30% of gesture types) while those with low frequency occurred for many gesture types. Also, of note, 1/3 of the gestures selected were chosen only by single individuals. These results seem to refute the hypothesis that there exist universal common gestures to express user intentions or commands.

We found evidence that the most popular gestures have complimentary counterparts, which where collectively matched with complimentary command pairs. This shows how important users intuitively associate gestures in a complementary manner, and this aspect should not be overlooked in a GV design.

A mathematical program was formulated to obtain a consensus GV with the objective of maximizing total intuitiveness. We also introduced the notion of complex consensus gesture vocabularies in which multi-gestures are associated with single commands and multi-commands are associated with single gestures. These complex gesture vocabularies may be found by reforming the constraints of the mathematical program.

Although, the intuitive indices resulted in a consensus based GV, this does not preclude the construction of individual based GVs. These can be obtained by using the same application screens designed for the consensus case. It was quite surprising that one third of the individuals used gestures that no one else did. This shows how highly individualized the cognitive force of gesture selection is. Furthermore, many of the consensus gesture vocabularies had tied solutions due to the fact that several commands were associated with a wide spread of gestures, all with low intuitive indices. It may be advisable for users to specify their own individual gestures for these commands. Based on these findings we suggest, for systems used by many users, the design of a *hybrid GV* where most of the vocabulary is fixed (decided upon

by consensus), but each user has the flexibility to select several gestures that are highly individualized. Exploration of this idea is left for future work

The methodology presented here to obtain intuitive indices fills the need to complement technical factors in the design of optimal hand gesture vocabularies. Future work includes a study of dynamic gestures, camera placement, and 3D stereo views. Additional research is suggested to determine the reasons for the high diversity of the GV. Is it due to the technical constraints of the static camera poses? Will there be more of a consensus on gesture use if the possibilities are opened up through the use of free form gestures? There is also a need for a rigorous cross cultural study on the use and selection of intuitive gestures for man – machine gestural communication.

10. Acknowledgment

11. References

[1] Wachs, J. P., Stern, H., & Edan, Y. (2005). Cluster labeling and parameter estimation for automated set up of a hand gesture recognition system. *IEEE Transactions on Systems, Man and Cybernetics, Part A, 35*(6), 932-944.

[2] Nielsen, M., Storring, M., Moeslund, T. B., & Granum, E. (2004). A procedure for developing intuitive and ergonomic gesture interfaces for HCI. *Gesture-Based Communication in Human-Computer Interaction: Lecture Notes in Computer Science,* 2915 Springer-Verlag, 409-420.

[3] Stern, H. I., Wachs, J. P., & Edan, Y. (2006). Optimal hand gesture vocabulary design using psycho-physiological and technical factors. *Proceedings of the Seventh International Conference on Automatic Face and Gesture Recognition* (pp. 257-262).

[4] Card S. K., Mackinlay J. D., & Robinson, G. G. (1990). The design space of input devices. *Proceedings of the CHI '90 Conference on Human Factors in Computing Systems* (pp. 117-124).

[5] Aboudan, R., & Beattie, G. (1996). Cross-cultural similarities in gestures: The deep relationship between gestures and speech which transcends language barriers. *Semiotica*, no. 111, 269-294.

[6] Archer, D. (1997). Unspoken diversity: cultural differences in gestures. Special Issue on Visual Sociology, *Qualitative Sociology, 1,* 3-137.

[7] Cohen, C. (1999). A brief overview of gesture recognition. http://www.dai.ed.ac.uk/CVonline/local_copies/cohen/gesture_overview.html

[8] Kirishima, T. Sato, K., & Chihara, K. (2005). Real-time gesture recognition by learning and selective control of visual interest points. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 27*(3), 351-364

[9] Munk, K. (2001). Development of a gesture plug-in for natural dialogue interfaces, gesture and sign languages. *Human-Computer Interaction, International Gesture Workshop*, London, UK.

[10] Kahol K., Tripathi, K., & Panchanathan, S. (2006). Documenting motion sequences with a personalized annotation system. *IEEE Multimedia, 13*(1), 37-45.

[11] Long, A. C., Landay, J. A., & Rowe, L. A. J. (1999). Implications for a gesture design tool. *Proceedings of Conference on Human Factors in Computing Systems CHI'99*, Pittsburgh, Pennsylvania

[12] Abe, K., Saito, H., & Ozawa, S. (2002). Virtual 3-D interface system via hand motion recognition from two cameras. *IEEE Transactions Systems, Man, and Cybernetics Part A, 32*(4), 536-540.

[13] Baudel, T., & Beaudouin-Lafon. (1993). Charade: remote control of objects using freehand gestures. *Communications of the ACM, 36*(7), 28-35.

[14] Kjeldsen, R., & Hartman J. (2001). Design issues for vision-based computer interaction systems. *Proceedings of the Workshop on Perceptual User Interfaces*, Orlando, Florida.

[15] Wagner, M. O., Yannou, B., Kehl, S., Feillet, D., & Eggers, J. (2003). Ergonomic modeling and optimization of the keyboard arrangement with an ant colony algorithm. *Journal of Engineering Design*, *14*(2): 187-208.

[16] Wachs, J. P. (2007). Ph.D. Thesis, Department of Industrial Engineering and Management, *Ben-Gurion University of the Negev*, Beersheva, Israel.

[17] Hashiyama, T., Sada, K., Iwata, M., & Tano, S. (2006). Controlling an entertainment robot through intuitive gestures. *Proceedings of 2006 IEEE International Conference on Systems, Man, and Cybernetics* (pp. 1909-1194).

[18] Lin J., Wu Y., & Huang T. S. (2000). Modeling human hand constraints In *Proceedings Workshop on Human Motion (Humo2000)* (pp. 121-126). Austin, TX.

[19] Zipf, G. K. (1932). *Selected Studies of the Principle of Relative Frequency in Language,* Cambridge, MA.