International Journal of Semantic Computing
Vol. 2, No. 1 (2008) 137–160

World Scientific
www.worldscientific.com

DESIGNING HAND GESTURE VOCABULARIES FOR NATURAL INTERACTION BY COMBINING PSYCHO-PHYSIOLOGICAL AND RECOGNITION FACTORS

HELMAN I. STERN*,‡, JUAN P. WACHS†,§
and YAEL EDAN*,¶

*Department of Industrial Engineering and Management
Ben-Gurion University of the Negev
Beer Sheva, 84105, Israel
†Department of Computer Science
Naval Postgraduate School
Monterey, CA, USA
‡helman@bgu.ac.il
§jpwachs@nps.edu
¶yael@bgu.ac.il

A need exists for intuitive hand gesture machine interaction in which the machine not only recognizes gestures, but also the human feels comfortable and natural in their execution. The gesture vocabulary design problem is rigorously formulated as a multi-objective optimization problem. Psycho-physiological measures (intuitiveness, comfort) and gesture recognition accuracy are taken as the multi-objective factors. The hand gestures are static and recognized by a vision based fuzzy c-means classifier. A meta-heuristic approach decomposes the problem into two sub-problems: finding the subsets of gestures that meet a minimal accuracy requirement, and matching gestures to commands to maximize the human factors objective. The result is a set of Pareto optimal solutions in which no objective may be increased without a concomitant decrease in another. Several solutions from the Pareto set are selected by the user using prioritized objectives. Software programs are developed to automate the collection of intuitive and stress indices. The method is tested for a simulated car — maze navigation task. Validation tests were conducted to substantiate the claim that solutions that maximize intuitiveness, comfort, and recognition accuracy performance measures can be used as proxies for the minimization task time objective. Learning and memorability were also tested.

Keywords: Hand gesture vocabulary design; multiobjective optimization; gesture interfaces; hand gesture recognition; human-computer interaction; semantic behavior; psycho-physiological; intuitiveness; comfort; memory; learning.

1. Introduction

Today's conventional machine interfaces to computers and robots use keyboards and mouse control devices. Hand gestures offer an alternative interface modality which allows a more flexible, natural and intuitively expressive form. In addition, they

provide redundant/complementary modalities, overcome noisy environments, provide non-contact interaction that are useful in sterile environments, and are particularly suitable for navigation commands. Today's technology allows hand gestures to be recognized with relatively good accuracy. However, recognition activity is not sufficient, if perceptual interaction is going to become a part of the user interface. A gesture vocabulary (GV) is defined as a set of matched pairs of verbal commands or semantic behaviors and their gestural expressions. If perceptual interaction is going to become a part of the user interface, both human factors (intuitiveness and comfort) and machine factors (recognition accuracy) should be considered [1] when designing a GV.

The main aim of this paper is to respond to the need of designing natural hand gestures control systems considering high learnability, usability, ergonomic design and comfort, and to provide a rigorous methodology that discusses formally how to obtain and evaluate gestures. Few researchers have considered the human psychophysiological aspects of gesture design. Nielsen *et al.* [2], investigated matching of gestures to commands empirically through user response queries, however, limited attention was given to the technical aspects. In their conclusion it is recommended that future work extend their benchmark procedure to include technical aspects. Inspired by Nielsen, the "Wizard of Oz experiment" is used in Preston *et al.* [3] to extract common gestures grouped in classes and subsequently converted to a vocabulary. To reduce the number of gestures and the mental load Kohler [4] suggests mapping every gesture to several similar tasks from different devices. Nevertheless, no methodology is presented by Kohler for this purpose. Kjeldsen and Hartman [5] discuss design issues for vision-based computer interaction systems and provide the beginnings of a framework for a more rigorous analysis of gestural interaction systems. However, the discussion does not address how to integrate the recognition factor into the analysis. In Munk [6] a set of gestures of interest is selected in cooperation with a group of linguists. He suggested that a future implementation of his methodology should provide a benchmark for the exploration of different gestures from two standpoints; computer recognizability and subjective naturalness of gestures experienced by the user. The same usability rules proposed by Nielsen are used in Cabral *et al.* [7] while emphasizing the importance of performance evaluation of hand gesture vocabularies. Cabral and his group compare mouse and gesture interfaces, but only for a single gesture vocabulary. In [8] 2D and 3D vocabularies are based on intuitiveness, ergonomics and easy of recognition criteria. However, the first two factors are limited to the author's own considerations. A graph-based approach to understand the meaning of hand gestures by associating dynamic hand gestures with known concepts and relevant knowledge enhances simple recognition techniques with the capability of understand the meaning of ambiguous sets of phrases consisting of three to five hand gestures is found in [9]. In essence it is a method to enhance the recognition capability of the gesture system to include meaning. It requires pre-specified hand "gesture–concept" associations and hence does not consider intuitiveness or stress.

Below we provide some background on various aspects of hand gestures, especially from the psycholinguistic and physiological points of view. In addition, hand gesture digital capture methods are discussed.

1.1. *Psycholinguistic aspects of hand gestures*

There are several ways to characterize human hand gestures [10]. From the psycholinguistic point of view, a gesture has four aspects, which are hand shape (configuration), position, orientation and movement [11]. These aspects are very useful for feature extraction in machine vision. Another way to characterize hand gestures is by temporal behavior [12]. A gesture with a fixed position, orientation and configuration over time is called a *static gesture*, or posture. A *dynamic gesture* is a gesture with variations in position, configuration (or orientation) over time [13]. Hand gestures can also be classified according to their purpose such as; communicative, control, conversational and manipulative gestures [14]. *Communicative* gestures are the basic form of human non-verbal interaction and are related to the psychological aspect of communication [15]. *Control* gestures are used to control real or virtual objects. Pointing gestures, for example, would command a robot to pick up an object [16], or the hand can be used like a three dimension directional input to navigate an object in a real or virtual reality environment. *Conversational* gestures are linguistic gestures that happen during conversation and refer to the content of the speech. They have traditionally been assumed to amplify or modulate the information conveyed by speech, and hence to serve a communicative function. *Manipulative* gestures are those used to effect object rotation, translation, etc. This study is mostly concerned with control gestures.

1.2. *Physiological aspects of hand gestures*

The recognition of hand gestures is challenging, because the human hand is highly articulated and deformable. It is able to form a variety of hand gestures, and each gesture consists of many different appearances. A simple model of the hand can have 20 degrees of freedom when considering the fingers alone. This is increased when considering the orientation of the palm and the wrist. Considering just the fingers, each finger has four degrees of freedom. Two of them are for the metacarpophalangeal (MP) joint, and abduction (ABD). The other two are for the proximal interphalangeal (PIP), and distal interphalangeal (DIP) joints. In addition, there are configuration constraints derived from the joint positional limitations between the hand elements. For example, if the first two fingers are bent in the next two fingers cannot be completely extended [17]. Any gesture recognition system must be cognizant of this physiology.

1.3. *Gesture capture methods*

Two common interfaces for capturing human hand gestures are digital gloved based and vision (or video) based interfaces [12]. Glove based gestures require the user to

be tethered to the computer. This reduces users comfort and constrains the working space of the user. Also, accurate digital gloves are expensive and hard to calibrate [18]. Primarily because of these difficulties, unencumbered vision based gestures will be used in this research. Vision based interfaces require sensors such as color or infrared cameras, and image processing algorithms to: (i) segment the hand from the background and correct for variable lighting [19], (ii) select features to represent gestures [20] and (iii) recognition algorithms to classify the gestures.

1.4. *Research objectives*

The objective is to design a gesture vocabulary that is both intuitive and comfortable on the one hand, and can be recognized with high accuracy, on the other. The first step is to decide on a task dependent set of commands or actions. Commands or actions are semantic behaviors such as; "move left", "increase speed", "stop". The second step is to decide how to express the command in gestural form, i.e., what physical expression to use such as "thumbs up" or making a "V" sign. The set of semantic behaviors associated with their gestural expressions is defined here as a gesture vocabulary (GV). The procedure is formulated as a multi-objective optimization problem (MOP) with three objectives, each to be maximized: recognition accuracy, comfort and intuitiveness. Our design is aimed at application domains in which users perform tasks for extensive periods of time, so that the gestures are not held in the air but are rested on a table to avoid excessive fatigue. Hence, the gestures considered are static in nature. It should be noted that the commands and gesture sets are very task dependent, and thus a specific task was selected to test the proposed methodology, i.e., that of the control of a car navigating a maze. A testing application is developed whereby a user evokes commands using hand gestures which are captured and recognized by a vision based fuzzy c-means classifier. The car is to be controlled from a start to finish position in the maze, with a special visit to a marked object.

It is postulated that a good GV should satisfy the following objectives: (i) intuitive (cognitive association) — the gesture corresponds to its meaning or intent (for example, extending one finger represents the number one), (ii) physically easy to perform — the gesture should not cause operator strain while holding a pose or during a transition between poses, (iii) easy to recognize — gestures should be sufficiently different so that there is enough discriminatory power between them for the recognition system to classify them, (iv) easy to learn — if a gesture is intuitive and can be presented with ease (physically), it will be easier to learn, and (v) easy to remember — if the gesture is natural and intuitive it will increase the rate of recall. Note that (iv) and (v) are a consequence of (i) and (ii). Ease of recognition refers to an algorithm that automatically recognises the gesture in a command interface. As a technological factor this depends on a good gesture recognition algorithm. With respect to comfort, care must be taken to select gestures that avoid muscle strain and fatigue. In the next section the problem is defined and is formulated as a

multi-objective optimization problem (MOP). In Sec. 3 its solution methodology is described. Several software applications and methods for determining human factor indices are described in Sec. 4. Results of these experiments for a car navigation task are presented in Sec. 5. Task performance experiments and results are presented in Sec. 6, followed by validation tests in Sec. 7. Conclusions and future work are detailed in Sec. 8.

2. Problem Definition and Formulation

An optimal hand GV is defined as a set of gesture-command pairs, such that it will minimize the time τ for a given user (or users) to perform a task, (or collection of tasks). A GV may be described in terms of an assignment function p where $p(i) = j$ indicates that the command i is assigned to gesture j, i.e. GV $= \{(i, p(i) \,|\, i = 1, \ldots, n\}$. Thus, a GV is a set of gesture-command pairs. The number of commands, n, is determined by the task(s), while the set of gestures, G_n, associated with the commands is selected from a large set of hand postures, called the gesture *master-set*, G_z. Since task completion time, as a function of GV, has no known analytical form, we propose three different performance measures as proxies: intuitiveness $Z_1(\text{GV})$, comfort $Z_2(\text{GV})$ and recognition accuracy $Z_3(\text{GV})$. The first two measures are human centered, while the last is machine centered.

2.1. *Performance measures*

The performance measures $Z_1(\text{GV})$, $Z_2(\text{GV})$ and $Z_3(\text{GV})$ are linear, quadratic, and unknown as shown in (a), (b) and (c), respectively.

(a) *Intuitiveness*: Intuitiveness is the naturalness of expressing a given command with a gesture. There are two types of intuitiveness, *direct* and *complementary*. For direct intuitiveness the value $a_{i,p(i)}$, represents the strength of association between command i and its matched gesture $p(i)$. Complementary intuitiveness, $a_{i,p(i),j,p(j)}$ is the level of association expressed by the matching of complementary gestures pairs $(p(i), p(j))$ (such as mirrored poses) to complementary command pairs (i, j) (semantically opposed). Total intuitiveness is shown in (1).

$$Z_1(\text{GV}) = \sum_{i=1}^{n} a_{i,p(i)} + \sum_{i=1}^{n} \sum_{j=1}^{n} a_{i,p(i),j,p(j)}. \tag{1}$$

(b) *Stress/comfort*: Stress/comfort is related to the effort needed to perform a gesture. Total stress is obtained by the stress index of a gesture, weighted by duration and frequency of use. The value s_{kl} represents the physical difficulty of a transition between gestures k and l, and s_{ii} represents the stress of repeating the same gesture over an interval of time. The duration to reconfigure the hand between gestures k and l is represented by d_{kl}. The symbol f_{ij} stands for the average frequency of transition between commands i and j for the task under consideration.

The value K is a constant and is used to convert stress into its inverse measure, comfort.

$$Z_2(\mathrm{GV}) = K - \sum_{i=1}^{n}\sum_{j=1}^{n} f_{ij} d_{p(i),p(j)} s_{p(i),p(j)}. \tag{2}$$

(c) ***Accuracy*****:** Accuracy is a measure of how well a set of gestures can be recognized. To obtain an estimate of gesture accuracy, a set of sample gestures for each gesture in G_n is required to train a gesture recognition system. The number of gestures classified correctly and misclassified is denoted as T_g and T_e, respectively. The gesture recognition accuracy is denoted by (3).

$$Z_3(\mathrm{GV}) = [(T_g - T_e)/T_g]100. \tag{3}$$

2.2. *Formulation of the multi-objective optimization problem*

Maximizing each of the measures over the set of all feasible GVs defines the multi-objective decision problem P_1. This MOP may have conflicting solutions as all the objectives cannot be maximized simultaneously. As with most multi-objective problems this difficulty is overcome by allowing the decision maker to select the best GV according to his own preferences.

Problem P_1: Multicriteria Problem

Using a binary variable $x_{ij} = 1$ to represent an assignment of gesture j to command i, and 0 otherwise (see (9)), it is possible to rewrite analytically the intuitive and comfort objectives. The accuracy objective, however, depends on the subset of gestures selected and not on the matching. Its form is unknown and must be determined by running a gesture recognition algorithm.

$$\max Z_1(\mathrm{GV}) = \sum_{i}^{n}\sum_{j}^{m} v_{ii} x_{ij} + \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{m}\sum_{l=1}^{m} v_{ijkl} x_{ik} x_{jl} \tag{4}$$

$$\max Z_2(\mathrm{GV}) = \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{m}\sum_{l=1}^{m} u_{ijkl} x_{ik} x_{jl} \tag{5}$$

$$\max Z_3(\mathrm{GV}) \tag{6}$$

$$\text{s.t.} \quad \sum_{j=1}^{m} x_{ij} = 1, \quad i = 1, \ldots, n \tag{7}$$

$$\sum_{i=1}^{n} x_{ij} \leq 1, \quad j = 1, \ldots, m \tag{8}$$

$$x_{ij} \in \{0, 1\}; \quad i = 1, \ldots, n, \;\; j = 1, \ldots, m. \tag{9}$$

The first and second terms of the objective in (4) contain intuitive indices for the direct gesture-command assignments and complementary matchings, respectively. Higher values of v_{ii} will force gesture-command pairings, which are more

intuitive. Similarly, higher values of the complementary intuitive indices, v_{ijkl} will force solutions in which these complementary command gesture pairs are matched. In (5) maximizing comfort tends to pair commands used with high frequency to less stressful gestures. The term u_{ijkl} represents the comfort of a matched pair of gestures (i, j) to a pair of commands (k, l). Constraints (7) and (8) ensure that each command i is assigned a unique gesture, and each gesture j is assigned to no more than one command, respectively. To evaluate (6), a recognition algorithm must be called, and solved for the particular GV represented by the 0–1 assignment variables. When there is more than one non-commensurable objective function to be maximized, solutions exist for which the performance in one cannot be improved without sacrificing performance in at least one other. Such solutions are Pareto optimal points [21], and the set of all such points form the Pareto frontier. A solution x^* is a Pareto point iff there does not exist another solution y such that; $f_i(y) \geq f_i(x^*)\ \forall\, i = 1, \ldots, D$, and $f_i(y) < f_i(x^*)$ for some i, where f_i is the ith objective function.

Given a candidate gesture set of size m and a command set of size n, there are $m!/((m-n)!n!)$ different gestures subsets G_n, each of size n. For each G_n, the total number of command-gesture matching is $n!$ Solving problems with such a large solution space provides a formidable challenge, especially with a non-analytical function such as $Z_3(\mathrm{GV})$. Consequently, in this study a heuristic approach is taken whereby the problem is decomposed into two sub problems; selection of subsets of gestures followed by matching them to commands. The complete method is described in the next section.

3. Solution Methodology

The solution methodology architecture includes the following four modules (Fig. 1): (1) determination of human psycho-physiological input factors (2) determination of gesture subsets, (3) matching gestures to commands and (4) finding a set of Pareto optimal solutions.

3.1. *Module 1: Hand gesture factor determination*

The task set T, a large gesture master set G_z and the set of commands C are the input parameters to Module 1. The objectives of Module 1 are: (a) to find the comfort matrix based on command transitions and fatigue measures, and to reduce the large set of gestures, to a reduced set G_m, and (b) to determine intuitive indices. The intuitiveness, V, comfort, U, and reduced gesture set, G_m, are determined empirically through experiments.

(a) *Task and command sets* (T, C): Associated with each task t_i of a set of tasks T, is a list of commands c_i. Let C be the union of all the task commands.

(b) *Command frequency matrix* (F): For a command set C of size n, a matrix $F_{n\times n}$ is constructed where; f_{ij} represents the frequency that a command c_j is evoked

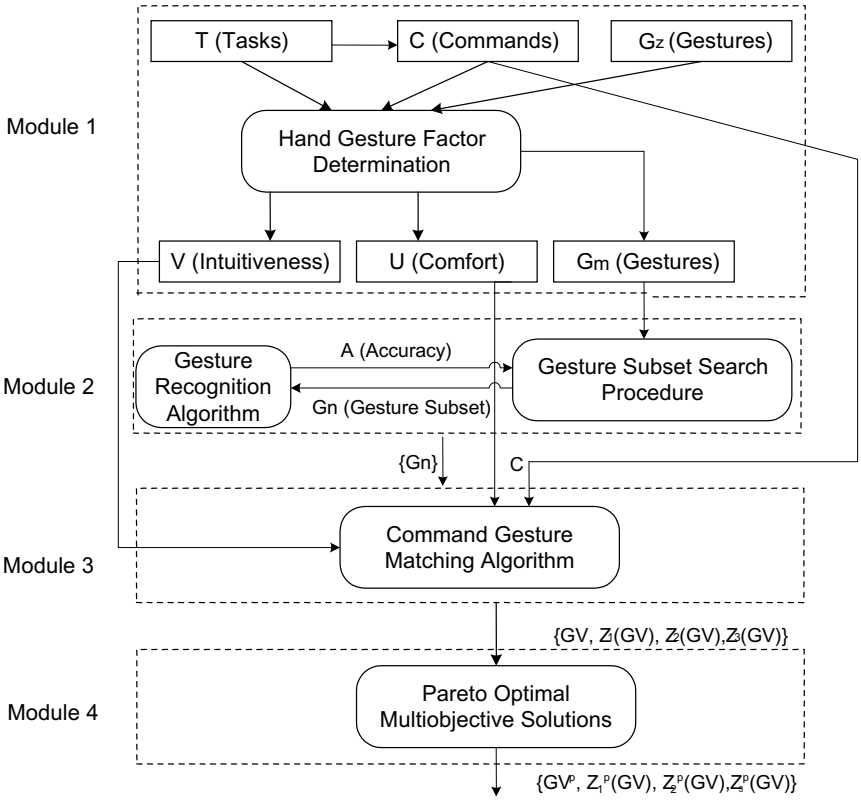


Fig. 1. Architecture of optimal hand gesture vocabulary solution procedure.

given that the last command was c_i. The frequency of command usage is determined by experiments using real or virtual models to carry out the tasks. This function resides inside the Hand Gesture Factor Determination block.

(c) *Large gesture master set* (G_z): Since the set of all possible hand gestures postures is quite large, we established a set of plausible gesture configurations based on a synthetic gesture generator using finger positions (extended, spread), palm orientations (up, down sideways), and wrist rotations (left, middle, right) as primitives. Examples of several synthetically generated hand gestures are shown in Fig. 2.

Fig. 2. Examples of synthetically generated hand gestures.

(d) *Direct intuitive matrix* ($\boldsymbol{I}$): The intuitiveness matrix, I_{mn} is composed of indices v_{ik} which represent the naturalness of using gesture i for command k.

(e) *Complementary intuitive matrix* ($\boldsymbol{I'}$): The naturalness of matching up a pair of complementary commands (i, j), such as up-down, with a pair of complementary gestures (k, l), such as thumb up–thumb down, is represented by a complementary intuitive index of the form v_{ijkl}. The matrix of complementary intuitive indices $I'_{n,m\times n,m,}$ can be quite large, but can be compacted considerably as most of the entries will be zero. Denote $V = [I, I']$ as the set of matrices including both the direct and complementary matrices.

(f) *Reduced gesture set* ($\boldsymbol{G_m}$): G_m represents the number of gestures used to represent candidates for the GV. This set is obtained by a two step reduction of the gestures in the G_z. The first step is based on eliminating gestures that are not visible from the camera due to occlusions, and also those that are impossible to form because of inter finger configuration constraints. This set is further reduced during the intuitiveness experiment whereby unpopular gestures are dropped as candidates.

(g) *Stress and comfort matrices* ($\boldsymbol{S,U}$): The stress indices are arranged in a matrix $S_{m\times m}$, whose common element s_{ij} represents the physical difficulty of performing a transition from gesture i to gesture j. Let the coefficients u_{ijkl} be the entries of a square matrix, $U_{nm\times nm}$, where the comfort index $u_{ijkl} = K - f_{ij}d_{ij}s_{kl}$ has a second term representing composite stress composed of the frequency of transition between commands i to j modified by the duration of the transition between commands i and j and by the stress of a command transition k to l given that i and j are paired with gestures k and l, respectively. This product reflects the concept that the total stress measure of GV depends on the frequency of use of a gesture or a gesture pair transition. The total comfort is the difference between the constant K and the total stress detailed above. Note, that the diagonal entries s_{ii} represent the total stress of using a gesture repeatedly to perform the same command.

3.2. *Module 2: Gesture subset selection*

The gesture subset selection problem is to find a set of gesture subsets each of size n selected from a large number of plausible gestures of size m $(n < m)$.

(a) *Gesture subset sampling*: Finding multiobjective solutions, GV, of Problem 1 can be done by sampling the solution space to obtain a representative set of gesture subsets, $\{G_n\}$. Each subset of gestures is of size n, selected from the reduced set G_m. As $\{G_n\}$ can be quite large, this requires a high computational load to determine the accuracies of each G_n in the set. Especially since each new gesture subset requires that the gesture recognition algorithm be retrained and recalibrated using a new gesture training set.

(b) ***Gesture recognition algorithm*:** The recognition process consists of: (a) extracting relevant features from the raw image of a gesture, and (b) using those image features as inputs to a classifier. Such an algorithm is described in [22], where the segmentation consists of the extraction of the hand gestures from the background using grayscale cues. The captured hand image is thresholded to a black/white segmented hand silhouette, and partitioned into block features. Using a distance metric, these features are compared to clusters obtained from a trained fuzzy c-means (FCM) clustering algorithm. The classification results are organized as a confusion matrix which allows the recognition accuracy to be computed using (4). For the gesture subset sampling procedure training the classifier and parameter calibration must be repeated many times. Since gesture classifiers such as a neural network, boosting method, etc. require large training sets, the gesture recognition algorithm selected is a fast FCM classifier which requires a relatively small training set. An automated method, based on an evolutionary parameter search procedure, is used to train and recalibrate the recognition algorithm every time it is called, providing a seamless operation (see [22]). The output of this module is a set of gesture subsets $\{G_n\}$ and associated accuracies $\{A(G_n)\}$. This accuracy is set equal to the value of the third objective $Z_3(\mathrm{GV})$, and it is understood that this accuracy function is dependent only on G_n and not on the success of matching to commands.

3.3. *Module 3: Command-gesture matching*

The inputs to the third module are the matrices; intuitiveness V, comfort U, command C, and the subset of gestures $\{G_n\}$. The goal of this module is to match the set of gestures G_n to the set of commands, C, such that the human measures are maximized, and to obtain values for the total comfort and total intuitiveness measures. A binary assignment variable x_{ij} is used to represent the inclusion or exclusion of each command — hand pose pair (i, j) in the final gesture vocabulary such that, when equal to 1 command i is assigned to gesture j, and zero otherwise. The resulting gesture-command assignment constitutes a gesture vocabulary, GV. Given a single set of gestures G_n found from module 2, the gesture-command matching can be represented as a quadratic integer assignment problem (QAP) [23] shown below as Problem P_2.

Problem P_2

$$\max \bar{Z}(G_n) = w_2 \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} u_{ijkl}x_{ik}x_{jl} + w_1\left[\sum_{i}^{n}\sum_{j}^{n} v_{ij}x_{ij} + \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} v_{ijkl}x_{ik}x_{jl}\right] \tag{10}$$

$$\sum_{j=1}^{n} x_{ij} = 1, \quad i = 1, \ldots, n, \tag{11}$$

$$\sum_{i=1}^{n} x_{ij} = 1, \quad j = 1, \ldots, n, \tag{12}$$

$$x_{ij} \in \{0, 1\}; \quad i = 1, \ldots, n, \;\; j = 1, \ldots, n. \tag{13}$$

Constraints (11) and (12) ensure that each command is paired with exactly one gesture, and each gesture is paired with exactly one command, respectively. A simulated annealing approach [24] is adopted to solve P_2. For each subset G_n found on Module 2, a set of *associated solutions* is found by varying the weights in the objective function of P_2 such that $w_1 + w_2 = 10$. This results in an expanded set of non-dominated GV solutions corresponding to each G_n in $\{G_n\}$. Their associated intuitiveness, Z_1, and comfort, Z_2, values can be determined from the unweighted second and first terms in (10) respectively after substituting the solution $G_n = \{g_i \ni x_{ik} = 1\}$, GV $= \{(i, k) \ni x_{ik} = 1\}$. The output of this module is then, the set of solutions {GV, Z_1(GV), Z_2(GV), Z_3(GV)}. The weights were varied in steps of one such that 11 associated solutions were obtained for each G_n tested.

3.4. *Module 4: Pareto optimal multiobjective solution*

Let each of the $\mathcal{N}$ solutions (gesture subsets G_n) from Module 2, have $\mathcal{M}$ associated solutions. This results in a total of $\mathcal{N} \times \mathcal{M}$ candidate GV's, each represented as a point in 3D space, (Z_1, Z_2, Z_3). The total set of multiobjective candidate solutions is then {Z_1(GV), Z_2(GV), Z_3(GV): GV $= \{1, \ldots, \mathcal{N} \times \mathcal{M}\}$. When there is more than one non-commensurable objective function to be maximized, as is the case here, our approach will be to determine the set of Pareto solutions from which a decision maker can select the GV that meets his/her internal preferences.

4. Determination of Hand Gesture Factors

4.1. *Description and experiments*

A car navigation task is used in the experiments to determine the human factor measures — comfort and intuitiveness. The task contains 'navigational' (directional) commands to control the direction of movement of the car and its speed. The three experiments needed to obtain these factors are: (i) intuitive selections of gesture poses to represent commands, (ii) the stress or effort to configure and hold static poses and transitions between poses (as well as transition duration) and (iiii) a command frequency. Figure 3 shows the place of these experiments in Module 1 for obtaining these factors.

4.2. *Subjects*

Subjects were undergraduate students in an Ergonomics course aged 20 to 35 with an equal number of male and female students. None had experience with the gesture interfaces. Thirty five subjects participated in the intuitiveness experiment, 19 in

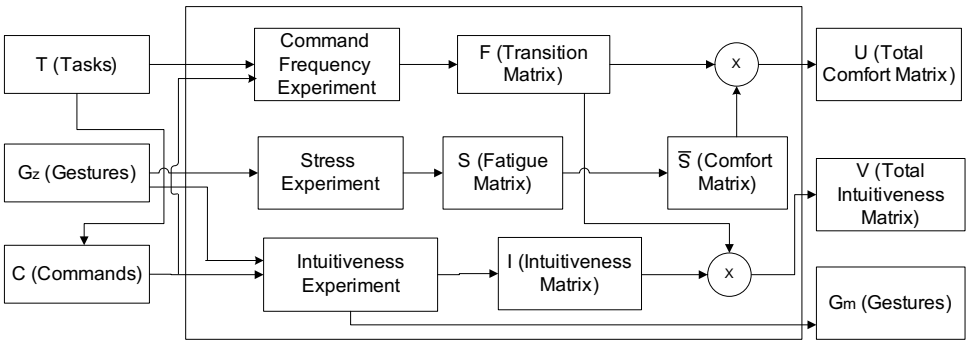
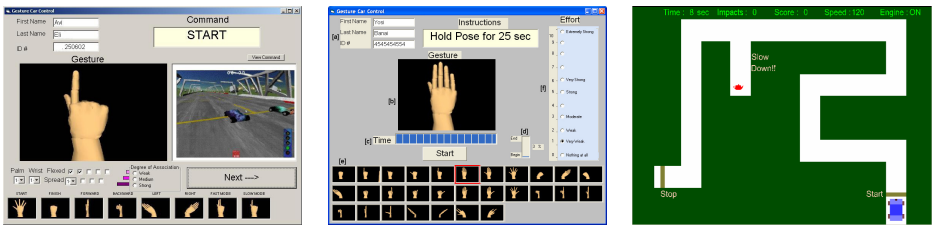


Fig. 3. Hand gesture factor determination stage.



(a) Intuitiveness (b) Stress (c) Command frequency

Fig. 4. Interfaces for the collection of intuitive, stress and command frequency data.

the effort experiment and a single experienced user in the frequency experiment. An experienced user is needed to obtain values as typical as possible when the application is used in practice.

4.3. *Equipment and software*

Subjects sit in front of a monitor, with a camera capture system to the right viewing the hand placed on a table. Three software applications automate the collection of subject responses using interface screens shown in Fig. 4 and briefly described below.

(a) ***Intuitive experiment*:** Figure 4(a) shows the interface for the collection of intuitive gesture — command data. The current command is shown at the top of the screen. On the left side is a view of a synthetic 3D hand model manipulated by check boxes under it. On the right side is a view of the car which allows the user to see an animation of the command. The bottom row shows thumbnails of the gestures selected by the user. Above this row, in the center of the screen, are radio buttons which allow the user to select the degree of belief of the association selected.

(b) *Stress experiment*: In the effort (stress) application pre-generated images are presented to each subject. The user, by button press, selects the perceived level of stress based the [0–10] Borg scale of perceived exertion [25]. Radio buttons also allow 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. Figure 4(b) shows the interface for static stress. A similar interface exists for the transition stress experiment. Additional gestures were added to the car master set of gestures to expand the range of stressful gestures.

(c) *Command frequency experiment*: A simulated car is controlled by the user to convey actions using the set of commands: $C = \{$'start', 'finish', 'backward', 'forward', 'left', 'right', 'fast', 'slow'$\}$. The scenario consists of a maze-like road, and a path to be traversed by the vehicle (Fig. 4(c)). The terminal ends of the road are marked with the start and stop lines. Also, an item, which must be visited by the vehicle, is placed in the middle of the path. Further details can be found in [26].

5. Hand Gesture Factor Results

5.1. *Intuitiveness experiment*

Only 59 out of the 280 possible gestural responses (35 respondents and 8 commands) were selected by the respondents (Table 1). Hence, there are 59 rows in the table, each corresponding to a different gesture type. If subject k associated gesture i with command j, then $v_{ij}^k = 1$, and 0 otherwise. The (i, j)th entry in Table 1 represents v_{ij}; the number of respondents selecting the gesture in row i to represent the command in column j.

$$\begin{aligned} v_{ij} = \Sigma v_{ij}^k, \quad q_i = \Sigma v_{ij}^k \\ k = 1, \ldots, K \quad j = 1, \ldots, 8. \end{aligned} \tag{14}$$

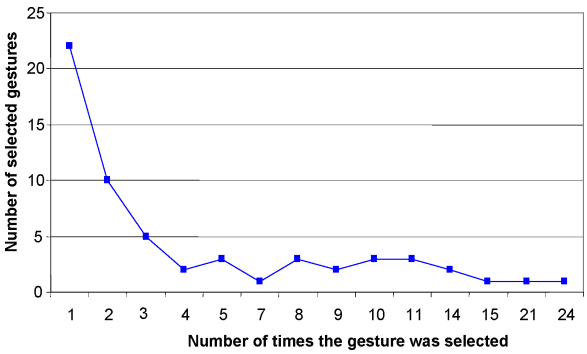
Values of q_i, located in the right-hand column represent the *popularity* of a gesture, i.e.; the number of respondents using a gesture configuration in row i. The value N_q represents the number of distinctly different gestures that had popularity value q. Note, that there were, $N_q = 22$, least popular gestures, i.e.; those selected by only one respondent ($q = 1$). The distribution of these responses according to popularity is shown in Fig. 5(a). 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(b). The reduced set, represented gestures where there was some level of agreement (consensus).

The three most popular gestures were selected by 24, 21 and 15 respondents. The next two most popular were tied with 14 respondent selections each. Images of these five most popular gestures are shown at the top of Fig. 5(b). As expected these gestures are very simple to compose, one being a closed fist and another possessing an open palm with all fingers extended. Also, from rows 7, 8, and 13 of Table 1, it can be seen that there are strong associations between these gestures and the right

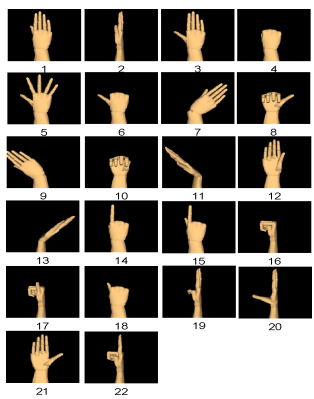
Table 1. Aggregate intuitive indices: gesture-command pairings (only a portion of the table is shown).

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
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				1	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

No	Command
1	START
2	FINISH
3	FORWARD
4	BACKWARD
5	LEFT
6	RIGHT
7	FAST MODE
8	SLOW MODE



(a) Popularity graph for car gestures



(b) Reduced gesture set (arranged by popularity)

Fig. 5. Popularity graph and reduced gesture set, G_m.

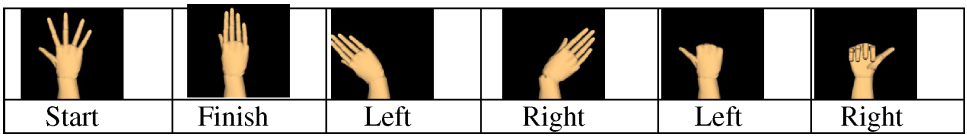


Fig. 6. Complementary command — gesture pairings.

command (col. 6). As can be seen in Fig. 5(b) these gestures are very intuitive for this command as they all tilt or point to the right.

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 complementary if they possess opposing configuration elements. Figure 6 provides three examples obtained from the experiments. The intuitiveness data was examined to extract frequency of complementary pairs. This formed the basis of the complementary intuitive indices v_{ijkl}. Further details can be found in [26].

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) were added to the set to widen the range of samples. Figure 7 shows an easy and hard pose. With regard to the effort results, lower values were obtained with means of 2 to 4.3 (2 — weak, 3 — moderate, 5 — strong, 10 — extremely strong). This substantiates a notion that subjects when selecting gestures in the intuitive experiment inadvertently filtered out difficult gestures.

An estimation function was used to fill the stress matrix $\mathrm{S}_{m\times m}$ with 462 transition values since the time to obtain these values experimentally is prohibitive. A linear function of dynamic stress s_{ij}, (the difficulty of performing a transition from gesture i to gesture j) was postulated to be a function of the static stress values of the start and end postures, denoted as s_i and s_j, respectively. The resultant regression function is shown in (15) below.

$$s_{ij} = 0.09s_i + 0.91s_j. \tag{15}$$

The F and T statistics revealed significance levels of 0.0, 0.01 and 0.0, with an r^2 of 0.98 indicating that about 90 percent of the transition effort is attributable

Fig. 7. Examples of easy and hard poses.

to the effort of forming the final posture. Similar results were found for transition duration.

5.3. *Frequency experiment*

During each trial the sequence of commands evoked by the user was recorded and parsed to obtain the frequency of pose changes (transitions). For example, given a sequence $S_1 = (1, 1, 1, 1, 2, 2, 2, 1, 1, 3, 3, 3, 2, 1, 3)$, the frequency matrix found is:

$$F = \begin{bmatrix} 4 & 1 & 2 \\ 2 & 2 & 0 \\ 0 & 1 & 2 \end{bmatrix}.$$

The entries in the F matrix include the frequencies, f_{ii}, of repeating the same command (diagonal), and the transition frequency, f_{ij} between commands i and j (off diagonal) summed over all the sequences. Results indicate the importance of experimental analysis of specific tasks. In the frequency matrix obtained from the experiment, the 'rest' command occurred at a far higher frequency than any other command in the task. This does not mean that there were long pauses while performing the task, but the pauses were very frequent. The matrix shows that between executions of any two commands, there was a short rest. Except for the 'start' command, and the 'finish' command, which occurred only at the beginning and end of the task, there were no transitions either to 'start' or from 'finish' registered in the sequence. In addition, duration of each command and the durations of breaks (intentional and unintentional) between commands were obtained. Further details may be found in [26].

6. Performance Experiments and Results

6.1. *Gesture subset selection*

For the multi-objective decision approach, a reduced complete search was adopted. Instead of inspecting $C_n^m = 1.2 \times 10^{10}$ solutions ($n = 8$ and $m = 22$), a sample of 600 solutions from the solution space was obtained. Each solution is a subset of gestures of size n.

6.2. *Command gesture matching*

For each G_n, Problem P_2 was solved using weight combinations such that, $w_1 + w_2 = 10$ for $w_1 = [0, 1, 2, \ldots, 10]$, to obtain 11 associated solutions {GV, Z_1(GV), Z_2(GV), Z_3(GV)}. This results in a total of 6600 MOP solutions.

6.3. *The Pareto set*

A 3D plot of all the solutions including 98 Pareto solutions are shown in Fig. 8. The Pareto solutions can be offered to the decision maker to select a GV according to his/her own preferences.

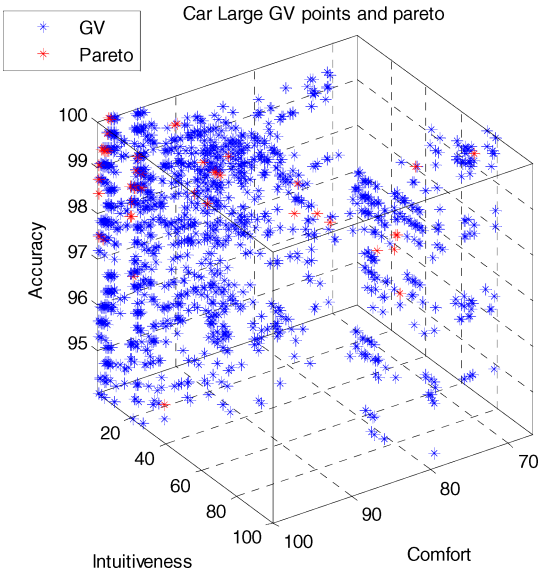


Fig. 8. 3D plot of the MOP GV and Perato solutions obtained using a semi-complete search.

6.4. *Sample user selected solutions*

The interface user may wish to consider the 1st, 2nd and 3rd priorities as accuracy, Z_3, comfort, Z_2, and intuitiveness, Z_1, respectively. Using these priorities, the following solution was obtained from the Pareto set: GV $= \{21, 16, 6, 18, 7, 8, 25, 10\}$. The associated indices to this solution are $Z_1 = 72$, $Z_2 = 4167$ and $Z_3 = 100\%$. If, alternatively, the interface user is willing to accept a lower comfort in turn for higher intuitiveness, he/she may pick the GV with a maximum intuitiveness of 2907 which has a comfort of 2992, without affecting the recognition accuracy, GV $= \{8, 6, 26, 27, 12, 10, 18, 7\}$. The images of the first and second solutions are shown Figs. 9(a) and 9(b), respectively.

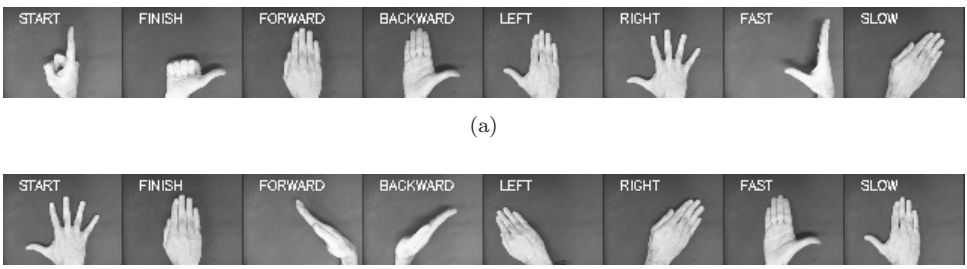


Fig. 9. Sample user selected solutions. (a) First priority is accuracy. (b) First priority is intuitiveness.

Clearly the solution in Fig. 9(a) is less intuitive compared to that of the solution in Fig 9(b). Note the lack of complementary intuitiveness in gesture-command pairings in solution (a) and their presence in (b). However, the comfort decreased significantly in (b). Slanted gestures cause ulnar deviation, extension and flexion at the wrist, and therefore, are harder to perform [27].

7. Validation

As stated in the introduction, it is postulated that minimization of performance time τ is represented by maximizing the three multi-objective proxy measures of intuitiveness, comfort and accuracy. To validate this claim the following hypothesis is tested, H_1: $\tau(V_G) < \tau(V_B)$. This states that using a vocabulary V_G will result in a shorter task completion time than using V_B where V_G is a vocabulary that is highly intuitive, comfortable, and easy to recognize, while V_B is a low intuitive, stressful, and hard to recognize vocabulary. A description of the testing vocabularies and a procedure for obtaining them is provided below. Also, in order to validate the claim that good GVs are easier to learn and remember, the following two hypotheses are tested; $H_2 : l(V_G) > l(V_B)$ and $H_3 : m(V_G) > m(V_B)$, where l and m represent learning and memory, respectively.

7.1. *Testing vocabularies*

The testing vocabularies are divided into good and bad sets according to a dominating principle which is used to guide their selection.

(a) *Dominate set partition: Good and bad GV solutions*: Given a set of n of multi-objective solutions, let $Z(i) = [z(i,1), z(i,2), z(i,3)]$ represent the ith such solution. For any pair (i,j) of solutions let the relation $i \succ j$ to denote that solution i dominates solution j, iff $z(i,k) > z(j,k), k = 1,2,3$. Let $[V_G, V_B]$ be a dominant pair partition of n solutions, if $V_G \cap V_B = \phi$, and $|V_G \cup V_B| = n$ AND for any two solutions $\forall i \in V_G, j \in V_B, \ni i \succ j$.

(b) *Determination of dominating sets*: Given a set of multi-objective solutions the dominate sets can be found by the following simple procedure. Create two lists L_G and L_B with each element on the list represented by a triple (arranged in 3 columns): Select a pair of solutions and test if they satisfy the dominating pair relation, if so place the dominating solution in L_G and the non dominating solution in L_B. Rearrange the values in each column of L_G and L_B in decreasing and increasing order, respectively.

(1) Let the first element in the lists be $z_G(k)$ and $z_B(k)$ for all k.
(2) Select a $Z(i)$ solution, if $z(i,j) > z_B(k)$, $\forall k \in L_B$, then place $Z(i)$ in the V_G list and reorder the column values. Otherwise, if $z(i,j) \leq z_B(k), \forall k \in L_G$ then place $Z(i)$ in the V_B list and reorder the column values and return to 1 until a sufficient number of solutions are obtained.

Table 2. Example of good and bad testing GVs.

	Commands								Criteria		
	1	2	3	4	5	6	7	8	z_1	z_2	z_3
V_B	5	1	2	22	4	10	26	3	204	3488	84.7
V_G	21	7	6	17	26	27	8	10	3020	3801	99.7

(c) *Good and bad testing vocabularies* (V_G and V_B): Using the procedure explained above, 16 GVs were obtained; eight dominating solutions (V_G) and eight dominated solutions (V_B). A sample dominating pair is given in Table 2. The order of the columns of GV follows the order of the command vector C. Thus, the qth column is associated with the qth command, and the table entries in this column indicate the gesture number assigned to it.

7.2. *Validation of performance hypothesis*

(a) *Overview*: This experiment is designed to test the claim that the multi-criteria performance measures Z_1, Z_2, Z_3 act collectively as proxies for task completion time. This implies that good vocabularies, indicated by high intuitiveness, comfort, and accuracy, correspond to reduced task completion time and vice versa. Restated in terms of a hypothesis (for a given task). H_1: V_G and V_B are sets of GVs where $V_G > V_B$, i.e. if $Z_i^G > Z_i^B$ $(i = 1, 2, 3) \Rightarrow \tau(V_G) < \tau(V_B)$. The validation experiment consists of operating a car through a maze. For each GV, a learning curve for repeated trials of carrying out the task is needed. The learning curves show steady improvements in performance as the number of repetitions are increased [28]. We selected the so called "standard times" of the learning curve to represent the run time performance of a given GV. Standard time occurs at the point where improvements become small. The form of the curve is:

$$Y_n = Y_1 n^{-b} \tag{16}$$

where Y_n is the estimated value of the completion time, in seconds, of the nth trial, n is the trial number, Y_1 is the time of the first trial, b is: $b = \log r / \log 2$, and r is the learning rate.

(b) *The experiment*: A software application is designed to automate the experiment (Fig. 10). The top left side of the screen contains a small window of continuous video images of the hand gesture acquired by a Panasonic Video Imager. A small label with the name of the command appears in the top left of this window when the gesture is recognized; otherwise, "Unrecognized" appears. Thumbnail images of each gestures and the associated command in the GV are displayed as a gesture posture reminders. Below the capture window, in the main area of the screen, the car maze appears. Six female and ten male subjects participated in the experiment. All were third year engineering students, and used a setup similar to the

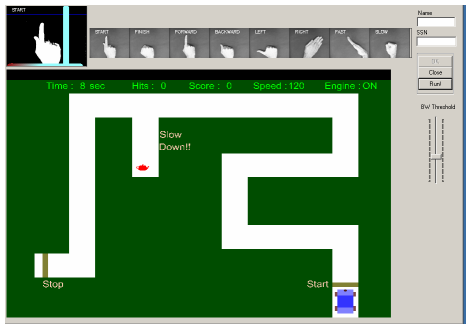


Fig. 10. Gesture car control task screen.

collection of intuitiveness data. The only difference is that the video capture stream was activated, using the WE-160 Panasonic Video Imager connected to the Matrox Meteor Standard frame grabber. The user controlled the actions in the application by evoking commands using hand gestures. Sixteen vocabularies were used in the experiment; eight good GV's and eight bad GV's, with each test repeated 15 times. Previous experiments showed that 15 trials were enough to reach standard times. At the start of each experiment it was explained to the user that the car was to be controlled from a start to finish position in the maze, with a special visit to a marked object.

(c) ***Performance validation — results and analysis*****:** The task completion time data is used to construct user learning curves. Each learning curve is based on fitting a curve of the form shown in (16) to a scatter plot of 120 points obtained by 8 users each running a different GV for 15 repetitions each. The learning equations for the V_G and V_B vocabularies are $Y_n = 229.9\ n^{-0.273}$ and $Y_n = 302.2\ n^{-0.260}$. To test the hypothesis, a t-test was performed between "standard completion times" for the V_G and V_B sets. The standard times are the average of the last three trials of the learning curves. The completion time using V_G was significantly shorter ($p = 0.00031$) than that of V_B ($\tau(V_G) = 114.67\,\text{sec} < \tau(V_B) = 153.04\,\text{sec}$). Therefore, the main hypothesis is true, indicating that the use of more natural vocabularies have a positive direct impact on the performance of the task, by reducing its completion time.

7.3. *Validation of learning hypothesis*

In the case of V_G, the first trial is much shorter than when using V_B, (230 sec $<$ 302 sec) which corroborates that V_G is easier to use for a beginner than V_B. Also, the learning rate was lower for V_G than for V_B ($0.827 < 0.835$). A smaller learning rate represents faster learning, supporting the posit that beginner users will learn faster, and thus achieve shorter standard times more quickly, when using V_G than when using V_B.

7.4. *Validation of memory hypothesis*

(a) ***An overview*:** To establish whether there is a relation between the naturalness of a GV and the memorability of the subject when using that GV, a post-validity experiment was conducted. This experiment was performed immediately after finishing the task completion time experiment. Using a software application users matched commands to gestures, reflecting the associations existing in the GV that were assigned to them. The goal here is to validate the hypothesis H_3: $m(V_G) > m(V_B)$, i.e., the V_G type vocabulary has greater memory recall than the V_B type vocabularies. The same sixteen subjects that participated in the performance time experiment continued here. The memorability experiment required a different computer station placed next to the computer with the Panasonic Video Imager, and hence, the user was not able to see the previous setup to avoid clues in the memory testing process. In this PC station, a software application automated the collection of recall data.

(b) ***Memory rates — Results and analysis*:** Memorability was determined by the experienced user's recall of the gesture-command associations. The score for this task was a measure of memorability based on the percent of correct associations. The average percent memorability scores were high at 96.7 and 95.00% for V_G (Good GV) and V_B (Bad GV), respectively. The t-test showed that the difference was far from being significant ($0.58 \gg 0.05$) at the 5% level. It was found that using the V_G, 5 subjects matched all commands to the correct gestures. In the worst case, two miss-matchings were done by one subject. Using V_B, four subjects found all the associations (0 mistakes), and in the worst performance one subject made 3 mistakes. However the results for this test were not statistically significant. It seems that the reason that there was no significant difference in memorability for good and bad vocabularies is that the car task included only eight commands-gestures associations. It was not difficult to remember a limited number of associations even when there was no correlation at all between the objects to be associated. It is hypothesized that when the number of associations grows, any clue that may help to find a correct association is highly valuable. Increased naturalness of gesture-command associations in larger vocabularies is a considerable clue that should result in increased memorability.

8. Conclusions and Future Work

This paper presents an optimal hand gesture vocabulary design methodology that considers both human factors and technical aspects. The first aspect includes the intuitiveness and comfort attributes of gesture vocabularies, while the last is related to the accuracy of the hand recognition algorithm. The most salient advantage of the approach is a structured formulation of the GV design problem in a rigorous manner so human psycho-physiological and technical aspects are combined in a unified approach. Lots of research has been devoted to increasing the accuracy of gesture recognition algorithms. However, 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 bottleneck in the design of optimal or near optimal gesture vocabularies. As such, we have developed software applications which enable the automated collection of intuitiveness and comfort indices. For intuitiveness indices the cognitive associations between commands and gestures were collected. For the comfort indices a similar method was employed in addition to methods for predicting stress and duration of transitions between pairs of gestures.

The problem of optimal design of a hand gesture vocabulary is formulated as a multiobjective decision problem, where the criteria are maximization of intuitiveness, comfort and recognition accuracy. A two stage decomposition approach is suggested for solving the optimal hand gesture vocabulary design problem. The first stage finds a feasible subset of gestures from the master set, given some recognition accuracy threshold. The second stage finds a set of gesture vocabularies, each obtained by finding the best match between commands and gestures so that a weighted sum of the total intuitiveness and comfort are maximized. After a large set of gesture vocabularies are generated, their Pareto set is constructed. Among these, we show how a user can select a gesture vocabulary according to his or her subjective priorities. Three main objectives (accuracy, intuitiveness, comfort) were included as proxies of task time performance using a GV. A hypothesis that these proxies can be used in lieu of a time performance objective was validated by an experiment which compared task completion times for sets of good and bad multiobjective solutions. Additionally, the hypothesis that good GV are easier to learn was validated. This was not the case for memorability due to the small vocabulary size, and the hypothesis that increased naturalness of gesture-command associations in larger vocabularies is a considerable clue for increased memorability will be tested in future work. Of interest is the relative effect of each of the individual objective factors on overall task time performance. As the validation of performance hypothesis was conducted for extreme cases using this data would warp any results. It is recommended that the significance of this affect should be studied in any future research.

Our comprehensive approach to the design of hand gesture machine interfaces results in improved task oriented hand gesture vocabularies that are cognizant of a cohort of users. Additionally, the human factors studies provide a data repository of intuitiveness and comfort measures, and an automated methodology for their collection. Our methods can be adapted to evaluate the design of other types of gestures such as those in free space, dynamic, two handed, encumbered, etc. For the case of dynamic gestures, although the basic methodology remains sound, it is expected that some retooling of the data collection procedures will be required (for example, segmenting out particular gestures from the video stream). In addition, because of the increased richness of dynamic gestures, it is expected that the number of distinct gestures selected will increase. The scale of such increased complexity is an issue for future research. However, from our study of static gestures

we have found a principle similar to Zipf's law where a small number of gestures, from that of the seemingly infinite number, were selected to represent most of the commands. It is anticipated that this principle will also hold for dynamic gestures although additional work is needed to verify this. Designing a gesture vocabulary by hand can be done for small vocabularies, especially for a single individual, but is not feasible for larger GVs. In contrast, our methods provide a consistent and reliable determination which may be repeated by other researchers. Also, our automated experimental methods provide measures for intuitiveness and comfort which account for differences between users. Obtaining such a "consensus" result for a user independent gesture interface would be difficult to obtain by hand.

Acknowledgments

This research was partially supported by the Paul Ivanier Center for Robotics Research & Production Management, and by the Rabbi W. Gunther Plaut Chair in Manufacturing Engineering, Ben-Gurion University of the Negev.

References

[1] T. Baudel and M. Beaudouin-Lafon, Charade: Remote control of objects using free-hand gestures, *Commun. ACM* **36**(7) (1993) 28–35.

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

[3] S. Preston, L. Matshoba and M. C. Chang, A gesture driven 3D interface, Technical Report CS05-15-00, Department of Computer Science, University of Cape Town, South Africa, 2005.

[4] M. R. J. Kohler, System architecture and techniques for gesture recognition in unconstraint environments, *Proceedings of the 1997 International Conference on Virtual Systems and MultiMedia* (VSMM'97), 1997, pp. 137–146.

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

[6] K. Munk, Development of a gesture plug-in for natural dialogue interfaces, Gesture and Sign Languages in Human-Computer Interaction, *International Gesture Workshop, GW 2001*, London, 2001.

[7] M. C. Cabral, C. H. Morimoto and M. K. Zuffo, On the usability of gesture interfaces in virtual reality environments, *Proceedings of the 2005 Latin American Conference on Human-Computer Interaction*, Mexico, 2005, pp. 100–108.

[8] A. A. Argyros and M. I. A. Lourakis, Vision-based interpretation of hand gestures for remote control of a computer mouse, *Proceedings of the HCI'06 Workshop* (*in conjunction with ECCV'06*), LNCS 3979 (Springer-Verlag, 2006), pp. 40–51.

[9] B. W. Miners, O. A. Basir and M. Kamel, Knowledge-based disambiguation of hand gestures, *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, October 6–9, 2002, pp. 201–206.

[10] D. Efron, *Gesture and Environment* (King's Crown Press, NY), 1941.

[11] W. C. Stokoe, *Semiotics and Human Sign Languages, Approaches to Semiotics Series*, C. Baker and R. Battison edition, Mouton, The Hague, 21, 1972.

[12] V. Pavlovic, R., Sharma and T. Huang, Visual interpretation of hand gestures for human-computer interaction: A review, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **19**(7) (1997) 677–695.
[13] W. Freeman and M. Roth, Orientation histograms for hand gesture recognition, *International Workshop on Automatic Face and Gesture Recognition,* Zurich, June, 1995.
[14] Y. Wu and T. Huang, Vision-based gesture recognition: A review, *Proceedings of the International Gesture Recognition Workshop*, 1999, pp. 103–115.
[15] A. Kendon, *Current Issues in the Study of Gesture, The Biological Foundations of Gestures: Motor and Semiotic Aspects*, Lawrence Erlbaum Associates, Hillsdale, NJ, 1986, pp. 23–47.
[16] R. Cipolla and N. J. Hollinghurst, Human–robot interface by pointing with uncalibrated stereo vision, *Image and Vision Computing* **14**(2) (1996) 171–178.
[17] K. N. An, E. Y. Chao, W. P. Cooney and R. L. Linsheid, Normative model of human hand for biomechanical analysis, *Journal of Biomechanics* **12** (1979) 775–788.
[18] J. J. LaViola, Whole-hand and speech input in virtual environments, Master's Thesis, CS-99-15, Brown University, Department of Computer Science, Providence, RI, 1999.
[19] J. Triesch and C. Malsburg, Robotic gesture recognition by cue combination, *Proceedings of the Informatik'98, 28th Annual Meeting of the Gesellschaft für Klassifikation*, Magdeburg, Germany, 1998, pp. 223–232.
[20] L. W. Campbell, D. A. Becker, A. Azarbayejani, A. F. Bobick and A. Pentland, Invariant features for 3-D gesture recognition, *Proceedings of Automatic Face and Gesture Recognition FG'96*, 1996, pp. 157–162.
[21] V. Pareto, Manuel, *D'Economie Politique*, 2nd ed. (Marcel Giard, Paris, 1927).
[22] H. I. Stern, J. P. Wachs and Y. Edan, Parameter calibration for reconfiguration of a hand gesture tele-robotic control system, *Japan–USA Symposium on Flexible Automation*, Denver, CO, July 19–21, 2004.
[23] T. C. Koopmans and M. J. Beckmann, Assignment problems and the location of economic activities, *Econometrica* **25** (1957) 53–76.
[24] D. T. Connolly, An improved annealing scheme for the QAP, *European J. Operational Research* **46** (1990) 93–100.
[25] G. A. Borg, Perceived exertion as an indicator of somatic stress, *Scandanavian Journal of Rehabilitation Medicine* **2** (1970) 92–98.
[26] J. P. Wachs, Optimal Hand Gesture Vocabulary Design Methodology for Virtual Robotic Control, Ph.D. Thesis, Department of Industrial Engineering and Management, Ben-Gurion University of the Negev, Beersheva, Israel, Oct. 2007.
[27] T. Griffins, Usability testing in input device design. On the web: http://tim.griffins.ca/writings/usability_body.html.
[28] T. P. Wright, Factors affecting the cost of airplanes, *Journal of Aeronautical Sciences* **3**(4) (2001) 122–128.