An analytical framework to measure effective human machine interaction

Juan Wachs[1], Brad Duerstock[2]

[1]School of Industrial Engineering, [2]Center for Paralysis Research
Purdue University, West Lafayette, IN 47907

ABSTRACT

Initially, human machine interface (HMI) was understood as the hardware and software through which human and machine could communicate. Gradually, is being recognized that many human factors such as usability, emotion, user's physical and cognitive characteristics, domain knowledge, contribute as much to the effectiveness and efficiency of HMIs as robust, reliable and sophisticated algorithms do. Clearly, both the human centered factors and the technical factors have direct or indirect relations with the effectiveness of the HMI. Nevertheless, the degree of influence of these factors on the effectiveness of human machine interaction is not well understood. Most of the work in the human machine interaction area is focused on creating and refining techniques and algorithms, application-driven efforts, or heuristic procedures, but there is a lack of basic or foundational work. In this work, we present a novel development of an accessible interface for the control of a robotic arm based on natural, effortless hand gestures designed for students with mobility impairments, and we provide a systematic framework to measure the effectiveness of this interface.

Keywords: human machine interfaces, hand gestures recognition, assistive technologies, robotics, intelligent wheelchairs.

INTRODUCTION

Physical access to classrooms, laboratories and learning resources is crucial for students with disabilities. Active participation once present, however, is also vital, encompassing interaction with teachers, other students, and engagement with course materials and equipment (Salend, 1998). In order for students with mobility impairments to gain educational experiences comparable to those of able-bodied students they must perform comparable tasks (Warger, 1998). Actively exploring and interacting with scientific concepts and practices grants a more thorough educational experience as a whole. This need for active learning, however, creates serious hurdles for students with disabilities.

Robotic assembly tasks and navigation planning and control are the most common tasks in automation and production labs. Able-bodied students are expected to personally control robots in assembly tasks, and analyze and design robot manipulations, in both undergraduate academic courses and high school science classes. In order to perform independent graduate or postgraduate research, or to pursue an engineering career, such as automation engineer, manufacturing engineer, or controls engineer, students with disabilities must be able to independently operate a robot in real-time.

There has been extensive research on the use of sensors that allow people with disabilities to interact with machines. Sensors allow the control of devices by eye-blinking, gaze, breathing, EEG and EMG signals, posture and gestures, lip reading and tongue movements. There are two main problems with these interfaces: (a) they are non-adaptive. Most of these methods leverage the strength of a single limb or body part that functions relatively well (Kim et al, 2006). Different solutions are needed, however, in cases of progressive illness where limb control skills decay gradually with time, or when the user is rehabilitating and hence has improving motor skills. In fact, technology permitting a single modality of interaction is appropriate only when the user's condition is stable. As most paralyzed people experience a change in their condition throughout their lives, a new paradigm is needed; (b) their design does not follow an analytic methodology.

The term "effective interface" in the context of human-robot interaction is relatively new and so far there is no universally accepted definition for this term. Most of the existing definitions are unstructured and only focus on one aspect of effectiveness, for example Olsen and Goodrich (Olsen and Goodrich, 2003) only focus on effectiveness as a function of task effort. In this proposal, the PI identifies a set of factors that influence the effectiveness of the interface, and attempts to organize then in a comprehensive and coherent framework. In order to evaluate the effectiveness of a given interface for robotic control, first performance measures need to be defined.

METHODOLOGY

STUDY THE FACTORS THAT INFLUENCE THE EFFECTIVENESS OF INTERACTION MODALITIES FOR ROBOTIC CONTROL

This section addresses foundational problems of the human-machine interaction area – how to define the effectiveness of a modality (or interface) used for interaction, and how to measure it? In order to evaluate the effectiveness of hand gestures over standard interface techniques for robotic control, performance measures must be defined. Interface effectiveness can be defined as a function (1), which is optimum when the interface used is the best among the options available. Different users may prefer a different interface according to their physical abilities (joystick, keyboard, hand gestures, sip-n-puff, EEG and EMG based signals, tongue control, etc).

$$\underset{I \in \Gamma}{Max}\; e(I) = f(T, U, M, E, L) \tag{1}$$

where:
e(I) = is the effectiveness for a given interface
f is some inverse function of *e,* including the following:
T = task completion time
U = user skills, expertise and knowledge domain
M = the number of discrete user expressions (physical or physiologic) required to complete a single operation.
E = number of user errors while completing the task.
L = learning rate (based on learning curve)
I = is an interface modality
Γ = the set of all feasible interface modalities (e.g. joystick, keyboard, hand gestures)

The function *e* defines the relationship between the interface and its effectiveness. For example, an interface that is easy to learn will improve task efficiency. The common measure of efficiency is the time to complete a task (T). Benchmark robotic tasks will be used in the user studies so that the results are comparable across different interfaces.

An interesting feature of this formulation is that it considers the user's knowledge and experience, since they affect the task performance, which in turn affects the interface's efficiency. The algorithm used to decode the signals into significant robot commands is not considered in this scheme, since it is extrinsic to the interface adopted. For example, there are several algorithms used for tongue movement recognition, and even that is possible to measure the superiority of some algorithms over the others, the success of these systems does not rely as much on the algorithm refinement, as it relies on the particular way the interface is used. The number of discrete movements required for the user to operate the robot is an indication of the cognitive load, the complexity, and the level of performance required in the task. The measurement of effectiveness involves the evaluation of (1), the analytical form of which is unknown. Therefore, a set of multiple objective performance measures are proposed to act collectively as proxies for (1):

task completion time Z_1, number of movements Z_2, number of errors Z_3, and learning rate Z_4. These proxies do not include U, since use experience is directly related to the learning rate. The recognition accuracy of the interface is not included, since it impacts the task performance indirectly. The goal is to use a formulation that is independent of the sensing technology. Since *f* is some inverse function of *e*, bringing the different objectives to minima will lead maximum interface effectiveness.

$$\begin{array}{l} Min\ Z_1(I), Min\ Z_2(I), Min\ Z_3(I), Min\ Z_4(I) \\ I \in \Gamma \end{array} \tag{2}$$

This multiobjective optimization problem may have conflicting solutions when all the objectives are minimized simultaneously. As with most multiobjective problems, this difficulty is overcome by allowing the decision maker (the user) to select the best *I* according to his preferences. Another method of overcome the conflicting multiobjective values is to adopt a goal programming approach: map the four performance measures into a single measure using weights w_i to reflect the relative importance of each objective.

$$\begin{array}{l} \underset{I \in \Gamma}{Min}\ f(I) \propto \underset{I \in \Gamma}{Max}\, Z(I) = w_1 Z_1(I) + w_2 Z_2(I) + w_3 Z_4(I) + w_4 Z_4(I) \\ s.t. \\ \sum_i w_i > 1 \\ w_i > 0\ \forall i \end{array} \tag{3}$$

where:
w_i = the relative importance of factor Z_i.

The weights in (3) can be found empirically by letting the decision maker assign importance to each factor according to his/her needs and preferences. Alternatively, the weights can be varied, and for each unique weighting scheme the corresponding solution can be presented to the user for acceptance or rejection. The objectives Z_i will be calculated by running simulations of a task using a virtual model of a robot (like the one in Figure 1), which each of the interfaces considered, for example: (a) standard joystick, (b) voice and gestures, (c) EEG signals. Then, (3) can be computed and comparisons can be made among the interfaces.

A NEW METHOD FOR GESTURE BASED ROBOTIC CONTROL

The concepts described in Tele-gest project (Wachs et al, 2005) to achieve a real-time implementation of a teleoperated control using static hand gestures can be extended to a highly adaptable and robust recognition system for users with mobility impairments. The main components of such as system are described in the following sections and their implementation is left for future work.

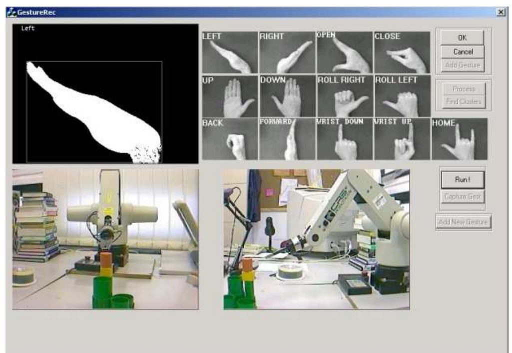

Figure 1. User interface for robot control

Mobility Capabilities - Hand Gesture and Face Movements:

The framework described here will allow the user to control a robot using a wide variety of user customized gestures. For instance, an able-bodied user would want to control the robot while moving his hand in the direction of the robot's intended movement. Moving his hand to the right causes the robotic arm to rotate to the right; moving the hand forward causes the robotic arm to move forward, and so on. While this mapping is perfectly natural to most users, some individuals with severe mobility impairments cannot move their hands in a straight line. Thus the logic used must be able to detect and recognize non-standard hand movements.

Hand and Facial Gesture Recognition

A software application is developed to enable the control of a robotic arm by hand and facial movements. The main hardware components (presented in Figure 2) are: an electric power wheelchair (EPW), a netbook running the recognition system, a Ladybug2 © spherical digital video camera system, and a 6-axis robotic arm. This video camera system has six digital cameras arranged in such a way that it can collect video from more than 75% of its full perimeter. The first and second cameras are oriented towards the hand and face, respectively. The netbook processes the images, recognizing the actions for wheelchair control, and provides visual feedback. The LCD display shows two windows: one for the hand gesture recognition feedback and the other for the facial movement and gesture recognition feedback (see Figure 3).

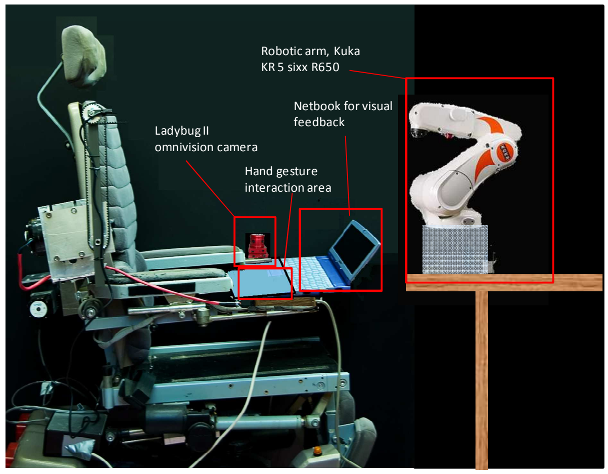

FIGURE 2. Prototype of the interface on the EPW

The data flow of the system is as follows: the user performs gestures with her hand. Sensors on the spherical digital camera system capture face and hand views. The images from the six sensors are ‘stitched’ together in one large image so the hand and face appear in a common system of coordinates.

The images from the sensors are sent to the netbook, where they are processed. The software searches the images for known movements, hand shapes, or facial expressions. If one of these patterns is recognized in the image, the type of movement is translated into a navigational command which was predetermined in an earlier stage of the system’s operation. A netbook display shows a feedback window. The window shows the area where the hand gesture was detected. On the left corner of the window, a caption with the name of the command associated with the recognized gesture is presented. At the same time, the action given by the command is carried out by a robotic arm for a parts assembly task. This type of task was chosen since the feasibility of this approach can be evaluated easily using the measures expressed in (1).

FIGURE 3. Two-view screen: left side: the hand; right side: the face.

When the system is used for the first time, a calibration process sets up the hardware and allows the student to interactively teach the gestures that will be used for robotic control. The calibration routine takes only a few seconds, and can be evoked again at any point during the robotic control. In calibration mode the user determines the neutral area and the interaction area. The interaction area is determined by the distance from the hand to the camera, the sensor's focal length, and the extent to which the hand can be moved. The focal length is fixed, while the distance to the camera can be modified by moving the camera further or closer to the user–this process takes place before the calibration.

The neutral area is determined automatically by detecting the user's hand position. It is a rectangular area around the hand's position with a size equal to the product of the minimum bounding box around the hand by a constant factor (to compensate for non precise hand movements, such as tremors) (see Figure 4). This process takes a few seconds.

The user teaches the gestures to the system by showing the same gesture multiple times when prompted. A vocabulary of 12 commands is designed for robot world coordinates control. The 'forward' and 'back' commands control the X-axis, the 'right' and 'left' commands control the Y-axis, and the 'up' and 'down' commands control the Z-axis of the robotic arm. The 'roll right' and 'roll left' commands rotate the wrist joint, and the 'open grip' and 'close grip' commands control the robot gripper. The 'stop' command stops any action the robot performs. The 'home' command resets all robot joints in the home position. Each of the robotic navigation commands are displayed one after the other with a delay of 30 seconds. When the command is displayed, the user must move her hand in any trajectory, leaving the neutral area, and then bring the hand back to the neutral area. The trajectory, velocity and shape are registered by the system and stored in a database for further use.

Each of the navigational commands is presented five times, and the user is prompted to show a gesture each time. Once the system has been calibrated it is ready be used in operation mode.

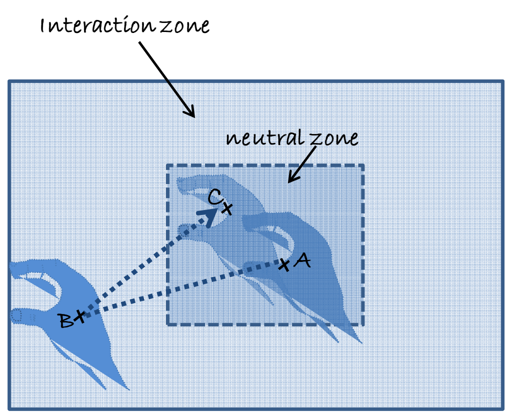

FIGURE 4. Performing a gesture

In operation mode, a graphical user interface (GUI) is presented on the netbook's LCD screen. The image displayed on the GUI is the camera system's view of the scene. On the image, two rectangles with different colors are plotted representing the interaction and the neutral area. Initially, the user's hand is placed inside the neutral area, so no action is carried out by the robot. If the user wants the robot to move, he moves his hand according to the type of action desired. Then the system tries to detect the hand and identify the gesture. If the gesture is recognized, the system displays a caption with the name of the recognized action printed on the screen, otherwise the 'try again' message is displayed.

The neutral area is dynamically updated according to the hand's preferred resting place. This is important because subjects with mobility impairments may find it difficult to move their hands back to the origin of the gesture after performing a movement. Suppose that the user performs a gesture which involves moving her hand from point A to point B. The user could then move her hand back to the origin of the gesture, point A, or move her hand to a point C (see Figure 4). In either case, the hand's final destination (point A or C) will be considered as the resting place (neutral area) for the hand until the next gesture is evoked. This requires only that points A and C be close to each other.

Continuous control is used to resemble joystick operation. With a joystick, the robotic arm continues moving in the indicated direction until the user tilts the handle back to the origin point or releases the handle. Analogously, in our system, as long as the user's hand is outside the neutral area, robot movement occurs. As in the previous example, if a gesture starts at point A (inside the neutral area) and ends at point B (outside the neutral area), then the action requested is continuously carried out until the subject returns to point A or any point inside the neutral area. Continuous control was selected because navigational actions are continuous; discrete operation is more appropriate when the actions required are discrete events in time, such as 'stop' or 'turn-on-engine' on a car.

When a person with mobility impairments is able to make different hand shapes (hand poses), this information can be used to discriminate between non-intended movements and navigational gestures. For example, the hand of a user holding an imaginary or virtual joystick will probably show the 'fist' shape. The user moves the fist left, right, upwards or downwards. Each of these movements can be translated to an action, as with standard joystick operation. The pointing pose could also be used. A subject pointing his index finger forward and tilting his hand in one of the four directions could indicate his desire to move the robot tip in that direction. Four different hand shapes are planned to be included that will be automatically recognized: fist, pointing gesture, palm up and palm down.

Facial gestures can be used as modifiers of the hand gestures. A 'modifier' alters how an action is interpreted by the recognition system. Certain facial expressions indicate that the preceding hand gestures will be interpreted as navigational actions, or that they should not be used (unintentional gestures). Facial expressions will change the operation mode from 'active' to 'sleep' or the other way around. For this switch only one facial expression is required, the open mouth.

EXPERIMENTS

A combination of qualitative and quantitative assessments and usability experiments will be used to measure the effectiveness of the interface with students with physical impairments. The repeated measures design approach, in which participants will serve as their own controls in rating the usability of the vision based interface, will be adopted. In the control condition, the subjects will perform an assembly task (Towers of Hanoi) with the robotic arm using a conventional isometric joystick to control the robot. The task will consist of moving the disks from one of the rods onto another rod, on top of other disks that may already be present on that rod. This task requires several manipulation tasks such as sliding, grasp, move, and release. These operations are very common in automation labs and thus this exercise provides a good case scenario for this system as a pedagogic tool. In the experimental condition, the students complete the assembly task using the interface. To control for extraneous variables such as practice effects across the two scenarios, the subjects will repeat the assembly task three times for each condition (standard joystick, hand gestures). The order of the control (conventional joystick) and experimental conditions will be counterbalanced for participants.

Performance measures: Six measures will be used to assess the user performance for the control and experimental conditions: four quantitative and two qualitative measures.

Quantitative measures:

1) Usability: User ratings of functionality, ease of use and additional human-centered measures for the vision interface and the conventional joystick will be collected using the Likert 5 point scale (1 = very hard, 5 = very easy). The subjects will rate several features for each of the two control cases, reflecting the level of suitability to the user. The questions will assess easiness of use and learn of the interface, memorability, comfort, intuitiveness, safety and precision.

2) Task completion time: The time required to complete the task from beginning to end will be recorded for each task.

3) Number of mistakes: The number of times that the robot gripper collided with an obstacle, released the object too early or imprecisely, or did not grasp the object at all.

4) Recognition accuracy: The number of gestures that were recognized correctly during the assembly task. For this ground truth annotations of the images used by the system are necessary.

Paired sample T-tests will be used to examine whether there is a significant difference in the user performance between the experimental condition and the conventional control (standard joystick).

Qualitative measures: After students complete the assembly task using the different control strategies, they will be interviewed individually to obtain feedback about their experiences using vision based interface compared to a conventional joystick. The interview questions will focus mostly on determining which features of the hand gesture interface were the most important and to identify which additional features would be particularly important. These responses will be tabulated and analyzed qualitatively to assess the perceived benefits and challenges of using each interface from the users' perspective. This information will be used to improve future versions.

CONCLUSIONS

The research presented in this paper is an attempt to offer a new methodology to assess effectiveness in interface design for human-machine interface and to implement a hand gesture based interface for robotic control. The main target population to use the propose system are students with physically and mobility impairments. The motivation for this choice is to develop an infrastructure to allow students with motor impairments to independently operate a robot, a fundamental piece of laboratory equipment in secondary and postsecondary automation classes.
Two fundamental problems are analyzed in this paper: – how to define effective human robot interaction and how to measure it? We offer an analytic framework based on maximization of multi-objectives to enhance the interface design. Finally, we suggest a procedure including quantitative and qualitative measures for the assessment and evaluation of the hand gesture based interface described. The implementation of the system is left for future work.

REFERENCES

Kim, K.H., Kim, H.K., Kim, J.S., Son, W., Lee, S.Y., (2006). A biosignal-based human interface controlling a powerwheelchair for people with motor disabilities. *ETRI Journal*, 28(1):111-114.

Olsen, D. R. Jr., and Goodrich, M. A. (2003) .Metrics for evaluating human-robot interactions. Proceedings *of Performance Metrics for Intelligent Systems (PERMIS)*, 2003.

Salend, S.J. (1998) Using an activities-based approach to teach science to students with disabilities. *Intervention in school and clinic*, 34(2), 67-72.

Wachs, J., Stern, H., Edan, Y. (2005) Cluster Labeling and Parameter Estimation for the Automated Setup of a Hand-Gesture Recognition System. *IEEE Trans. on Systems, Man and Cybernetics. Part A. v*ol. 35, no. 6, Nov. 2005, pp. 932-944.

Warger, C. (1998) Integrating assistive technology into the standard curriculum. ERIC/OSEP Digest E568. ERIC Clearinghouse on Disabilities and Gifted Education, Office of Educational Research and Improvement, Washington, DC.