Automatic Facial Feature Extraction for Predicting Designers' Comfort with Engineering Equipment during Prototype Creation

Shruthi Bezawada
Industrial and Manufacturing Engineering
The Pennsylvania State University
srb321@psu.edu

Qianyu Hu
Industrial and Manufacturing Engineering
The Pennsylvania State University
qzh5042@psu.edu

Allison Gray
Human Development and Family Studies
The Pennsylvania State University
axg5562@psu.edu

Dr. Timothy Brick
Human Development and Family Studies
The Pennsylvania State University
tbrick@psu.edu

Dr. Conrad Tucker[1]
Industrial and Manufacturing Engineering
The Pennsylvania State University
ctucker4@psu.edu

ABSTRACT

Designers frequently utilize engineering equipment to create physical prototypes during the iterative concept generation and prototyping phases of design. Currently, evaluating designers' efficiency during prototype creation is a manual process that either involves observational or survey based approaches. Real-time feedback when using engineering equipment, has the potential to enhance designers' efficiency or mitigate potential injuries that may result from incorrect use of equipment. Towards an automated approach to addressing these challenges, the authors of this work test the hypotheses that i) there exists a *difference in designers' comfort levels before and after they use a piece of engineering prototyping equipment and ii) a machine learning model predicts the level of comfort a designer has while using engineering prototyping equipment with accuracies greater than random chance. It has been shown that the level of comfort that an individual has while completing a task, impacts their performance. The authors investigate whether automatic tracking of designers' facial expressions during prototype creation, predicts their level of comfort. A study, involving 37 participants using various engineering equipment, is used to validate the approach. The Support Vector Machine regression model yielded a range of R squared values from 0.82 to 0.86 for an equipment-specific model. A general model built to predict comfort level across all engineering equipment, yielded an R squared value of 0.68. This work has the potential to transform the manner in which design teams utilize engineering equipment, towards more efficient concept generation and prototype creation processes.*

1 INTRODUCTION

Typically, a designer has an architectural view of how they wish a product or design should evolve [1]. The initial phase of design involves making decisions about the main characteristics of the product. Then, designers develop conceptual solutions by forming functional structures [2], an important step towards a successful product development process [3]. The key to successful prototyping involves the ability of designers to quickly translate their ideas into real world solutions [4]. Prototyping helps design teams explore multiple approaches and ideas, hereby reducing risk and ensuring that all the design requirements are met. Prototyping typically involves iterations that include feedback pertaining to either the prototype itself or the process by which a designer creates the prototype. In order to achieve faster and more efficient prototyping processes, designers must understand the interactions between humans and other elements in a particular work system. Humans must be comfortable operating all elements of the work system, such as power operated machines [5]. These machines help the workforce achieve the required level of productivity and efficiency [6]. Stiff global competition, tighter schedules, and low budgets demand higher accuracy with respect to performance of work [7][8]. Decision makers are therefore motivated to adopt technologies that have the potential to increase productivity and efficiency through the quantification of factors (e.g., affective states) that impact designers' performance.

From designers' perspective, real time feedback pertaining to their performance during design-related tasks has the potential to enable them to learn about their creative strengths/weaknesses. *Comfort* influences performance because designers must be comfortable with the tools and equipment that they use on a regular basis in order for them to efficiently perform their tasks [6][9]. In this work, *comfort* is defined *as a state of an individual wherein they are in control of a task at hand and perceive themselves at being at minimal risk from errors and injuries.* While comfort may not link directly to ability, increased comfort during the design process leads to more openness for learning with a potential to increase efficiency while working with heavy machinery. In a prototyping experiment using mixed reality techniques, Bordegoni discovered that among good prototyping practices, people were found to be comfortable during the prototyping process [10].

As work task performance often determines successful engineering design projects, real time information concerning performance, has great value for real-time, or near real-time decision making [7]. In an attempt to address these issues, semi-automated/automated technologies have been developed to capture designers' internal representations using text [11], speech [12] or body language [13]. For example, individuals' internal representation can be captured and mined through textual data in order to model individuals' responses to external stimuli [14]. However, analyzing designers' internal representations using text may be impractical in a design workshop environment because it may interfere with the required task at hand. Speech analysis may also be a challenging solution, as the noise levels of the engineering equipment may interfere with the audio frequencies projected by designers' voices. Other data modalities may be explored to mitigate the limitations of using audio or textual information to communicate designers' comfort with engineering machines. For example, Behoora and Tucker explored the use of non-wearable sensing systems to capture designers' body language during design team interactions in order to infer designers' affective states [13]. However, body language exhibited during the prototyping phase of design, is dependent on the task, rather than the designers' particular comfort with a given engineering equipment.

To mitigate the aforementioned challenges, the authors of this work propose to model designers' comfort levels with engineering equipment by capturing their facial expressions while engaged in product prototyping. Facial expressions involve different cognitive processes. Emotional expressions elicit rapid responses which an individual might not be aware of. They are not just reflexive, but also have a communicative component [15].

Facial data is also explored for other reasons such as:

- Unlike other body parts such as hands and feet, a designers' face is typically unobstructed during the prototype creation phase
- Facial expressions exhibited by a designer can be mapped to internal affective states
- Facial key points can be automatically captured and mined using machine learning algorithms

The field of Affective Computing aims to reduce the gap between the human and the computer by developing computational systems that recognize and respond to affective states [16]. Systems that recognize affect are appearing in a number of domains, including gaming, mental health, and learning technologies [17]. It has been shown that automatically recognizing and responding to a user's affective states, enhances the quality of the interaction between the user and the computer, thereby making the whole system more effective [18]. In automatic facial expression detection, accurate registration of the face is required which can be achieved using a deformable model approach where 60-70 points on the face is used. The mean shift algorithm used in this paper extracts 66 key points for automatic comfort detection [19].

Advancements in the field of Affective Computing have the potential to enhance certain engineering design activities that benefit from automation and real time performance feedback. In this work, the authors hypothesize that comfort is reliably detected using analysis. The method presented in this paper has the potential to improve the design prototyping process through enhanced efficiency and safety feedback.

2 LITERATURE REVIEW

2.1 Concept Generation and Prototyping in Design

Design concept generation is a critical step of the engineering design process. If this step is not well understood, it leads to undesirable outcomes such as design fixation that refers to designers' inability or reluctance to establish and solve a design need in multiple ways [18]. Also, design concepts affect the quality and efficiency of production, as they lead to different manufacturing processes (i.e., mass production, lean production or manual production) [20]. Methods such as brainstorming, parallel thinking and technology probes are used to create design concepts [21] [22]. A significant correlation exists between the quantity of brainstormed ideas and the quality of design outcomes [23]. After design conceptualization, design prototypes are created (often in an iterative manner) to test the feasibility of the design concepts. A design prototype is an abstraction of the schema that results from similar design cases [24].

A design prototype consists of three factors: function, structure, and behavior [25]. The specific level of a factor depends on the stage of the design [26]. In many cases, equipment such as power saws, drilling machines etc., are used to create prototypes that represent a physical manifestation of design concepts and ideas. In one study, the ability to create visual prototypes enabled designers to perceive failure as an opportunity for learning, which ultimately strengthened their beliefs about their creative ability [27]. It was found that successful design teams use more physical prototypes throughout the design process [28]. Therefore, given the iterative nature of prototyping, it is important that designers feel comfortable while using engineering equipment in order to minimize risks associated with incomplete prototypes or injuries during their creation. Proper equipment performance is difficult to guarantee without the careful focus on human factors and safety [29]. Furthermore, the ergonomics of the machines themselves could be assessed towards more user-friendly designs [30].

However, there is a knowledge gap in terms of how to provide designers with personalized feedback during the concept generation and prototyping phases of design. The method presented in this work aims to mitigate these challenges by exploring the use of automated facial feedback capture systems that model and predict designers' comfort with engineering equipment during the physical prototype process. This work has the potential to enhance designers' performance and efficiency.

2.2 Modeling Designers' Performance on Tasks

Engineering systems design require design teams to use design as well as decision-making skills [31]. In order to acquire these skills, novice designers require feedback to improve. The timing and consistency of feedback are crucial to the design learning process [32]. More feedback during the design phase leads to a better prototype and, ultimately, a better product. For example, students who received systematic feedback and encouragement concerning their sketching abilities were more likely to sketch, and show improvement in sketching skills over the course of a semester [33]. Similarly, it has been shown that self- reflection on the quality of team work during the design process, leads to a more positive rating of team satisfaction [34]. Feedback and prompts for self-reflection ensures that individuals consistently reflect on their own work. Continual reflection leads to better team work among designers and increased confidence within individual designers [35].

The four main criteria for a good design practice are 1) control over the process, 2) clear and available information, 3) feedback, and 4) support from management [36]. Balancing supervision with autonomy can be difficult for management with limited time resources, especially during the prototype development stage [11]. Automating this process could mean that designers receive appropriate amounts of feedback while also maintaining a sense of control. Computers are able to offer feedback and encouragement during the design process [37]. In addition, computers can be adapted to the specific needs of the designer. As the timing of feedback becomes more adaptive to the situation, the entire design process has the potential to become more efficient. Evaluating efficiency of the entire prototyping process is necessary when evaluating designers. Performance can be evaluated using computational methods, observational methods, or a combination of the two. For example, Latent Semantic Analysis (LSA) is a computational method used to evaluate design team performance [34]. Other observational methods capture the psychological experience of the designer because emotional states relate to productivity of the designer’s process [38]. However, there has been a push in design research to automatically collect and analyze data [39]. Facial expressions allow for a noninvasive, yet computationally robust assessment approach to providing designers with feedback during the prototyping phase of design. In addition, researchers are able to collect more objective data as more technology is integrated into the designers’ processes.

2.3 Affective State Recognition and Classification

Automatic affect analysis has attracted much interest from researchers in various research fields. The most current approaches to computer-based analysis of human affective states are limited to either face images or the speech signals [40]. Ekman conducted various experiments on human judgment of deliberately posed facial expressions and concluded that there were six affective states that could be recognized universally: anger, happiness, disgust, sadness, and surprise [39]. Previous work on affective state classification and affective state detection focused on these six expressions. A complete review of recent affective state recognition systems based on facial expressions is provided by Pantic et al. [40]. Instead of proposing a new affective state class, the authors of this work propose to use facial key points to predict designers' comfort level with engineering machinery.

In terms of using affective states to predict real world outcomes, Zeng and colleagues explored ways for a computer to recognize users' affective states (e.g. interest, boredom, frustration and confusion) and to apply the corresponding feedback strategy [41]. Khan et al. presented a concept to identify and integrate learning styles and affective states of a learner into web-based learning management systems. This system provides learners with adaptive courses and additional guidance that is tailored to their learning styles and affective states [42].

Previous work mainly consists of classification of the six universal emotions such as happiness, anger, sadness, fear, disgust and surprise using traditional classification models. The method presented in this paper focuses on predicting the level of comfort, as opposed to the classification of the six basic emotions and paves the way for automated and real time personal feedback during design prototype creation.

3 METHOD

The authors of this work propose a machine learning model that predicts the level of comfort a designer has while using a given engineering equipment. However, before a machine learning model is developed that predicts the level of comfort a designer has while using a piece of engineering equipment, it must be quantitatively shown that comfort levels do indeed change while using a piece of engineering equipment. Therefore, the authors first test the hypothesis that designers' comfort levels change while using a given piece of engineering equipment. If comfort levels do not change during actual prototyping, then only the initial comfort level could be used to infer designers' performance when using a given piece of engineering equipment. On the other hand, if it can be quantitatively shown that comfort levels during prototyping vary, then this demonstrates the need for real time feedback to measure these changes. The authors investigate if a set of facial movements of a designer, will be able to predict a designer's level of comfort with a given piece of engineering equipment. In this work, facial key points are captured automatically by video recording systems and mined using machine learning algorithms. The detailed steps of the method are presented in Figure 1. The method involves four major steps: Data Acquisition, Facial Key Point Extraction, Data Normalization and Model Building and Evaluation, with the design context presented below for further clarity.

The Design Context

Given a design objective (e.g., redesign a shopping cart in just five days), designers will embark on design conceptualization and iterative physical prototyping, often under tight time constraints. There are a wide range of engineering tools that can be used to create physical prototypes, ranging from power drills to electric band saws. For example, Figure 2 represents a designer from the company, IDEO, creating a physical shopping cart prototype that will be compared against other prototypes being simultaneously built by other design team members [43]. In this design context, the designer is focused on ensuring that the ideas presented during the design conceptualization phase are communicated accurately by the physical prototype. Furthermore, given the stringent time constraint that designers typically face, it is imperative that prototypes are created in a timely and efficient manner, while ensuring safety. Therefore, the proposed method has the potential to capture designers' facial expressions (even if they are wearing safety glasses) that are exhibited during the prototype creation phase, in order to quantify their comfort level with different pieces of engineering equipment. Such capabilities would enable design teams to discover the engineering equipment that are better suited for creating a specific prototype or what design team member(s) is better suited to use a specific piece of engineering equipment.

3.1 Data Acquisition of Designers' Facial Expressions

The objective of the data acquisition step is to capture the facial expressions of a designer while using a given piece of engineering equipment. To accomplish this, it is assumed that video data pertaining to a designer's facial expressions, can be acquired for each piece of engineering equipment in use. Formally stated, given designer d, $d \in D$ and a piece of engineering equipment *m*, $m \in M$, the objective of the data acquisition step is to capture the facial expressions **E**, expressed by designer *d*, when utilizing a piece of engineering equipment *m*. To accomplish this, it is assumed that video data pertaining to a designer's facial expressions **E**, can be acquired for each engineering equipment *m*. Standard resolution (640x480 pixels) video recording equipment are assumed to be the minimum requirements to capture facial expressions. For each designer, the video data captured, represents a given designer performing a certain task using a given piece of engineering equipment. In this study, engineering equipment refers specifically to prototyping equipment that is used to generate a prototype of a design concept. Power saw, drill machines and scissors are typical types of prototyping equipment. Ground truth data is needed to train the machine learning model for future unseen instances of designers using machines. Questionnaires, which capture initial affective states and the affective states after the completion of the required task, establish ground truth data. From the questionnaire data collected, the Wilcoxon Signed Rank T test can then be performed to test whether there is a statistically significant difference in the level of comfort before and after performing each task at a given piece of engineering equipment. The data acquisition protocol can be applied in a prototyping workshop when the designers are at work using the different engineering equipment. It is a non-invasive form of data capture, enabling designers to perform their tasks with minimal disruption.

3.2 Facial Key Point Extraction

Each video clip *v* acquired from a designer, is comprised of a set of mutually exclusive frames, $f \in F$, where frame f_1, represents the first frame acquired by the video recording system and frame f_n represents the last frame acquired by the video recording system. These facial key points are reported in the two-dimensional space of the image, and hence are in the form of (x, y). The facial key points are shown in Figure 3.

For each frame f in Figure 3, facial key points or facial features are extracted using the regularized mean-shift algorithm [19]. Deformable model fitting refers to identifying a parametrized shape model to a frame such that its landmarks correspond to consistent locations on the object of interest. For a detailed explanation about the regularized mean shift algorithm, please refer to [19]. In this work, the mean and standard deviation of each key point in the frames of the video sequences are modelled as a single feature for simplicity.

3.3 Data Normalization

The location, size, and orientation of the facial image in the frame are likely to be characteristics of a designer’s pose and location relative to the camera, and not aspects of the designer’s affective state. Therefore, Ordinary Procrustes Analysis is performed on the facial key points obtained from each frame in order to standardize the facial location and orientation while retaining the unique facial expression information from each frame. Ordinary Procrustes analysis matches all the faces in the frames as closely as possible to a specified reference frame by uniform rotation, translation and scaling, as shown in Figure 4. The reference frame is chosen such that the face of the individual is approximately in the center of the frame and is looking directly into the camera.

3.4 Model Building and Evaluation

In order to validate the predictive ability of face-tracking to assess the comfort level of designers, nonlinear regression was conducted using a Support Vector Machine (SVM) algorithm. SVMs were chosen because they have been shown to perform well with continuous variables in high dimensional spaces [44]. In this case, the predictors for the model are the means and the standard deviations of the normalized key points across all the frames in a particular video clip. The mean and standard deviation of the rotation parameter of the Procrustes Analysis are also used as predictors along with the means and standard deviations of the facial key points. The dependent variable for the model is the level of comfort, which is rated by the designers on a scale of 1- 10. It is important to note that this rating is only performed once by each designer to establish ground truth data for the machine learning algorithm.

For the SVM model, hyper parameter optimization chooses the best parameters for the model by training multiple models. The two parameters to be estimated are Epsilon (ε) and Cost(C). Epsilon (ε) controls the width of the ε-insensitive zone used for fitting the training data. The Cost(C) determines the tradeoff between the model complexity and the degree to which deviations larger than ε are tolerated in the optimization formulation with different epsilon (ε) and cost (C) values. The mathematical formulation of the SVM is presented below.

$$\textbf{Minimize } \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{N}(\xi_i + \xi_i^*) \tag{1}$$

$$\textbf{subject to } \quad y_i - \boldsymbol{w}g(x_i) - b \le \varepsilon + \xi_i \tag{2}$$

$$\boldsymbol{w}g(x_i) + b - y_i \le \varepsilon + \xi_i^* \tag{3}$$

$$\xi_i, \xi_i^* \ \ge 0 \tag{4}$$

where,

x_i	Training sample of facial key points acquired from designers, $i = 1, \dots, N$
y_i	The dependent variable *comfort* that is based on a scale of 1-10, where 1 represents "least comfortable" and 10 represents "most comfortable", $i = 1, \dots, N$
$\boldsymbol{w}$	Normal Vector to the Hyperplane
ε	Tolerable Error
C	Cost of error
ξ_i, ξ_i^*	Deviation outside epsilon intensive band, i=1,…,N
$g(x\quad)$	Radial Basis Kernel Transformation, $i = 1, \dots, N, j = 1, \dots, N$ $g(x_i) \cdot g(x_j) = \exp(-\gamma \lVert x_i - x_j \rVert^2), \gamma > 0$
b	Bias Term

SVM regression uses an epsilon intensive loss function to allow deviation from the true value within distance and at the same time, reach global minimum. Specifically, the points within the epsilon intensive band have no cost of errors and the cost of error outside the band are measured by parameter C. Slack variables ξ_i, ξ_i^* are introduced to measure the deviation outside the epsilon intensive band. Figure 5 is a conceptual representation of a one dimensional nonlinear SVM regression model with an epsilon intensive band. The epsilon intensive band boundaries are determined by the parameter ε.

The SVM regression model is employed to determine the relationship between the facial key points acquired from a designer (*d*), and their comfort levels with different pieces of engineering equipment (*m*). The following section presents a case study that demonstrates the feasibility of the proposed method in a real world engineering design workspace.

4. APPLICATION

4.1 Experimental Setup

The data for the study was collected in an engineering workshop at the Pennsylvania State University. For this study, three engineering equipment stations were set up. Station 1 contained a power saw, Station 2 contained a drilling machine, and Station 3 contained a pair of scissors. Video recording equipment was set up at each engineering equipment station such that the face of the participant performing the task could be seen clearly (Figure 6). Cardboard pieces of uniform size (21cm x 9 cm) were used for designing two tasks at each engineering equipment station, one simple task and one more complicated.

The first task at Station 1 (power saw) and Station 3 (scissors) consisted of cutting along a straight line drawn along the middle of the piece of cardboard. The second task consisted of cutting a figure ‘8’ in a piece of cardboard. At Station 2 (drill), the first task consisted of drilling a hole in the center of the piece of cardboard, and the second task consisted of drilling two holes 1 cm apart along a line parallel to the edge of the cardboard. The participants had to ensure that the two holes did not join together while drilling. Approximate task instructions are shown in Figures 7 and 8.

4.2 Questionnaire Design

An initial questionnaire asked participants to rate their current affective state based on a series of emotional words. Participants rated the degree to which they were currently experiencing the emotion on a scale of 0-10. Participants also completed the Personality Minimarker, which evaluates personality according to the Big Five traits [45] . Participants were also asked about their knowledge and comfort of workshop machinery and laboratory tasks. Lastly, participants provided their gender, age, and race (Table 1). Once participants provided signed consent, they completed the initial questionnaire. Upon completing the initial questionnaire, an experimenter assigned each participant to a randomized order of machines.

Participants completed one task specific questionnaire after each task. The task specific questionnaire consisted of the same emotional items as the initial questionnaire. In addition, questions asked participants about perceived danger of the specific machine. Lastly, the task specific questionnaire asked participants to evaluate their performance on the task and report their level of focus. Each participant provided six task-specific questionnaires: two (one for each task) per station for three stations.

4.3 Participant Recruiting

The participants in the study were all freshman engineering students of 18-19 years of age and enrolled in EDGSN 100 *Introduction to Engineering Design*. All participants provided informed consent. Participation in the study was voluntary. A total number of 40 students participated in the study.

4.4 Data Acquisition

The participants entered the machine shop in groups of three, with a different participant assigned to each workstation. There was no interaction or communication between participants. The participants were recorded while performing the two tasks at each station. After completion of both tasks, participants were rotated to the next machine so that each participant completed all six tasks. Videos were edited such that the final clips used for analysis consisted of the participants performing the two tasks (Figure 9 shows an example of the video recording).

4.5 Facial Key Point Extraction

A total of 60 clips were used for the study. The clips were chosen such that the face was not hidden by the equipment. The clips consist of a diverse mix of participants performing the two tasks at each of the three engineering equipment stations. Some of the clips were omitted, as the participants in most of the frames were blocked by the equipment or were out of range of the video recording equipment. Therefore, while the data collection started with 40 participants, only 37 were used in the hypothesis test and model generation, as explained in the results section. The number of frames in each clip varied as the time taken for each task by the participant differed. 66 facial key points were extracted using the CSIRO Face Analysis SDK and the Face Modeling [19]. Each feature had an *x* and *y* coordinate, resulting in 132 features per frame (see Figure 10). Ordinary Procrustes analysis was performed to align all faces to a canonical orientation and scale centered at the origin and scaled to unit variance, as explained Section 3.3.

4.6 Model Building and Evaluation

SVM Regression was used to predict self-reported level of comfort from mean and standard deviations of the 66 facial key points and Procrustes alignment parameters using the R package 'e1071' [46]. As the facial keypoints had an *x* and *y* component, the total number of facial features were 132. The means and standard deviations of all 132 features, along with the procrustes parameters were used. Grid search determined that $C = 1$ and $\varepsilon = 0.01$ were the optimal parameters to minimize the cost function of the SVM model (see equations 1-4 in section 3.4). The parameters were chosen so that they yielded the maximum model accuracy. Box-Cox transformations [47] were performed to equalize the variance for both the positively and negatively skewed predictors. The model performance was tested using a leave-one-out cross-validation approach. In this approach, one video is left out of the training process and used for testing. This process is repeated until all the videos have been left out exactly once.

5. RESULTS AND DISCUSSIONS

5.1 Hypothesis Test Exploring Whether Designers' Comfort Levels Change While Using a Given Piece of Engineering Equipment.

As stated at the start of the method section, before the value of the SVM model can be evaluated, it must first be shown that designers' level of comfort does indeed vary, as they perform a given task on a specific piece of engineering equipment. The questionnaire data collected is used to compute the summary statistics of the level of comfort of the participants. Formally stated, the hypothesis to be tested is as follows (at α=0.05):

$H_0: \mu_A = \mu_B$

H_a: $\mu_A \neq \mu_B$

Where, H_0 is the null hypothesis which indicates there is no significant difference between the mean initial level of comfort at an engineering equipment station (μ_A) and that after completion of the tasks (μ_B).

H_a is the alternate hypothesis which indicates there is a significant difference between the mean initial level of comfort at an engineering equipment station (μ_A) and that after completion of the tasks (μ_B).

The results of the hypotheses test indicate that the comfort level before and after each task at the three engineering equipment stations varied across the different machines. Overall, the participants were mostly comfortable with the equipment in the machine shop with the level of comfort at all engineering equipment stations (sampled mean ($\bar{x}$) = 6.8). This may be attributed to the fact that the participants had been working with the same equipment for a semester for their course EDGSN 100. Participants were more comfortable with the power driven equipment, the power saw $\bar{x}_{Task\ 1}$ = 6.9, $\bar{x}_{Task2}$ = 8.02) and the drilling machine ($\bar{x}_{Task\ 1}$ = 8.05, $\bar{x}_{Task2}$ = 8.2), compared to the scissors ($\bar{x}_{Task\ 1}$ = 5.05, $\bar{x}_{Task2}$ = 4.88). Surprisingly, participants were least comfortable with scissors ($\bar{x}_{Task1}$ = 5.05, $\bar{x}_{Task2}$ = 4.88). Participants reported highest comfort at the drill station ($\bar{x}_{Task1}$ = 8.05, $\bar{x}_{Task2}$ = 8.2). Participants were slightly more comfortable with the second task ($\bar{x}_{Task2}$= 8.02) at the power saw station than the first task ($\bar{x}_{Task1}$ = 6.9) despite the increase in complexity of task two. The comfort increase may stem from the knowledge about the task processes and machine operations gained from the previous task. However, there is a comfort decrease at the scissors station between two tasks. According to the video, participants exerted considerably more effort and strength to cut the cardboard using scissors. Furthermore, because the figure “8” is more intricate than a straight line, more effort and strength is needed in task two. The increase of comfort by knowledge gained from the previous task may be offset by the heavy demands of strength and effort at the scissors station. Figure 11 presents a visual summary of the comfort levels at different engineering workstations.

A series of Wilcoxon Signed Rank tests were conducted to determine whether there were statistically significant differences in participants’ comfort level before and after completing a particular task at a given piece of engineering equipment. The results indicate that there was no significant difference between the initial levels of comfort and that after the completion of task one at the power saw station ($p = 0.054$). However, there was a significant difference between the initial level of comfort and that after the completion of task two at the power saw station ($p = 0.04$). According to Table 2 and Table 3, comfort level significantly increased in task two at the power saw station. In terms of all other four tasks, under the significant level of 5%, there are significant differences between the initial levels of comfort and that after the completion of these four tasks respectively (i.e., two tasks at the drill station and two tasks at the scissors stations). There is a significant comfort increase at the drill station for both tasks, while there is significant comfort decrease at the scissors station. Table 3 summarizes the results of the t- tests for both tasks completed at each of the three engineering equipment stations. As a reminder, while the researchers originally captured data of 40 participants, only data pertaining to 37 participants was used in the hypothesis test due to the fact that the facial data of some participants, was hidden from the view of the video recording system, and hence, represented noise in the data set.

As discussed in the literature review section, the comfort level of a designer when using a piece of engineering equipment, has a direct impact on their performance and safety. Furthermore, as showed in Tables 2 and 3, comfort levels vary when using different machines and performing different tasks (i.e., five experiments out of six have significant comfort level difference). Therefore, it is of great importance to measure the real-time comfort level while designers' are at work. Based on real-time comfort level feedback, designers may adjust the process of prototyping, hereby improving design performance and workshop safety. Low or decreasing level of comfort may act as an alarm or stop signal for assistance or timely feedback. On the other hand, positive feedback may act as an encouragement or safety signal while comfort level is increasing or high. The following section provides evidence of the ability of machine learning algorithms to predict comfort levels, based on participants' facial expressions while performing tasks using engineering equipment.

5.2 Evaluation of the SVM Regression Model in Predicting Designers' Comfort Levels

5.2.1 Evaluation of the Equipment-Specific SVM Regression Model in Predicting Designers' Level of Comfort

Numeric results in the previous section demonstrated the existence of statistically significant variation of comfort levels before and after the tasks at different engineering equipment stations. Five null hypotheses out of the six stated in the previous section were rejected. Given the variation of comfort level, without feedback, designers will not be able to gain knowledge of the comfort level during the design prototyping process. This may decrease the productivity of the prototyping process or even increase the risk of getting injured during the process. A machine-specific SVM regression model was trained for each equipment station to predict participants' comfort level from automatically-extracted facial expression data. The clips were segregated based on the work station the participant was working at. The R squared values for each work station is shown in Table 4 with the R^2 range from 0.82 to 0.86. Different fitness of data may result from individual discrepancy such as frame quality or self-perception. As seen in Table 4, the SVM regression model is quite accurate at predicting what a designer's level of comfort will be when using a given piece of engineering equipment (that has been used before and that has ground truth data associated with it). With such a model, designers working to iteratively create physical prototypes, will be able to determine quantitatively what pieces of engineering equipment are most/least comfortable to use to design creative and efficient prototypes. The comfort level can be used as a quantitative measure to determine the level of comfort that results in increased creativity or safety.

While equipment-specific models are quite accurate at predicting the designers' comfort levels, prior ground truth is needed (i.e., comfort level of the participant and facial key points at the specific equipment station) to generate future predictions of comfort levels while using a specific machine. However, prior data about the designer or the given piece of equipment may not always be available. For example, if a design team attempts to use a new piece of engineering equipment for which ground truth data has yet to be acquired. Therefore, this study further explores the accuracy of a general model that could predict comfort level without prior knowledge of the specific machine being used. Details are demonstrated in the next section.

5.2.2 Evaluation of the General SVM Regression Model in Predicting Designers' Level of Comfort

In the general SVM regression model, the objective is to predict comfort level, regardless of the piece of engineering equipment or designer in the engineering workspace. As can be imagined, this is a more complex objective, given the variations that exist in machine complexity, designer experience, etc. However, in the absence of engineering-equipment-specific ground truth data (as needed in the previous section), the general model could be used as a baseline model, while additional ground truth data is being acquired for the new design prototype scenario. The results in Table 5 show that the SVM regression model has a Coefficient of Determination (R^2 value) of 0.68, which indicates that 68% of the variation in level of comfort is explained by the facial key points. Moreover, the levels of comfort were rated by the participants themselves on a continuous scale. This indicates that the level of comfort used as "ground truth" is subjective, resulting in high variance across individuals. Another factor that may have contributed to lower accuracy (compared to section 5.2.1) is missing facial key point data for some engineering tasks. During data collection, the participants' face might have been hidden by the equipment while performing the task for a short period of time or may have moved away from the focus of the video recording equipment. In future setups, the angle of the video camera should be adjusted to each individual so that the equipment does not block the face. In addition, the use of smaller wireless cameras will allow for more flexible and customizable camera angles.

For a new design scenario involving equipment or designers where ground truth data may not be yet available, using the general model during the prototyping process would provide an initial prediction of the comfort level, albeit with lower accuracy, compared to the engineering-specific-equipment model in section 5.2.1. However, in the absence of any feedback system, the general model is still better than random chance. Furthermore, given that comfort is on a continuous scale, designers could discretize the comfort level range (e.g., “low comfort”, “medium comfort”, “high comfort”) so that if the general SVM regression model over or under estimates a level of comfort, it still maps to one of the discretized comfort categories. In terms of feedback, efficiency and productivity of the designers may benefit, based on the ability to provide assistance as and when they need it (i.e., when they are least comfortable with the equipment they are using).

The results indicate that an automatic classifier operating on videos of natural facial expression is able to extract the level of comfort of an individual. The results show that the data collected through non-invasive data capture techniques while designers are at work, can be used to effectively model and predict their comfort level. During design conceptualization and prototyping, designers would not need to be disturbed and could continue with the task at hand, while the proposed system modeled and predicted their comfort level. Design teams could choose if/when intervention feedback would be provided that enhanced designer efficiency, while minimizing distraction and injury.

It is expected that with a larger data set and more advanced modelling algorithms, the accuracies of the proposed method will improve. Future experiments will explore these research questions by including dynamic information in a traditionally static system using procedures such as time-delay embedding, for example, in a Support Vector Machine.

5.2.3 Base Facial Expressions

It is possible that individuals may have very different base facial expressions. For example, one person may naturally have an angry facial expression. However, this does not necessarily mean that they are uncomfortable or angry. Another scenario worth exploring is that the variation of each individual's facial expressions, may cause our prediction model to predict inaccurately, as the model described in this paper does not account for interpersonal differences in facial structure. The accuracy of this model reflects the ability to predict level of comfort without explicit base-expression information about the individual in view. We could improve our prediction if baseline measurements were incorporated into the prediction model. Future research can include explicit control for person-specific neutral facial expressions for better prediction accuracy so that a model is tailored to an individual.

6. CONCLUSION

Developing prototypes is an essential step in the design process. The current design landscape calls for enhanced productivity and efficiency. Therefore, designers must be comfortable with the equipment that they utilize during design conceptualization and prototyping. The questionnaires in this study indicated that comfort levels were higher after the use of a machine ($\bar{x}_{power\ saw}$ = 8.459, *SD* = 1.626; $\bar{x}_{drill}$ = 8.135, *SD* = 2.668) compared to pre task comfort levels ($\bar{x}_{pre\ task}$ = 6.838, *SD* = 2.339). However, comfort levels differed from task to task. To properly track the dynamics of comfort levels, designers must be monitored throughout the prototyping process. Although real time observation and feedback may be costly for enterprise decision makers, it has the potential to minimize safety hazards and increase productivity. Automating this process may enable enterprise decision makers to provide the same benefits as manual observation and feedback at a fraction of the cost. In this work, a predictive model for engineering-specific equipment had quite promising results, with a R^2 range from 0.82 to 0.86. However, such a model may not be applicable to all design scenarios, especially if new engineering equipment or new designers are introduced to a design project, where no ground truth is yet available. To mitigate these challenges, a generalized SVM model was also explored in this work. The general model showed the comfort level in individuals can be predicted by using facial expressions with a reasonable accuracy (R^2 = 0.68) using standard video equipment. Introducing video monitoring equipment to the design process is a low cost, yet beneficial, adjustment to 1) streamline the delivery of feedback, 2) enforce safety guidelines, and 3) intervene if a designer is distracted while using power equipment.

Though this study lays preliminary groundwork, the findings advocate for building intelligent feedback systems to help enhance productivity and efficiency. The system will be able to detect if an individual is comfortable or distracted while using potentially dangerous equipment. Automatic feedback systems also benefit workforce training, engineering laboratory teaching, and other domains. Intelligent feedback systems can increase efficiency and productivity of learning processes by providing assistance as needed. In addition, in workshop or laboratory settings, more monitoring will ensure that individuals follow safety protocols at all times. Feedback from these systems will build confidence in individuals as they perfect their design and prototyping skills. More self-esteem in engineering and designers has the potential to induce more confidence in projects and influence productivity.

ACKNOWLEDGEMENTS

This research is funded in part of by the National Science Foundation NRI #1527148: *Real Time Observation, Inference and Intervention of Co-Robot Systems towards Individually Customized Performance Feedback Based on Students' Affective States*, and by National Science Foundation (BCS–1030806) and the National Institute on Aging (1R21-AG041035). Any opinions, findings, or conclusions found in this paper are those of the authors and do not necessarily reflect the views of the sponsors.

REFERENCES

[1] M. Ó. Cinnéide and P. Nixon, "A methodology for the automated introduction of design patterns," in *Software Maintenance, 1999.(ICSM'99) Proceedings. IEEE International Conference on*, 1999, pp. 463–472.

[2] A. Römer, M. Pache, G. Weißhahn, U. Lindemann, and W. Hacker, "Effort-saving product representations in design—results of a questionnaire survey," *Design Studies*, vol. 22, no. 6, pp. 473–491, 2001.

[3] B. Camburn, B. Dunlap, T. Gurjar, C. Hamon, M. Green, D. Jensen, R. Crawford, K. Otto, and K. Wood, "A Systematic Method for Design Prototyping," *Journal of Mechanical Design*, vol. 137, no. 8, p. 81102, 2015.

[4] Y.-K. Lim, E. Stolterman, and J. Tenenberg, "The anatomy of prototypes: Prototypes as filters, prototypes as manifestations of design ideas," *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 15, no. 2, p. 7, 2008.

[5] S. Zuboff, *In the age of the smart machine: The future of work and power*. Basic Books, 1988.

[6] J. Dul and W. P. Neumann, "Ergonomics contributions to company strategies," *Applied ergonomics*, vol. 40, no. 4, pp. 745–752, 2009.

[7] J. Teizer, M. Venugopal, and A. Walia, "Ultrawideband for automated real-time three-dimensional location sensing for workforce, equipment, and material positioning and

tracking," *Transportation Research Record: Journal of the Transportation Research Board*, no. 2081, pp. 56–64, 2008.

[8] S. D. Eppinger and D. E. Whitney, "Accelerating product development by the exchange of preliminary product design information," *Journal of Mechanical Design*, vol. 117, p. 491, 1995.

[9] B. W. Field, "Visualization, intuition, and mathematics metrics as predictors of undergraduate engineering design performance," *Journal of Mechanical Design*, vol. 129, no. 7, pp. 735–743, 2007.

[10] M. Bordegoni, U. Cugini, G. Caruso, and S. Polistina, "Mixed prototyping for product assessment: a reference framework," *International Journal on Interactive Design and Manufacturing (IJIDeM)*, vol. 3, no. 3, pp. 177–187, 2009.

[11] A. Dong, A. W. Hill, and A. M. Agogino, "A document analysis method for characterizing design team performance," *Journal of Mechanical Design*, vol. 126, no. 3, pp. 378–385, 2004.

[12] J. Stempfle and P. Badke-Schaub, "Thinking in design teams-an analysis of team communication," *Design studies*, vol. 23, no. 5, pp. 473–496, 2002.

[13] I. Behoora and C. S. Tucker, "Machine learning classification of design team members' body language patterns for real time emotional state detection," *Design Studies*, vol. 39, pp. 100–127, Jul. 2015.

[14] D. Munoz and C. S. Tucker, "Modeling the Semantic Structure of Textually-Derived Learning Content and its Impact on Recipients Response States," *Journal of Mechanical Design*, 2015.

[15] C. Frith, "Role of facial expressions in social interactions," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 364, no. 1535, pp. 3453–3458, Dec. 2009.

[16] A. Kapoor and R. W. Picard, "Multimodal affect recognition in learning environments," in *Proceedings of the 13th annual ACM international conference on Multimedia*, 2005, pp. 677–682.

[17] R. A. Calvo and S. D'Mello, "Affect detection: An interdisciplinary review of models, methods, and their applications," *Affective Computing, IEEE Transactions on*, vol. 1, no. 1, pp. 18–37, 2010.

[18] J. S. Linsey, I. Tseng, K. Fu, J. Cagan, K. L. Wood, and C. Schunn, "A Study of Design Fixation, Its Mitigation and Perception in Engineering Design Faculty," *J. Mech. Des*, vol. 132, no. 4, pp. 041003–041003, Apr. 2010.

[19] J. M. Saragih, S. Lucey, and J. F. Cohn, "Deformable Model Fitting by Regularized Landmark Mean-Shift," *International Journal of Computer Vision*, vol. 91, no. 2, pp. 200–215, 2010.

[20] L. Wang, W. Shen, H. Xie, J. Neelamkavil, and A. Pardasani, "Collaborative conceptual design—state of the art and future trends," *Computer-Aided Design*, vol. 34, no. 13, pp. 981–996, 2002.

[21] M. F. Powers and J. Jones-Walker, "An interdisciplinary collaboration to improve critical thinking among pharmacy students," *American Journal of Pharmaceutical Education*, vol. 69, no. 1–5, p. 516, 2005.

[22] H. Hutchinson, W. Mackay, B. Westerlund, B. B. Bederson, A. Druin, C. Plaisant, M. Beaudouin-Lafon, S. Conversy, H. Evans, and H. Hansen, "Technology probes: inspiring design for and with families," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, 2003, pp. 17–24.

[23] M. C. Yang, "A study of prototypes, design activity, and design outcome," *Design Studies*, vol. 26, no. 6, pp. 649–669, 2005.

[24] J. S. Gero, *Prototypes: a new schema for knowledge-based design*. University of Sydney, Architectural Computing Unit, 1987.

[25] L. Qian and J. S. Gero, “Function–behavior–structure paths and their role in analogy-based design,” *Artificial Intelligence for Engineering, Design, Analysis and Manufacturing*, vol. 10, no. 4, pp. 289–312, 1996.
[26] E. Gerber and M. Carroll, “The psychological experience of prototyping,” *Design studies*, vol. 33, no. 1, pp. 64–84, 2012.
[27] E. Fadier and C. De la Garza, “Safety design: Towards a new philosophy,” *Safety Science*, vol. 44, no. 1, pp. 55–73, 2006.
[28] S. Pavlovic-Veselinovic, “Ergonomics as a missing part of sustainability,” *Work*, vol. 49, no. 3, pp. 395–399, 2014.
[29] J. Jang and C. D. Schunn, “Physical design tools support and hinder innovative engineering design,” *Journal of Mechanical Design*, vol. 134, no. 4, p. 41001, 2012.
[30] J. S. Busby, “The neglect of feedback in engineering design organisations,” *Design Studies*, vol. 19, no. 1, pp. 103–117, 1998.
[31] J. W. Booth, E. A. Taborda, K. Ramani, and T. Reid, “Interventions for teaching sketching skills and reducing inhibition for novice engineering designers,” *Design Studies*, vol. 43, pp. 1–23, 2016.
[32] M. A. Busseri and J. M. Palmer, “Improving teamwork: the effect of self-assessment on construction design teams,” *Design Studies*, vol. 21, no. 3, pp. 223–238, 2000.
[33] R. S. Adams, J. Turns, and C. J. Atman, “Educating effective engineering designers: the role of reflective practice,” *Design Studies*, vol. 24, no. 3, pp. 275–294, May 2003.
[34] J. Austin-Breneman, T. Honda, and M. C. Yang, “A study of student design team behaviors in complex system design,” *Journal of Mechanical Design*, vol. 134, no. 12, p. 124504, 2012.
[35] K. Lauche, “Job design for good design practice,” *Design studies*, vol. 26, no. 2, pp. 191–213, 2005.
[36] M. Bernal, J. R. Haymaker, and C. Eastman, “On the role of computational support for designers in action,” *Design Studies*, vol. 41, pp. 163–182, 2015.
[37] M. Dinar, J. J. Shah, J. Cagan, L. Leifer, J. Linsey, S. M. Smith, and N. V. Hernandez, “Empirical studies of designer thinking: Past, present, and future,” *Journal of Mechanical Design*, vol. 137, no. 2, p. 21101, 2015.
[38] D. McDuff, R. El Kaliouby, T. Senechal, D. Demirdjian, and R. Picard, “Automatic measurement of ad preferences from facial responses gathered over the internet,” *Image and Vision Computing*, vol. 32, no. 10, pp. 630–640, 2014.
[39] P. Ekman, “An argument for basic emotions,” *Cognition & emotion*, vol. 6, no. 3–4, pp. 169–200, 1992.
[40] M. Pantic and L. J. Rothkrantz, “Automatic analysis of facial expressions: The state of the art,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, no. 12, pp. 1424–1445, 2000.
[41] Z. Zeng, J. Tu, M. Liu, T. Zhang, N. Rizzolo, Z. Zhang, T. S. Huang, D. Roth, and S. Levinson, “Bimodal HCI-related affect recognition,” in *Proceedings of the 6th international conference on Multimodal interfaces*, 2004, pp. 137–143.
[42] F. A. Khan, E. R. Weippl, and A. M. Tjoa, “Integrated approach for the detection of learning styles and affective states,” in *World conference on educational multimedia, hypermedia and telecommunications*, 2009, vol. 2009, pp. 753–761.
[43] Alfonso Neri, *ABC Nightline - IDEO Shopping Cart*. 2009.
[44] A. E. Dimitriadou, K. Hornik, F. Leisch, D. Meyer, A. Weingessel, and M. F. L. Friedrichleischcituwienacat, “The e1071 Package,” 2006.
[45] J. Block, “Going beyond the five factors given: Rejoinder to Costa and McCrae (1995) and Goldberg and Saucier (1995),” *Psychological Bulletin*, vol. 117, no. 2, pp. 226–229, 1995.

[46] D. Meyer, E. Dimitriadou, K. Hornik, A. Weingessel, F. Leisch, C.-C. C. (libsvm C++-code), and C.-C. L. (libsvm C++-code), *e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien*. 2015.
[47] R. M. Sakia, "The Box-Cox transformation technique: a review," *The statistician*, pp. 169–178, 1992.

LIST OF TABLES

LIST OF FIGURES

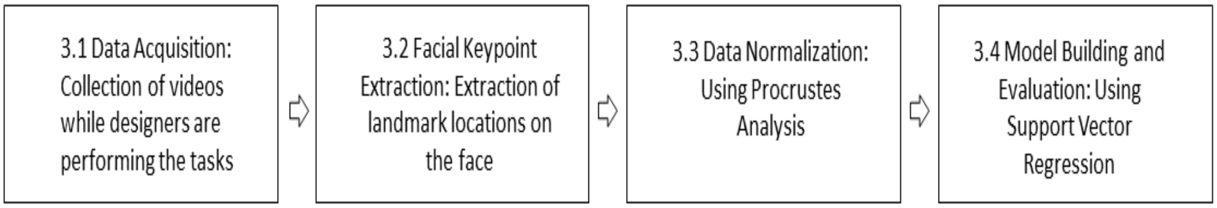


Figure 1: Overview of Method

Figure 2: A Designer at IDEO Creating a Shopping Cart Prototype Using a Power Tool [43]

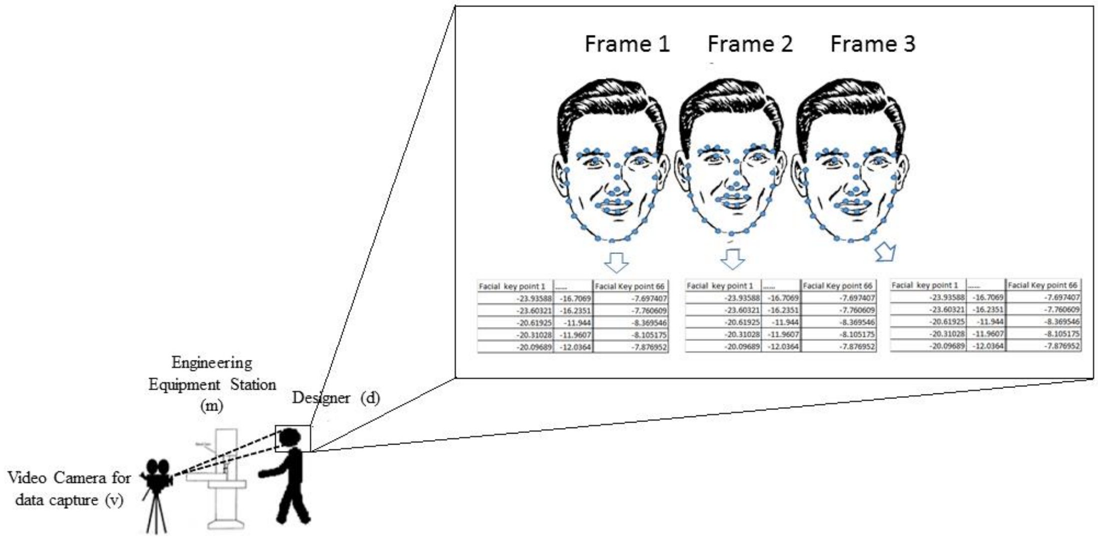


Figure 3: Extraction of Facial Key Points from Video Sequences

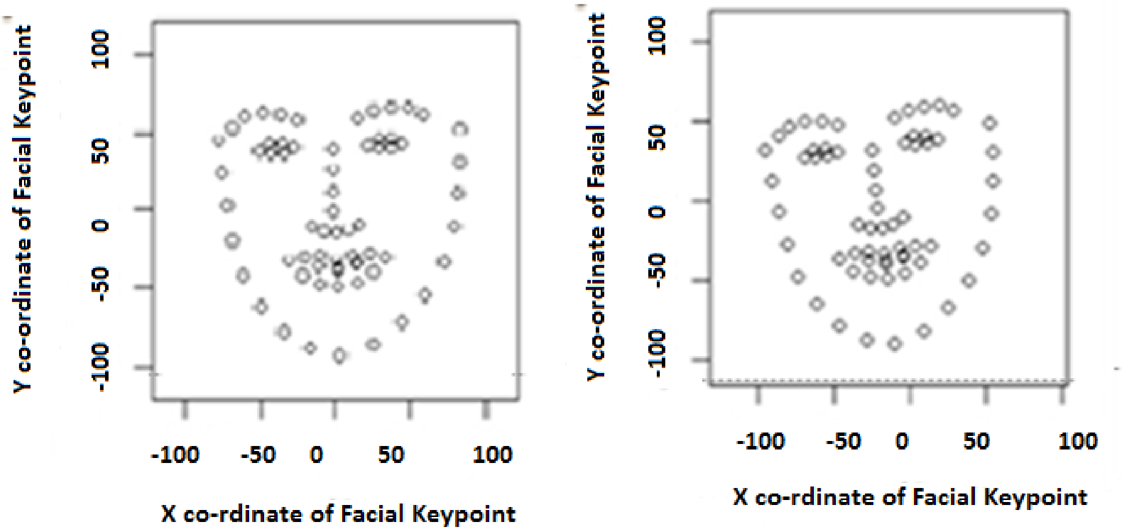


Figure 4: Procrustes Analysis: The left figure is the pre-aligned figure and the right figure is the aligned figure

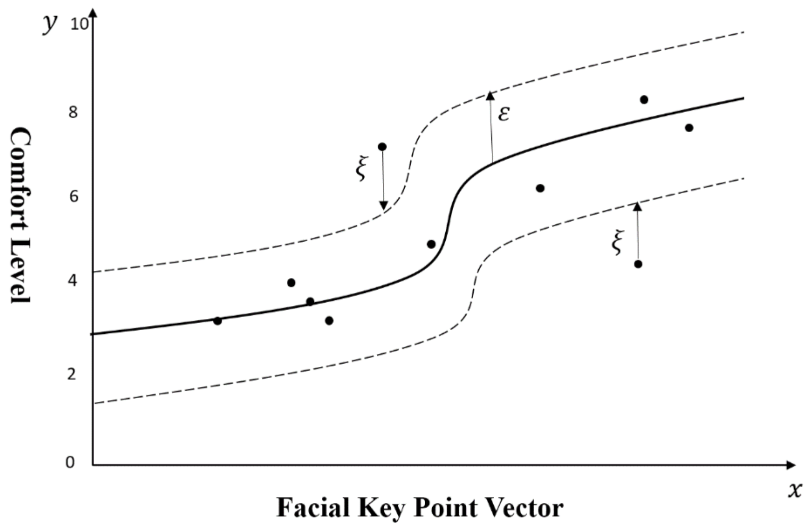


Figure 5: One Dimensional Nonlinear Regression with Epsilon Intensive Band

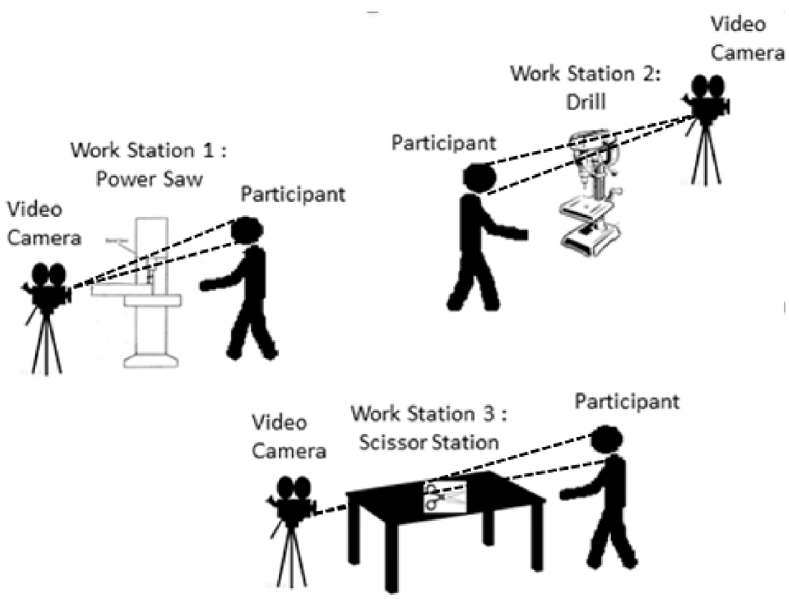


Figure 6: Experimental Layout

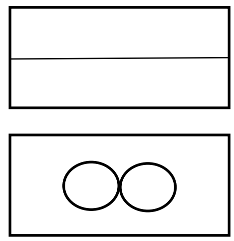


Figure 7: Tasks to be performed with the Power Saw as well as Scissors Stations

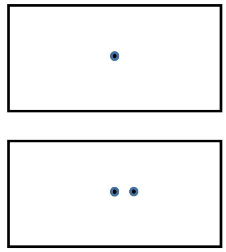
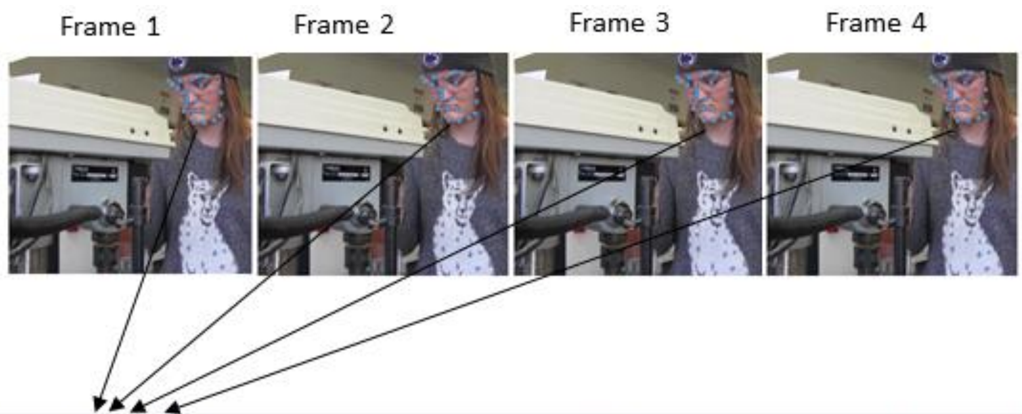


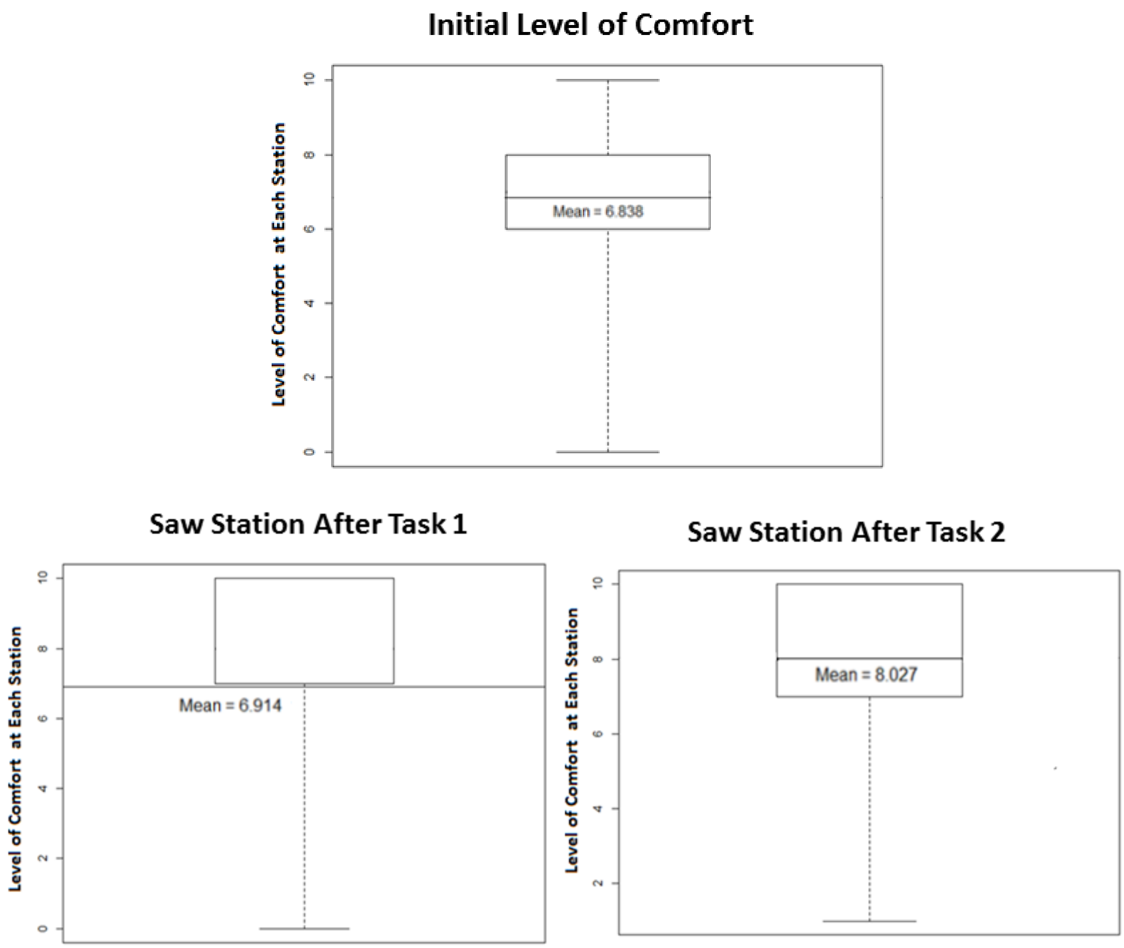
Figure 8: Tasks to be performed at the Drill Station

Figure 9: Snapshot of Video Sequence



Mean of Facial Keypoint 1	Std. Dev of Facial Keypoint 1	....	Std. Dev of Facial Keypoint 132	Target(Level of Comfort)
-23.93588	-16.70689		11.5637	10
-23.60321	-16.23509		0	6
-20.61925	-11.944		9.102761	3
-20.31028	-11.96066		9.03498	7
-20.09689	-12.03642		9.010708	5
-19.99634	-12.22651		9.053423	8

Figure 10: Conversion of Raw Facial Key Point Data into Data ready for Analysis


Initial Level of Comfort
Level of Comfort at Each Station
Mean = 6.838
0
2
4
6
8
10
Saw Station After Task 1
Level of Comfort at Each Station
Mean = 6.914
0
2
4
6
8
10
Saw Station After Task 2
Level of Comfort at Each Station
Mean = 8.027
2
4
6
8
10

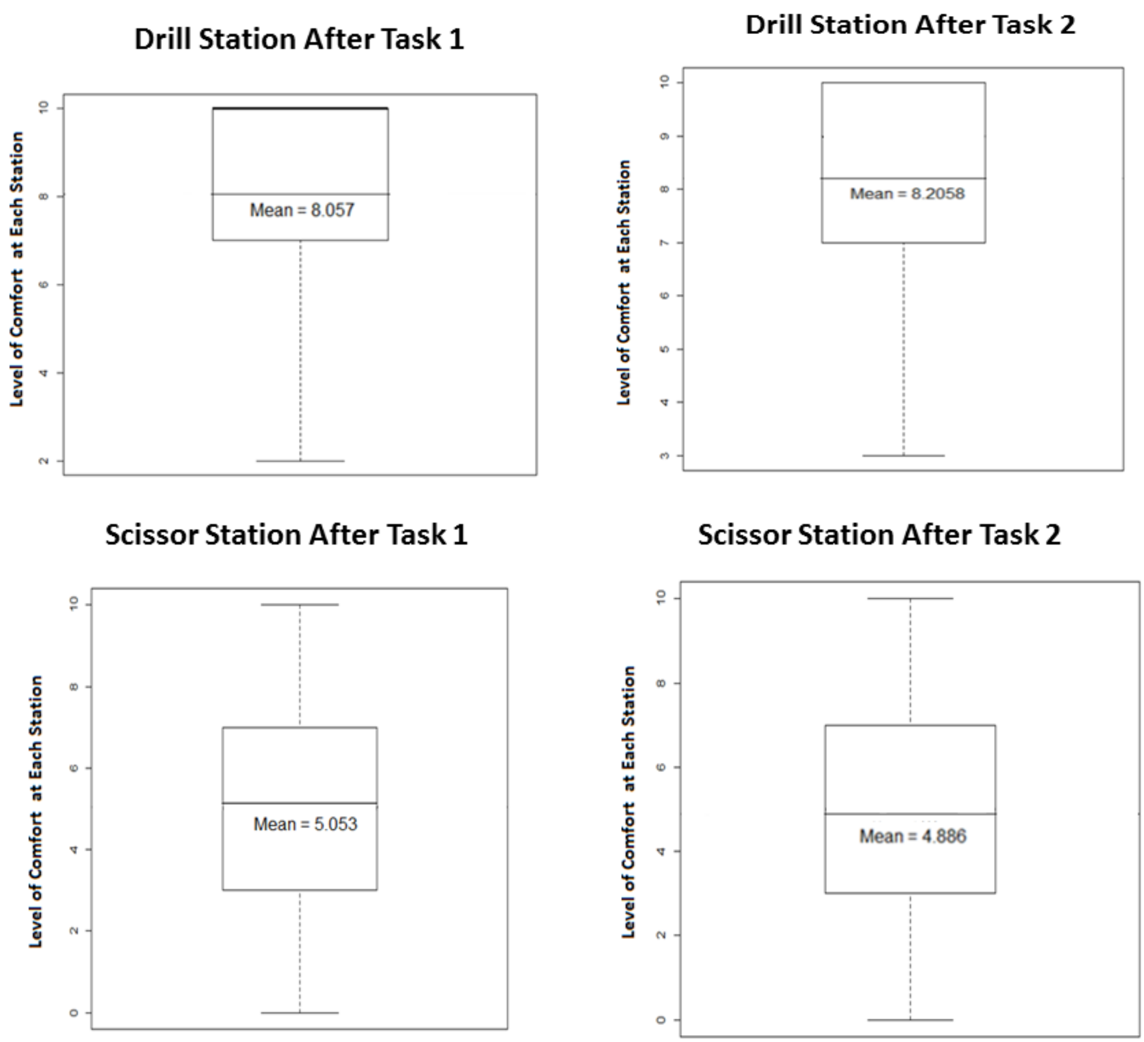


Figure 11: Statistical Summary of Level of Comfort at the Three Workstations

Table 1: Participant Summary

Participant Summary	
Number of Paricipants	40
Age of Participants	18-19 years
Undergraduate Year	Freshman
% of Female Participants	27.5%

Table 2: Analysis of Comfort Level at the Three Workstations

Level of Comfort	Minimum	1st Quartile	Median	Mean	3rd Quartile	Maximum
Initial	0	6	7	6.838	8	10
Saw Station after Task 1	0	7	8	6.914	10	10
Saw Station after Task 2	1	7	8	8.027	10	10
Drill Station after Task 1	2	7	10	8.057	10	10
Drill Station after Task 2	3	7	9	8.2058	10	10
Scissor Station after Task 1	0	3	5	5.053	7	10
Scissor Station after Task 2	0	3	5	4.886	7	10

Table 3: Wilcoxon Signed Rank Wilcoxon Signed Rank test to Assess Comfort as Baseline and After the Tasks at the Equipment Stations

Comfort Level	N	Mean	Standard Deviation	Standard Error	V-Value	P value
Baseline	37	6.838	2.339	0.385	-	-
Power Saw after Task 1	37	6.703	3.045	0.501	161	0.054
Power Saw after Task 2	37	8.459	1.626	0.267	255.5	0.0408
Drill after Task 1	37	8.459	1.789	0.294	105.5	0.002
Drill after Task 2	37	8.135	2.668	0.439	130	0.012
Scissors after Task 1	37	8.324	2.199	0.362	169.5	0.012
Scissors after Task 2	37	8.14	2.134	0.541	141.5	0.021

Table 4: R squared values for the Equipment Specific Models

Engineering Equipment Work Station	R squared Value
Power Saw Station	0.835
Drill Station	0.863
Scissors Station	0.824

Table 5: Evaluation of SVM Regression Model

Metric	Value
Correlation Coefficient	0.83
Adjusted R squared Value	0.68
Mean Squared Error	1.157
Root Mean Squared Error	1.32
Relative Absolute Error	0.41
Root Relative Squared Error	0.23