PREDICTING SUBJECTIVE RESPONSES FROM HUMAN MOTION: APPLICATION TO VEHICLE INGRESS ASSESSMENT (MSEC2014-4039)

Hadi I. Masoud[1]
Integrative Systems and Design
University of Michigan–Ann Arbor
Ann Arbor, MI, USA
hadimas@umich.edu

Industrial Engineering
King Abdul-Aziz University
Jeddah, Saudi Arabia

Matthew P. Reed
University of Michigan Transportation Research Institute
Ann Arbor, MI, USA
mreed@umich.edu

Industrial and Operations Engineering
University of Michigan–Ann Arbor
Ann Arbor, MI, USA

Kamran Paynabar
Industrial Systems Engineering
Georgia Institute of Technology
Atlanta, GA, USA
kamran.paynabar@isye.gatech.edu

Jionghua (Judy) Jin
Industrial and Operations Engineering
University of Michigan–Ann Arbor
Ann Arbor, MI, USA
jhjin@umich.edu

Ksenia K. Kozak
Vehicle Design Department
Ford Motor Company
Dearborn, MI, USA
kkozak3@ford.com

Nanxin Wang
Vehicle Design Department
Ford Motor Company
Dearborn, MI, USA
nwang1@ford.com

Jian Wan
Vehicle Design Department
Ford Motor Company
Dearborn, MI, USA
jwan1@ford.com

Gianna Gomez-Levi
Vehicle Design Department
Ford Motor Company
Dearborn, MI, USA
ggomezle@ford.com

[1] Corresponding Author.

ABSTRACT

The ease of entering a car is one of the important ergonomic factors that car manufacturers consider during the process of car design. This has motivated many researchers to investigate factors that affect discomfort during ingress. The patterns of motion during ingress may be related to discomfort, but the analysis of motion is challenging. In this paper, a modeling framework is proposed to use the motions of body landmarks to predict subjectively reported discomfort during ingress. Foot trajectories are used to identify a set of trials with a consistent right-leg-first strategy. The trajectories from 20 landmarks on the limbs and torso are parameterized using B-spline basis functions. Two group selection methods, group nonnegative garrote (GNNG) and stepwise group selection (SGS), are used to filter and identify the trajectories that are important for prediction. Finally, a classification and prediction model is built using support vector machine (SVM). The performance of the proposed framework is then evaluated against simpler, more common prediction models.

Keywords: ingress discomfort, human motion, dimension reduction, group variable selection, support vector machine

1 INTRODUCTION AND LITERATURE REVIEW

The discomfort associated with the ingress motion has increasingly become one of the important ergonomic factors considered in the vehicle design (Wegner et al., 2007). Many methods have been proposed in the literature to assess ingress discomfort. For example, earlier studies (Bottoms, 1983) (Petzall, 1995) used time required for entry as the main discomfort measure, under which optimal vehicle design parameters were recommended to reduce entry duration. Kim and Lee (2009) developed a method that uses muscle forces to

predict discomfort during ingress, in which fuzzy logic was used to establish the relationship between muscle forces and discomfort. Other studies were conducted to understand the relationship between vehicle design parameters and ingress discomfort. Giacomin and Quattrocoloo (1997) analyzed the discomfort of ingress/egress into the rear car seat under different design parameters of the doorframe and seat, where discomfort was assessed using subjective responses. Causse et al. (2012) assessed the effects of roof height on ingress/egress discomfort using subjective responses. Although these studies took into account vehicle design parameters, they did not consider the effect of participants' movement variability.

Dufour and Wang (2005) proposed the concept of "neutral movement," which uses joint angles to assess discomfort during ingress/egress. The joint angles in this study were calculated from ingress/egress motion data obtained using motion capture systems. In addition, recent advancements in human motion simulation technology have provided vehicle engineers the ability to simulate drivers' ingress/egress motion before physical prototypes are made. The simulations can efficiently generate motion data of participants with a wide range of body size. However, limited research has been conducted to model the relationship between ingress discomfort and human motion data.

Understanding the relationships between drivers' discomfort and their ingress/egress motion has several purposes. First, it helps guide vehicle design to reduce ingress difficulty. Second, it provides the potential capability of using computer-based simulations of ingress movements to predict subjective responses. This will reduce the need to conduct human subjects' tests, hence reducing the cost and time required to assess the ingress discomfort of new vehicle designs. Third, subjective responses could be predicted in situations in which motion data can be obtained but soliciting subjective responses is not feasible or desirable

This paper presents a modeling framework that predicts subjective discomfort responses using ingress motion data, described by the Cartesian trajectories of body landmarks, known as *motion curves*. The biggest challenge in constructing such a model is the high dimensionality of motion curves. The underlying assumption in most statistical regression modeling methods is that the number of samples is higher than the number of variables (Bellman, 1961; Donoho, 2000; Fan and Li, 2006). In the current case, as in many situations with high-dimension data, the number of potential predictors greatly exceeds the number of trials (samples or observations). This may lead to overfitting and unstable parameter estimates (Vapnik, 1998; Jian, Duin, and Mao, 2000). Moreover, if only a few variables are actually important for prediction, the remaining variables act as noise that increases the classification error (Fan and Fan, 2008).

Another challenge in analyzing motion curves is their misalignments among different trials. This issue is inevitable because different participants may have different start/end locations and perform trials at different paces. Directly using such misaligned curves for analysis will produce misleading results (Ramsay and Li, 1998). Therefore, it is crucial to develop an effective modeling approach that can integrate curve alignment and data dimension reduction methods into the prediction model development.

The remainder of this paper is organized as follows. Section 2 describes the data, the framework, and the methods used in this paper. Section 3 presents the results of applying the framework to the ingress experiment data. Finally, Sections 4 and 5 present discussions and conclusions.

2 METHODS

Data Source.

Ford Motor Co. performed a laboratory experiment to capture the motion of participants during the ingress and egress process. The 32 participants whose ingress and egress motions were captured are representative of the general population spanning from a 3% female to a 98% male in height (stature). The participants also vary in weight and amount of body fat from very thin (BMI = 19) to obese (BMI = 52). Numerous anthropometric measures were taken for each participant to ensure that a digital human manikin could be created in the future to accurately represent the participants.

The Ingress and Egress motions were captured for 17 vehicle package dimensions. The experiment was designed to ensure that 7 key design variables were varied to include values that not only are representative of vehicles currently in production, but also go beyond those in existing vehicles to allow greater ranges for future models. These 7 design variables are shown in Figure 1 and are defined as per the SAE J1100 standard. Moreover, The Cartesian coordinate system used for this analysis was based on the SAE J187 standard coordinate system used in vehicle engineering. The programmable vehicle buck, the Human Occupant Package Simulator (HOPS), was utilized to create the vehicle packages. The HOPS includes integrated Vicon motion capture cameras that were strategically located to avoid obstructions between the cameras and body markers.

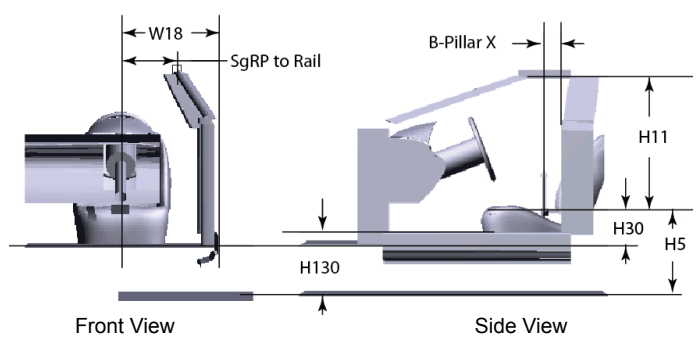

Figure 1 Car seat design variables as per SAE J1100

A unique set of reflective markers was designed to capture whole-body motion. The markers were attached to both anthropometric landmarks as well as in clusters to some segments (such as the thigh) to ensure that motion of the joints

could be reconstructed accurately. Most markers were placed directly on the participant's skin to ensure that the motion of the bony structure and not that of the clothes was captured. Using custom software, the trajectories of these markers were used to estimate the locations over time of 20 joints that define a kinematic linkage of the body. The 20 joints calculated are right ankle, right knee, right toe, right shoulder, right elbow, right wrist, right clavicle, left hip, left ankle, left knee, left toe, left shoulder, left elbow, left wrist, left clavicle, head, neck, and the spinal vertebrae T12L1, T1T2, S1L5. These joints were calculated because they were sufficient to define the whole-body kinematic linkage. An example of the motion trajectories calculated from one ingress trial is shown in Figure 2. Following the standard automotive industry convention, the trajectories were expressed in a coordinate system with the z axis (vertical), x axis (oriented fore and aft along the vehicle longitudinal axis), and y axis ("cross car"). For the current analysis, only the data of the ingress motion trajectories were analyzed. The methodology developed in this paper can also be used on egress motion trajectories.

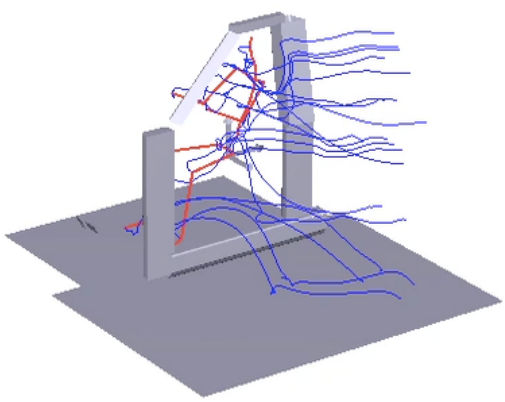

Figure 2 Motion trajectories calculated from one ingress trial

After participants completed an ingress/egress trial they were asked to rate the trial. The participants were asked: "Please rate the ease of getting in and out of this vehicle configuration"; of which the possible responses on a 10-point scale were 1-2 (unacceptable), 3-5 (average), 6-8 (outstanding), and 9-10 (truly exceptional). For the current analysis, these responses were transformed into 0/1 responses using cut points, such as transforming a response to 1 if the rating was greater than 5.

Modeling Framework

The proposed modeling framework consists of three major steps (Figure 3). The first step is to register the motion curves. This includes selecting motion trials that demonstrate a consistent strategy (Chateauroux, 2009) and aligning/ normalizing the motion curves. At the second step, the dimensionality of motion curves is reduced by using a hierarchical two-level approach. The first level is to fit each motion curve by using B-spline basis functions, which can effectively represent functional data with a small number of spline coefficients (Ramsay and Silverman, 2005). At the second level, the number of joints is reduced by integrating two commonly used variable selection approaches in designing a classifier: the filtering approach and the wrap approach (Kohavi and John, 1997). In the paper, the filtering approach is based on group nonnegative garrote (GNNG), a group variable selection method that is used to filter out joints that are not important for predicting the subjective response. Afterward, the wrap approach, based on the stepwise group selection (SGS) method, is used to identify more critical joints that closely affect the classification performance. In the third step of the proposed framework, the support vector machine (SVM) classifier model is trained and validated via the cross-validation method. The resultant model can then be used to predict subjective responses of future trials.

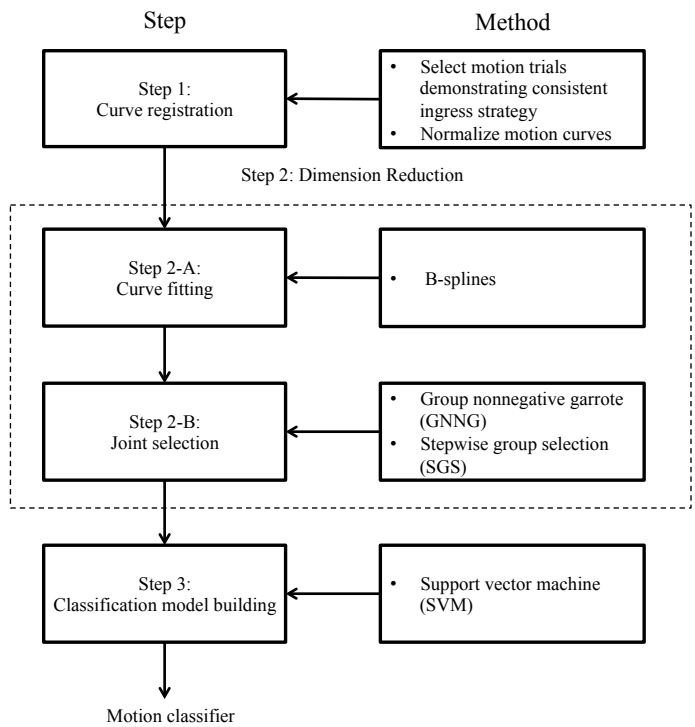

Figure 3 Framework for developing a classifier based on human motion

Step 1: Curve Registration

Curve registration aligns data in space and time so that particular features of the curves are colocated and that points on curves are compared at similar states (Ramsay and Li, 1998). In this paper, curve registration is conducted using two steps: defining the start/end points of each trial and normalizing the motion curves to have the same time range of [0, 1]. We have left the analysis of motion duration for future work.

At the first step of curve registration, the right-leg-first strategy (Chateauroux, 2009) was used to define the start/end points for ingress trial, that is, each ingress trial starts when the right foot leaves the ground to enter the vehicle and ends when

the left foot subsequently enters the vehicle. On the basis of this strategy, the speed signals of the right and left ankles can be used to identify the start/end point of ingress motion trials using the following rule: each trial starts at the beginning of the largest right ankle motion and finishes at the end of the largest subsequent left ankle motion. Figure 4 shows an example of using the right-leg-first strategy to define the trial's start/end points.

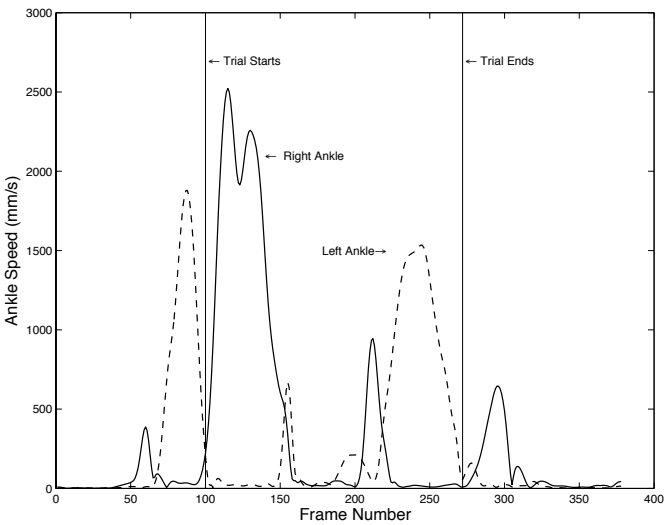

Figure 4 Define start/end points using the right-leg-first strategy

At the second step of curve registration, all motion curves are normalized to have the same time duration [0,1] so that signal shapes with different sampling durations (i.e., different number of frames) can be analyzed and compared more effectively (Ramsay and Silverman, 2005). The duration of each trial was normalized by linear interpolation to the interval [0, 1] such that each curve contains exactly 200 frames. Linear interpolations were conducted as follows (Davis, 1975):

Given that the number of frames in the original curve is (l_{old}) and the number of frames in the new curve is ($l_{\text{new}} = 200$) and

$$X_{\text{old}} = \left[0, \frac{1}{l_{\text{old}}-1}, \frac{2}{l_{\text{old}}-1}, \dots, 1\right], \quad (1)$$

$$X_{\text{new}} = \left[0, \frac{1}{l_{\text{new}}-1}, \frac{2}{l_{\text{new}}-1}, \dots, 1\right]. \quad (2)$$

The new curve is defined as follows:

$$y_{\text{new},i} = y_{\text{old},j} + \left(y_{\text{old},j+1} - y_{\text{old},j}\right)\frac{X_{\text{new},i} - X_{\text{old},j}}{X_{\text{old},j+1} - X_{\text{old},j}}, \text{for } 1 < i < l_{\text{new}}, \quad (3)$$

$$y_{\text{new},1} = 0, \quad (4)$$

$$y_{\text{new},l_{\text{new}}} = 0, \quad (5)$$

where $y_{\text{old},j+1}$ and $y_{\text{old},j}$ are points on the original curve, and $y_{\text{new},i}$ is an interpolated point of the new normalized curve. The value of $X_{\text{new},i}$ is between $x_{\text{old},j+1}$ and $x_{\text{old},j}$.

Figure 5 and Figure 6 show an example of motion curves before and after registration. It can be observed that features are not aligned in Figure 5, whereas features are well aligned for comparison in Figure 6.

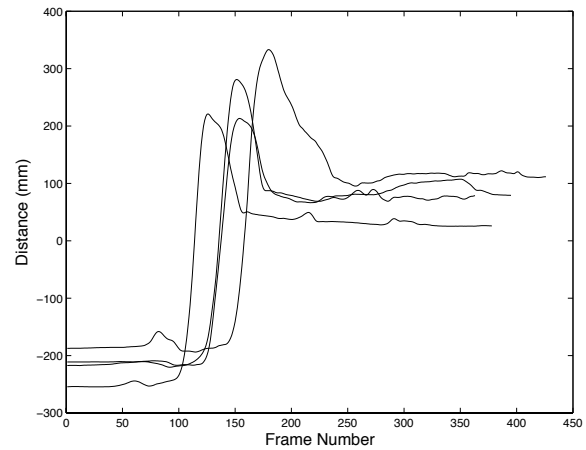

Figure 5 Example of right ankle curves (z direction) before registration

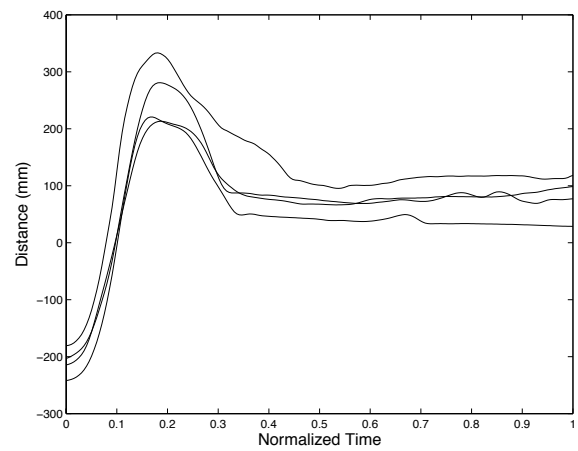

Figure 6 Example of right ankle curves (z direction) after registration

Step 2: Dimension Reduction

Step 2A: Individual Curve Fitting Using B-splines

In this methodology, a B-spline basis function (de Boor, 2001) is used to fit the curves. A B-spline is a smooth polynomial function that can represent complex curves by a small number of coefficients while preserving the basic features of the curve. B-splines are widely used in statistics to represent complex data (Cardot et al., 2004; Ramsay and Silverman, 2005). The B-spline method was chosen in this framework to fit motion curves due to its simplicity of construction, its accuracy, and its ability to represent complex motion curves. Because the B-splines fit the trajectory curves closely, alternative curve fitting techniques with similar accuracy would likely have

produced equivalent results. B-spline functions are mathematically defined as follows:
Let $T = \{t_0, t_1, t_2, \dots, t_{m-1}\}$ be a nondecreasing sequence of m real numbers called *knots*. Then the linear combination,

$$C(t) = \sum_{i=0}^{n} P_i N_{i,p}(t) \quad , \tag{6}$$

is a B-spline curve with degree $p = m - n - 2$, where $P_0, P_1, \dots, P_n$ are referred to as *control points* and the ith B-spline basis function $N_{i,p}(t)$ is defined by the following recurrence relations:

$$N_{i,1(t)} \begin{cases} 1 & \text{if } t_i \le t \le t_{i+1} \\ 0 & \text{otherwise} \end{cases}, \text{ for } p = 1 \tag{7}$$

and

$$N_{i,p}(t) = \frac{t - t_i}{t_{i+p} - t_i} N_{i,p-1}(t) + \frac{t_{i+p+1} - t}{t_{i+p+1} - t_{i+1}} N_{i+1,p-1}(t) \quad , \quad for\ p > 1. \tag{8}$$

When the independent variable is not spatial (e.g., time normalized, as in the current case), the control point coordinates are interpreted as spline coefficients. Using this parameterization, each Cartesian trajectory is represented by $3m$ spline coefficients, where m is the number of knots. For the current analysis, nine uniformly spaced knots were used. The spline coefficients were fit using a least-squares fitting procedure in R software.

Step 2B: Identification of Important Trajectories Using Group Variable Selection

Filtering Stage: Group Nonnegative Garrote (GNNG). Motions described by the spline coefficients of trajectories still have a high dimension when many trajectories are considered as potential predictors. GNNG (Yuan and Lin, 2006; Paynabar et al., 2014) is a group variable selection method that can reduce data dimensionality by grouping predictors into certain categories and only selecting those important for prediction. GNNG is mathematically defined as follows:

$$\min_{d_k} \frac{1}{2} \left\| \mathbf{y} - \sum_{k=1}^{K} \mathbf{X}_k \hat{\beta}_k^{\text{ols}} d_k \right\|^2 + \lambda \sum_{k=1}^{K} d_k, \quad \text{subject to: } d_k \ge 0,\ k = 1,2,\dots,K, \tag{9}$$

where $\mathbf{y}$ is the vector of observed ingress discomfort responses, $\mathbf{X}_k$ is the matrix of predictor variables for joint k (i.e. the B-spline coefficients associated with joint k), $\hat{\beta}_k^{\text{ols}}$ is the vector of the estimated coefficients using the ordinary least square estimation method corresponding to group k, d_k represents the importance of each group, and λ is a tuning parameter. Conceptually, GNNG optimizes the number of groups selected by penalizing the addition of new groups to the model. The performance of the prediction model highly depends on the choice of λ. An optimal value of λ that minimizes the prediction error is found using the k-fold cross-validation method (Hastie, Tibshirani, and Friedman, 2008).

For the current analysis, all trajectories associated with a certain joint were considered a group. GNNG is used in this methodology as a filtering method to exclude joints that are not important for prediction rather than choosing the optimal set of joints. GNNG therefore substantially reduces the computational cost of the SGS method.

Wrapping Stage: Stepwise Group Selection (SGS). SGS is the last stage of the dimension reduction step where the critical trajectories for classification are identified. SGS uses the results obtained from GNNG to identify the optimal trajectory combination that will maximize the classification prediction accuracy (i.e., the number of correct predictions divided by the total number of trials).

The SGS method is an alteration to the well-established stepwise regression method (Hocking, 1976; Draper and Smith, 1998). In stepwise regression, variables are added to the regression model one at a time based on the p value. Similarly, in SGS, groups (the B-spline coefficients of the three Cartesian trajectories from a single joint) are added to the classification model one at a time based on the classification prediction accuracy. In this paper, the classification model used for choosing groups was the SVM, which will be explained in the next section. The method can be explained is the pseudo code shown in Figure 7. The main advantage of SGS over other methods is that it directly uses the classification model prediction accuracy as a criterion to select joints rather than relying conventional variable selection criteria such as mean square error (MSE), Akaike information criterion, and Mallows's C_p. This ensures that a group will be added to the model only if it increases its prediction accuracy.

SGS, as is the case with traditional stepwise methods, is a computationally expensive method. The number of iterations required by SGS to obtain a result is $\frac{g\,(g+1)}{2}$, where g is the number of groups. It can be observed that the computation cost of SGS increases rapidly as the number of groups increase. It is for this reason that SGS was preceded by a filtering stage.

Step 4: Classification Using Support Vector Machines (SVM)

The focus in the previous steps was to reduce the dimensionality of the data and to choose the variables that will most accurately predict the subjective response in future trials.


```
Read Joints_Input

Initialize SGS = Ø, SGS_Accuracy = Ø, j = 1

While Joints_Input ≠ Ø

        Let n = number of joints in Joints_Input
        Initialize i = 1, k = 1, max_accuracy = 0, accuracy = 0,
                G = Ø, SVM_input = Ø

        For i ≤ n
                SVM_input = Joints_Input(i) + SGS
                Run SVM model with SVM_input
                Accuracy = Calculate SVM model cross-
                validated prediction accuracy
                If accuracy > max_accuracy
                    max_accuracy = accuracy
                    k = i
                End

        i = i + 1

        End

        G = Joint_Input(k)
        Remove G from Joints_Input
        SGS = SGS + G
        SGS_Accuracy(j) = max_accuracy
        j = j + 1

End
```

Figure 7: SGS pseudo code

In this step, these variables are used to train a classification model using SVM (Cortes and Vapnik, 1995). SVM is a supervised learning model for classifying binary data that have gained popularity in the classification community in recent years because it produces accurate classifiers, avoids overfitting data, and is able to separate data that are not linearly separable (Cherkassky and Ma, 2004; Vapnik, 1999; Pal and Foody, 2010).

The main goal of an SVM is to construct an optimal hyperplane that separates two classes of data. An optimal hyperplane maximizes the distance between the classes and the hyperplane, called *margin* (Figure 8). It can be observed that the optimal hyperplane depends on a small number of data points, also called *support vectors*, which define the optimal margin.

In the simplest case where classes are linearly separable, SVM is defined as follows:
Given some training Ddata,

$$\mathcal{D} = \{(x_i, y_i) \mid x_i \in \mathbb{R}^p, y_i \in \{-1, 1\}\}_i^n. \quad (10)$$

Find the hyperplane $w \cdot x - b = 0$, which will maximize the margin; thus,

$$\min_{w,b} ||w|| \qquad \text{Subject to } y_i\,(w \cdot x - b) \geq 1. \quad (11)$$

Because cases are not linearly separable in many applications, SVM uses the so-called kernel trick to perform nonlinear classification. The kernel trick uses a kernel function to map the original data into a higher dimensional space (inner product space), where the data are linearly separable (Boser, Guyon, and Vapnik, 1992).

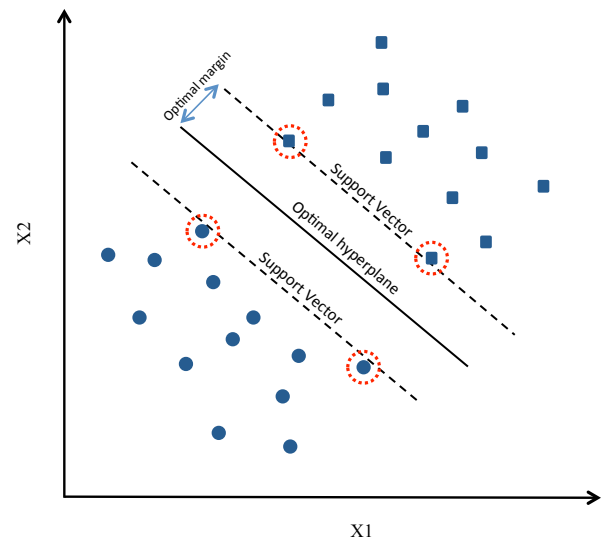

Figure 8 SVM illustration—example of a linearly separable case with two variables (adapted from Cortes and Vapnik, 1995)

In the current analysis, a Gaussian kernel (RBF Kernel) was used to train the SVM model. The parameters of the RBF kernel were optimized using grid search to obtain the maximum SVM prediction accuracy (Matheny et al., 2007). Moreover, because SVM is a binary classifier, the discomfort response was transformed from a 1 to 10 scale to binary using a range of cut points. For example, if a cut point of 5 was used to train the SVM model, responses that have a score more than 5 would be labeled as 1 and responses that have a score less than or equal to 5 would be labeled as 0. To assess the performance of an SVM model at a given cut point, the model prediction accuracy was calculated using a 10-fold cross validation.

During the process of developing this framework, many different classification methods in addition to SVM were examined. These include logistic regression, Boosting (Freund and Schapire 1996), and Random Forests (Breiman, 2001). SVM consistently outperformed these methods. However, this observation cannot be generalized to all human motion data without further investigation."

3 RESULTS

In this section, the results of applying the proposed framework to the ingress experiment are presented. Results will include the joints selected using GNNG and SGS and will also include the prediction accuracy obtained from using the final SVM model.

Joints Selected

As discussed previously, trajectories that are important for prediction were identified using a two-stage process; filtering (GNNG) and wrapping (SGS). GNNG was performed on the data set with the discomfort response modeled as a binary response using cut point 5. Using GNNG, 12 of the initial 20 joints were identified as important for prediction. Figure 9 shows the results obtained from GNNG. GNNG shrinks the importance factor (d_k) of joints that are not important for prediction to zero; therefore, they are not selected. By doing so, GNNG ensures that only the joints that minimize the MSE are selected.

The results obtained by GNNG are the main input to SGS. By reducing the number of joints from 20 to 12, GNNG significantly reduces the computational cost of SGS from 210 iterations to 78 iterations, a 63% reduction in the computational cost.

SGS was then performed on these selected joints to choose the optimal set of joints that will maximize the prediction accuracy. Figure 10 shows the results of the SGS at each step. The prediction accuracy increases until five joints are added. Afterward, as more joints are added, results gradually deteriorate due to overfitting.

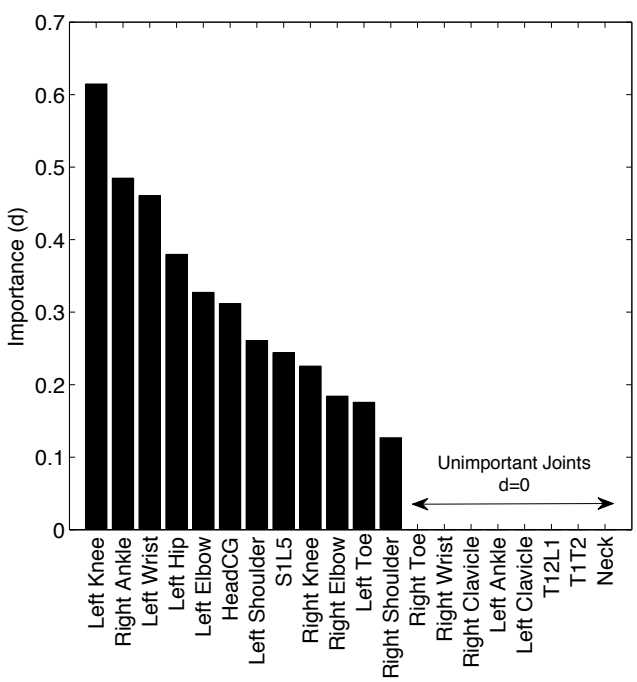

Figure 9 Importance of joints using GNNG

The final joints selected for the SVM model are the following five joints: left hip, right shoulder, right elbow, S1L5, and head. A cut point of 5 was used when performing SGS. Different runs sometimes resulted in different sets of joints. Moreover, using different cut points also resulted in different sets of joints. This occurs because of the high correlation among the joint motions and because SGS uses prediction accuracy (a random variable) as criterion to select joints. However, regardless of the specific joint set chosen, the number of joints selected was always 4 or 5, and the prediction accuracy obtained using the selected joints did not vary more than 1% per a specific cut point.

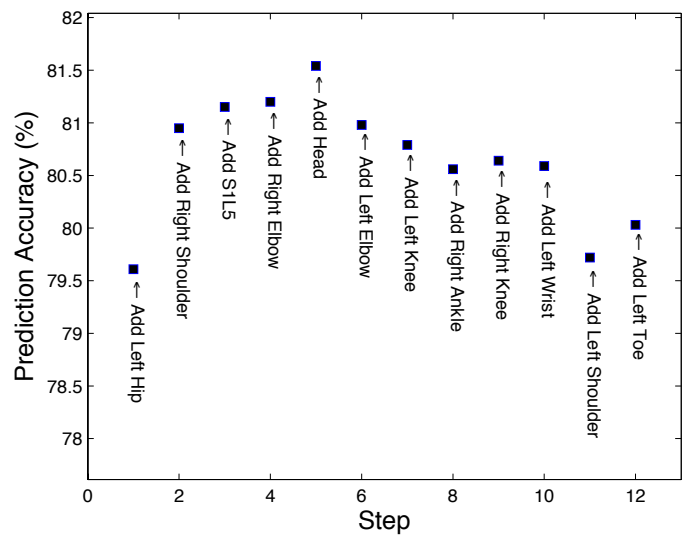

Figure 10 SGS results at each step

Prediction Accuracy

SVM models were created for cut points 2 to 7 using the joints selected by SGS. Table 1 shows the SVM prediction accuracy using the methodology proposed in this paper and the models created using Anthropometric and vehicle design variables. The model created using the proposed methodology outperformed the other model in all cut points. Depending on the cut point, the results were improved between 0.8% and 6.2%. In addition to the improvement in prediction accuracy, the motion-based model proposed in this paper is a more robust prediction model than one that is based solely on anthropometric and vehicle design variables. Such a model could capture important nonlinearities that a simpler model would not. Table 1 also shows the prediction accuracy of the model that includes both human motion curves and anthropometric and design variables. The prediction accuracy of these models is very similar to the models based on only human motion curves. This confirms our intuition that the trajectory data encode information about the effect of anthropometric and design variables

Table 2 shows the prediction accuracy for the SVM models created using 5 joints (SGS), 12 joints (GNNG), and all the 20 joints. Although SGS only uses five joints to create the SVM

models, it produces models that are on a par with and sometimes better than the models with a higher number of joints. Future experimenters can therefore use the proposed methodology to track a significantly smaller number of joints without risking the loss of important information.

Table 1 Comparison between prediction accuracy using human motion curves and prediction accuracy using anthropometric and design variables

Cut Point	Prediction Accuracy		
	Human motion curves (Using joints selected by SGS)	Anthropometric and design variables	Human Motion curves + Anthropometric and design variables
2	83.2	77.0	84.6
3	74.6	73.8	75.3
4	79.6	75.7	79.2
5	81.5	77.6	80.9
6	81.2	79.0	81.2
7	85.5	80.2	86.2

Table 2. Prediction accuracy of SVM models using different numbers of joints

Cut Point	Prediction Accuracy		
	SGS (5 joints)	GNNG (12 joints)	All Joints (20 joints)
2	83.0	83.8	82.6
3	75.6	75.3	73.6
4	79.8	78.0	76.9
5	81.9	80.0	79.6
6	81.2	80.6	81.3
7	85.5	84.9	84.9

4 DISCUSSION

The analysis in this paper showed that subjective ratings of ingress difficulty could be better predicted using motion data than solely with information about the vehicle layout and driver body dimensions. This suggests that the use of simulated human motion data to analyze candidate vehicle designs has promise for providing more accurate assessments.

At a broader level, this paper is the first to use a functional analysis of motion data to predict a subjective response. Coupled with accurate human motion simulation that takes into account vehicle and anthropometric factors, such as body dimensions and age, the method could have substantial utility, reducing the need for prototype builds and physical testing.

The extension of this methodology to other situations would seem to be limited by the need for detailed motion capture data. However, rapid advances in markerless motion capture using ordinary video cameras (Cheung et al., 2005; Corazza et al., 2006) are reducing the investment needed to obtain good-quality data. Furthermore, the analysis in this paper shows that, at least for this application, only a relatively sparse set of data is needed to obtain good prediction accuracy. This suggests potential applications in domains in which obtaining even subjective responses would be intrusive.

However, broader applications of this methodology can be anticipated. At a high level, the method extracts features from grouped time series data that provide good predictions of binary covariates. By providing an application-focused method of identifying the ideal sparsity in the input data set, the method can improve the efficiency of building classifiers in many domains.

This study has important limitations based on the nature of the underlying data. In particular, a relatively small number of participants were studied in laboratory conditions. The presence of motion capture markers may have altered the participants' motions, and the range of mockup conditions presented could have influenced the ratings. A study with actual vehicles and a different range of conditions might have produced different results.

The utility of the feature-selection procedures depends considerably on the available data and their relation to the outcome variables. For example, the potential inputs could be largely unrelated to the output, in which case only poor classifiers could be produced. In the current case, most of the variance in the output variable was generated by the experimental manipulation (changing the vehicle geometry over a large range). If all of the data were obtained from a single vehicle, most of the variance would have been due to between-subject variability, and the creation of a good classifier would not be assured.

Additional work is needed to improve our understanding of the relationships between human motion and subjective response, with particular attention to partitioning the sources of variation. Understanding the variations between individuals, and the effects of vehicle variables on these sources of variation, will help develop better prediction models.

5 CONCLUSION

This paper presented a framework that enables users to predict subjective response using human motion curves. Several tools were used to overcome the challenges presented by the complexity of motion curves, including curve registration, curve fitting, group variable selection, and binary classification. The results obtained from using this method on predicting ingress discomfort were significantly more accurate than a model using anthropometric and design variables as inputs. The framework also enables the use of a small number of joint trajectories to predict discomfort without loss of prediction accuracy.

REFERENCES

[1] Wegner, D., Chiang, J., Kemmer, B., Lamkull, D., Roll, R., 2007. "Digital Human Modeling Requirements and Standardization." *SAE International Conference of Digital Human Modeling,* 2007-01-2498.

[2] Bottoms, D., 1983. "Design Guidelines for Operator Entry-Exit Systems on Mobile Equipment." *Applied Ergonomics*, 14.2, pp. 83–90.

[3] Petzall, J., 1995. "The Design of Entrances of Taxis for Elderly and Disabled Passengers." *Applied Ergonomics* 5, pp. 343–352.

[4] Kim, S. H., Lee, K., 2009. "Development of Discomfort Evaluation Method for Car Ingress Motion." *International Journal of Automotive Technology* 10, no. 5, pp. 619−627.

[5] Giacomin, J., Quattrocolo, S., 1997. "An Analysis of Human Comfort When Entering and Exiting the Rear Seat of an Automobile." *Applied Ergonomics* 28, nos. 5–6, pp. 397−406.

[6] Causse, J., Wang, X., Denninger, L., 2012. "An Experimental Investigation on the Requirement of Roof Height and Sill Width for Car Ingress and Egress." *Ergonomics* 55, no. 12, pp. 1596−1611.

[7] Dufour, F., Wang, X., 2005. "Discomfort Assessment of Car Ingress/Egress Motions Using the Concept of Neutral Movement." *SAE International*, paper 2005-01-2706.

[8] Bellman, R., 1961. *Adaptive Control Processes: A Guided Tour*. Princeton University Press.

[9] Donoho, D. L., 2000. "High-Dimensional Data Analysis: The Curses and Blessings of Dimensionality." Aide-memoire of the lecture in AMS conference Math Challenges of 21st Century.

[10] Fan, J., Li, R., 2006. "Statistical Challenges with High Dimensionality: Feature Selection in Knowledge Discovery." *Proceedings of the International Congress of Mathematicians,* Madrid, Spain.

[11] Vapnik, V. N., 1998. *Statistical Learning Theory*. New York: Wiley.

[12] Jian, A. K., Duin, R. P. W., Mao, J., 2000. "Statistical Pattern Recognition: A Review." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22, no. 1.

[13] Fan, J., Fan, Y., 2008. " High-Dimensional Classification Using Features Annealed Independence Rules." *Annals of Statistics*, 36, no. 6, pp. 2605–2637.

[14] Ramsay, J. O., Li, X., 1998. "Curve Registration" *Journal of the Royal Statistical Society: Series B* 60, pp. 351−363.

[15] Chateauroux, E., 2009. "Analyse du mouvement d'accessibilité au poste de conduite d'une automobile en vue de la simulation—Cas particulier des personnes âgées." PhD Thesis INSA Lyon.

[16] Ramsay, J. O., Silverman, B. W., 2005. *Functional Data Analysis.* Berlin: Springer.

[17] Kohavi, R., John, G. H., 1997. "Wrappers for Feature Subset Selection" *Artificial Intelligence 97*, pp. 273−324.

[18] Davis, P. J., 1975. *Interpolation and Approximation*. New York: Dover.

[19] De Boor, C., 2001. *A Practical Guide to Splines*. Berlin: Springer.

[20] Cardot, H., Crambes, C., Sarda, P., 2004. "Spline Estimation of Conditional Quantiles for Functional Covariates." *Comptes Rendus Mathematique* 339, no. 2, pp. 141–144.

[21] Yuan, M., Lin, Y., 2006. "Model Selection and Estimation in Regression with Grouped Variable" *Journal of the Royal Statistical Society: Series B* 68, pp. 49–67.

[22] Paynabar, K., Jin, J., Reed, M. "Informative Sensor and Feature Selection via Hierarchical Non-negative Garrote (under review).

[23] Friedman, J., Hastie, T., Tibshirani, R., 2008. *The Elements of Statistical Learning.* Berlin: Springer.

[24] Hocking, R. R., 1976. "The Analysis and Selection of Variables in Linear Regression." *Biometrics,* 32.

[25] Draper, N., Smith, H., 1998. *Applied Regression Analysis*. New York: Wiley.

[26] Cortes, C., Vapnik, V., 1995. "Support-Vector Networks." *Machine Learning* 20, pp. 273–297.

[27] Cherkassky, V., Ma, Y., 2004. "Practical Selection of SVM Parameters and Noise Estimation for SVM Regression." *Neural Networks* 17, pp. 113–126.

[28] Vapnik, V., 1999. *The Nature of Statistical Learning Theory*. Berlin: Springer.

[29] Pal, M., Foody, G. M., 2010. "Feature Selection for Classification of Hyperspectral Data by SVM." *IEEE Transactions on Geoscience and Remote Sensing* 48, no. 5, pp. 2297–2307.

[30] Boser, B. E., Guyon, I. M., Vapnik, V., 1992. "A Training Algorithm for Optimal Margin Classifiers." *Annual ACM Workshop on COLT*, pp. 144–152.

[31] Matheny, M. E., Resnic, F. S., Arora, N., Ohno-Machado, L., 2007. "Effects of SVM Parameter Optimization on Discrimination and Calibration for Post-procedural PCI Mortality." *Journal of Biomedical Informatics* 40, no. 6, pp. 688–697.

[32] Freund, Y., & Schapire, R. E. ,1997. "A decision-theoretic generalization of on-line learning and an application to boosting." *Journal of computer and system sciences*, 55, no. 1, pp. 119-139.

[33] Breiman, L. ,2001. "Random forests". *Machine learning*, 45, no.1, pp. 5-32

[34] Cheung, K. M., Baker, S., Kanade, T., 2005. "T. Shape-from-Silhouette across Time. Part II: Applications to Human Modeling and Markerless Motion Tracking." *International Journal of Computer Vision* 63, no. 3, pp. 225–245.

[35] Corazza, S., Mündermann, L., Chaudhari, A. M., Demattio, T., Cobelli, C., Andriacchi, T. P., 2006. "A Markerless Motion Capture System to Study Musculoskeletal Biomechanics: Visual Hull and Simulated Annealing Approach." *Annals of Biomedical Engineering* 34, no. 6, pp. 1019–1029.