

Human performance assessment using fNIR

Il-Young Son, Markus Guhe, Wayne D. Gray,
Birsen Yazici, Michael J. Schoelles
Rensselaer Polytechnic Institute

ABSTRACT

We explore the utility of functional Near Infra Red (fNIR) technology in providing both empirical support and a basis for assessing and predicting dynamic changes in cognitive workload within the theoretical context of computational cognitive modeling (CCM). CCM has had many successes and in recent years has expanded from a tool for basic research to one that can tackle more complex real-world tasks. As a tool for basic research it seeks to provide a model of cognitive functionality; as a tool for cognitive engineering it seeks applications in monitoring and predicting real-time performance. With this powerful theoretical tool we combine the empirical power of fNIR technology. The fNIR technology is used to non-invasively monitor regional hemodynamic activities, namely blood volume changes and oxygenation dynamics.

We examined a simple auditory classification task in four different workload conditions. We monitored the blood activity in the prefrontal cortex region of the frontal lobe during the performance of the task to assess the patterns of activity as workload changes. We associated patterns of model activity with patterns of the hemodynamic data. We used ACT-R for creating the computational cognitive model. For the fNIR analysis, we used a generalized linear regression model along with time series clustering. We found that in the highest workload condition the model predicts a cognitive ‘overload’, which correlated well with the fNIR cluster and classification analysis, as this condition differs significantly from the other three conditions. Linear regression on a subset of the data where workload increases monotonically shows that apart from the overload condition, there was a positive relationship between increase in workload and increase in blood volume activation. In addition, individual variations in hemodynamic response suggest that individuals differ in how they process different workload levels.

Keywords: fNIR, Computational Cognitive Modeling, Workload

1. INTRODUCTION

Workload is a comprehensive concept with many aspects. A simple definition of workload would be the demand placed upon humans.¹ A still simple but theoretically more interesting definition is that workload refers to that portion of the operator’s limited capacity that is required to perform a particular task.²

Measuring workload or predicting peaks of workload are important to a number of real-time, safety-critical domains such as aviation or traffic. Many measurements for quantifying workload have been proposed. These measures can be classified into three categories: *subjective measurement*, such as self-report, e.g. the NASA-TLX, *performance measurement*, where the operator’s performance in the tasks is evaluated, e.g. Multiple Resource Theory,³ and *physiological measurement* like Galvanic Skin Response. However, these approaches suffer from limitations in predictive power and general applicability.⁴

For this paper, we define workload based on external demands; namely the rate at which a given set of operations must be performed. Our assumption is that increasing the rate of external demands correlates with workload to the extent that user performance keeps pace. Our goal is to develop a real-time system to measure and detect differences in external demands. The work presented here has three aspects. First, we study the relationship between task demand and two performance measures; namely, outcome and response time. Second, we introduce and discuss the physiological response detected by a *functional Near InfraRed* (fNIR) brain imaging device that measures the hemodynamic response in the prefrontal cortex. Third, in order to determine how task demands stress functional cognition, we build a computational cognitive model, whose performance matches that of human participants in the face of varying task demands.

The fNIR technology used in this study measures blood volume and oxygenation changes in the participants to infer information about neural activation as participants are subjected to series of stimuli. The idea behind fNIR is that the response to stimulus results in increased regional cerebral blood flow (rCBF) in activated areas due to increased energy demand. The hypothesis is that by measuring the changes in oxy- and deoxy-hemoglobin (Hb and HbO₂ respectively),

one can infer relative changes in activation as the external demand is varied. The fNIR takes advantage of the fact that light in the near infrared spectrum (wavelengths 750 to 1000 nm) is weakly absorbed by biological tissue. The primary absorbers at these wavelengths are water, oxy-hemoglobin and deoxy-hemoglobin. Furthermore, the absorption spectra of oxy- and deoxy-hemoglobin differ significantly at these wavelengths.^{5,6} This last characteristic gives fNIR technique an advantage over fMRI which cannot image oxy-hemoglobin information concurrently with deoxy-hemoglobin information. Also, the fNIR technology is non-ionizing, non-invasive, non-constraining and affordable. A portable wireless fNIR system has already been developed and is discussed by Hoshi.⁷

Additionally, we consider detecting individual differences an important part of our research. In particular the physiological measures under different workload conditions differ strongly with the individual. This is analogous to the differences in corresponding affective states, such as stress.⁸ This is necessary to understand and predict a user's behavior, e.g. for an estimation of how likely it is that the user will make an error. We build an computational cognitive model in the ACT-R framework⁹ and compare the model's predictions with the participants' physiological response for validation.

2. TASK DESIGN

We use a simple classification task to set the participants' workload. The simplicity of the task makes it possible to concentrate on the effects of workload, i.e. we can be certain that the changes in our measurements are due to the differences in task demand. This sets us apart from approaches involving complex tasks.¹⁰

In our task the participants hear a sequence of letters and have to compare the current letter with the previous one. They have to determine whether the current letter comes before or after the previous one in the alphabet. Thus, for the sequence M–G the correct response is *before*, for the sequence C–K *after*. We use inter-stimulus intervals (ISIs) of 1 s, 2 s, 4 s, and 6 s to define the workload. Tasks are presented in four 10-minute blocks, where each block consists of eight intervals of 72 s. Between the intervals we enforce a pause of 10 s, between blocks a pause of 2 min minimum. After the 2 min pause the participant has to press a button to begin a new block.

The workload is the number of problems the participant has to respond to in one interval (72, 36, 18, 12, respectively). The participants' physiological responses to stimuli are recorded in real-time using the fNIR equipment.

3. FUNCTIONAL BRAIN IMAGING USING NIRS

Near infrared spectroscopy (NIRS) was introduced in 1977 by F. F. Jobsis in a seminal paper¹¹ as a noninvasive way to monitor oxygen sufficiency. Since then NIRS and diffuse optical tomography using near infrared light has been explored for use in breast cancer diagnosis and various clinical application.^{12,13} More recently, there's been a surge of research examining potential for use of NIRS technique in functional brain imaging.^{14–16} Since the nineties, it has been recognized that NIRS is well suited for brain imaging due to the fact that light penetrates the skull well at near infrared range. The sensitivity of NIR light to local oxygenation changes makes it possible for direct measurement of hemodynamics.¹⁷ This is the main advantage of applying NIRS to functional brain imaging over magnetic resonance imaging (MRI) techniques which calculate blood volume changes indirectly, the so-called BOLD (Blood Oxygenation Level Dependent) response. The nature of NIRS technology makes it completely noninvasive. Also, the use of optical light in probing allows the equipment to be less restrictive, safer, and more portable than that of fMRI or PET. A very good review of NIR technology is provided by Boas et al. in.⁵

As the source most fNIR equipments use either continuous wave (CW) light or pulsed light at picosecond resolution. The latter method is called time-resolved spectroscopy (TRS). The equipment used in this study is a CW type instrument. The CW-type method can detect dynamic changes in rCBF by measuring concentration changes in hemoglobin, both Hb and HbO₂, relative to a reference. With CW-type equipment, it is not possible to determine the absolute value of hemoglobin concentration changes due to its inability to obtain optical path length information. TRS equipments were develop to address these issues.⁷

The CW-type equipments rely on the Beer-Lambert law to calculate the relative hemoglobin concentration changes. The Beer-Lambert law can be succinctly expressed as a linear relationship between absorbance and concentration.

$$A = -\log I/I_0 = \epsilon^\lambda CL \quad (1)$$

where A is the absorbance, I the intensity of received light, I_0 the intensity of transmitted light, ϵ^λ the extinction coefficient at wavelength λ , C the concentration, and L the path length. For a highly scattering medium, a modified Beer-Lambert law is used.^{18, 19} Change in absorbance at a specific wavelength λ can be expressed as

$$\Delta A^\lambda = -\epsilon \Delta C L \quad (2)$$

In terms of the concentration of two chromophores of interest, namely Hb and HbO₂, equation (2) becomes

$$\Delta A^\lambda = (\epsilon_{\text{HbO}_2}^\lambda \Delta[\text{HbO}_2] + \epsilon_{\text{Hb}}^\lambda \Delta[\text{Hb}]) L^\lambda \quad (3)$$

Assuming we let $L = 1\text{cm}$, we have two unknowns, the concentration changes in the two chromophores. By taking measurements at a second wavelength and using the known ϵ at these wavelengths for each chromophores, we can solve for these concentration changes.

$$\Delta[\text{Hb}] = \frac{\epsilon_{\text{HbO}_2}^{\lambda_2} \Delta A^{\lambda_1} - \epsilon_{\text{HbO}_2}^{\lambda_1} \Delta A^{\lambda_2}}{\epsilon_{\text{Hb}}^{\lambda_1} \epsilon_{\text{HbO}_2}^{\lambda_2} - \epsilon_{\text{Hb}}^{\lambda_2} \epsilon_{\text{HbO}_2}^{\lambda_1}} \quad (4)$$

$$\Delta[\text{HbO}_2] = \frac{\epsilon_{\text{Hb}}^{\lambda_1} \Delta A^{\lambda_2} - \epsilon_{\text{Hb}}^{\lambda_2} \Delta A^{\lambda_1}}{\epsilon_{\text{Hb}}^{\lambda_1} \epsilon_{\text{HbO}_2}^{\lambda_2} - \epsilon_{\text{Hb}}^{\lambda_2} \epsilon_{\text{HbO}_2}^{\lambda_1}} \quad (5)$$

4. OVERVIEW OF FNIR SYSTEM

The fNIR equipment used in this study consists of the following components:

1. the probe
2. a data processing unit for pre-filtering,
3. a power supply,
4. a laptop computer (with a DAQ card) to send control signals to the probe,
5. a computer to host the cognitive task.

Figure 1 shows a labeled picture of the overall fNIR system used in this study. The participant wears a flexible headband that contains an array of four photodiodes and ten sensors (see figure 2). Each diode is coupled with its four neighboring sensors, yielding a total of 16 available channels. The raw data is acquired from the probe which is pre-filtered and processed in the data processing unit. The data is then sent to the DAQ card to be digitized and read by the computer. The computer either saves the data for off-line analysis or analyzes the received data in real-time. Signals are also sent from the computer through the DAQ to the probe in order to control the photodiodes and sensors.

5. PERFORMANCE MEASURES

We use two performance measures:

1. *Outcome*: percent of correct responses over the problems presented
2. *Reaction time*: the time between the onset of the auditory stimulus and the time of the first response – regardless of whether it was correct.

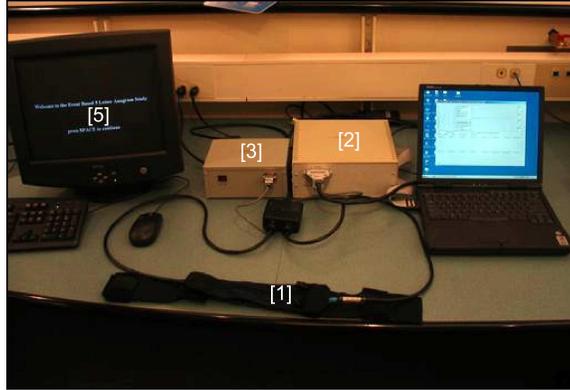


Figure 1. Data flow through the fNIR system.

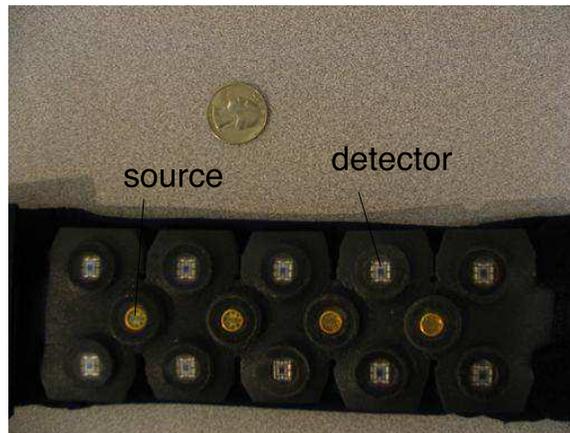


Figure 2. Headband probe used to collect participant's hemodynamic response as the participant engages in a cognitive activity.

Figure 3 shows both measures for 11 participants including the standard error. Since the error is reasonably small, we can assume that the same workload leads to comparable performance across participants. The most apparent effect is that the outcome in the ISI-2 to ISI-6 conditions is almost identical: $mean(ISI-2) = 0.947$, $mean(ISI-4) = 0.951$, $mean(ISI-6) = 0.950$. Only the outcome in the ISI-1 condition is considerably lower: $mean(ISI-1) = 0.793$. Based on outcome, therefore, only one distinction is possible, namely the ISI-1 condition from the other three.

In contrast to outcome, reaction time shows a steady increase over all four conditions. The mean values are 633 ms, 959 ms, 1284 ms, and 1385 ms, respectively. Inter-individual variability for the reaction time increases with longer ISIs, cf. figure 4. The standard deviation is 51 ms, 90 ms, 154 ms, and 269 ms, respectively. Intra-individual variability consistently increases with longer ISIs. (Only one of the 11 participants shows a decrease in variability from ISI-1 to ISI-2.) This analysis is supported by the fact that we find the same overall pattern in a related study, cf. reference 20.

The data show that despite decreasing workload the outcome is constant in the ISI-2, ISI-4, and ISI-6 conditions. Thus, the participants need decreasing effort to achieve the same outcome. It also means that outcome shows a ceiling effect at approximately 0.95. For this task, outcome alone is, therefore, not sufficient to measure a user's workload, because the reaction time data suggests that not only the task demand but also the workload (defined as the number of mental operations per time unit) differs in these three conditions. Reaction time also is a better measure of workload than outcome, but the large overlap of the standard errors shows that no reliable classification into the four workload conditions is possible.

6. ACT-R MODEL

We built a computational model to simulate the auditory classification task in ACT-R 5.0.⁹ The model's performance is shown in figure 5. The model interacted with the same software as the participants and ran for five blocks (i.e. ten intervals

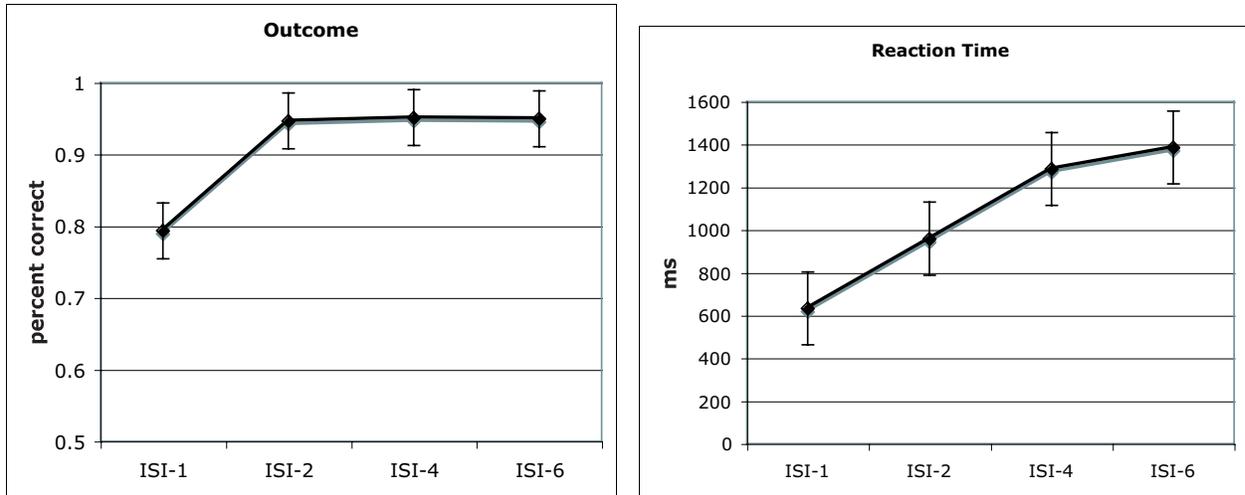


Figure 3. Performance measures for 11 participants in the auditory letter classification task. Error bars show the standard error

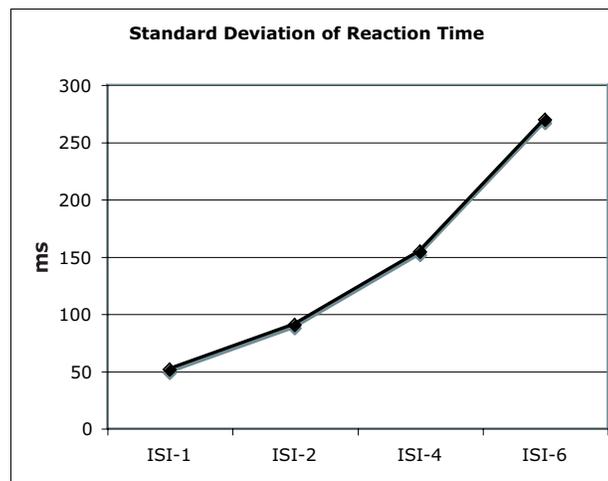


Figure 4. The mean of the standard deviation of the reaction time increases with longer ISIs

of each type). The model was reinitialized after each interval. We used ACT-R's default settings, apart from two usual changes, namely turning on subsymbolic processing and base-level learning (BLL=0.5). This already showed the effects discussed in the following. In order to better fit the data we slightly changed the Latency Factor (parameter LF) to 0.85 (default is 1.0) and the Activation Noise (parameter ANS) to 0.05 (default is 0.0), which is the data shown in figure 5.

The overall pattern of the model's performance simulates the participants' data in the two most salient aspects: ISI-1 can be distinguished reliably from the other three conditions in outcome, and reaction time is increasing over all four conditions. Although the reaction time data shows an increase similar to the participants' data, it is not as steep. It allows to distinguish the ISI-1 condition from the other three, but not ISI-2 from ISI-4 and ISI-6. (The model could be refined and extended to make this distinction.)

A further point supporting the validity of the model is that a number of participants show the same pattern as the model, even if it cannot be found in the average of participants. This is an important point if we were to use our methodology in an assistance system, which has to determine the user's workload: if the user does not experiences a meaningful difference (as, for example, between the ISI-4 and ISI-6 condition), this should not be the case in the model either. (However, there are settings that benefit from detecting such differences in workload, if not by a cognitive model, cf. reference 20 for such a system.)

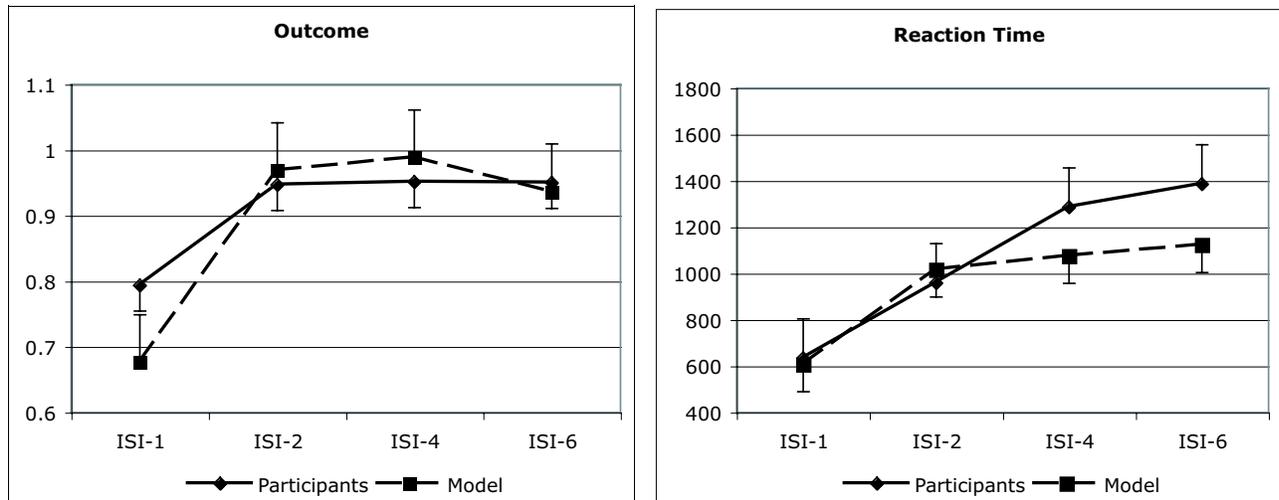


Figure 5. Comparison of the performance measures of 11 participants and a simple ACT-R model with turned on subsymbolic processing, $BLL=0.5$, $LF=0.85$, $ANS=0.05$. Error bars show the standard error

Figure 6 shows two typical responses to stimuli. It gives the major processing steps of the model when it reacts to a stimulus. The first example shows a successful response, the second one a failure. The analysis shows that the model needs around 1000 ms to respond, and in the failure example the experimental environment registers the model's key-press after the next stimulus is already presented. Thus, ISI-1 must be regarded as an 'overload' condition: a condition humans cannot cope with without additional effort, and a condition in which it is (almost) impossible for them to achieve top performance. This explains the relative difficulty that humans and the model have with the ISI-1 condition. (The other major source of errors in similar models is usually a retrieval of an incorrect chunk from declarative memory. Since the activation noise in the simulations shown was very low, however, this is not the case here.) Both, model and humans, however, perform above chance level, so the task is not so difficult that it cannot be performed. Obviously, humans can speed up computing a response to some extent. This acceleration is not captured by the model. Since the model parameters would have to be changed while the model is running, this would require an extension of the ACT-R architecture.

7. FNIR ANALYSIS

7.1. Analysis Methodology

The analysis of the fNIR data was done in three stages. Initially, the first data block was examined which consisted of increasing workload. The data was linearly fit between each ISI levels. A generalized linear model was used to regress the data piecewise between each increase in ISI levels. Next, a time series cluster analysis was performed in an effort to see if the ISIs could be 'naturally' differentiated. Thirdly, we performed a supervised classification using the ISI levels as the class labels.

7.1.1. Data Description

The raw data used in the following analysis consists of time series data extracted from the hemodynamics calculated from fNIR readings. Time series are defined as a random process $X(t)$ where $t \in T$ is taken from a countable set and usually denote some scale of time. Usually the time scale is at a regular interval. In fact, many instances of time series are sampled from a continuous random process. In this study data is collected at 3 samples per second with each time series being a 72 s interval in which the workload is held constant.

7.1.2. Preprocessing

It is not always ideal to use raw data when performing analysis. The analysis on the raw data may be unduly influenced by the noise present in the data. To minimize the influence of noise, preprocessing is usually performed on the raw data to extract more useful features. Another disadvantage of using the raw time series is its usually high dimensionality.

```

success: LF=0.85 ANS=0.05 four-boosts ISI=2
-----
Rel.  Abs.  Log
0000  30.273 problem presented (letter "C")
0350  30.623 audio stimulus detected
0450  30.723 respond to audio stimulus
0656  30.929 previous letter retrieved (letter "G")
0706  30.979 letter pair retrieved ("C"/"G")
0806  31.079 key-press initiated ("backward")
0956  31.229 System records letter press
1106  31.379 motor movement finished

```

```

failure: LF=0.85 ANS=0.05 four-boosts ISI=1
-----
Rel.  Abs.  Log
0000  4.693 problem presented (letter "J")
0350  5.043 audio stimulus detected
0450  5.143 respond to audio stimulus
0602  5.295 previous letter retrieved (letter "P")
0652  5.345 letter pair retrieved ("P"/"J")
0752  5.445 key-press initiated ("backward")
0952  5.645 NEXT LETTER PRESENTED ("N")
1052  5.745 System records letter press
1202  5.895 motor movement finished

```

Figure 6. A typical successful response to a key-press in the ISI-2 condition (from the middle of a block) and an error in the ISI-1 condition due to limitations of processing time. The times given are the relative time (in ms) to stimulus onset and the absolute time (in s) in the log-file.

Sometimes, there may be features of the time series that may provide more information than the raw amplitudes might, such as statistical information, embedded patterns (e.g. seasonal patterns), etc. The time series may also be affected by phase shifts.

To deal with these difficulties, numerous preprocessing techniques have been developed and applied. Cross-correlation functions between the time series sequences and appropriate signal representing the task paradigm (usually represented as pulses) have been suggested for preprocessing fMRI data.^{21,22} Cross-correlation reveals signals hidden in noise by measuring its correlation with a known signal, in our case the task paradigm. In addition, the periodic nature of the resulting cross-correlation function allows to truncate it, reducing the dimensionality. For these reasons, we found cross-correlation to be an attractive preprocessing technique and have employed it to our data.

7.1.3. Regression Analysis

Linear regression was applied piecewise to each transition between increasing ISI levels. The idea of a general linear regression is to fit a linear line through a given set of observations as a function of corresponding set of predictor variables or features. Generally, the model is described by the multivariate linear equation

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e} \tag{6}$$

where \mathbf{y} is the time series data, \mathbf{X} is a matrix with each column representing a predictor variable, \mathbf{b} is the vector of coefficients, and \mathbf{e} is the noise term. It is assumed that noise is independent and identically distributed with an expectation of zero. The data is fit in the least square sense using the Moore-Penrose generalized inverse.²³

$$\mathbf{b} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \tag{7}$$

7.1.4. Time Series Clustering

The idea behind clustering is simple: given a set of features and a similarity measure on this set, the data set with the smallest deviations from each other are grouped together. Thus, it is useful in identifying datasets with similar features. As mentioned before, the most widely used time series clustering method is K -means. This method is attractive for several reasons, the main one being the simplicity of the method, in both the sense of implementation and of computational complexity. The major drawback is that the value of K must be specified beforehand.

One method of clustering that avoids explicitly setting K is hierarchical clustering.²¹ The method can be thought of as a tree where the leaves are the individual feature vectors. At each successively higher level, two clusters with the smallest distance (in feature space) are merged together into one larger cluster. This is done repeatedly until there is only one cluster left. We can then examine the constructed tree at any level we desire giving us the freedom to choose any K . In the present study, the hierarchical method is used, as it was easier to examine the clustering process rather than re-running the K -mean clustering with a different K each time. This helped in our analysis, as the number of clusters did not align naturally with the number of ISI conditions. The method's main drawback is that it is only locally optimal in the sense that at each level of the hierarchy, the K clusters are optimally separated given the previous $K + 1$ cluster arrangement.

Assuming there are N number of data vectors $\mathbf{x}_i \in \mathbb{R}^P$, the method proceeds as follows. First, each data vector is assigned to its own cluster resulting in N initial clusters. An $N \times N$ symmetric distance matrix is constructed with each entry equal to

$$\delta_{i,j} = \delta_{j,i} = \frac{1}{N} d^2(\mathbf{x}_i, \mathbf{x}_j) \quad (8)$$

the distance measure between the i - and j -th time series data. Two clusters with the smallest distance are merged into a new cluster. Suppose clusters A and B are joined. Then for each cluster C_i that is different from A or B , new distances are calculated and updated in the distance matrix. The new distance is

$$\delta_{C_i, A \cup B} = \frac{(|A| + |C|)\delta_{A,C} + (|B| + |C|)\delta_{B,C} + |C|\delta_{A,B}}{|A| + |B| + |C|} \quad (9)$$

where $|\cdot|$ is the cardinality of the cluster, i.e. number of the members in the cluster. The process is repeated $N - 1$ times until there is only one cluster.

The algorithm depends heavily on the choice of the distance metric. Usually a form of generalized distance due to Mahalanobis is used

$$d^2(\mathbf{x}, \mathbf{y}) = (\mathbf{x} - \mathbf{y})^T \mathbf{D}(\mathbf{x} - \mathbf{y}) \quad (10)$$

where \mathbf{D} is a positive definite matrix. With $\mathbf{D} = \mathbf{I}$ we get the usual Euclidean distance. $\mathbf{D} = \Sigma^{-1}$ where Σ is the covariance matrix of the data vectors is the normal Mahalanobis distance. In the present case, $d^2(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|_2^2 = \mathbf{x} \cdot \mathbf{y}$ and so $\mathbf{D} = \mathbf{I}$.

7.1.5. Supervised Classification

Clustering is a type of unsupervised classification method. It is called unsupervised because the data do not come attached with class labels. Instead, the data is separated into 'natural' groupings. In contrast, supervised classification uses labeled data to train the classifier to minimize the error rate. Using Bayes's rule, one can derive that minimizing the error rate is the same as maximizing the posterior probability, where the posterior probability, $p(c_i|\mathbf{x})$, is the probability of having class c_i given the feature data \mathbf{x} . If we assume that class conditional distributions are normal then the following discriminant function can be used following Bayes's rule.

$$g_i(\mathbf{x}) = -\frac{1}{2}(\mathbf{x} - \mu_i)^T \Sigma_i^{-1}(\mathbf{x} - \mu_i) - \frac{1}{2} \ln |\Sigma_i| + \ln p(c_i) \quad (11)$$

7.2. Experimental Results

7.2.1. Regression Analysis

We performed a regression on unprocessed data for segments where workload was increased gradually from lowest to highest for each participant. Three linear segments were fitted for oxygen saturation (StO_2) data, one for each ISI transition. StO_2 is a ratio between HbO_2 and total hemoglobin concentration. For each participant, a channel with best overall fit was chosen as the representative case. F-statistics on each of these regression was generally very high (> 100) suggesting that the linear regression was significant. The R^2 measures were generally low, averaging around 46%. This suggests that linear model isn't quite sufficient to represent StO_2 as a function of ISI. Even still, examining the linear trends in oxygen saturation as ISI is decreased gives us useful insights. Results of four participants that summarizes the results of all other participants well are shown graphically in figure 7 along with their performance, both plotted against ISI.

The graphs show that for lower performance at ISI-1, oxygen saturation decreased from ISI-2, while at higher performance (those with an outcome better than 0.9) the hemoglobin concentration increased. This behavior was typical and was seen in nine out of eleven participants. We speculate that lower performance for ISI-1 reflects a decrease in effort put in by the participant. Such a relationship is not clear for lower workloads. The graphs also show that it is not generally possible to directly compare hemodynamic responses between participants. Each participant produces different hemodynamic responses to the same stimuli.

7.2.2. Cluster Analysis

We applied the clustering algorithm to the pre-processed dataset. Each participant and each channel was analyzed separately. The pre-processing was done by cross-correlating the data set with pulses representing the stimuli. The pulse duration was determined by the average reaction time for the participant. Examining the clustering for a K of 4, 3 and 2, it was found that the algorithm had difficulties separating the dataset according to the grades of ISI in each channel. However, all channels showed signs of being able to differentiate ISI-1 much better than any of the other ISI conditions. Only two out of the eleven participants tested showed no such pattern. A summary of the best results in isolating ISI-1 for each participant is shown in figure 8 with $K = 2$.

As was discussed above, the ACT-R model predicts that the participants need approximately 1000 ms to respond to a stimulus. This helps explain why ISI-1 can be isolated much better than the others. The model predicts that ISI-1 'overloads' the participant by giving barely enough time for the participant to register a response. This also sheds some light on the negative trend observed in the transition between ISI-2 and ISI-1 in the linear regression model discussed above. The stimulus may be arriving too quickly for the participants to even consider the problem.

The clustering patterns were generally non-uniform between participants. As was noted before, this confirms that it is generally not possible to directly compare hemodynamic responses between different participants. One of the reasons for this is that calculation of hemodynamic responses using a CW-type fNIR equipment limits it to a relative measurement. Another reason is that individuals differ in the way they process the fluctuations in workload. Assessing workload is therefore specific to each individual and a system built for such a task should be trained on an individual basis.

7.2.3. Classification result

Preprocessed data was used for the classification task. The preprocessing was performed in the same manner as was done in clustering. The data was split into two sets, containing training and test data. The training data consisted of two examples from each ISI level with rest being the test set. The classification results generally showed that the classifier was able to correctly label the ISI-1 case much better than the other ISIs. This trend agrees with the results that were seen in the clustering case. Figure 9 summarizes the best classification result (the best overall percentage) for the same 9 subjects as shown in figure 8. The most confusion in classification seemed to occur between the ISI-2 and ISI-4 cases with the two extreme cases performing relatively better than the middle two cases and ISI-1 performing better than all others.

8. CONCLUSION

Comparing the workload that is inferred by analyzing the fNIR data with the reaction time data shows that the reaction time decreases when the workload increases. Reaction time often is a good indicator of workload.⁴ However, it is apparent that this is not the simple 'higher workload causes longer reaction time'. Rather, the inverse is true. A more plausible explanation is that the participants assess the time they have to respond to the stimuli. For shorter ISIs this increases their

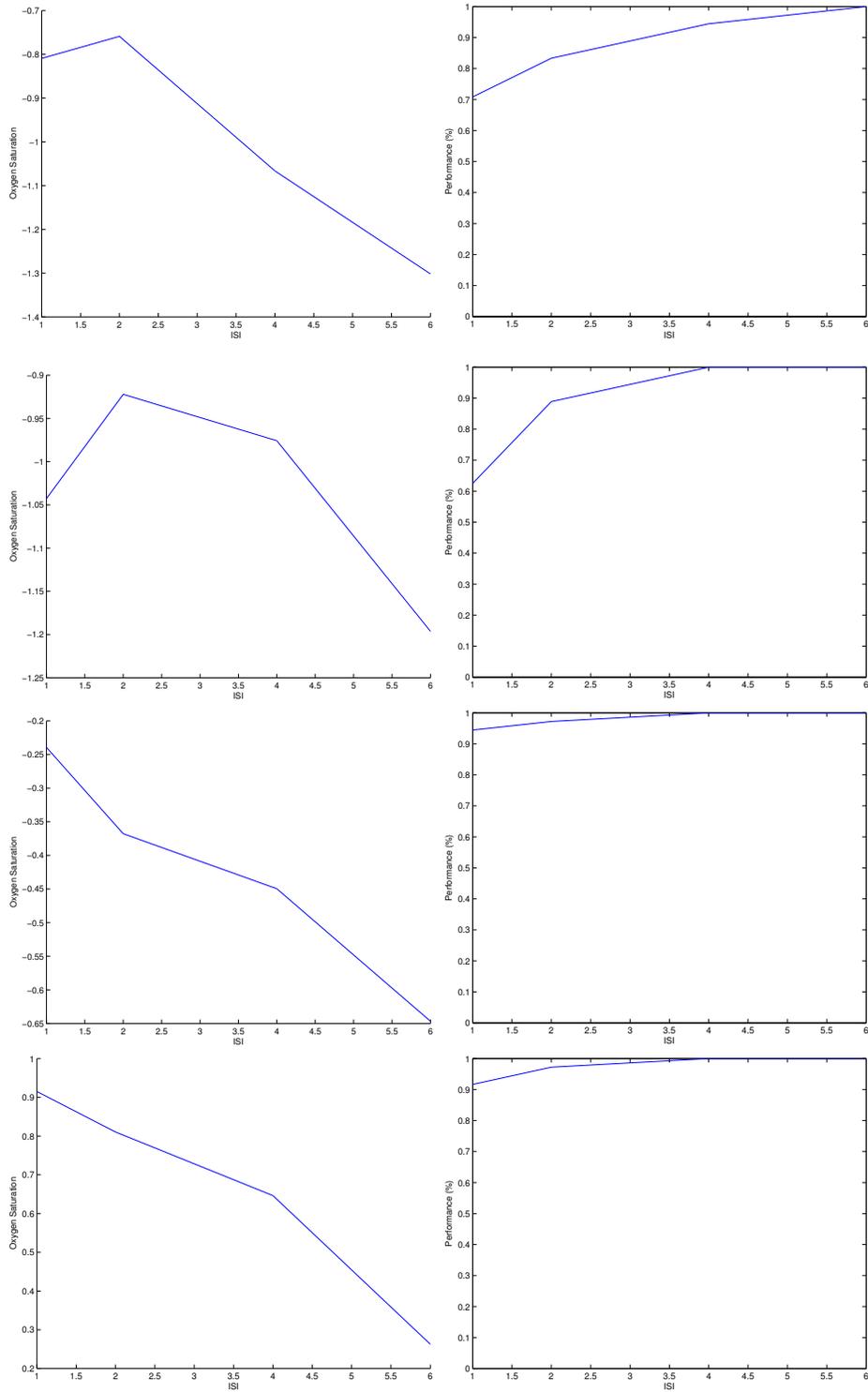


Figure 7. (Left) Regression results with best overall fit for 4 different participants on oxygen saturation fit against ISI. (Right) Performance results on 4 different participants plotted against ISI. Notice that with lower performance on ISI-1, there is typically a negative trend in transition from ISI-2 to ISI-1.

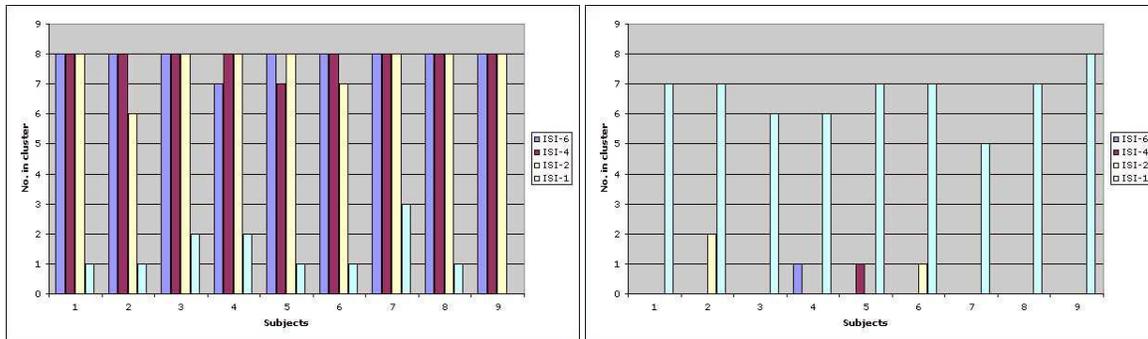


Figure 8. Tally of ISI levels belonging to each of the two clusters for 9 out of 11 participants. For two of the participants was not possible to isolate the ISI-1 condition as for the others.

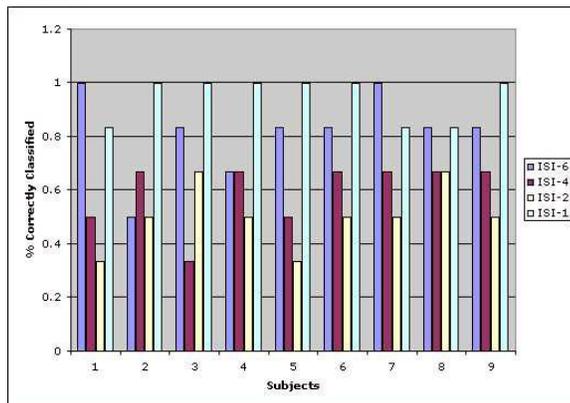


Figure 9. Results of the supervised classification. Percentage of correctly classified cases for each ISI is shown for the same 9 subjects examined in the clustering analysis.

level of arousal, which allows them to respond faster. The underlying cognitive mechanism is that a decrease in arousal increases the time for memory retrievals and slows down the other processes involved in performing this task.⁹

Figure 7 illustrates a direct relationship between task demand and the amount of cortical activation in the case of monotonically increasing workload except in the case of ISI-1. This behavior is understandable when compared with the predictions of the ACT-R model, which indicates that ISI-1 does not allow for enough time between stimuli for the participants to respond properly. However, such a relationship is generally not present when the task demand is varied more sporadically. In such a case, it was found that ISI-1 could be identified amongst other ISI conditions further supporting the model's suggestion that ISI-1 is significantly different from other ISI conditions.

Also, the fact that hemodynamic responses are generally quite different between individuals suggest that individuals differ in the way they internally assess and handle the variations in workload. However, even in the midst of such disparities, there is a general pattern in the way individuals respond physiologically to differing workloads, namely the isolation of the ISI-1 condition. This proves to be a promising lead for further analysis as it suggests that generalization is possible.

In summary, we present a new method to measure workload that offers several advantages. First, it uses a device that is only slightly intrusive: a lightweight headband. Second, the workload is measured in real-time. Third, the setup is cheap compared to the very expensive fMRI technology. Fourth, we go beyond simply measuring performance (outcome and reaction time) and demonstrate that just using such measures (outcome in particular) does not suffice to measure workload, because the same outcome can be achieved despite differences in workload.

9. ACKNOWLEDGEMENT

Various portions of this research were supported by the Center for Subsurface Sensing and Imaging Systems, under the Engineering Research Centers Program of the National Science Foundation (Award Number EEC-9986821), and Rensselaer

Polytechnic Institute. Support was also provided by the Defense Advanced Research Projects Agency (DARPA) through ONR grant #N00014-03-1-1003.

REFERENCES

1. D. de Waard, *The measurement of drivers' mental workload*. PhD thesis, University of Groningen, 1996.
2. R. D. O'Donnell and F. T. Eggemeier, "Workload assessment methodology," in *Handbook of perception and human performance*, K. R. Boff, L. Kaufman, and J. P. Thomas, eds., **2**, pp. 1–49, Wiley, New York, 1986.
3. C. D. Wickens, "Multiple resources and performance prediction," *Theoretical Issues in Ergonomics Science* **3**(2), pp. 159–177, 2002.
4. W. W. Wierwille, M. Rahimi, and J. G. Casali, "Evaluation of 16 measures of mental workload using a simulated flight task emphasizing mediational activity," *Human Factors* **27**(5), pp. 489–502, 1985.
5. D. Boas, D. Brooks, E. Miller, C. DiMarzio, M. Kilmer, R. Gaudette, and Q. Zhang, "Imaging the body with diffuse optical tomography," *IEEE Signal Processing Magazine* **18**(6), pp. 57–75, 2001.
6. A. A. Baird, J. Kagan, T. Gaudette, K. A. Walz, N. Portman, and D. A. Boas, "Frontal lobe activation during object permanence: Data from near-infrared spectroscopy," *NeuroImage* **16**, pp. 1120–1126, 2002.
7. Y. Hoshi, "Functional near-infrared optical imaging: Utility and limitations in human brain mapping," *Psychophysiology* **40**, pp. 511–520, 2003.
8. R. W. Picard, *Affective Computing*, MIT Press, Cambridge, MA, 1997.
9. J. R. Anderson and C. Lebiere, *The Atomic Components of Thought*, Lawrence Erlbaum Associates, Mahwah, NJ, 1998.
10. K. Izzetoglu, S. Bunce, B. Onaral, K. Pourrezaei, and B. Chance, "Functional optical brain imaging using near-infrared during cognitive tasks," *International Journal of Human-Computer Studies* **17**(2), pp. 211–227, 2004.
11. F. F. Jobsis, "Noninvasive, infrared monitoring of cerebral and myocardial oxygen sufficiency and circulatory parameters," *Science* **198**, pp. 1264–1267, 1977.
12. X. Intes, C. Maloux, M. Guven, B. Yazici, and B. Chance, "Diffuse optical tomography with physiological and spatial a priori constraints," *Physics in Medicine and Biology* **49**, pp. N155–N163, 2004.
13. V. Ntziachristos, A. H. Hielscher, A. G. Yodh, and B. Chance, "Diffuse optical tomography of highly heterogeneous media," *IEEE Transactions on Medical Imaging* **20**(6), pp. 470–478, 2001.
14. B. Chance, E. Anday, S. Nioka, S. Zhou, L. Hong, K. Worden, C. Li, T. Murray, Y. Ovetsky, D. Pidikiti, and R. Thomas, "A novel method for fast imaging of brain function, non-invasively, with light," *Optical Express* **2**(10), pp. 411–423, 1998.
15. A. Villringer and B. Chance, "Non-invasive optical spectroscopy and imaging of human brain function," *Trends Neurosci.* **20**(10), pp. 435–442, 1997.
16. T. Kusaka, K. Kawada, K. Okubo, K. Nagano, M. Namba, H. Okada, T. Imai, K. Isobe, and S. Itoh, "Noninvasive optical imaging in the visual cortex in young infants," *Human Brain Mapping* **22**, pp. 122–132, 2004.
17. A. Yodh and B. Chance, "Spectroscopy and imaging with diffusing light," *Physics Today* **48**(3), pp. 34–40, 1995.
18. M. Cope and D. Delpy, "System for long-term measurement of cerebral blood flow and tissue oxygenation on newborn infants by infra-red transillumination," *Med. Biol. Eng. Comput.* **28**, pp. 289–294, 1988.
19. D. A. Boas, T. Gaudette, G. Strangman, X. Cheng, J. J. Marota, and J. B. Mandeville, "The accuracy of near infrared spectroscopy and imaging during focal changes in cerebral hemodynamics," *NeuroImage* **13**, pp. 76–90, 2000.
20. M. Guhe, W. Liao, Z. Zhu, Q. Ji, W. D. Gray, and M. J. Schoelles, "Non-intrusive measurement of workload in real-time," *subm. Submitted*.
21. C. Goutte, P. Toft, E. Rostrup, F. Nielsen, and L. Hansen, "On clustering fmri time series," *Neuroimage* **9**(3), pp. 298–310, 1999.
22. C. Goutte, L. K. Hansen, M. G. Liptrot, and E. Rostrup, "Feature-space clustering for fmri meta-analysis," *Human Brain Mapping* **13**, pp. 165–183, 2001.
23. M. H. Kutner, C. J. Nachtsheim, and J. Neter, *Applied Linear Regression Models*, McGraw-Hill Irwin, 4 ed., 2004.