The AAAI 2017 Spring Symposium on
Computational Context: Why It's Important, What It Means, and Can It Be Computed?
Technical Report SS-17-03

# Heterogeneous Hidden Markov Models for Context Modeling Through Eye Gaze Observations

**Pujitha Mannaru,[1] Balakumar Balasingam,[1]**
**Krishna Pattipati,[1] Ciara Sibley,[2] Joseph Coyne[2]**
[1]Department of Electrical and Computer Engineering, University of Connecticut, Storrs, CT, USA
[2]Warfighter Human Systems Integration Lab, Naval Research Laboratory, Washington DC, USA
{pujitha.mannaru, balakumar.balasingam, krishna.pattipati}@uconn.edu, {ciara.sibley, joseph.coyne}@nrl.navy.mil

**Abstract**

Eye gaze patterns are known to have correlation with cognitive context, such as cognitive understanding, difficulty, fatigue and inattention. Traditional eye-gaze metrics that are developed for such analyses, such as fixation duration, saccade length, the nearest neighbor index (NNI), fail to accommodate the dynamic nature of mental states. In a recent work, a hidden Markov model-based observation model was suggested, where the gaze-patterns on the Cartesian plane, which correspond to each cognitive state, are modeled as Gaussians. However, we recognized that a single observation model is not sufficient to represent diverse gaze patterns that correspond to different cognitive states of the brain. In this paper, we assume a heterogeneous hidden Markov model to represent such observations and demonstrate a modified Baum-Welch approach to train such a model. The effectiveness of our approach is demonstrated using eyetracking data collected from volunteers engaged in a simulated task that required varying levels of cognitive inputs.

## 1 Introduction

Context-awareness forms the basis for proactive decision support tools in today's complex human-machine systems. Recent advances in context-aware computing and availability of low-cost unintrusive eyetrackers are encouraging researchers to employ eyetrackers in modeling user behavior and develop context-aware subsystems. In this paper, we employ the eye-gaze patterns to characterize the individual's cognitive state of the brain (or cognitive context). Cognitive context-awareness[1] should be considered as an integral part of adaptive automated systems in order to understand the human operator's cognitive state and optimize the performance of the human-machine system.

Eye gaze patterns are known to have correlation with cognitive context, such as cognitive understanding, difficulty, fatigue and inattention. Several statistical eye movement metrics, such as average fixation duration, saccade length and the nearest neighbor index (NNI), have been studied as indicators of cognitive context (Jacob and Karn 2003; Di Nocera, Camilli, and Terenzi 2007). Often, these metrics are computed either on certain areas of interest (AOIs) that are predefined by the researcher or on a convex hull surrounding the fixations; however, we believe that the AOIs have to be modeled adaptively over time. In our recent work, we suggested a model-based approach to identify the AOI by modeling the gaze patterns as 2D Gaussians (Mannaru et al. 2016b). The Markov property of eye gaze transitions (Pieters, Rosbergen, and Wedel 1999) encouraged us to model eye movements using hidden Markov models (HMMs). HMMs have been successfully employed in the past to model gaze patterns in several applications, such as user task classification, scanpath modeling and measuring attention switching (Hayashi 2003; Coutrot, Hsiao, and Chan 2017; Grobelny and Michalski 2017; Hayashi, Beutter, and McCann 2005). However, we recognized that a single observation model is not sufficient to represent diverse gaze patterns that correspond to different cognitive states of the brain. Indeed, the entropy studies of gaze data suggest that focused gaze and uniform gaze in a certain AOI may be interpreted as corresponding to over-loaded and under-loaded cognitive states of the brain, respectively (Tole et al. 1983; Harris Sr, Glover, and Spady Jr 1986). Therefore, we suggest the use of a heterogeneous hidden Markov model (HHMM) to represent gaze patterns as Gaussians and random gazes as a uniform distribution.

The goal of this paper is to demonstrate a general approach to model and quantify eye-gaze patterns in a human-computer interactive system. The rest of this paper is organized as follows: In Section 2, we present a HHMM-based approach to represent gaze patterns. In Section 3, two distinct methods to apply the model-based approach to cognitive context classification are described. The results of modeling simulated and experimental data by the suggested approach are presented in Section 4, and the paper is concluded in Section 5.

## 2 Heterogeneous HMMs for Context Modeling

A gaze at a computer screen is represented by the coordinates $\mathbf{x}_i = [x_i, y_i]$ on the screen. Given a time series of gaze point data $\mathbf{x}_t, \quad t = 1, 2, \ldots, T$, where $T$ is the number of



[1]The notion of cognitive context is defined as the state of the person's cognitive apparatus and the characteristics of the cognition that is observed/that occurs are dependent on that state (Mannaru et al. 2016a).

time samples, its distribution on the screen is modeled as a *Gaussian mixture model (GMM)*, i.e.,

$$p(\mathbf{x}_t|\boldsymbol{\lambda}^k) = \begin{cases} \sum_{i=1}^{M_k} w_i \mathcal{N}(\mathbf{x}_t; \mathbf{m}_i^k, \boldsymbol{\Sigma}_i^k) & \text{AOI focused} \\ 1/\mathcal{A} & \text{random gazes} \end{cases} \tag{1}$$

where

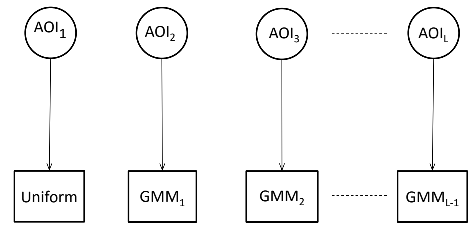
$$\boldsymbol{\lambda}^k = \left\{\mathbf{m}_1^k, \boldsymbol{\Sigma}_1^k.\right\}, \ldots, \left\{\mathbf{m}_{M_k}^k, \boldsymbol{\Sigma}_{M_k}^k\right\} \tag{2}$$

are the parameters corresponding to each Gaussian in the mixture, and the weights $w_i$ sum to unity. Here, $k$ is used as the index of the GMM; as indicated in Figure 1, there can be up to $L-1$ GMMs in the data, where $L$ is a parameter that depends on the application and needs to be estimated online using the observations. Further, the number of Gaussians in the $k^{\text{th}}$ GMM, $M_k$, is also a design parameter to be estimated; in this paper, we fix $M_k = 1$ for all $k = 1, 2, \ldots L-1$. The area of the screen is denoted by $\mathcal{A}$ and the probability density of random gazes is uniformly distributed as $1/\mathcal{A}$.

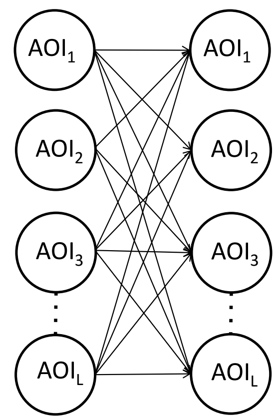
Figure 1 denotes the structure of the HHMM model that is proposed in this paper to represent the gaze patterns (similar to the ones shown in Figure 2(b)). Rather than modeling all the gaze patterns as GMMs, the proposed HHMM allows one to incorporate different models that realistically represent the nature of human gaze on computer screens.

The probability of gaze transitions from one mode (or AOI) to another (represented by one of the models in Figure 1(a)) is given by a *mode transition matrix*, as indicated by the trellis diagram in Figure 1(b). The elements of the mode transition matrix can be either estimated online, or they can be assigned based on subject matter expertise.

The Gaussian mixtures are proposed to capture different AOIs that the human is likely to focus on the screen; the uniform model is designed to capture randomly dispersed gaze patterns on the screen. There are conflicting reports of how random gazes relate to cognitive context; some researchers claim random gazes are associated with increased cognitive difficulty (Di Nocera, Camilli, and Terenzi 2007), whereas some claim otherwise (Harris Sr, Glover, and Spady Jr 1986). Regardless, the purpose of this paper is to present a model that can capture gaze patterns; we leave detailed interpretations for future studies.




(a) Heterogeneous HMM (HHMM)




(b) Probability of mode switching

Figure 1: Heterogeneous hidden Markov Model. Each AOI is represented as a Gaussian (with mean $m_i = [m_{x_i}, m_{y_i}]^T$ and covariance matrix $\Sigma_i$). The $L \times L$ state transition probability matrix (illustrated through the trellis in (b)) describes the likelihood of transition of gazes between different AOIs, as well as "random gazes" that relate to inattention. The prior probabilities of $L$ models are given by $\mu_0, \mu_1, \ldots, \mu_{L-1}$.

## 3 Approaches to Cognitive Context Classification

### Using the NNI Metric

The nearest neighbor index (NNI) is used to measure the randomness in spatial data. The NNI is defined as (Di Nocera, Camilli, and Terenzi 2007; Di Nocera, Terenzi, and Camilli 2006)

$$\text{NNI} = \frac{d_{\text{NN}}}{d_{\text{RND}}} \tag{3}$$

where $d_{\text{RND}} = 0.5\sqrt{\frac{A}{N}}$, $d_{\text{NN}} = \sum_{i=1}^{N}\left(\min\frac{(d_{ij})}{N}\right), 1 \leq j \leq N, j \neq i$, $d_{ij}$ is the distance from the $i^{\text{th}}$ gaze point to the $j^{\text{th}}$ gaze point, $N$ is the total number of gazes at that region, and $A$ is the area of the region. NNI ranges in value from 0 for a distribution with maximum aggregation to 2.1491 for a distribution, which is as evenly and widely spaced as possible. It is less than, equal to, or greater than 1 according to whether the distribution pattern of the individual gazes in the population is more aggregated, the same as, or more uniform (regular) than would be expected in an infinitely large random distribution of the same density(Clark and Evans 1954).

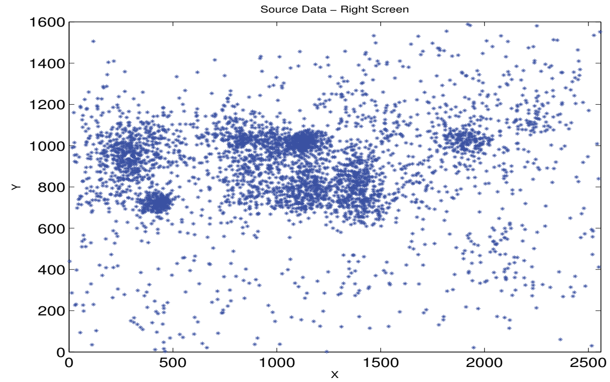
Given that a cluster of gaze data is distributed according to a Gaussian model with standard deviation $\sigma$ (in both directions), the corresponding model-based Gaussian NNI (GNNI) can be computed as

$$\text{GNNI}(\sigma) = \left(\frac{\text{NNI}(\sigma_0)}{\sigma_0}\right)\sigma \tag{4}$$

where $\text{NNI}(\sigma_0)$ is computed using (3) on Gaussian distributed data with standard deviation $\sigma_0$ in both directions.

Now, assuming that there are $L$ different AOIs, $L-1$


(a) Experimental setup




(b) Sample eye-gaze measurements

Figure 2: Sample gaze measurements corresponding to one of the SCOUT™ interface screens.

AOIs distributed as circular Gaussians with standard deviation $\sigma_i$, $i = 1, \ldots, L-1$, the overall GNNI is given by

$$\text{GNNI} = \mu_0 + \mu_1 \text{GNNI}(\sigma_1) + \ldots + \mu_{L-1}\text{GNNI}(\sigma_{L-1}) \quad (5)$$

where

$$\mu_0 + \mu_1 +, \ldots, +\mu_{L-1} = 1 \quad (6)$$

Here, $\mu_0$ is the prior probability of the HMM corresponding to the uniform gaze, $\mu_i$ is the computed prior probability of the HMM corresponding to the $i^{\text{th}}$ AOI.

Alternatively, the NNIs corresponding to a particular AOI can be computed based on (4) and used for NNI-based classification. Using NNI for cognitive context classification is popular among human factors researchers and this method is an alternative automated way of computing the NNI of an AOI that avoids random gazes and out-of-sequence gazes from consideration in computing the NNI. We believe that the NNI computed through the proposed approach will be more realistic than the one computed through traditional methods.

### Using the HHMM Classifier

As an example, the application considered in this paper involves three levels of cognitive difficulty: easy, medium and hard. For cognitive context classification purposes, we propose that the data corresponding to these three difficulty levels be used to train three different HHMMs. Once the three models are adequately trained, the trained HHMM can be used to monitor the cognitive context of a new subject.

Figure 3 illustrates a model-based approach to online classification of cognitive states. Below, we describe the change detection algorithm in one of the detection modules of Figure 3. Consider the following log-likelihood ratio

$$S_{k^*} = \sum_{k=1}^{k^*} \ln \frac{P_{\boldsymbol{\lambda}^j}(\mathbf{x}(k))}{P_{\boldsymbol{\lambda}^i}(\mathbf{x}(k))} \quad (7)$$

where $S_{k^*}$ can be incrementally updated as new data arrives and the model change is declared when

$$S_{k^*} - s_{k^*} > h \quad (8)$$

where

$$s_{k^*} = \min_{1 \le k \le k^*} S_k \quad (9)$$

is the current minimum value of $S_{k^*}$ and $h$ is a predefined threshold value.

Formally, the model change detection time is written as

$$\hat{k}^* = \arg\min_{k^*} \{k^* : S_{k^*} - s_{k^*} > h\} \quad (10)$$

where we have omitted the dependence of $S$, $s$ and $h$ on $i$ and $j$ for notational simplicity. Based on the key idea from Page's CUSUM algorithm (Page 1954), eq. (10) can be recursively computed as

$$\text{CUSUM}_k = \max\left\{0, \text{CUSUM}_{k-1} + T_k\right\} \quad (11)$$

where

$$T_k = \ln \frac{P_{\boldsymbol{\lambda}^j}(\mathbf{x}(k))}{P_{\boldsymbol{\lambda}^i}(\mathbf{x}(k))} \quad (12)$$

and a change from model ($i$) to model ($j$) is declared when $\text{CUSUM}_k$ exceeds the threshold $h$.

## 4 Results

In this section, the HHMM training approach is demonstrated using both simulated and experimental data.

### Simulated Data

Figure 4 summarizes the simulated data and the result of the experiment on it. The simulated data consists of three Gaussians that are sufficiently separated from each other. A randomly generated $4 \times 4$ mode transition matrix is used to switch the model among the three Gaussians and a uniform gaze distribution.

The objective of the learning algorithm is to use the simulated gaze data to estimate the model parameters $\mathbf{m}$ and $\boldsymbol{\Sigma}$ of the three Gaussians, along with the mode transition matrix. Figure 4(a) shows the result of training this data using a Gaussian HMM with three modes (this is a special case

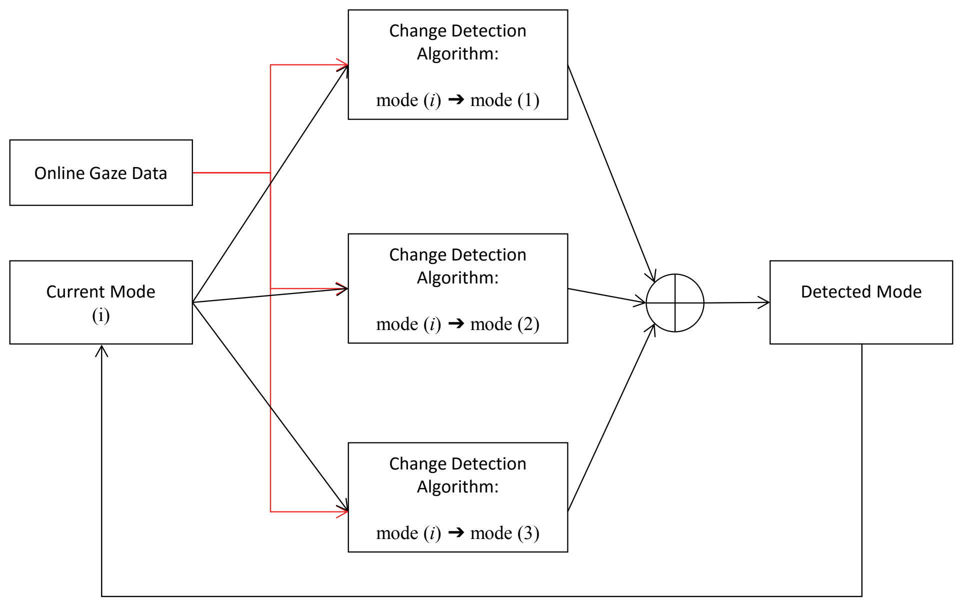



Figure 3: Model-based online classification scheme. Current mode $(i)$, $i = 1, 2, 3$, denotes the detected present state. At a given time, there needs to be 3 different online change detection algorithms running in parallel. Whenever a change in mode is detected, the current mode $(i)$ is updated to the new mode. The three models in this application correspond to the cognitive difficulty levels - $\{\mathrm{easy}, \mathrm{medium}, \mathrm{hard}\}$.

of GMM where there is only one Gaussian in the mixture and $w_1 = 1$.). The blue asterisks (*) represent the simulated gaze points, the red point shows the estimated mean $\mathbf{m}$ and the red ellipse indicates the estimated covariance $\mathbf{\Sigma}$ of the model. In other words, the learning algorithm was *forced* to estimate only three Gaussians from the data. As shown in Figure 4(a), two Gaussians were accurately estimated. The remaining data was forced into a third Gaussian model, resulting in a large covariance matrix. This is also the motivation for having a heterogeneous HMM for representing eye-gaze data, because real world datasets of gaze patterns are not necessarily Gaussian distributed.

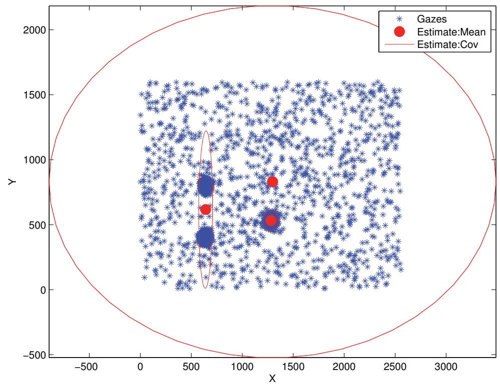
Figure 4(b) shows the results of training using the proposed HHMM architecture. As expected, all three Gaussian parameters were accurately learned by assigning the uniform gaze data to the fourth (uniform) model in the HHMM. It must be noted that learning the number of GMMs (or AOIs) is another learning objective; however, this is not tested in this paper. Information theoretic approaches, such as minimum descriptive length (MDL) and Bayesian Information Criteria (BIC), are possible tools to achieve this objective.

## Experimental Data

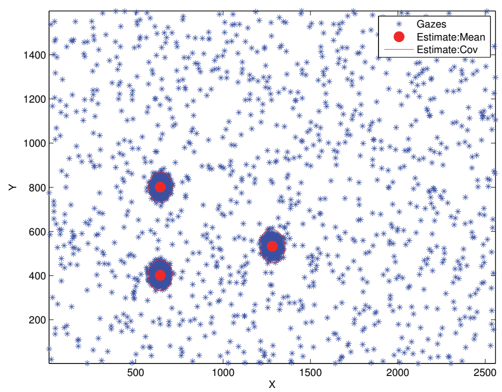
The SmartEye Pro 6.1 was used to record gaze patterns of 20 participants while they were engaged in a simulated unmanned aerial system operation on SCOUT[2] for approximately 30 minutes. The participants searched for targets of varying worth and also responded to chat messages that comprised requests for information and commands to update flight parameters (refer to (Mannaru et al. 2016a) for a detailed description of the experiment). The experimental setup and a sample set of gaze patterns are shown in Figure 2. This data was used to train the HHMM described in Section 2.

Figure 5 demonstrates the results of the HHMM training of the eye tracking data described previously. The number of Gaussians in each mixture is set to one and the number of GMMs is set to 4 ( $= L - 1$). Figure 5 shows the result of training the eye tracking data using a HHMM. Since the number of GMMs is set to 4, and the number of Gaussians in each of these GMMs is set to 1, it can be seen that the HHMM captures four visible "gaze clusters". The next objective of testing these models is to relax these assumptions and to devise an automated system to select the number of mixture components in each GMM and the number of GMMs in a HHMM. Also, the HHMM training demonstrated in this paper was done using data from one user of a specific cognitive difficulty. Future training stages will involve three HHMMs, each trained by using data from multiple test subjects undergoing three different difficulty levels. Once such an HHMM is trained, it will be tested using the approach illustrated in Figure 3.

[2]The Supervisory Control Operations User Testbed (SCOUT) was developed by researchers at the Naval Research Laboratory (NRL) for purposes of exploring UAV-operator (i.e., system) performance in a multi-UAV, supervisory control context(Sibley, Coyne, and Thomas 2016).




(a) Gaussian HMM. Only two Gaussians were accurately estimated and the remaining data was forced into a third Gaussian model, resulting in a large covariance matrix.




(b) Heterogeneous HMM (HHMM). All three Gaussian parameters were accurately learned by assigning the uniform gaze data to the fourth (uniform) model in the HHMM.

Figure 4: Demonstration of HHMM on simulated data.

## 5 Conclusions

In this paper, we presented a new approach to cognitive context modeling and classification. In a deviation from traditional approaches that compute a single metric, such as entropy and NNI, we proposed to have a model, viz., heterogeneous hidden Markov models (HHMMs), to capture the patterns of eye gazes. There are several advantages to using a model to represent the cognitive context rather than a single metric: these models are able to capture very complex features and, once trained, they can be used for online classification. However, the inability to infer cognitive context directly from the models is a limitation.

In this paper, we demonstrated the training stage of this newly proposed HHMM for cognitive context modeling. We showed that by introducing appropriate models of differing statistical properties, hence the name "heterogeneous" HMM, higher accuracies in model training can be achieved. The demonstration was done on selected data from a single participant where the number of Gaussians is fixed a priori. The next step is to implement model selection criteria to select the appropriate number of components in the Gaussian mixture, which forms the basis of the proposed HHMM, and to train using data from multiple participants for a realistic performance evaluation.

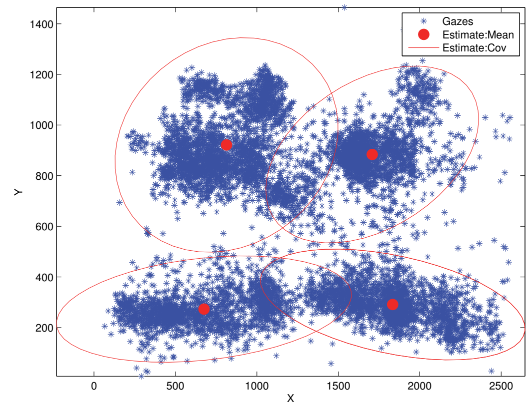



Figure 5: Demonstration of HHMM on Eye Tracking Data. The HHMM captures four visible "gaze clusters".

## 6 Acknowledgement

The authors would like to thank Dr. Jeffrey Morrison and the Command Decision Making (CDM) program at the U.S. Office of Naval research and Department of Defense High Performance Computing Modernization Program for supporting this work. In addition, the authors would like to thank the symposium organizers for their encouragement of this work. This research was funded by the U.S. Office of Naval Research and the Department of Defense under contracts #N00014-12-1-0238, #N00014-16-1-2036 and #HPCM034125HQU.

## References

Clark, P. J., and Evans, F. C. 1954. Distance to nearest neighbor as a measure of spatial relationships in populations. *Ecology* 35(4):445–453.

Coutrot, A.; Hsiao, J.; and Chan, A. 2017. Scanpath modeling and classification with hidden markov models. *Behavior Research Methods*.

Di Nocera, F.; Camilli, M.; and Terenzi, M. 2007. A random glance at the flight deck: Pilots' scanning strategies and the real-time assessment of mental workload. *Journal of Cognitive Engineering and Decision Making* 1(3):271–285.

Di Nocera, F.; Terenzi, M.; and Camilli, M. 2006. Another look at scanpath: distance to nearest neighbour as a measure of mental workload. *Developments in human factors in transportation, design, and evaluation* 295–303.

Grobelny, J., and Michalski, R. 2017. Applying hidden markov models to visual activity analysis for simple digital control panel operations. In *Information Systems Ar-*

*chitecture and Technology: Proceedings of 37th International Conference on Information Systems Architecture and Technology–ISAT 2016–Part III*, 3–14. Springer.

Harris Sr, R. L.; Glover, B. J.; and Spady Jr, A. A. 1986. Analytical techniques of pilot scanning behavior and their application. *NASA Technical Paper*.

Hayashi, M.; Beutter, B.; and McCann, R. S. 2005. Hidden markov model analysis for space shuttle crewmembers' scanning behavior. In *Systems, Man and Cybernetics, 2005 IEEE International Conference on*, volume 2, 1615–1622. IEEE.

Hayashi, M. 2003. Hidden markov models to identify pilot instrument scanning and attention patterns. In *Systems, Man and Cybernetics, 2003. IEEE International Conference on*, volume 3, 2889–2896. IEEE.

Jacob, R. J., and Karn, K. S. 2003. Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. *Mind* 2(3):4.

Mannaru, P.; Balasingam, B.; Pattipati, K.; Sibley, C.; and Coyne, J. 2016a. Cognitive context detection for adaptive automation. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 60, 223–227. Sage Publications.

Mannaru, P.; Balasingam, B.; Pattipati, K.; Sibley, C.; and Coyne, J. 2016b. On the use of hidden markov models for eye-gaze pattern modeling. In *SPIE Defense+ Security*, 98510R–98510R. International Society for Optics and Photonics.

Page, E. 1954. Continuous inspection schemes. *Biometrika* 100–115.

Pieters, R.; Rosbergen, E.; and Wedel, M. 1999. Visual attention to repeated print advertising: A test of scanpath theory. *Journal of Marketing Research* 424–438.

Sibley, C.; Coyne, J.; and Thomas, J. 2016. Demonstrating the supervisory control operations user testbed (scout). In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 60, 1324–1328. SAGE Publications Sage CA: Los Angeles, CA.

Tole, J.; Stephens, A.; Vivaudou, M.; Ephrath, A.; and Young, L. 1983. Visual scanning behavior and pilot workload. *NASA Contractor Report*.