Correlation of Cognitive Scores and the Onset of Hypoxia

Arjun Rajasekar; Sayandeep Acharya; Barry S. Shender; Chris Rorres; Leonid Hrebien; Moshe Kam

INTRODUCTION: The negative effects of hypoxia on human cognitive function have been well documented. In this study we assess the correlation of performance in the SynWin cognitive Multi-Task Battery (MTB) and the onset of hypoxia and describe the use of cognitive assessment scores for real-time hypoxia detection.

METHODS: We performed a correlation analysis between MTB scores (Arithmetic, Memory, Audio Monitoring, Video Monitoring tasks) and blood oxygen saturation levels to discover if the scores are good candidates to detect hypoxia. Since this analysis showed positive correlation, we proceeded to develop a parallel decision fusion system that uses these cognitive scores for real-time hypoxia detection using the Neyman-Pearson criterion.

RESULTS: We demonstrate that MTB scores have considerable hypoxia detection potential and can be used (if measurable passively) in a real-time detection framework. Analysis of receiver operating characteristic (ROC) curves established a hierarchy of importance of the various MTB modules. The Arithmetic task module had the most significant contribution toward correct hypoxia detection (improvement of ~13.5% and ~13.9% in detection accuracy under global false alarms of 0.1 and 0.05, respectively), followed by the Memory and Audio Monitoring modules. Fusion of multiple cognitive assessment scores resulted in significantly higher detection accuracy (>86%) than using any one of the scores by itself.

DISCUSSION: When available, cognitive assessment scores can be a useful tool for real-time hypoxia detection. Since these MTB tests also assess neuropsychological functioning, study of distributed detection systems based on MTB scores could help in designing tests that are more useful for detecting hypoxic symptoms.

KEYWORDS: cognitive performance, high altitude, blood oxygen saturation, decision fusion, pulse oximeter.

Rajasekar A, Acharya S, Shender BS, Rorres C, Hrebien L, Kam M. *Correlation of cognitive scores and the onset of hypoxia*. Aerosp Med Hum Perform. 2019; 90(5):429–439.

Hypoxia is the diminished availability of oxygen to the cells of the body.[14] Hypoxia in mammals can occur due to inadequate oxygenation of the lungs for extrinsic reasons, such as deficiency of oxygen in the environment, or due to intrinsic reasons such as venous to arterial shunts (intrapulmonary or intracardiac), inadequate transport and delivery of oxygen, or inadequate tissue oxygenation or oxygen use. Decreased oxygen availability to brain tissues can cause mild to severe deterioration in cognitive functional abilities, leading to impairment and, eventually, incapacitation and in severe cases death. Decreased cognitive abilities or incapacitation in an operational environment can lead to disastrous consequences, making accurate techniques of detection of the onset of hypoxia highly valuable.

The common ways to detect hypoxia in humans involve various physiological and environmental sensors that measure different attributes about the person being monitored and his/her surrounding environment. Over the years, there have been several studies that offered designs of systems for hypoxia monitoring using such sensors. A hypoxia detection and warning system was patented by Richardson as an aviation hypoxia monitor[21] using a single pulse oximeter attached to the ear. It provided visual and audio signal feedback if the wearer's blood oxygen saturation (S_pO_2) had a significant decrease. The hypoxia

From the Department of Electrical and Computer Engineering and the Department of Mathematics, Drexel University, Philadelphia, PA, USA; the Human Systems Department, Naval Air Warfare Center Aircraft Division, Patuxent River, MD, USA; and the Department of Electrical and Computer Engineering, Newark College of Engineering, New Jersey Institute of Technology, NJ, USA.

This manuscript was received for review in November 2017. It was accepted for publication in February 2019.

Address correspondence to: Moshe Kam, Ph.D., Newark College of Engineering, New Jersey Institute of Technology, 323 Martin Luther King Jr. Blvd., Newark, NJ 07103, USA; kam@njit.edu.

DOI: https://doi.org/10.3357/AMHP.5040.2019

detection and warning system described by Kelly and Pettit[17] is composed of an electrochemical sensor located within a breathing mask which was used to detect if the partial pressure of oxygen dropped below a set point corresponding to 11,000 ft; it provided a vibratory warning. The study by Simmons et al.[24] presented a forehead-mounted reflectance oximetry system as a possible in-cockpit hypoxia early detection and warning system. Pribil et al.[20] suggest monitoring the lactate content in undiluted sweat for noninvasive diagnosis of hypoxia related symptoms. A distributed detection system to assign a severity level of hypoxia using multiple pulse oximeters and an altimeter was proposed by Acharya et al.[2]

Most of the systems proposed so far monitor the causes of hypoxia directly (through low S_pO_2 levels) or indirectly (through low oxygen partial pressure in the environment or ascension in altitude). Some detection methods also rely on the common effects of hypoxia (changes in heart rate, respiratory rate, lactate content in sweat). None of these systems assess the cognitive state of the monitored person, though degraded cognitive capability as a function of changing environmental stressors is often the real objective behind hypoxia monitoring. We know, for example, from a simulated climb of Mount Everest,[1] that the mental efficiency and psychomotor performances of the climbers progressively deteriorated with increased altitude. Studies on 24-h exposure to hypoxia[9] showed that vigor, attention, visual and working memory, concentration, executive functions, inhibitory control, and speed of mental processing worsened due to hypoxia exposure. A study by Adam et al.[3] showed that multitask performance declined during initial, unacclimatized high altitude exposure compared to sea level. These studies[1,3,9] used cognitive assessment as a post processing tool for studying the effects of hypoxia, not as a real-time sensory output that could detect hypoxia before its onset. Cognitive scores for real-time detection have not been suggested yet, possibly because the current nature of cognitive assessment requires active subject participation at frequent intervals, which is slow and may not be feasible in many environments. With the increased interest in passive monitoring of the human cognitive state,[25,28,31] the investigation of usefulness of the cognitive state for hypoxia monitoring may now have practical implications. In this study, we aimed to analyze the nature of variation of the cognitive state of a human under the conditions that cause hypoxia and also investigate the associated correlations between physiological signals, such as blood oxygen saturation level (S_pO_2), and cognitive test scores. The analysis was performed on the scores collected from the SynWin[30] (Activity Research Services, Chula Vista, CA) Multi-Task Battery (MTB), in which subjects are scored based on performance during various memory, arithmetic, visual monitoring, and auditory monitoring tasks. For our dataset, we included measurements of respiration rate and in our volunteers there were only a few instances of hyperventilation (increased respiration rate). In future studies, as we add datasets from exposures under hypobaric conditions in which varying respiration activities and respiratory gas measurements are recorded, the model presented in this study may be reassessed. However, when there is minimal hyperventilation, S_pO_2 is considered a good indicator of hypoxia. A significant correlation between cognitive test scores and S_pO_2 could possibly justify the use of cognitive scores for hypoxia detection if scores can be obtained automatically in real time (previous studies such as Whitley and Shender[29] have demonstrated a promising relationship between predicted reduction in cerebral function, degraded math performance, and S_pO_2). Therefore, as a preliminary step we carried out correlation analysis between the S_pO_2 level and the cognitive assessment scores [obtained from the modules in the test battery: Memory, Arithmetic, Visual Monitoring (VM), and Auditory Monitoring (AM) tasks]. The analysis showed noticeable correlation. Based on the measured correlation, we also designed a real-time hypoxia detection system that uses the cognitive assessment scores as inputs.

We used data from an experiment conducted by the Naval Air Warfare Center Aircraft Division (NAWCAD).[23] During this experiment, human volunteers continually performed the MTB while exposed to simulated altitudes ranging from ground level to 18,000 ft or 25,000 ft [5486 or 7620 m; in this paper ground level = 650 ft (198 m) above sea-level]. The Reduced Oxygen Breathing Device was used to provide normobaric simulated altitude exposures by changing the breathing air mix administered via an aviator oronasal mask and helmet.[12] Using these data, we propose a real-time processing of the scores from the MTB through a hypoxia detection architecture consisting of multiple detectors associated with the different modules in the test battery. The individual hypoxia/no-hypoxia decisions of these detectors are integrated through a decision fusion algorithm to synthesize a global estimate of the hypoxic state of the monitored individual. Both local detectors and the fusion algorithm use a Neyman-Pearson criterion for decision making. They fix the upper bound on the false alarm rate (probability that hypoxia is declared when it is absent) and maximize the true detection rate (probability that hypoxia is declared when it is present). The designed system demonstrates significant hypoxia detection power at low false alarm rates. It also provides an assessment of the contribution of each MTB module to the overall performance. For our dataset we observed that the Arithmetic task module provided the largest contribution, followed by the Auditory Monitoring and Memory task modules. The Visual Monitoring module contributed least to the hypoxia detection task.

METHODS

We posed hypoxia detection of a monitored subject as a binary hypothesis problem with hypothesis H_0 = the subject is nonhypoxic and hypothesis H_1 = the subject is hypoxic. Input is provided by N sensors, each attached to a different module in a cognitive assessment battery (in our study $N = 4$). These N inputs are processed independently to synthesize N local decisions u_i ($i = 1, \ldots$), where $u_i = +1$ for accepting hypothesis H_1, and $u_i = -1$ for accepting hypothesis H_0. The fusion center combines these N decisions to a global decision u_0, where $u_0 = +1$ if hypothesis H_1 is accepted, and $u_0 = -1$ if hypothesis H_0 is

accepted. The local and global detector performances are assessed through receiver operating characteristics (ROC) curves (probability of detection vs. probability of false alarm). Both local detectors and the fusion center use the Neyman-Pearson criterion. They set an upper limit to $P_f = P$ (accept H_1/H_0 is true) and maximize $P_d = P$ (accept H_1/H_1 is true). The designed system was tested using experimental data (see following section on data collection).

Subjects

In an experiment conducted in 2008 by NAWCAD,[23] 45 datasets from 26 subjects (4 women and 22 men) were collected. (The study was approved by NAVCAD IRB, protocol NAWCAD.2008.0001, original approval date: 13 March 2008. U.S. Navy IRB protocols comply with SECNAVINST 3900.39D and DoD Directive 3216.02, and Title 45, CFR 46. These protocols include compliance with the Declaration of Helsinki Revision 6, 2008.) All test subjects were physically fit military or civil service personnel between the ages of 18 and 50. Subjects were nonsmokers and successfully underwent the physical testing in accordance with NAWCAD test subject qualification standards, and had those results reviewed by the medical monitor to ensure that they were physically fit to participate. Datasets from 25 of these subjects were available for the current study.

Procedure

The subjects were exposed to varying normobaric altitude profiles using the reduced oxygen breathing device. One profile ranged from 0 to 18,000 ft (0 to 5486 m), and the other from 0 to 25,000 ft (0 to 7620 m; severe hypoxia). The experimental profiles were the following:

Profile 1. Ascend at 1000 ft · s^{-1} (304.8 m · s^{-1}) to 10,000 ft (3048 m); hold for 10 min; ascend to 18,000 ft (5486 m) at the same rate and hold for up to 20 min; and descend at the same rate to ground level.
Profile 2. Ascend at 1000 ft · s^{-1} (304.8 m · s^{-1}) to 10,000 ft (3048 m); hold for 10 min; ascend to 25,000 ft (7620 m) at the same rate and hold for up to 20 min; and descend at the same rate to ground level.

Test scores from the MTB were recorded when subjects were exposed to one or two repetitions of both profiles with at least 48 h between trials. Profile 1 was not used for the analysis presented in this study since none of the subjects developed hypoxic symptoms in this case. Only those subjects that required 100% oxygen administration due to cognitive, muscular, visual, or respiratory symptoms were considered hypoxic. Blood oxygen saturation (S_pO_2) and heart rate were also recorded using a transmittance type pulse oximeter used on the finger (515B, Novametrix Medical Systems Inc., Wallingford, CT), and two reflectance type oximeters used on the forehead (9847, Nonin, Plymouth, MN; and Rad-87, Masimo, Irvine, CA). Early termination criteria of the experiment included subject request, finger S_pO_2 below 60% for 10 s, or nonresponsiveness of the subject.

The MTB used was a version of the Synwork1,[11] a PC-based tool for assessment of performance in a simulated work environment. The test was presented on a screen consisting of four modules as shown in **Fig. 1**. Test takers were instructed to pay equal attention to each module and switched between them, providing an answer in one module before proceeding to another. No specific module order was specified. Results were recorded after a session lasting 20 s and the complete test lasted about 100 sessions (some test takers terminated the test earlier).

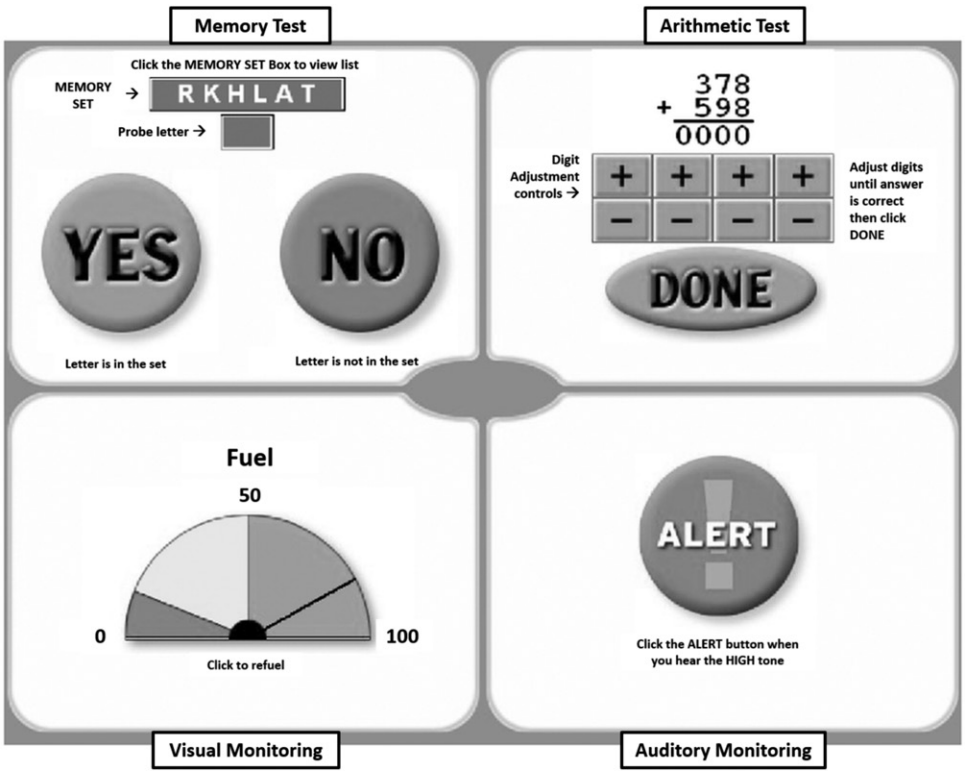

Fig. 1. Synwork1 MTB test screen; images presented clockwise from top left: Memory task, Arithmetic task, Audio Monitoring, Visual Monitoring.

The detailed descriptions for the four modules are given below.

Module 1. Memory (Sternberg) task: the subject is shown a list of six letters for 5 s, and then shown a single letter and asked to answer 'YES' or 'NO' as to whether the letter belongs to the previously shown list. The single letter shown changes once a response is recorded, continuing till the end of trial time (5 s in our case). After this period a new list is shown and the process is repeated. Points are awarded for correct responses

and deducted for each error (as well as for when there is no response).

Module 2. Arithmetic task: a three-column addition task presents two randomly selected integers greater than 99 and less than 1000, with their sum displayed as 0000. The task is to adjust the answer for the sum by clicking on scroll buttons below each character of the answer. Clicking on the DONE button results in points being awarded for correct answers, deducted for errors, and the presentation of a new problem. The task has no set time limits, thus it is completely self-paced.

Module 3. Visual Monitoring: a meter-style pointer moves gradually from '100' to '0'. Clicking the mouse on the meter resets the pointer to 100. The task is to prevent the pointer from reaching 0 while maximizing the distance from 100 at the time of reset. Points are awarded for each reset, with their number being proportional to the distance of the pointer from 100 at the time of reset. Points are deducted for each second the pointer remains at 0.

Module 4. Auditory Monitoring: a sound is periodically played every 5 s. The sound consists of pure tones of two possible frequencies: low (1000 Hz) or high (2000 Hz). One of the sounds is assigned as positive (+) while the other as negative (−). The task is to click on an ALERT button on hearing the positive tone before the next tone is heard. Points are awarded for each correct detection (a click after a positive tone), and deducted for each incorrect detection (a click after a negative tone). If the subject provides no response there is neither an award nor a penalty.

Besides recording test scores, several additional metrics of subject performance were recorded. These include the following:

Reaction time: the time needed for the subject to respond to the memory task, audio monitoring, or visual monitoring cues. No reaction time was recorded for the arithmetic task as it was self-paced.

Dwell time: the time spent on any one of the four modules; the dwell time was calculated by aggregating the total time the cursor was on a specific module on the screen during a given 20-s period.

Dataset labeling: In the NAWCAD study, each subject was exposed to the simulated altitude profiles as described above. **Fig. 2** shows the timeline of the entire experimental run which lasted for 45 min for each subject. Interval A represents the time period at the beginning of the experiment when the subjects were at varying altitudes but did not report any abnormal symptoms. Most subjects reported cognitive, muscular, visual, or respiratory symptoms at some stage during the 25,000-ft (7620-m) altitude run. It is assumed that the transition from a healthy state to a hypoxic state begins at the start of interval B since most subjects reported discomfort at the beginning of interval B. When experimental termination criteria were reached for any subject (i.e., $S_pO_2 < 60\%$ for 10 s, a maximal subjective symptom rating, or nonresponsiveness), 100% oxygen was administered to this subject, which reverted the simulation altitude to the ground level and subjects were followed for a 15-min recovery period. In Fig. 2, the beginning of interval D marks the time when oxygen administration was initiated. The duration of 100% oxygen administration (interval D) differed for subjects. The data points collected in the 1-min period prior to interval D (denoted by interval C) were labeled as hypoxic. Sometimes, hypoxic symptoms may linger on even after 100% oxygen administration is completed; this may be due to the oxygen paradox.[22] Additionally the revival period differs from person to person. Due to ambiguity in the actual state of the person, data points collected in intervals B, D, and E (1-min interval after 100% oxygen was administered) were not included in the analysis. Interval F depicts the time period when the subject had completely revived due to oxygen administration. Points collected in intervals A and F were labeled as nonhypoxic. During the entire experimental run pulse rate was measured by pulse oximeter and heartrate via ECG and behaved as expected—it became elevated during simulated altitude exposure. The higher the altitude, the higher the heart rate. Heart rate recovered along with S_pO_2 upon return to ground level.

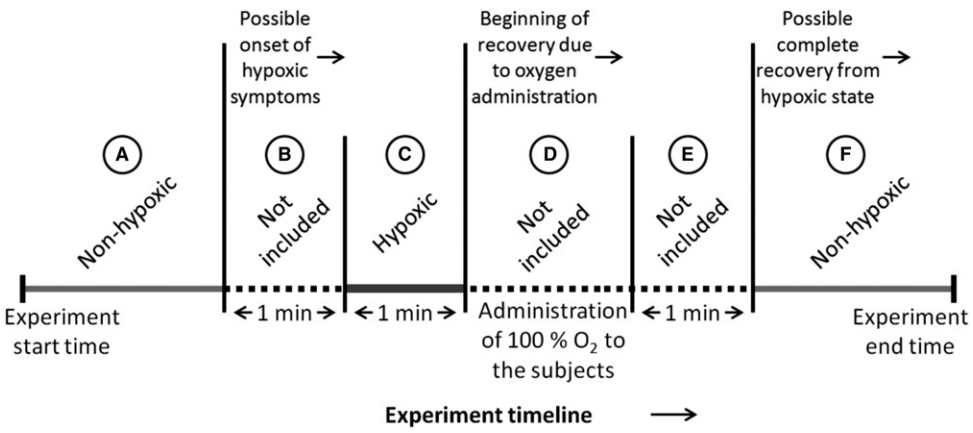

Fig. 2. Experiment timeline indicating the regions used for dataset labeling (time progresses left to right).

Statistical Analysis

Physiological signals such as blood oxygen saturation (S_pO_2) are directly affected by conditions causing hypoxia. As a preliminary step, we analyzed the nature and degree of correlation between cognitive scores and S_pO_2 levels under hypoxic and nonhypoxic conditions. A high correlation value with S_pO_2 would corroborate the usefulness of cognitive test battery scores in detection of hypoxic conditions in an individual.

We used two correlation measures, namely the Pearson correlation coefficient and the Spearman rank correlation coefficient.[4] While the Pearson coefficient assumes a linear relationship between the

variables (in our case the scores and S_pO_2), the Spearman rank coefficient only assumes a monotonic (linear or nonlinear) relationship among the variables. Other correlation measures such as mutual information can also be used. The correlation coefficients between two random variables X and Y are defined as follows:

Pearson correlation coefficient:

$$r_P = \frac{cov(X,Y)}{\sigma_X \sigma_Y}. \quad \text{Eq. 1.}$$

Spearman rank correlation coefficient:

$$r_S = \frac{cov(R_X, R_Y)}{\sigma_{R_X} \sigma_{R_Y}}. \quad \text{Eq. 2.}$$

Here $X = \{x_1, ..., x_M\}$ and $Y = \{y_1, ..., y_M\}$ are sets of M values of two variables, *cov* is the covariance, $\sigma_{(\cdot)}$ is the standard deviation. R_X and R_Y are ranks of the observations X and Y when sorted in ascending order. In this study X represents the S_pO_2 levels and Y represents the cognitive module scores coming from the MTB. The Arithmetic task and Memory task had the highest correlation scores followed by the Auditory Monitoring module. **Fig. 3** shows the sample scores for a randomly chosen subject. The top trace shows the altitude profile. The remaining traces (from second to fifth trace: Memory, Arithmetic, Visual Monitoring, Auditory Monitoring) show the raw MTB assessment scores, s_1–s_4 (triangles), and the corresponding 5-point moving averaged signals, z_1–z_4 (solid lines). Here the abscissa for all five traces show sampling instance, being sampled every 5 s.

To further investigate the usefulness of cognitive scores for hypoxia detection, we developed the distributed decision fusion architecture shown in **Fig. 4** that can use cognitive scores as inputs and generate "yes/no" decisions on hypoxia in real-time. The system consists of the four modules of the MTB, each providing subject scores from the corresponding tasks (s_1–s_4). The scores are smoothed using a 5-point moving average filter to create the observations z_1–z_4. The processed observations z_i, $i = 1, \ldots, 4$, are forwarded to the associated local detectors for decision making. Each local detector computes a binary decision u_i, $i = 1, \ldots, 4$, supporting H_1 ($u_i = +1$) or H_0 ($u_i = -1$). The local detector decisions u_i, $i = 1, \ldots, 4$, are sent to a fusion center, where they are combined using a likelihood ratio test, to generate a global decision $u_0 = +1$ supporting hypothesis H_1 (hypoxic) or $u_0 = -1$ supporting hypothesis H_0 (nonhypoxic).

Both the local detectors and the fusion center use the Neyman-Pearson criterion[27] for decision making that fixes the false alarm rate (probability of accepting H_1 given H_0 is true) at a prespecified level, $0 < \alpha < 1$ and then attempts to achieve the maximum probability of detection (probability of accepting H_1 given H_1 is true). The local detector false alarm and true detection rates are used by the Decision Fusion Center (DFC) for generating global decisions. A detailed mathematical exposition of the distributed detection system is available as **Appendix A** online; https://doi.org/10.3357/amhp.5040.2019.

Local decision rule. The observations z_i, $i = 1, ..., N$ (in our study $N = 4$, see Fig. 4) received by the local detectors are the moving averaged cognitive scores. These observations are assumed to have continuous probability distributions conditioned on the hypothesis. Each local detector fixes the false alarm rate at a specific level (α_i, $i = 1, ..., N$) and uses the Neyman-Pearson test to compute the local true detection rates.

Global decision rule. The observations for the fusion center are the local detector decisions u_i, $i = 1, ..., N$. Once the global false alarm is fixed to a specified level (denoted by α_0), a randomized Neyman-Pearson rule (for more details see Van Trees,[27] p. 43) is implemented to compute the global (DFC) detection rate.

Cognitive score density estimates. We estimated the distributions of the cognitive scores z_i using the technique of kernel density estimation[19] with a Gaussian kernel. Kernel density estimation is a nonparametric technique that uses the available data points to construct a smooth probability density estimate of the form:

$$f(x;h) = \frac{1}{M}\sum_{i=1}^{M} K\left(\frac{x - x_i}{h}\right). \quad \text{Eq. 3.}$$

Here the chosen kernel is parameterized by h, often referred to as the kernel bandwidth; M is the number of independent and identically distributed example data points each represented as x_i; $K(\cdot)$ is the kernel. In this study we chose the Gaussian kernel,

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}. \quad \text{Eq. 4.}$$

For each local detector, the corresponding collective set of data points over all subjects (25 in total) was used to estimate the densities. The kernel bandwidth was chosen based on the algorithm described in Botev et al.[7] (Algorithm 1, page 2932). In **Fig. 5**, the kernel probability density function estimates under each hypothesis are shown superimposed on the histograms (probability mass function) of each component score. The continuous density estimates were used in the local decision-making process. Sweeping the local false alarms over the interval [0, 1] provides the local ROC curves shown in **Fig. 6**. Based on inspection of the ROC curves, the detectors associated with the arithmetic task and memory task modules seem to perform the best (the more the ROC curve is toward the top left corner for the same false alarm rate, the higher the detection power), followed by the detector associated with the audio monitoring module. The visual monitoring module has the lowest detection power of the four modules.

RESULTS

Correlation Analysis

Since physiological measurements like S_pO_2 are useful in detecting the onset of hypoxia, we investigated the relationship between the cognitive score components and S_pO_2 measurements through two correlation measures. These are the Pearson

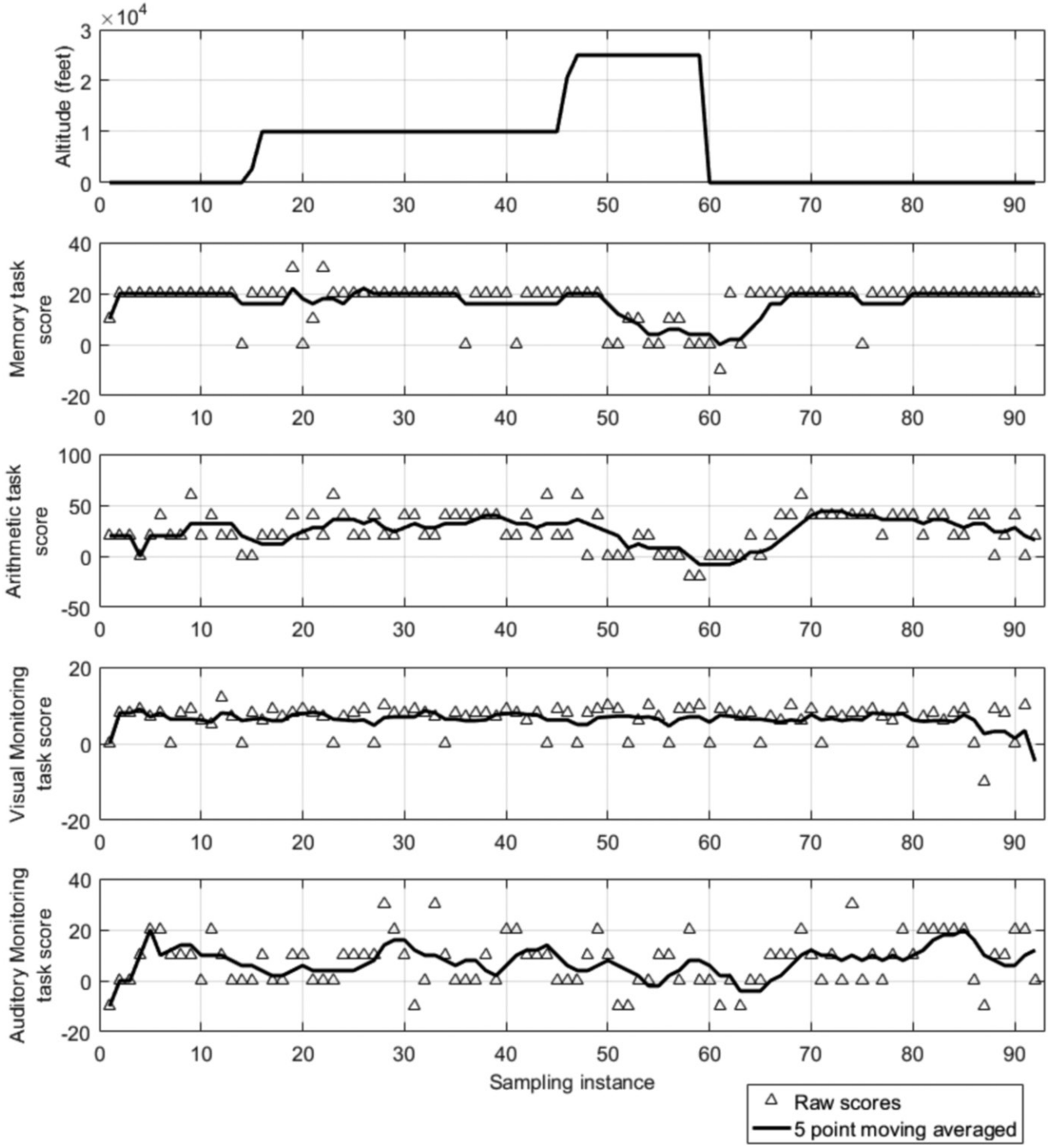

Fig. 3. Simulated altitude profile, raw, and moving averaged MTB assessment scores.

correlation coefficient and the Spearman rank correlation coefficient.[4] The mean (computed over 25 subjects) Pearson correlation coefficient between S_pO_2 levels and the Arithmetic task, AM task, Memory task, and VM task were 0.437, 0.256, 0.246, and 0.139, respectively. The mean Spearman rank correlation coefficients in the same order were 0.121, 0.094, 0.091, and 0.044. Both correlation measures highlight the order of correlation of the cognitive scores with S_pO_2 as Arithmetic task $\geq$ AM $\sim$ Memory task $>$ VM ($\sim$ refers to similar). Existence of correlation with S_pO_2 underlines the usefulness of cognitive scores for hypoxia detection.

Fusion of Cognitive Scores

We studied the detection performance of the system shown in Fig. 3 through the ROC curves. A detector A is said to be more powerful than another detector B, if $P_D^A\left(P_f\right) \geq P_D^B\left(P_f\right)$, where

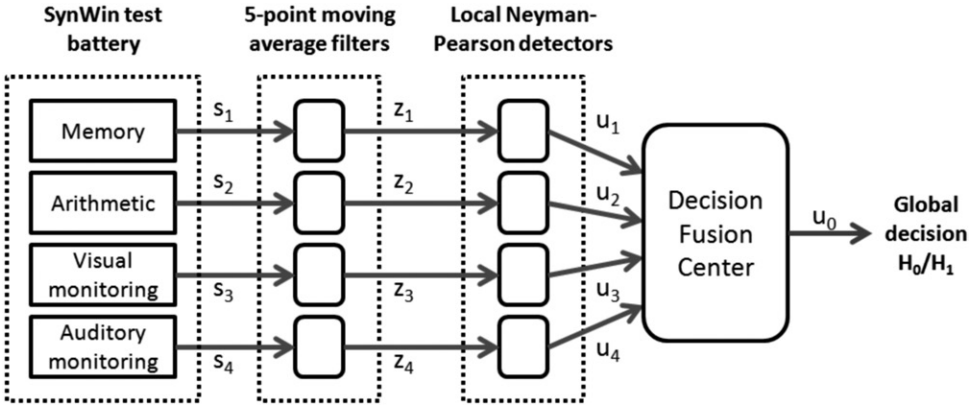

Fig. 4. Binary decision fusion setup.

$P_D^A(\cdot)$ and $P_D^B(\cdot)$ are the detection rates of detectors A and B, respectively, as functions of an arbitrary false alarm rate of $P_f \in [0,1]$.

Data from all 25 subjects were compiled together to model the local observation densities under each hypothesis. The local detector false alarm rates α_i, $i = 1, ..., N$ ($N = 4$ in our study) were fixed. The corresponding local true detection rates were computed from the local ROC curves (see Fig. 6). The local detector operating points (false alarm and true detection pair) were used to compute the global false alarm and true detection rates (see Thomopolous et al.,[26] section II-B, p.648). Varying the global false alarm rate over the interval [0, 1] generated the global ROC curves.

In **Fig. 7** we show the fusion center ROC curve (diamond markers) for two cases; the local false alarm rates α_i, $i = 1, ..., N$, were fixed at 0.1 (case 1, as shown in Fig. 7A) and 0.05 (case 2, as shown in Fig. 7B). The local detector detection rates obtained from local ROC curves (see Fig. 6) for case 1 were P_{D1} (Memory task) = 0.706, P_{D2} (Arithmetic task) = 0.732, P_{D3} (VM) = 0.383, P_{D4} (AM) = 0.514. The local detector detection rates for case 2 were P_{D1} (Memory task) = 0.644, P_{D2} (Arithmetic task) = 0.643, P_{D3} (VM) = 0.352, P_{D4} (AM) = 0.466. Comparing Figs. 7A and 7B, a certain deterioration in the fusion center performance for case 1 is observed compared to that in case 2, especially in the low false alarm region ($\alpha_0 < 0.05$). This performance degradation can be attributed to the fact that in case 2 local detector performances were superior (in terms of probability of error).

By removing a sensor (corresponding to a module) one at a time, we analyze the individual contributions of each sensor. The global performance is presented in **Table I** for case 1 and **Table II** for case 2.

Inspection of Table I and Table II shows that the Arithmetic task score makes the largest contribution toward accurate

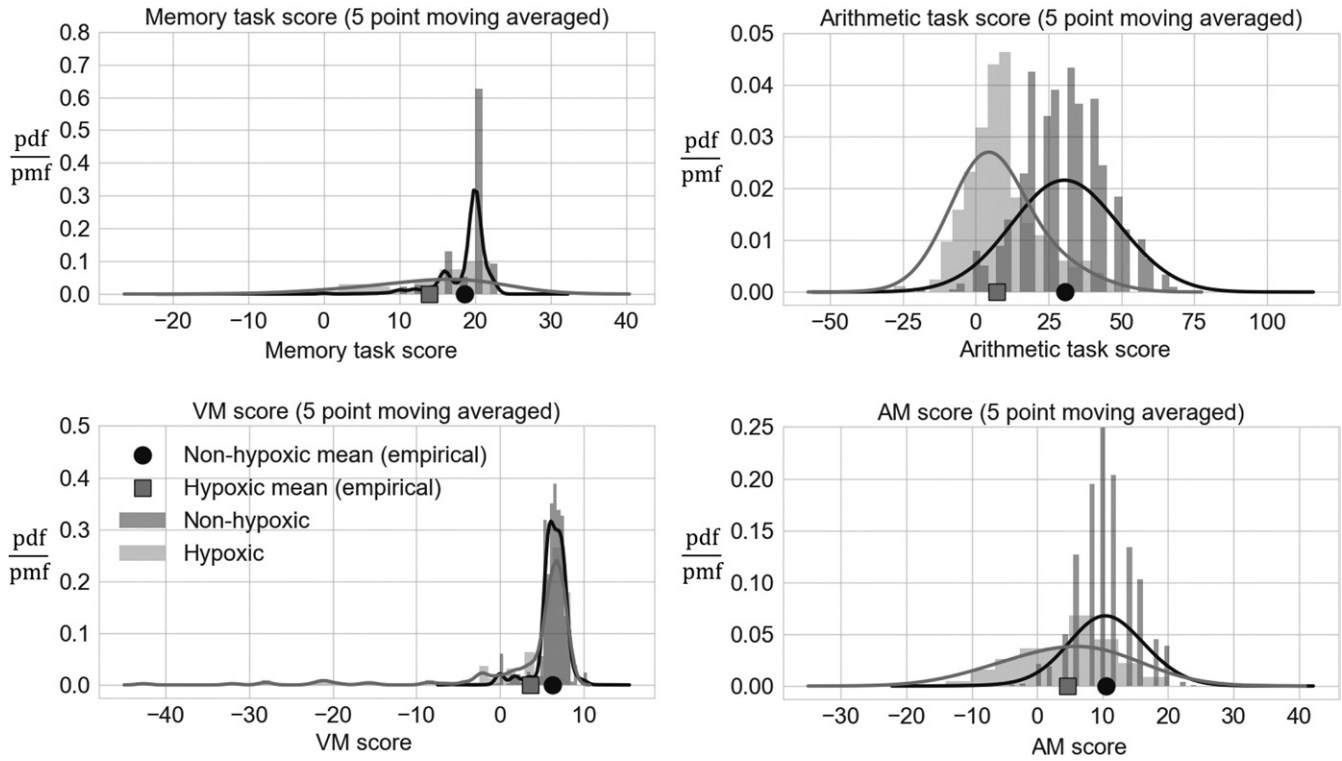

Fig. 5. Empirical distributions and kernel density estimates for cognitive scores. Clockwise from top left: Memory task score, Arithmetic task score, Audio Monitoring (AM) score, Visual Monitoring (VM) score.

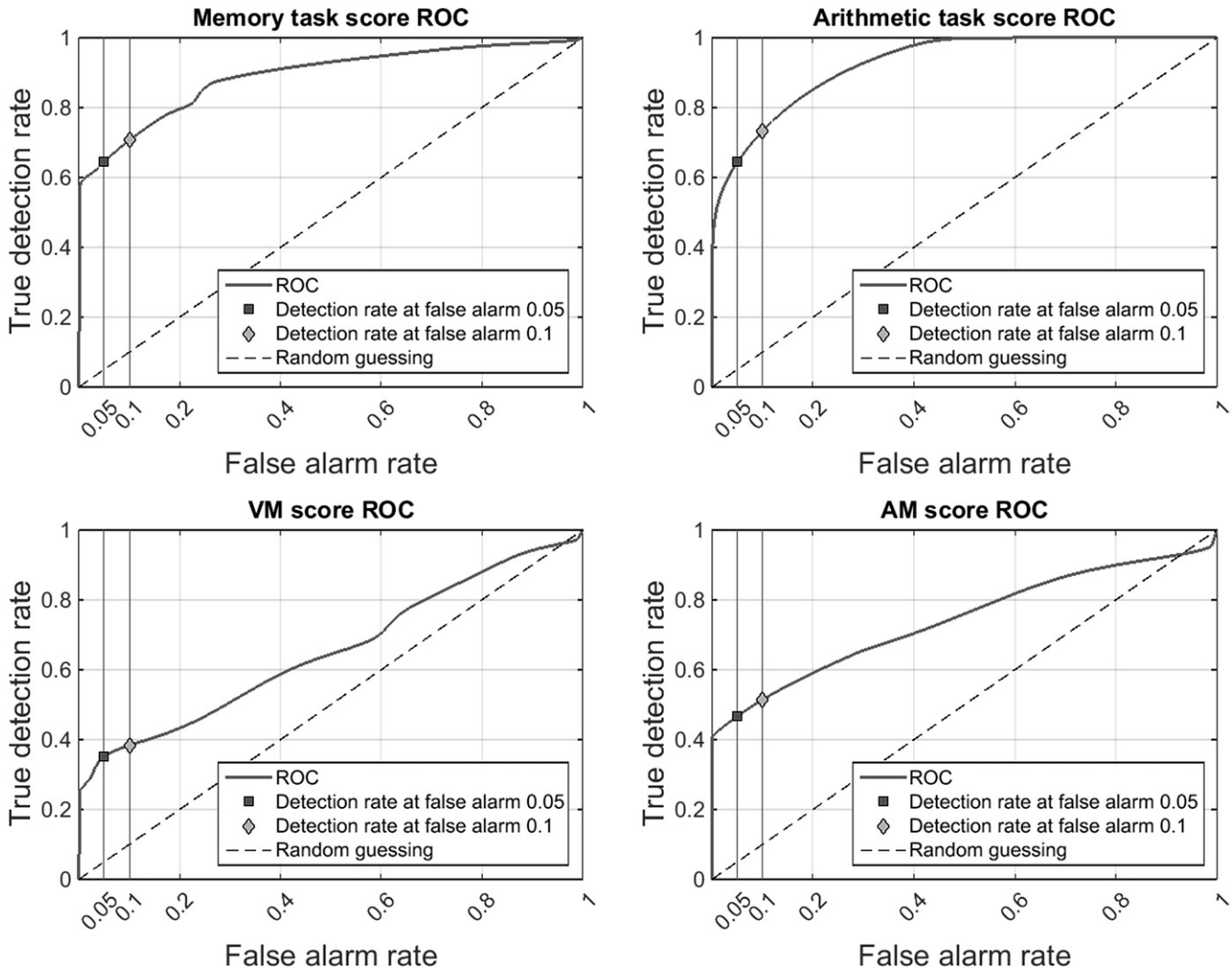

Fig. 6. ROC curves for local detectors. Clockwise from top left: Memory task score, Arithmetic task score, Audio Monitoring (AM) score, Visual Monitoring (VM) score.

hypoxia detection. This is because in case 1, removal of the corresponding module from the fusion architecture resulted in approximately 13.5% and 24.9% drops in the overall detection accuracy compared to when all four modules were used for fusion under global false alarms of 0.1 and 0.05, respectively. For the same global false alarms, for case 2, on removal of the Arithmetic task module, the accuracy dropped by approximately 8.9% and 13.9%, respectively, compared to

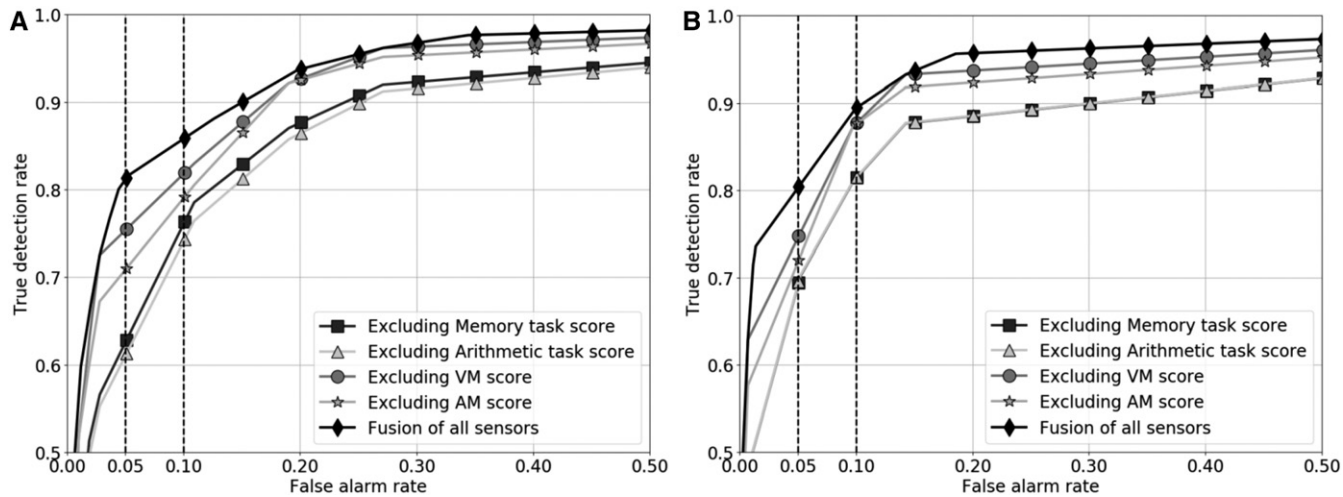

Fig. 7. Global ROC curves with local detector false alarm at A) 0.1 and B) 0.05.

Table I. Global Detection Rates with Sensor Modules Removed One at a Time for Case 1.

	GLOBAL FALSE ALARM = 0.1		GLOBAL FALSE ALARM = 0.05	
MODULE REMOVED	DETECTION RATE	% CHANGE	DETECTION RATE	% CHANGE
None	0.857	-	0.811	-
Memory	0.761	11.2	0.625	22.9
Arithmetic	0.741	13.5	0.609	24.9
VM	0.817	4.6	0.753	7.1
AM	0.790	7.8	0.708	12.7

when all four modules were used in fusion. A similar result is observed in Figs. 7A and 7B, where the ROC curve with the Arithmetic module removed lies at the bottom (furthest away from the top left corner). The performance degradation at the fusion center due to the removal of the Memory module has a very similar effect as can be visualized in Figs. 7A and 7B, where the ROC curve with the Memory module removed lies close to that obtained by removing the Arithmetic module. The Visual Monitoring module made the least contribution toward hypoxia detection since removal of this module resulted in the lowest shift of the fusion center ROC curve (diamond markers) in the bottom right direction (Figs. 7A and 7B).

The higher contribution of the Arithmetic and Memory task modules are in agreement with the trend observed in the correlation analysis. The highest correlation of the Arithmetic component with S_pO_2 is consistent with the highest contribution of this sensor toward hypoxia detection.

Leave-One-Out Cross Validation

To investigate the generalizability of the fusion architecture, we performed a leave-one-out analysis. The process involves leaving out each subject's data and using the data from the remaining subject pool to design the detection system. The leave-one-out process was repeated 25 times (each time leaving out the data from 1 of the 25 subjects). The detection performance was examined for three cases; namely global false alarm α_0 fixed at 0.05, 0.1, and 0.2.

The global detection rates were 0.81 ± 0.011, 0.86 ± 0.009, and 0.938 ± 0.005, respectively. For all cases, the local false alarms were fixed at $\alpha_i = 0.1$, $i = 1, \ldots, N$. The corresponding local detection rates were P_{D1} (Memory task) $= 0.707 \pm 0.013$, P_{D1} (Arithmetic task) $= 0.732 \pm 0.014$, P_{D2} (VM) $= 0.391 \pm 0.022$, P_{D4} (AM) $= 0.513 \pm 0.011$. Small standard errors show that the detection performance was not affected considerably by changing the subject pool and therefore the system can have generic applicability.

Table II. Global Detection Rates with Sensor Modules Removed One at a Time for Case 2.

	GLOBAL FALSE ALARM = 0.1		GLOBAL FALSE ALARM = 0.05	
MODULE REMOVED	DETECTION RATE	% CHANGE	DETECTION RATE	% CHANGE
None	0.894	-	0.802	-
Memory	0.813	9.06	0.690	13.9
Arithmetic	0.814	8.9	0.690	13.9
VM	0.876	2.01	0.744	7.2
AM	0.875	2.1	0.716	10.7

Incorporation of Reaction and Dwell Times

In the experimental data collection process, reaction time was recorded for the memory and auditory tasks. The arithmetic task recorded response time. For all four modules (Memory task, Arithmetic task, Visual Monitoring, Auditory Monitoring) dwell times were recorded. Reaction/response time was the interval time between the presentation of the stimulus and the submission of response from the test taker. Dwell time measured the time interval for which the mouse pointer was within any one of the four task module quadrants.

Fig. 8 shows the fusion center's ROC curves when the reaction times and dwell times were incorporated in the fusion framework. The trace with diamond markers shows the fusion center's ROC when only the reaction and dwell times were fused. The trace with circle markers is the ROC obtained by combining only the cognitive task scores. We observe that by themselves the reaction and dwell time measurements do not have significant detection power. However, when combined with the task score measurements, the fusion center's ROC (square markers) achieves some improvement, especially in the 0.05 to 0.2 false alarm range. For example, when we combine the four task scores (Memory task, Arithmetic task, Visual Monitoring, Auditory Monitoring) then for a false alarm rate of 0.1 we get a detection rate of approximately 0.86. When we add to the four task scores the seven reaction/dwell times, then for the same false alarm rate of 0.1 the detection rate jumps to 0.91.

DISCUSSION

We studied the usefulness of cognitive scores from the MTB (Memory, Arithmetic, Visual Monitoring, and Auditory Monitoring tasks) in detecting hypoxia. The cognitive scores were shown to be correlated with S_pO_2 measurements based on both Pearson and Spearman correlation coefficients. Since S_pO_2 is a good indicator of hypoxia, the correlation analysis supports the idea of using cognitive state estimates for hypoxia detection. Furthermore, in this particular battery of tests, correlation coefficients showed that the Arithmetic task scores have the highest correlation with S_pO_2 levels as compared to the other cognitive component scores.

We further used cognitive assessment scores from multiple modules as real-time inputs to a distributed detection system for hypoxia detection. Local detectors assigned to each assessment module process the scores and generate "yes/no" decisions on hypoxia. A fusion center

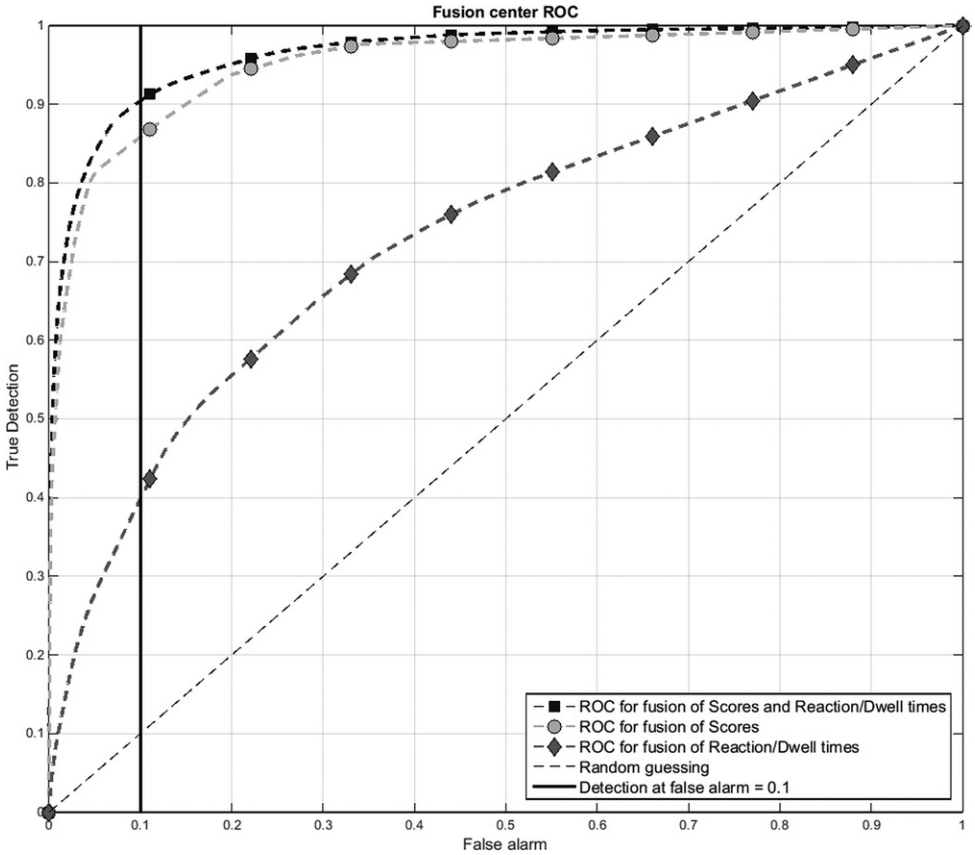

Fig. 8. Effect of incorporation of reaction and dwell times in the global fusion performance.

combines local decisions to generate a global decision. The ROC curves show that there is improvement in detection performance of the fusion center compared to the local detectors working independently with any of the component assessment scores.

We analyzed the degree of contribution of each component score of the MTB toward the global decision-making process using the ROC curves. The Arithmetic task module seemed to contribute the most. The Memory task and AM modules followed next with similar contributions. The VM module had relatively the least contribution. In general this approach of using ROC curves for sensor analysis can help in creating a hierarchy of sensor importance such that for a known tolerable global false alarm, the appropriate groups of sensors could be identified and used (and some sensors of only marginal contribution can be dropped). Nevertheless, scores from all four modules had reasonable detection power (see local ROC in Fig. 6). A system that combines physiological sensor observations such as oximeter readings and heart rate with cognitive scores might increase hypoxia detection performance as compared to a system dependent on physiological sensors or cognitive scores exclusively.

There is a growing interest in doing cognitive tests and recent studies have used such tests to assess variations in human physiological state under various conditions. For example, Brebeck et al.[8] and Germonpre et al.[13] use cognitive tests to study the effects of narcosis in divers breathing different gas mixtures. However, in general, real-time data collection from such MTB cognitive tests is not easy because of the requirement of active subject participation. Passive and fast test interfaces are required, such as the cognitive assessment based on the critical flicker fusion frequency test[15] and the use of eye metrics to assess the cognitive state of the subject.[18] An alternative would be the use of physiological measurements with high correlation to cognitive functioning such as electroencephalography[6] (EEG) and functional near-infrared spectroscopy[16] (fNIR) to estimate cognitive state. Currently, many studies attempt to undertake cognitive tests in a passive way. Yet there is no feasibility study that justifies such efforts to do passive tests. The current study indeed shows that if passive cognitive test scores are collected and available, they are informative enough to be used for hypoxia detection under certain circumstances. Based on the findings about the contributions of each module toward hypoxia detection, newer passive test batteries focusing on analytical, auditory, and/or memory-related activities can be designed.

The study presented here considers short duration unacclimatized exposure to altitude. Hence, remodeling (density estimation and threshold computation) might be needed to address hypoxia detection under long duration exposures. Furthermore, the system was tested on data collected under normobaric conditions. Normobaric datasets are likely to be different from data collected in an altitude chamber (hypobaric conditions), where the variation of atmospheric pressure is taken into account as well. Recent studies[5,10] have shown that hypobaria can have effects on the onset and severity of hypoxia. Even though the modeling and decision making framework remain identical to that presented in this paper, the detection results may vary if a hypobaric data set is used.

ACKNOWLEDGMENTS

This work was partially sponsored by Office of Naval Research (ONR) under Grant No. N00014-13-1-0733.

Authors and affiliations: Arjun Rajasekar, B.Tech., M.S., Sayandeep Acharya, M.S., Ph.D., and Leonid Hrebien, M.S., Ph.D., Department of Electrical and

Computer Engineering, and Chris Rorres, M.S., Ph.D., Department of Mathematics, Drexel University, Philadelphia, PA, USA; Barry S. Shender, MSBME, Ph.D., Human Systems Department, Naval Air Warfare Center Aircraft Division, Patuxent River, MD, USA; and Moshe Kam, M.S., Ph.D., Department of Electrical and Computer Engineering, Newark College of Engineering, New Jersey Institute of Technology, Newark, NJ, USA.

REFERENCES

1. Abraini JH, Bouquet C, Joulia F, Nicolas M, Kriem B. Cognitive performance during a simulated climb of Mount Everest: implications for brain function and central adaptive processes under chronic hypoxic stress. Pflugers Arch. 1998; 436(4):553–559.
2. Acharya S, Rajasekar A, Shender BS, Hrebien L, Kam M. Real-time hypoxia prediction using decision fusion. IEEE J Biomed Health Inform. 2017; 21(3):696–707.
3. Adam GE, Fulco CS, Muza SR. Multi-task performance at sea-level and high altitude. Natick (MA): Army Research Institute of Environmental Medicine; 2008.
4. Bailey DE. Probability and statistics: models for research. Hoboken (NJ): John Wiley & Sons; 1971.
5. Beidleman BA, Fulco CS, Staab JE, Andrew SP, Muza SR. Cycling performance decrement is greater in hypobaric versus normobaric hypoxia. Extrem Physiol Med. 2014; 3(1):8.
6. Berka C, Levendowski DJ, Lumicao MN, Yau A, Davis G, et al. EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. Aviat Space Environ Med. 2007; 78(5, Suppl.):B231–B244.
7. Botev ZI, Grotowski JF, Kroese DP. Kernel density estimation via diffusion. Ann Stat. 2010; 38(5):2916–2957.
8. Brebeck AK, Deussen A, Schmitz-Peiffer H, Range U, Balestra C, et al. Effects of oxygen-enriched air on cognitive performance during SCUBA-diving–an open-water study. Res Sports Med. 2017; 25(3):345–356.
9. de Aquino Lemos V, Antunes HK, dos Santos RV, Lira FS, Tufik S, de Mello MT. High altitude exposure impairs sleep patterns, mood, and cognitive functions. Psychophysiology. 2012; 49(9):1298–1306.
10. DiPasquale DM, Strangman GE, Harris NS, Muza SR. Hypoxia, hypobaria, and exercise duration affect acute mountain sickness. Aerosp Med Hum Perform. 2015; 86(7):614–619.
11. Elsmore TF. SYNWORK1: a PC-based tool for assessment of performance in a simulated work environment. Behav Res Methods. 1994; 26(4): 421–426.
12. Environics. Reduced oxygen breathing device 2 (ROBD2) [Internet]. c2014 [Accessed 2017 March 18]. Available from: http://www.environics.com/product/reduced-oxygen-breathing-device-robd-0.
13. Germonpré P, Balestra C, Hemelryck W, Buzzacott P, Lafère P. Objective vs. subjective evaluation of cognitive performance during 0.4-MPa dives breathing air or nitrox. Aerosp Med Hum Perform. 2017; 88(5):469–475.
14. Guyton AC. Textbook of medical physiology, 4th ed. Philadelphia (PA): WB Saunders Co; 1971:512–513.
15. Hemelryck W, Rozloznik M, Germonpré P, Balestra C, Lafère P. Functional comparison between critical flicker fusion frequency and simple cognitive tests in subjects breathing air or oxygen in normobaria. Diving Hyperb Med. 2013; 43(3):138–142.
16. Izzetoglu K, Bunce S, Onaral B, Pourrezaei K, Chance B. Functional optical brain imaging using near-infrared during cognitive tasks. Int J Hum Comput Interact. 2004; 17(2):211–227.
17. Kelly ME, Pettit DR. inventors. Method and apparatus for monitoring oxygen partial pressure in air masks. U.S. Government as Represented by Administrator of NASA, assignee. U.S. patent 7,040,319B1. 2006 May 9.
18. Marshall SP. Identifying cognitive state from eye metrics. Aviat Space Environ Med. 2007; 78(5, Suppl.):B165–B175.
19. Parzen E. On estimation of a probability density function and mode. Ann Math Stat. 1962; 33(3):1065–1076.
20. Pribil MM, Laptev GU, Karyakina EE, Karyakin AA. Noninvasive hypoxia monitor based on gene-free engineering of lactate oxidase for analysis of undiluted sweat. Anal Chem. 2014; 86(11):5215–5219.
21. Richardson JW. inventor. Richarson, Joseph W., assignee. Aviation hypoxia monitor. U.S. patent 5,372,134. 1994 Dec. 13.
22. Sharp GR, Ernsting J. Hypoxia and hyperventilation. In: Dhenin G, editor. Aviation medicine, vol. 1. London: Tri-Med Books Ltd; 1978.
23. Shender B, Mattingly C, Warren M, Coleman S, Askew G, Tucker A. Relating the time complex cognitive performance degrades to physiologic response during moderate and severe normobaric hypoxia. [Abstract]. Aviat Space Environ Med. 2013; 84(4):340.
24. Simmons RG, Chandler JF, Horning DS. Forehead-mounted reflectance oximetry for in-cockpit hypoxia early detection and warning. Aviat Space Environ Med. 2012; 83(11):1067–1076.
25. Spüler M, Walter C, Rosenstiel W, Gerjets P, Moeller K, Klein E. EEG-based prediction of cognitive workload induced by arithmetic: a step towards online adaptation in numerical learning. ZDM. 2016; 48(3):267–278.
26. Thomopoulos SC, Viswanathan R, Bougoulias DC. Optimal decision fusion in multiple sensor systems. IEEE Trans Aerosp Electron Syst. 1987; AES-23(5):644–653.
27. Van Trees HL. Detection, estimation, and modulation theory. Part I. Detection, estimation, and linear modulation theory. Section 2.2, Simple binary hypothesis tests. Hoboken (NJ): John Wiley & Sons; 1968:23–34.
28. Wang S, Gwizdka J, Chaovalitwongse WA. Using wireless EEG signals to assess memory workload in the n-back task. IEEE Trans Hum Mach Syst. 2016; 46(3):424–435.
29. Whitley PE, Shender BS. Use of a neurological model to predict cognitive deficit during normobaric and hypobaric hypoxia. [Abstract]. Aerosp Med Hum Perform. 2016; 87(3):289.
30. Wong J. Software: SynWin version 1.2. Ergon Des. 2016; 13(4):30–32.
31. Zander TO, Kothe C. Towards passive brain-computer interfaces: applying brain-computer interface technology to human-machine systems in general. J Neural Eng. 2011; 8(2):025005.

APPENDIX A. MATHEMATICAL FORMULATIONS FOR THE DISTRIBUTED DETECTION SYSTEM

We provide a detailed discussion of the mathematical formulations for the various decision-making components in the distributed detection system shown in Fig. 4 in the article.

Neyman-Pearson Test

The Neyman-Pearson test[27] fixes the false alarm rate [P (accept H_1 given H_0 is true)] at a prespecified level, $0 < \alpha < 1$ and then attempts to achieve the maximum probability of detection [P (accept H_1 given H_1 is true)]. Let the observations at a Neyman-Pearson detector be denoted by y. We solve for the threshold t^* in:

$$\int_{t^*}^{\infty} P(\Lambda(y)|H_0)d\Lambda = \alpha, \qquad \text{Eq. A.}$$

and make a decision based on a likelihood ratio test of the form:

$$\Lambda(y) = \frac{P(y|H_1)}{P(y|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} t^*. \qquad \text{Eq. B.}$$

The threshold t^* computed from Eq. A determines the corresponding optimal detection rate q^*:

$$\int_{t^*}^{\infty} P(\Lambda(y)|H_1)d\Lambda = q^*. \qquad \text{Eq. C.}$$

Local Decision Rule

The observations z_i, $i = 1, ..., N$ (in our study $N = 4$, see Fig. 4 in the article), received by the local detectors are the moving averaged cognitive scores. These observations are assumed to have continuous probability distributions conditioned on the hypothesis. In a following section, we describe how the continuous distributions $P(z_i|H_j), i = 1,..,N, j = 0,1$ were estimated. Each local detector fixes the false alarm rate at a specific level $(\alpha_i, i = 1,\ldots,N)$ and solves for the corresponding threshold t_i^* such that:

$$\int_{t^*}^{\infty} P(\Lambda(z_i)|H_0)d\Lambda = \alpha_i. \qquad \text{Eq. D.}$$

The local detectors perform a likelihood ratio test to accept one of the hypotheses, namely:

$$\Lambda(z_i) = \frac{P(z_i|H_1)}{P(z_i|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} t_i^*. \qquad \text{Eq. E.}$$

The corresponding true-detection rate of the i^{th} local detector is:

$$P_{D_i} = P(u_i = 1|H_1) = \int_{t^*}^{\infty} P(\Lambda(z_i)\,|\,H_1)d\Lambda. \qquad \text{Eq. F.}$$

Global Decision Rule

The observations for the fusion center are the local detector decisions u_i, $i = 1, ..., N$. Once the global false alarm is fixed to a specified level (denoted by α_0), the likelihood ratio at the DFC takes the form:

$$\Lambda(u) = \frac{P(u_1,\ldots,u_N|H_1)}{P(u_1,\ldots,u_N|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} t_0^*. \qquad \text{Eq. G.}$$

The threshold t_0^* is computed such that the global false alarm is equal to α_0. Assuming that the local decisions are independent conditioned on the hypotheses, we have:

$$\Lambda(u) = \prod_{i=1}^{N} \frac{P(u_i|H_1)}{P(u_i|H_0)} = \prod_{i=1}^{N} \Lambda(u_i) \underset{H_0}{\overset{H_1}{\gtrless}} t_0^*. \qquad \text{Eq. H.}$$

Since the local decisions are Bernoulli random variables, the conditional probability distributions $P(\Lambda(u_i)|H_0)$ and $P(\Lambda(u_i)|H_1)$ for the i^{th} sensor likelihood ratio are discrete. Therefore, to attain all possible values of global false alarm, a randomized Neyman-Pearson rule is implemented (for more details see Acharya et al.,[2] p. 43). Let the set of values of $\Lambda(u)$, at which the probability masses of $P(\Lambda(u)|H_j)$, $j = 0, 1$ are located be denoted by L. The chosen arbitrary α_0 is obtained using a convex combination:

$$\gamma \sum_{\Lambda(u)>t_l} P(\Lambda(u)|H_0) + (1-\gamma) \sum_{\Lambda(u)>t_u} P(\Lambda(u)|H_0) = \alpha_0. \qquad \text{Eq. I.}$$

Here $\gamma \in [0, 1]$ is the convex combination parameter. The threshold $t_l \in L$ is computed such that it is the largest value of $\Lambda(u)$ satisfying:

$$\sum_{\Lambda(u)>t_l} P(\Lambda(u)|H_0) > \alpha_0. \qquad \text{Eq. J.}$$

The threshold $t_u \in L$ is computed such that it is the smallest value of $\Lambda(u)$ satisfying:

$$\sum_{\Lambda(u)>t_u} P(\Lambda(u)|H_0) < \alpha_0. \qquad \text{Eq. K.}$$

The corresponding global true detection rate is calculated as:

$$\gamma \sum_{\Lambda(u)>t_l} P(\Lambda(u)|H_1) + (1-\gamma) \sum_{\Lambda(u)>t_u} P(\Lambda(u)|H_1). \qquad \text{Eq. L.}$$

The probability mass functions $P(\Lambda(u)|H_j)$, $j = 0, 1$ are completely defined by the local detector operating points (P_{D_i}, P_{F_i}).[26]