Policies to Optimize Work Performance and Thermal Safety in Exercising Humans

Mark Buller[1,2], Eric Sodomka[1], William Tharion[2], Cynthia Clements[2], Reed Hoyt[2], Odest Chadwicke Jenkins[1]

mbuller@cs.brown.edu, sodomka@cs.brown.edu, william.tharion@us.army.mil, cynthia.clements1@us.army.mil, reed.hoyt@us.army.mil, cjenkins@cs.brown.edu

[1]Brown University
Providence RI 02912

[2]US Army Research Institute of Environmental Medicine
Natick MA 01760

Abstract

Emergency workers engaged in strenuous work in hot environments risk overheating and mission failure. We describe a real-time application that would reduce these risks in terms of a real-time thermal-work strain index (SI) estimator; and a Markov Decision Process (MDP) to compute optimal work rate policies. We examined the thermo-physiological responses of 14 experienced U.S. Army Ranger students (26±4 years 1.77±0.04 m; 78.3±7.3 kg) who participated in a strenuous 8 mile time-restricted pass/fail road march conducted under thermally stressful conditions. A thermoregulatory model was used to derive SI state transition probabilities and model the students' observed and policy driven movement rates. We found that policy end-state SI was significantly lower than SI when modeled using the student's own movement rates (3.94±0.88 vs. 5.62±1.20, P<0.001). We also found an inverse relationship between our policy impact and maximum SI (r=0.64 P<0.05). These results suggest that modeling real world missions as an MDP can provide optimal work rate policies that improve thermal safety and allow students to finish in a "fresher" state. Ultimately, SI state estimation and MDP models incorporated into wearable physiological monitoring systems could provide real-time work rate guidance, thus minimizing thermal work-strain while maximizing the likelihood of accomplishing mission tasks.

Introduction

In this paper we describe an emerging real-time decision making application for emergency workers whose jobs often demand long periods of work while wearing personal protective equipment in challenging environments. Examples of teams who could benefit from this include the National Guard Civil Support Teams (CST) who respond to incidents involving chemical and biological risks (e.g. figure 1), wild land firefighters, or military personnel.

Figure 1: National Guard CST Members Conducting a Casualty Evacuation Exercise While Wearing Chemical Protection.

We address the problem that hot environments combined with heavy workloads and cumbersome protective equipment pose both an acute risk of heat strain/stroke (Sawka and Young 2006; Bouchama and Knochel 2002); and a longer term degrading effect upon performance (Cheuvront et al. 2010). Current techniques to manage the long term thermal safety and performance of these teams are based upon work-rest tables (OSHA 1985, TBMED 507, 2003). These tables prescribe alternating periods of work and rest, with the duration of each phase based upon the environmental conditions, proposed work rates, and protective equipment being worn. However, this approach does not take into account the actual state of the human. Real time physiological monitoring has been proposed as a way to manage the risk of acute thermal injury (Bernard and Kenny 1994), but has often lacked a well validated means to assess thermal work strain.

Physiological monitors that track activity and/or heart rate are being used with some virtual training applications (Chi-Wai et al. 2011). Applications can provide guidance to improve training based on exercise at the right heart rate intensity, or activity profiles that over time will meet Center for Disease Control (CDC) guidlines. A more advanced system proposed by Lopez-Matencio et al. (2010) uses a k-nearest neighbors approach to advise runners in real time which training track to take. Track advice is based upon their current heart rate, desired training heart rate, track footing, and ambient temperature. However, these applications focus on open ended goals versus the time, safety and performance constraints placed upon emergency workers.

Our application develops a real-time non-invasive thermal work strain state estimator for use with physiological monitors and combines this with a Markov decision process (MDP) model of the goals and constraints of a mission. Using this framework the intent of our application is to optimize the accomplishment of mission goals, individual safety, and long term team performance potential. To realize this system, both health state and policy estimation research questions need to be answered.

Health State Estimation

A simple index of thermal-work strain (Moran et al. 1998), can be calculated from heart rate (HR) and internal temperature (IT). This thermal-work strain index (SI) is a weighted combination of HR and IT (see eq. 1) that ranges from 0 (no strain) to 10 (very high strain). An SI of 10 is achieved at a HR of 180 beats/min and an IT of 39.5 °C (103.1 °F) corresponding to a level of thermal-work strain associated with a 50% likelihood of becoming a heat casualty (Sawka and Young 2006).

$$\begin{aligned} SI = {} & 5(IT - IT_{rest})(39.5 - IT_{rest})^{-1} \\ & +5(HR - HR_{rest})(180 - HR_{rest})^{-1} \end{aligned} \tag{1}$$

Typically, routine estimation of SI in field environments has been hampered by the need to invasively measure IT using rectal or esophageal probes or thermometer pills. Fortunately, recent work suggests IT can be estimated non-invasively from time-series HR data (Buller et al. 2010), simplifying ambulatory assessment of SI in free-living individuals.

From classic exercise and thermoregulatory physiology, we know HR reflects both metabolic heat production and the amount of blood shunted to the skin to dissipate excess metabolic heat. Recognizing that HR is a "noisy" reflection of IT, we used a simple Kalman filter (KF) to infer IT, and thus SI. The result was IT estimates superior to other current state-of-the art methods. In addition, recent analysis of 12 laboratory and field studies (unpublished) found that the KF estimates had a bias of −0.03 ± 0.32 °C indicating that ~95% of all estimates fell within ± 0.63 °C of the observed IT. Using direct measures of HR and inferred IT we have been able to estimate and display accurate estimates of SI in real-time during field exercises (e.g. see Android tablet display in figure 1).

Policy Estimation

We focused on two questions: 1) is it possible to adequately express the tasks, risks and goals surrounding thermal-work strain management in terms of a simple MDP that would provide a realistic policy?, and 2) would any derived policy perform better than the solution provided by the human agents themselves, responding to their own self-perceived physiology? To answer these questions we developed an MDP to model the U.S. Army Ranger Training Brigade course selection road march, and compared our optimal policy and its effect on SI to real world data. The road march had several attractive features that relate to our overall systems goals. First Ranger school has a series of arduous tasks that must be completed for students to graduate. Thus finishing the road march with as low an SI as possible is an advantage, conserving physiological reserves needed for subsequent events. Second as the Ranger students are very fit, motivated, and experienced in extreme conditions, finding a policy that improves upon their performance is non-trivial. Finally, the SI safety constraints we wish to place upon our policy are needed in this training environment, as one student from our data collection was medically withdrawn from the road march due to hyperthermia with an SI of 10.

Experiment

The U.S. Army Ranger students needed to complete an eight-mile road march while carrying 32 kg (70 lbs) within 130 min or be dropped from the Ranger School. The road march was one of a series of demanding tasks scheduled for the week. The road march was conducted at night in temperatures of ~25°C and 85% relative humidity. The required march pace was such that students often needed to run parts of the course. Students applied different strategies to completing the march. Some started quickly and then reduced speed, while others started slowly then increased speed to complete on time, yet others kept a consistent pace. Our experimental goal was to demonstrate: 1) we can learn a policy that allows students to complete the course on-time and avoid hyperthermia; and 2) that the policy also allows students to complete the course with lower final SI scores than without using our policy.

Subjects and Measures

Fourteen male U.S. Army Ranger students who averaged 26±4 years of age; 1.77±0.04 m in height; weighed 78.3±7.3 kg; who carried loads of 31.5±1.1 kg and had

14.4±3.8% percent body fat (mean ± standard deviation (SD)) were used for this analysis.

Direct measures of HR (Equivital I heart rate monitor, Hidalgo Inc. Cambridge UK), and IT (Jonah Ingestible Thermometer Pill, Respironics, Bend OR) were collected in 1 minute intervals. Times to complete each mile of the course were derived from location data collected from GPS units worn by each student (Foretrex 101, Garmin, Olathe, KS). SI scores were computed according to equation (1) using observed IT and HR and resting values of 71 beats/min and 37.1 °C (Moran et al. 1998). Tri-axial accelerometry data were collected at the chest at 25.6 Hz.

Markov Decision Process

An MDP describes an environment where by a set of states (e.g. S:={SI, distance to goal, time}) an agent can assume, and a set of possible actions (A:={movement speeds}).

An in-depth description of an MDP can be found in Russell and Norvig (2010). In our environment the goals and health constraints are described by assigning rewards (R) and penalties (negative rewards) for being in various states at certain times. Our Ranger training road march has a finite horizon. Thus the utility (U) of a sequence of states can be computed from the sum of rewards (R) for being in each state over time (equation 2):

$$U(s_0, s_1, s_2, \dots, s_{n-1}, s_n) = \sum_{t=0}^{n} R(s_t) \qquad (2)$$

The transition from one state to another is determined by the current state, the chosen action and the transition probabilities to the new state (*P(S'|S, A)*). A policy (π) is a mapping from states to actions that prescribes an action to be taken in each state. For any policy we can compute a utility function over states for that policy starting in state *s* as the sum of expected rewards over time (equation 3).

$$U^{\pi}(s) = E\left[\sum_{t=0}^{n} R\,(S_t)\right] \qquad (3)$$

At each time point there will be an optimal policy that for each state will determine the optimal action to be taken which provides the most utility from that point on until the end goal is reached:

$$\pi^{*}(s) = \mathrm{argmax}_{\pi} U^{\pi}\,(s) \qquad (4)$$

With a constrained state space, known transition probabilities, and finite horizon this optimal policy can be computed using dynamic programming.

Ranger Road March MDP Definition

The goals of the actual road march were simple. Complete the 8 mile road march in 130 min or less or be dropped from the course. Given that thermal state (IT) changes relatively slowly, a time interval of 5 min was selected enabling a more thorough search of the transition probabilities state-action space.

State Definition

For states we use SI, in integer units; distance (D), completed in units of 0.0417 miles or the fraction of a mile that can be completed at 0.5 miles per hour (mph) within 5 minutes; and time. Thus: S_t:={{SI},{D}} Where: SI:={1,…,14}, D:={0, 0.0417, … , 8.9583, 9}.

Action Definition

Actions were constrained to just movement speeds from 0 to 7 mph in 0.5 mph increments, thus A:={0, 0.5, 1,…, 6.5, 7.0}. Figure 2 shows the resulting directed acyclic graph that represents our state-action space.

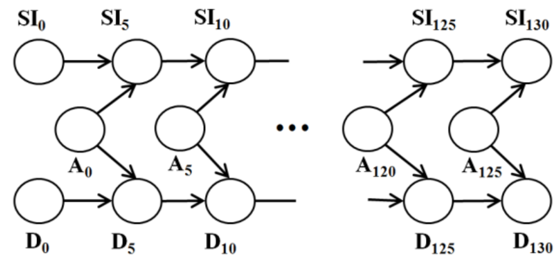

Figure 2: Graph Representation of our State-Action Space.

Reward Definition

Two types of rewards were used in the definition of our MDP, (a) immediate rewards for SI at each time point to model safety limits; and (b) end state rewards for D and SI to model the course completion requirement, and the goal to finish with as low an SI as possible. The end state reward (t=130) for D was represented by a reward of 0 for completing the course on time (D≥8 miles), and a penalty of −1000 (D<8 miles) for not. Immediate and goal reward functions for SI are presented in table 1.

SI	1 to 8	9	10	11	≥12
$R_{t<130}$	0	−100	−500	−2000	−5000
$R_{t=130}$	100−10(SI−1)		−100	−2000	−5000

Table 1: Immediate and End-State Reward Functions for SI.

The end-state SI rewards are designed to allow the students to complete the course with the lowest possible SI. $R(SI)_{t=130}$ shows that finishing with a lower SI is better than finishing with a higher SI. However, finishing with SI's > 9 is not good. The −100 reward for an SI of 10 indicates that it is acceptable to push to complete the course on time. However, an SI > 10 is an unacceptable end state, hence the large negative penalties. The immediate reward function $R(SI)_{t<130}$ shows that it is equally fine for students to have an SI between 0 and 8 during the race but above 8 penalties will accrue. The negative rewards for both SI's of 9 and 10 allow for one or several steps to be taken at these high SI's and still receive higher utility than not completing the race on time. However, the very large negative rewards of SI's > 10 are designed to indicate that stopping the race for health is better than completing.

Transition Probabilities

For distance traveled we placed a small amount of uncertainty (*N(0,1)*) around the distance travelled in 5 minutes for a given movement speed. The transition probabilities are shown in table 2 where d=D+A(5/60).

D'	−.126	−.084	−.042	d	+.042	+.084	+.126
P	.01	.05	.24	.4	.24	.05	.01

Table 2: P(D'|D,A), where d= D+A(5/60).

In general, the SI transition probabilities are complex, and are dependent on a large number of factors such as work rate, personal characteristics (body surface area, fat mass, fitness, acclimation), environmental conditions (ambient temperature, relative humidity, wind speed, solar load), and clothing characteristics (insulation properties and vapor permeability). While these dynamics are complex they have been captured to a high fidelity in physics- and physiology- based thermoregulatory models. The SI transition probabilities were learned by Monte Carlo approximation using the SCENARIO (Kraning and Gonzalez, 1997) thermoregulatory model to simulate the responses of humans under similar conditions to the Ranger training road march. For the model runs the mean personal characteristics of our students were used. Mean environmental conditions obtained from a nearby airport weather station were air temperature = 24.4 °C, relative humidity = 85.3%, black globe temperature = 24.4 °C, and wind speed 2.75 m/s. Clothing insulation and vapor permeability parameters for the modeling were measured from copper manikin tests of the uniform used by the students (clothing insulation factor = 1.3 CLO, and vapor permeability (im) = 0.42). Metabolic rate was computed from movement speed, height, weight and load (assuming an average course grade of 0 and movement over hard top for a terrain factor of 1) using the equation developed by Pandolf, Givoni and Goldman (1977) with the Givoni and Goldman (1971) running correction factor. With starting SI's ranging from 1 to 9 all combinations of actions in our action set were run over six 5 minute intervals. Each conditional transition probability space had at least 10^4 samples.

MDP Learning

We wish to learn a policy function that for any state (Time=t, SI=si and D=d) provides us with an action that maximizes the expected utility until our goal is reached. An optimal policy for any time point on our road march course can be defined as follows:

$$\pi_t^*(s) = \text{argmax}_{a \in A(s)} \text{R(s)} + \sum_{s'} P(s'|s,a) U_{t+5}(s') \quad (5)$$

A set of optimal policies can be solved iteratively using dynamic programming. Starting at the end of the race the policy at time point 125 ($\pi_{125}^*(s)$) is easily computed, as the utility function (U_{130}(s)) is defined by the goal rewards. Next $\pi_{120}^*(s)$ can then be computed using the previously computed U_{125}(s) function and so on, where the current U is computed as:

$$U_t(s) = \max_{a \in A(s)} \text{R(s)} + \sum_{s'} P(s'|s,a) U_{t+5}(s') \quad (6)$$

Analysis

For this analysis, since we were unable to provide real-time guidance to Ranger students during the road march, we utilized the SCENARIO model to both simulate the individual SI responses of the students using their self paced movement (SI_{model}), and simulate the SI responses when conforming to the learned policy (SI_{policy}). Observed SI (SI_{obs}) was compared to SI_{model} by examining the mean root mean square error (RMSE) and bias to verify that the model provided an accurate simulation of the Ranger student's responses. We then compared SI_{model} and SI_{policy} values at the end of the road march using a paired t-test. We examined the relationship between the degree of impact ($SI_{model}-SI_{policy}$, at t=130) of the learned policy to the maximal SI_{model} and SI_{obs} reached during the march by Pearson correlation. Finally, we examined the actual movement profile (presented as stride frequency spectograms for the road march) of four students; two where the policy had the least impact and two where the policy had the most. Stride frequencies (proportional to movement speed) were found by applying fast Fourier transforms (FFTs) to the vertical axis accelerometry data. The alpha level for all hypothesis testing was set at 0.05.

Results

SI Transition Probabilities

Figure 3 shows three sets of SI transition probabilities in a gray scale map where black = 0 and white = 1.

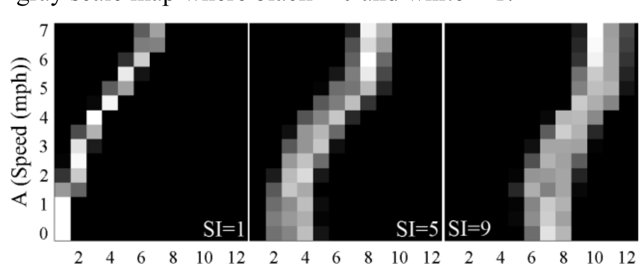

Figure 3: Transition Probabilities for SI of {1, 5, 9}. Grid Shade=P(SI'|SI,A) where White =1 and Black=0.

Policy Efficacy

Figure 4 shows the mean group responses for the SI_{obs} (gray), $SI_{modeled}$ (black), and SI_{policy} (dashed). $SI_{modeled}$ differs from SI_{obs} with a bias of −0.26 and RMSE of 1.34 ± 0.45 SI units.

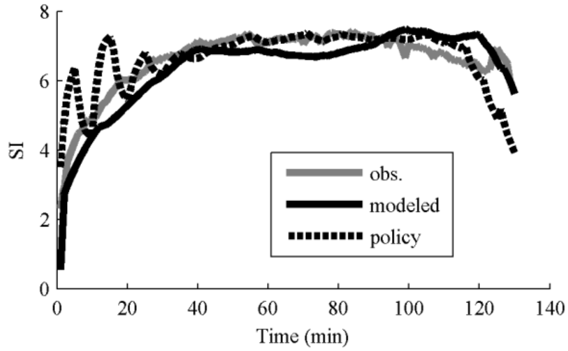

Figure 4: Mean Group Responses for the SI_{obs} (gray), $SI_{modeled}$ (black), and SI_{policy} (dashed).

The SI_{policy} responses have a significantly lower end point at 130 minutes than SI_{model} with a mean of 3.94 ± 0.88 versus 5.62±1.20 (t=2.16, P<0.001). Thus the policy had an overall mean impact of 1.67 SI units, and allowed the students to end with a "Low" thermal strain compared to a "Moderate" thermal strain (Moran et al. 1998). For all students the end-point SI_{policy} was lower than the end-point $SI_{modeled}$ and the maximum SI_{policy} score reached by any student was < 8.6.

The mean of each student's movement speeds when following our policy are shown in figure 5.

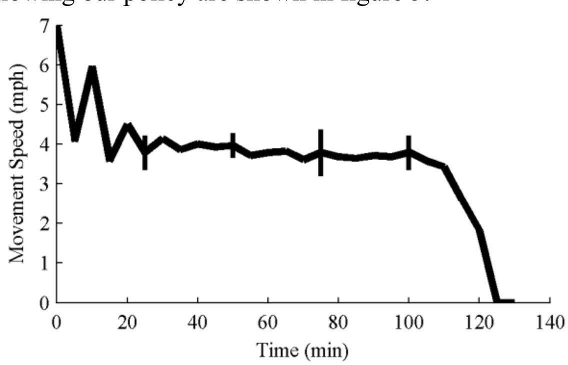

Figure 5: Mean of Speeds Taken for Each Student According to our Optimal Policy ± SD. Running is at Speeds > 4.5 mph.

Figure 6 shows the stride frequency spectograms for the whole road march course for four students. Panel A and B show the movement profiles where the policy had the least impact with differences in end point SI of 0.63 and 0.75 units. Panel C and D show the movement profiles where the policy had the most impact with end point differences in SI scores of 3.14 and 2.18 units. The movement rates of the students can be seen as highlights around ~2 Hz (walking) and ~3 Hz (running). These charts show that where our policy had least effect these students were already following our optimal policy (start fast, end slow). For the two students where the policy had the most affect it can be seen that one continually transitioned between walk and run (C), while the other starts and ends with walk run transitions (D).

The correlation between the policy impact and maximum $SI_{modeled}$ is 0.635 (P < 0.05); and maximum $SI_{observed}$ is 0.352 (not significant, N=14) but 0.622 (P<0.05, N=13).

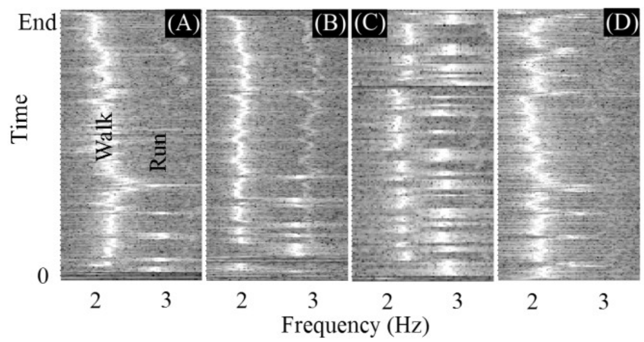

Figure 6: Stride Frequency Spectrograms of Students where the Policy had the Least (Panel A and B) and Most (Panel C and D) Impact. Light shades indicate more energy. Stride frequency is proportional to movement speed. Walking speeds are around 2Hz, and running speeds around 3Hz. Movement patterns for the whole road march run from the bottom of the chart to the top.

Discussion and Conclusions

The SCENARIO human thermo-regulatory model was able to accurately estimate the SI responses of the Ranger students with a small bias and RMSE close to one. Thus, SCENARIO provided a means to generate valid estimates of the student's thermoregulatory responses to our policy. Using this modeling approach, we found that the learned policy allowed all students to complete the course on time, with a lower SI, and without hyperthermia. Even though our learned policy is conservative with respect to avoiding high SI's, the policy allowed the students to finish in a significantly less thermally stressed state. Thus we conclude that this road march task can be modeled as a simple MDP which can generate a policy that is likely to improve the performance of these experienced students. This suggests that other MDP policies could be developed for other physically-demanding Ranger School events which could help students finish tests with the least amount of work and thermal stress possible.

While the modeling suggests that our learned policy is effective at reducing the end state SI score is it a reasonable policy that could be followed by people? The initial fast run and walk transitions appear, at first glance, peculiar. However, when we examine the actual movement rates of students (see figure 6) many adopt this same movement pattern. This reflects the fact that the needed early steady state pace would force an unnatural, and energy inefficient gait that is between a walk and a run (Paroczai and Kocsis, 2006). To avoid this awkward gait the policy we learned instead alternates between an energy-efficient walk and an energy-efficient run. The fact that this pattern is learned by the model where only thermal work strain is a factor is notable. When we examined the movement profiles of students where the policy had minimal impact we found that their actual movement profiles were similar to the learned policy (start fast, end slow). Conversely where the policy had a large impact the student's movement profiles were quite different. Additionally, we found a positive relationship between the degree of impact and the maximal SI obtained during the road march. This relationship held for SI from both modeled mile times and the observed data (albeit a student with highest SI had to be removed for the relationship to hold for N=13), suggesting that our learned policy was realistic and achievable and likely to result in an overall less thermally stressful road march.

Deploying this as a real system for the Ranger students would be fairly simple. Our thermal work strain state estimator is already implemented in an Android tablet that receives data from a wearable physiological status monitor (see figure 1). The tablet is GPS enabled and thus distance could be calculated. A policy would need to be generated for the environmental conditions of each prospective road march. Then, given the real time estimates of SI the tablet would be able to prompt the student with the optimum pace for each 5 minute segment.

These results also suggest that this technique shows promise for other areas such as marathon races where an athlete may desire to finish with the best time possible but avoid hyperthermia. Similarly, in cycling a safe and effective pacing strategy is desired. Atkinson et al. (2003) suggest "More research, using models and direct power measurement, is needed to elucidate fully how … pacing strategy might save time in a real race and how much variable power output can tolerated by a rider."

In conclusion this study has shown that it is possible to adequately express the tasks, risks and goals of an arduous physical activity in terms of a simple MDP. Solving the MDP for an optimal policy provided a realistic policy that allowed humans to perform their task according to pre-set goals and finish in a state of less thermal strain than would occur if they were left to follow their own training and instincts. These results suggest that live physiological state estimation, when coupled with MDP models of constrained real-world tasks, can optimize work rate policies to improve safety and reduce overall thermal-work strain burdens.

Disclaimer

The views expressed in this paper are those of the authors and do not reflect official policy of the Department of the Army, Department of Defense, or the U.S. Government.

References

Atkinson, G.; Davison, R.; Jeukendrup, A.; Passfield, G. 2003. Science and cycling: current knowledge and future directions for research. *Journal of Sports Sciences* 21:767-787.

Bernard, T. E.; and Kenny, W. L. 1994. Rationale for a personal monitor for heat strain. *Am. Ind. Hyg. Assoc. J.* 55(6): 505-514.

Bouchama, A.; and Knochel, J. P. 2002. Heat stroke. *N. Engl. J. Med.* 346(25):1978-1988.

Buller, M. J.; Tharion, W. J.; Hoyt, R. W.; and Jenkins, O. C. 2010. Estimation of human internal temperature from wearable physiological sensors. *22nd Conference on Innovative Applications of Artificial Intelligence (IAAI)*. 1763-1768.

Cheuvront, S. N.; Kenefick, R. W.; Montain, S. J.; and Sawka, M. N. 2010. Mechanisms of aerobic performance impairment with heat stress and dehydration. *J. Appl. Physiol.* 109:1989-1995.

Chi-Wai, R., K.; So-Ning, T., M.; Wing-Kuen, K., L.; Sai-Chuen, S., H.; Ka-Shun, P., W.; and Choi-Ki, C., W. 2011. Can Mobile Virtual Fitness Apps Replace Human Fitness Trainer? *5th Internaltional Conference on New Trends in Information Science and Service Science.* 56-63.

Givoni, B.; and Goldman, R. F. 1971. Predicting metabolic energy cost. *J. Appl. Physiol.* 30(3):429-433.

Kraning, K. K.; and Gonzalez, R. R. 1997. A mechanistic computer simulation of human work in heat that accounts for physical and physiological effects of clothing, aerobic fitness, and progressive dehydration. *J. Therm. Biol.* 22(4/5): 331-342.

Lopez-Matencio, P.; Alonso, J. V.; Gonzalez-Castano, F. J.; Sieiro, J. L.; and Alcaraz, J. J. 2010. Ambient intelligence assistant for running sports based on k-NN classifiers. IEEE 3rd Conference on Human Systems Interactions. 605-611.

Moran, D. S.; Shitzer, A.; and Pandolf, K. B. 1998. A physiological strain index to evaluate heat stress. *Am. J. Physiol. Regulatory Integrative Comp. Physiol.* 275: 129-134.

OSHA. 1985. *Occupational Safety and Health Guidance Manual for Hazardous Waste Site Activities.* Occupational Safety and Health Administration, Washington DC. 8-37 - 8-39.

Pandolf, K. B.; Givoni, B.; and Goldman, R. F. 1977. Predicting energy expenditure while standing and walking very slowly. *J. Appl. Physiol.* 43(4):577-581.

Paroczai, R.; and Kocsis, L. 2006. Analysis of human walking and running parameters as a function of speed. *Technology and Health Care* 14:251-260

Russell, S.; and Norvig, P. 2010. *Artificial Intelligence A Modern Approach (Third Edition)*. Prentice Hall, New York NY.

Sawka, M. N.; and Young, A. J. 2006. Physiological systems and their responses to conditions of heat and cold. In: Tipton, C. M. (ed). *Advanced Exercise Physiology*, American College of Sports Medicine. Lipcot, Williams & Wilkins. New York, NY pp535-63.

TBMED 507. 2003. *Heat Stress Control and Heat Casualty Management*. Department of the Army and Technical Bulletin. Washington DC.