A Study on Pervasive Health Monitoring for Heat Stroke Prevention Using Wrist-worn Devices

January 2018

Takashi HAMATANI

A Study on Pervasive Health Monitoring for Heat Stroke Prevention Using Wrist-worn Devices

Submitted to
Graduate School of Information Science and Technology
Osaka University

January 2018

Takashi HAMATANI

List of Publications

Related Journal Articles

(1) Takashi Hamatani, Akira Uchiyama, Teruo Higashino, "On-line Core Temperature Estimation Using Wearable Sensors During Sport with Variable Exercise Intensity" *IPSJ Journal*, vol. 58, no. 11, pp. 1818–1831, November 2017. (in Japanese)

(2) Takashi Hamatani, Akira Uchiyama, Teruo Higashino, "Estimating Core Body Temperature Under Hot Environment Based on Human Thermal Model Using Wearable Sensors" *IPSJ Journal*, vol. 56, no. 10, pp. 2033–2043, October 2015. (in Japanese)

Related Conference Papers

(1) Takashi Hamatani, Akira Uchiyama, Teruo Higashino, "HeatWatch: Preventing Heatstroke Using a Smart Watch", in *Proceedings of 2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom 2017 Workshops)*, pp. 661–666, March 2017.

(2) Akira Uchiyama, Takashi Hamatani, Teruo Higashino, "Estimation of Core Temperature Based on a Human Thermal Model Using a Wearable Sensor", in *Proceedings of IEEE 4th Global Conference on Consumer Electronics (GCCE 2015)*, pp. 605–609, October 2015.

(3) Takashi Hamatani, Akira Uchiyama, Teruo Higashino, "Real-Time Calibration of a Human Thermal Model with Solar Radiation Using Wearable Sensors", in *Proceedings of the 1st ACM/SIGAPP Workshop on Wearable Systems and Applications (WearSys 2015)*, pp. 45–50, May 2015.

(4) Takashi Hamatani, Akira Uchiyama, Teruo Higashino, "Estimating Core Body Temperature Based on Human Thermal Model Using Wearable Sensors", in *Proceedings of the 30th ACM/SIGAPP Symposium On Applied Computing (SAC 2015)*, pp. 521–526, April 2015.

Other Conference Papers

(1) Takashi Hamatani, Yudai Sakaguchi, Akira Uchiyama, Teruo Higashino, "Player Identification by Motion Features in Sport Videos Using Wearable Sensors", in *Proceedings of Ninth International*

Conference on Mobile Computing and Ubiquitous Networking (ICMU 2016), pp. 25–30, October 2016.

Abstract

Threat of global warming and catastrophic heat waves has recently triggered immense interest in heat stroke. Guidelines on exercise in a hot environment are widely shared for heat stroke prevention, but are still insufficient because of individually different resistance to heat stress. For reducing the number of heat stroke patients, core body temperature monitoring is essential so as to individually apply appropriate treatment such as fluid intake and intermission. The major precursor of heat stroke is known as rise of core body temperature. When core temperature rises, a human body prevents it by increasing sweat and skin blood flow. However, this thermoregulation capability may fail because of dehydration under severely hot environment and/or heavy exercise.

To measure or estimate core temperature, there have been many devices and studies. Core temperature can be measured as either rectal, tympanic, esophageal or pulmonary artery temperature. Some wearable or ingestible devices are now on the market, which can measure core temperature by various sensors (e.g., ingestible pills, thin probes with infrared thermometer and skin-mounted probes). These devices are applicable to clinical environment, however they are usually disallowed in the wild environment due to its high invasiveness or restriction of user activity. Motivated by the need of core temperature estimation in the wild environments (e.g., outdoor environments and exercise), several research efforts propose the estimation of core temperature. For instance, a study uses measured heart rate and another combines human thermoregulation simulation and environmental simulation. The major drawback of them is to require preliminary calibration by measuring actual core temperature or comprehensive simulation in terms of air flow, solar radiation and shading. The other concern is that these method may also require calibrations in each situation (e.g., fitness level and clothing), leading to considerable cost. These challenges spark the need of seamlessly applicable and light-weight calibration for any environment.

Dehydration is also a key precursor of heat stroke, and studied by many researchers. To ensure user's hydration, there have been emerging devices to keep track of fluid intake such as smart bottles which connect to smartphones. Also bountiful research efforts use different sensing modalities (e.g., image, audio and motion sensors) for the purpose of drinking activity recognition. Despite these efforts, a pervasive and mobile manner to unobtrusively keep record of daily fluid intake with scalability, does not still emerge.

Our research goal is to make a novel paradigm for heat stroke prevention which satisfies wide

applicability, little interference and scalability. For this, we leverage commercial-off-the-shelf wrist-worn gadgets (forecasted to be shipped 161 millions in 2021) and propose two important methods to prevent the risk of heat stroke in this dissertation.

Firstly, we design and evaluate a method to estimate core body temperature which can be concurrently optimized for individuals with little effort. The method employs Gagge's two-node model to calculate changes of core and skin temperature based on physical heat equations by inputs given by wrist-worn sensors and environmental sensors. The major drawback of the two-node model is lacking consideration of individually different thermoregulation function, leading to incomplete accuracy for core temperature estimation. To address this challenge, our method optimizes four parameters representing differences of human thermoregulation capability (e.g., increasing speed of sweat and blood flow on skin surface) by exhaustively comparing actually measured skin temperature and each of 3,200 simulated skin temperatures. This realizes the estimation of core temperature which can be seamlessly deployed in any environment with minimal interference in user activity as well as wide scalability owing to requiring only common devices on the market. Despite limited information was available during exercise, we confirmed the method successfully estimated core temperature and it could choose an effective set of four parameters in moderate exercise.

Secondly, we propose essential extensions for Gagge's two-node model to realize core temperature estimation in the wild. We have found Gagge's two-node model can be optimized for exercises with fixed workload, however its accuracy becomes worse in the exercise with variable workload. This is mainly due to unexpected responses of core temperature during transient period (e.g., taking a break and increase/decrease exercise workload). To explain this, we introduce two delay parameters which approximate actual human responses. Our method is applicable to calibrate the two parameters as well as the other parameters by observing actual human responses against variable exercise workload. We also propose some extensions for the two-node model to appropriately consider the effect of solar radiation, wind, and water ingestion. The method is evaluated through over 120 hours of walking, running, biking, and tennis and achieves up to 0.30 degree Celsius error in each exercise. Further, we developed a prototype of heat stroke caution to show its effectiveness in a real application.

Thirdly, we propose automated estimation of fluid intake by leveraging inertial sensors in wrist-worn gadgets. Although Gagge's two-node model assumes the human thermoregulation (i.e., sweat and blood flow increase) properly operates, it may fail in the severely hot environment and/or heavy exercise because of dehydration. This severely drops reliability of core temperature estimation, accordingly dehydration must be prevented. To prevent dehydration, our approach leverages human hydration context estimated from drinking episode and amount of fluid intake. We collected data in both of laboratory and wild experiments and designed a novel algorithm to capture drinking episode from noisy motion traces of the wrist. The algorithm recognizes human activities with macro and micro scale classifiers so as to firstly separate a drinking activity from the other activities and then recognize small gestures constituting a series of fluid intake. We also propose a robust model for estimating

amount of fluid intake which uses not only the duration of sipping gesture, but also the posture of the arm. The method has been validated through rigorous datasets and shown 15% error for overall fluid intake amount.

Through these contributions, commercial-off-the-shelf devices on the wrist will be utilized to assess risk of heat stroke and appropriately encourage individuals to avoid it. The complete system will appropriately notify the timing of taking breaks and fluid intake to users in real time, and keep all the core temperature levels of users safe. This dissertation has established the foundation of pervasive method for preventing heat stroke for all mankind.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In recent decades, the threat of global warming and catastrophic heat waves has triggered immense interest in heat stroke. Heat stroke causes severe health disorders such as heat cramps, heat syncope, and heat exhaustion, worstly leading to death. World Health Organization (WHO) globally posts a health manual to ensure how to deal with heat environment [4], where it suggests referring environmental heat indices such as the wet bulb globe temperature (WBGT) and standard effective temperature (SET) to assess the heat stress. These indices are composed of not only air temperature, but also humidity, radiant heat and solar radiation, and provide reliable decisions to prevent over-exercise especially for outdoor work and exercise. For example, a coach uses the combined heat index to organize exercise duration, and the interval of breaks and water intake. However, these heat indices remain insufficient for heat stroke prevention, as inferred by the number of heat stroke patients not decreasing in recent years. The major drawback of the environment-based instruction is the lack of the consideration of differently rising core temperature inside individual human bodies. High core body temperature is a fundamental precursor of heat-related illness [5], but shows various individual differences caused by a variety of human body characteristics such as speed of core temperature rise, resistance against heat stress, and water supplement level. Therefore, deep insight of individual body context by human-centric sensing technology is essential for reducing the number of heat stroke patients all over the world.

Ideally, heat stroke is preventable by grasping the heat context which differs from person to person, and then appropriately suggesting timings of breaks and water supplement. When core body temperature rises according to the heat environment, a human body emits heat by increasing blood flow on skin surface and sweating rate, then the temperature is usually reduced. However, core temperature can be higher in the case of severe heat environment and heavy exercise, thus heat stroke can occur. Hence, measuring core temperature is the most effective way to assess heat stroke risk. Sufficient water supplement is also important to maintain human's thermoregulation function (i.e., sweating and blood flow increase). Dehydration symptoms arise when the body loses water in a rate higher than its intake rate. Some people can sense the dehydration by feeling thirsty, dry or sticky mouth, but others (e.g.,

children and elderly people) are sometimes not aware of their dehydration. Accordingly, it is essential to capture individually different contexts of dehydration (i.e., fluid intake and water loss level) and suggest appropriate timing and amount of fluid intake.

To address these challenges, many systems and studies have been proposed. For core temperature assessing, some wearable and ingestible sensors are available [6, 7, 8], and some studies [9, 10] propose methods of core temperature estimation leveraging non-invasive sensors. Core temperature is directly monitored by either the ingestible pill-like sensor [6] or the infrared thermometer inserted into the ear canal [7], and indirectly measured by the thermal probe attached on the skin [8]. These systems provide individual monitoring of core temperature, however they have unacceptable drawbacks: ingestible sensors require highly invasive measurement and involve considerable cost for disposable pill sensors; measurement probes are disallowed in situations such as contact sport, in which players collide with each other; a skin-mounted probe does not scale well in the wild environment (e.g., daily life and outdoor exercises) since it involves a wired control unit. To assess core temperature in outdoor environments, Reference [9] proposes its estimation using heart rate observation. Although it requires preliminary data collection to calibrate estimation model parameters, it achieves core temperature estimation with 0.21 degree Celsius error. In Reference [10], the human thermal model is used to assess the risk of heat stroke for outdoor workers. It provides the prediction of core temperature rise based on the integrative use of comprehensive simulation of environment, solar radiation and human thermal regulation. As suggested by these efforts, human core body temperature can be predicted by observable contexts captured by environmental and/or wearable sensors. Meanwhile, these techniques involve some preliminary calibration for considering each environment or each human's thermoregulation dynamics. This severely limits their wide applicability.

Supported by emerging devices such as sensorized gadgets and smarter wearables, tracking fluid intake is also in deep interest recently. Smart bottles (e.g., Trago smart bottle [11] and Hidrate Spark [12]) have been widely available which are capable to automatically keep record of user's fluid intake. The bottles are appropriate to track water intake in designated environments. However, they are neither scalable nor ubiquitous since users have to buy them and bring to every drinking context. Additionally, many researchers combine different sensor modalities for not only recognizing ingestive activities, but also assessing amount of fluid intake. They use either image [13] , audio [14] or wearable sensors [15, 16], however they have some limitations. For instance, vision-based systems require capturing photos with constraints in terms of the view angle and quality. Audio analysis may raise privacy concern and background noise may overlap with the collected audio. In other words, these methods are not appropriate for continuous and daily tracking of fluid intake in terms of ease of deployment, scalability and interference in user's life.

Despite these considerable research efforts, there has not been a method for core temperature estimation and fluid intake assessment which satisfies all the following requirements: (1) Ease of deployments: leveraging existing infrastructure without installing special and expensive sensors. (2)

Minimal interference: minimal burden and cost for calibration. (3) Scalability: seamlessly applicable to different environments. To meet these criteria, we employ wrist-worn sensors as pervasive and unobtrusive sensing modality, and propose methods for estimating both of core temperature and fluid intake. In this dissertation, three primary contributions will be made toward preventing heat stroke by wrist-worn devices.

Firstly, we design and evaluate a method to estimate core body temperature which can be concurrently optimized for individuals with little effort. The method employs Gagge's two-node model to calculate changes of core and skin temperature based on physical heat equations by inputs given by wrist-worn sensors and environmental sensors. The major drawback of the two-node model is lacking consideration of individually different thermoregulation function, leading to incomplete accuracy for core temperature estimation. To address this challenge, our method optimizes four parameters representing differences of human thermoregulation capability (e.g., increasing speed of sweat and blood flow on skin surface) by exhaustively comparing actually measured skin temperature and each of 3,200 simulated skin temperatures. This realizes the estimation of core temperature which can be seamlessly deployed in any environment with minimal interference in user activity as well as wide scalability owing to requiring only common devices on the market. Despite limited information was available during exercise, we confirmed the method successfully estimated core temperature and it could choose an effective set of four parameters in moderate exercise.

Secondly, we propose essential extensions for Gagge's two-node model to realize core temperature estimation in the wild. We have found Gagge's two-node model can be optimized for exercises with fixed workload, however its accuracy becomes worse in the exercise with variable workload. This is mainly due to unexpected responses of core temperature during transient period (e.g., taking a break and increase/decrease exercise workload). To explain this, we introduce two delay parameters which approximate real human responses. Our method is applicable to calibrate the two parameters as well as the other parameters by observing real human responses against variable exercise workload. We also propose some extensions for the two-node model to appropriately consider the effect of solar radiation, wind, and water ingestion. The method is evaluated through over 120 hours of walking, running, biking, and tennis and achieves up to 0.30 degree Celsius error in each exercise. Further, we developed a prototype of heat stroke caution to show its effectiveness in a real application.

Thirdly, we propose automated estimation of fluid intake by leveraging inertial sensors in wrist-worn gadgets. Although Gagge's two-node model assumes the human thermoregulation (i.e., sweat and blood flow increase) properly operates, it may fail in the severely hot environment and/or heavy exercise because of dehydration. This severely drops reliability of core temperature estimation, accordingly dehydration must be prevented. To prevent dehydration, our approach leverages human hydration context estimated from drinking episode and amount of fluid intake. We collected data in both of laboratory and wild experiments and designed a novel algorithm to capture drinking episode from noisy motion traces of the wrist. The algorithm recognizes human activities with macro and micro

scale classifier so as to firstly separate a drinking activity from the other activities and then recognize small gestures constituting a series of fluid intake. We also propose a robust model for estimating amount of fluid intake which uses not only the duration of sipping gesture, but also the posture of the arm. The method has been validated through rigorous datasets and shown 15% error for overall fluid intake amount.

Our study overcomes the challenge of collecting huge amounts of data in modeling core temperature estimation by integrating a model-based estimation method and a light-weight calibration method. We also designed a novel method for calibration which is applicable in severely restricted environments (i.e., during exercise). Further, our study makes a contribution for heat stroke prevention from the perspective of information science through sensing individual contexts by lightweight sensing modalities and suggesting individually optimal care. Through these contributions, commercial-off-the-shelf devices on the wrist will be utilized to assess risk of heat stroke and appropriately encourage individuals to avoid it. This dissertation has established the foundation of pervasive method for preventing heat stroke for all mankind.

The rest of this dissertation is organized as follows. Chapter 2 reviews related work on core temperature and fluid intake estimation. Chapter 3 explains the design and performance of core temperature estimation method in the basic environment and exercise. Chapter 4 proposes the method of lightweight and on-site calibration, and the extension of Gagge's two-node model to suit complex environment and exercise. Chapter 5 describes the novel method to estimate amount of fluid intake by observing inertial sensor measurement, and the evaluation through laboratory and wild experiments. Finally, Chapter 6 summarizes and concludes this dissertation.

Chapter 2

Related Work

2.1 Heat Stroke Prevention

In recent decades, heat stroke has received considerable attention because of global warming and catastrophic heat waves [17]. The mechanism of heat stroke is described in [5], where core temperature rise is a fundamental factor. When core temperature rises, a human body emits heat by increasing blood flow on skin surface and sweating rate. These thermoregulation functions usually decrease core temperature and keep safe. However, severely heat environments and/or heavy exercise can overwrite these resistance function, then core temperature exceeds beyond normal range and finally heat stroke arises. In healthy adults under 40 years old, the upper limit of the normal oral (core) temperature range is 37.7 $^{\circ}C$ [18]. In contrast, the measured oral temperatures of heat stroke patients exceed 39.5 $^{\circ}C$ [19]. Therefore, heat stroke is indicated when the core temperature exceeds 39.5 $^{\circ}C$, and can be monitored by core temperature measurements. Another major factor of heat stroke is dehydration which makes the cardiovascular system difficult to suppress rise of core temperature. A study [20] revealed that core temperature rises $0.3^{\circ}C$ for every 1% of fluid lost during exercise. Severe dehydration also causes a failure of sweating, the most important heat emission function.

Accordingly, core temperature is the most important precursor of heat stroke. There have been many thermometers to measure core temperature, however they have some limitations to be used in the wild environments. Probe measurements of oral, rectal, tympanic, and esophageal temperatures (as conducted in [21]) are generally used as core temperature in clinical situations, but they are not available in almost all exercise situations due to their invasiveness. Tympanic thermometer such as [7] has a potential to be used in daily life, yet it is still inadequate for core temperature monitoring during interpersonal sports. Brain temperature can be estimated by measuring the forehead temperature [22] due to its correlation with core temperature in indoor environments. However, the correlation deteriorates under environmental effects such as solar radiation and wind. In addition, there are also clinically-approved gadgets to measure core body temperature such as Cortemp [6] and 3M SpotOn System [8]. The former provides continuous and reliable measurement of in-body temperature by

Table 2.1: Guidance for athletic trainers [1]

WBGT($^\circ C$)	**Alert Level**	**Recommendation**
> 31	Stop Exercise	No outdoor exercise
> 28	Strict Warning	Avoid hard exercise
> 25	Warning	Frequent breaks and water intake
> 21	Attention	Make sure water intake in the intervals
< 21	None	Usually safe

leveraging pill-like sensors. The latter provides indirect measurement of core temperature through a heater unit attached on the skin surface by equalizing its temperature to the inner body temperature. However, these sensors are not appropriate for monitoring core temperature in daily and outdoor exercise owing to its considerable cost for disposable and ingestible sensors, or slow response to changing body temperature. In summary, there are still no sensors which can easily measure core temperature in outdoor work and exercise despite many sensors are developed to measure it.

Therefore, environmental factors are now used for heat stroke prevention. WHO provides health manuals for summer exercise [4], where the risk level of heat stroke is explained by heat-stress indicators such as SET [23] and WBGT [24]. These indices are composed of multiple factors regarding ambient environment (e.g., air temperature, humidity, solar radiation, etc.) and used as the measures for decision making in outdoor working and exercise. For instance, Japanese Ministry of the Environment introduces recommendations based on WBGT as shown in Table 2.1. However, many Japanese people work or participate in sports during daytime even though many days in summer reach higher or the highest alert level (e.g., Osaka observed 43 days with alert "Stop Exercise" or "Strict Warning" (WBGT>28 ℃) in 2015). Even in those severe situation, some people can continue their activities as they are accustomed to heat environment, but many others should stop their activities. Individual difference of resistance to heat and fatigue degree also affect how the core temperature rises, and they may vary everyday. Thus, these heat indices are useful for avoiding high-risk activities, yet they are not suitable for individual risk assessment since they cannot account for the heterogeneous changes in core temperature among individuals.

To estimate core temperature, Buller et al. proposes core temperature estimation with heart rate observation and Kalman filter [9]. They designed and evaluated their method with multiple datasets collected by soldiers wearing personal protective equipment with different exercise, and showed $0.21^\circ C$ root mean squared error. The method realizes core temperature estimation relying on only heart rate which can be captured by commercial-off-the-shelf devices, but it must involve calibration of a model by measuring heart rate and actual core temperature with ingestible pills in advance. They also infer the model parameter has to be optimized according to a variety of fitness level, ages, environments and clothes. Yabuki et al. proposes a heat stroke alert system for outdoor workers that integrates computational fluid dynamics (CFD), sun and shade simulations, and a core temperature estimation model [10]. The method provides the prediction of core temperature rise based on the integrative use of com-

prehensive simulation of environment, solar radiation and human thermal regulation. Their method is appropriate to forecast the risk of heat environment for designated workload and environment, yet it always requires 3-D model building and comprehensive simulation for each environment. Thus, the method does not scale well in terms of deployment cost for different environments and workloads.

Despite these research efforts, to our knowledge, there are no methods to measure or estimate core temperature in exercise with low cost and minimal interference in user's life. We address this problem by employing the wrist-worn devices forecasted to be shipped 161 millions in 2021 [25], and introducing a lightweight method for individual calibration.

2.2 Human Thermal Models for Core Temperature Estimation

The heat dynamics of human body can be represented by equations of thermodynamics, accordingly a variety of human thermal models [26, 27, 28, 29] are proposed. In those models, human body is approximated as multiple nodes, and then the heat production, transfer and emission are calculated between neighboring nodes per unit time. Despite different complexity of those model, each model can reproduce standard thermoregulation function of humans such as increasing blood flow on skin surface and sweat production as response to rising core temperature. Gagge's two-node model [26] regards a human body as a combination of core and skin nodes. In Stolwijk's 25-node model [27], a human body is composed of 25 nodes arranged in four layers (core, muscle, fat, and skin). Smith also proposes 15-node model with blood vessels along with almost all borders of nodes [28]. These nodes can be further divided for a detailed thermal simulation [29], although accurate simulations of such detailed models require a large number of inputs and a long computational time. Subsequently, there have been other advanced models such as Fiala model [30] and JOS-2 model [31]. Fiala et al. propose 15-node models which consists of passive system to consider thermal energy exchange in the body and active system to explain vascular dilation and constriction, sweating, and shivering [30]. Kobayashi et al. propose JOS-2 model considering detailed vascular system [31]. Yabuki et al. use Gagge's two-node model to predict core temperature rise of outdoor workers in hot environment [10]. Their method is supported by comprehensive simulation of air temperature distribution, solar radiation, sun shading and metabolic heat production. Nevertheless, their method does not focus on the precise estimation of individual subject's core temperature, but assesses heat risks in multiple working scenarios.

To represent individual difference of thermal responses in hot and cold environments, Takada et al. [32] introduced six parameters into the Gagge's two-node model. The method calibrates these parameters by comparing the measured rectal temperature and skin temperature (i.e., a mean skin temperature calculated by Hardy-Dubois formula [33]) with the simulated core and skin temperatures over 120 minutes. Their approach achieves more accurate simulations than the original Gagge's model, but requires long and invasive measurements of the core temperature, suggesting the method is unsuit-

able for most exercises. Another concern is the diversity of human conditions which alters the optimal parameter set even in the same subject. For example, sweating and blood flow work efficiently under normal conditions, but may be less effective if the subject is exhausted.

Compared to their research effort, this dissertation focuses on developing practical calibration method of Gagge's two-node model. In other words, we design and evaluate the method to launch human thermal simulation on the model with available inputs even in the exercise scenario, and individually optimize with use of additional context. Additionally, we introduce the reasonable modifications for Gagge's model to suit the model to complex exercise and real environment.

2.3 Assessment of Energy Consumption in Exercise

In this dissertation, we use heart rate as a clue for assessing metabolic heat production. The major advantage of using heart rate is that our method can seamlessly suit to wide variety of environments by using commercial-off-the-shelf devices. In contrast, its drawback is that heart rate may involve errors caused by not physical load, but psychological load (e.g., getting nervous in competitions). To assess the calorie consumption in exercise, there have been many studies as follows. Choi et al. utilize accelerometers to estimate energy expenditure in physical activities [34, 35]. Lester et al. combine accelerometer, barometer and geographic information for assessing calorie expenditure with use of only one accelerometer [36]. Zhan et al. propose a method of calorie expenditure estimation in biking which combines cadence sensor, heart rate and geographic information [37]. Tsou et al. leverage Microsoft Kinect [38] to capture skeleton for estimating calorie burned in aerobics [39]. These research efforts ensure our method for core temperature estimation to improve its accuracy even in wide variety of exercise (e.g., biking and aerobics) due to more precise input (i.e., heat energy production and energy loss by exercise). Also the advanced device such as Kinect makes our method free from wearing accelerometers and heart rate sensors, leading to high scalability.

2.4 Advanced Activity Recognition Considering Hidden Contexts

Activity recognition has been deeply studied in recent decades. Its fundamental idea is to extract meaningful contexts from the sensor readings (e.g., waveform generated by accelerometer). Machine learning algorithm has made great contributions for activity recognition due to its capability to firstly identify features of potential mechanisms and then predict new data using the identified pattern. Basically, many of machine learning algorithms simply find the potential relationships between given feature values and true states. Further, advanced machine learning models such as Hidden Markov Model (HMM) [40] and Conditional Random Field (CRF) [41] are widely utilized and show promising results for activity recognition since they can consider the relationship among hidden states as well as the relationship between the observed feature and a true state behind the feature. CRF showed higher

performance than HMM in the use case of human activity recognition [42], sentence recognition [43] and human mobility recognition [44]. Accordingly, we employ CRF to infer the potential relationship among ingestion-related activities in this dissertation.

2.5 Activity Recognition for Capturing Ingestive Contexts

To log daily dietary records, there have been many research works to recognize the eating moments and detect the food type as well as estimate the amount of food consumed. These systems leverage either dedicated body-worn sensors or smartwatches-equipped sensors. Some systems leverage acoustic sensors to detect eating-related activities [45, 46, 47]. For instance, [45, 46], employ an earbud to capture characteristic sounds such as chewing, thus classifying some food types based on the chewing pattern. Another system [47] leverages a throat microphone to recognize the swallow pattern which can be used to classify 4 food types (cookie, apple, bread, banana). Additionally, wrist-worn inertial sensors embedded in smartwatches can recognize eating moments (i.e., meals time [48, 49, 50]). These inertial sensors can also be used to estimate the biting count [51] and predict the medical adherence [52] by applying machine learning to distinguish higher order gestures from raw motion sensor waves. Inertial sensors on ears also showed high performance in the detection of eating episodes in very challenging unconstrained environments [53]. Finally, the fusion of different sensing modalities proved successful to provide deep context recognition of the eating-related activities. For example, the fusion of acoustic and motion features can estimate the family meal time [54]; eating, drinking activities and the exercise workload [55]; and food type and the consumed amount of food [15, 56].

2.6 Techniques for Hydration

In recent years, applications for human hydration have been on increasing interest. There are some applications released to the market [57, 58], in which the users manually register their fluid intake to the app, then the application generates daily records of overall fluid intakes. The manual application is suitable for enthusiastic users, however the motivation of the other users easily decreases due to the burden to measure or guess each weight of fluid intake. Hence, the automated method of fluid intake tracking is essential for continuously monitoring human health.

For this, several studies propose the automated fluid intake estimation using different sensing modalities: accelerometer [59, 60], capacitive sensor [61] and RFID tags [62] are attached on the container and they can catch drinking event when they are used. These sensors can provide rich context regarding fluid intake due to direct attachment on a container used for drinking and help to recognize drinking episodes. Further, there have been nowadays commercial smart bottles such as [11, 12, 63] which automatically track fluid intake by leveraging fluid level sensor embedded in the bottle. These sensor-assisted containers are appropriate for tracking of fluid intake in a designated environment (e.g., using same bottle during sports), yet they do not scale well as they require buying

a special device or attaching accelerometer on the container, and always using the sensor-assisted container. In addition, mobile sensing modalities have been also used to keep track of fluid ingestion [16, 15, 13]. Amft et al. [16] combines a wrist-worn motion sensor and a magnetic coupling sensor on the shoulder and the arm to capture wrist motion, 3-D orientation of the wrist and the distance between the shoulder and the wrist. They can recognize nine types of containers with 75% average accuracy and three levels of fluid in each container achieving an accuracy that ranges from 58% to 83% on average. However, their method does not provide a fine-grained estimation of the fluid intake amount while using dedicated wearable sensors, and all experiments are conducted in the laboratory only. By combining acoustic features with motion features, Mirtchouk et al. [15] can estimate the fluid type as well as the consumed amount. They concluded that it is still difficult to recognize the fluid type, but they can estimate the weight of consumed fluid with 47.2% mean absolute error. However, they fused many sensors including an earbud, two smartwatches, and a smart eyeglass. This hinders the scalability of their solution. Mengistu et al. [55] also suggest that drinking activities can be captured by the fusion of acoustic signals from throat microphone and motion signals from smartwatch sensors. Similarly, their method does not scale well and it does not provide an estimation of the fluid intake amount. Image processing based techniques have also been proposed to recognize the hand grasping posture when a person is drinking by applying Haar-like feature on the input image [13]. However, it is intrusive, raises privacy issues, and the estimation of fluid intake amount is still limited.

These seminal systems and works have demonstrated feasibility of fluid intake monitoring by pervasive sensing modalities, however they are still far away from eligibility to keep record of daily fluid intake in terms of cost for deployment and interference in user's life. Our novel idea breaks through the challenges of usability and deployment cost by leveraging inertial sensors embedded in state-of-the-art smartwatches.

Chapter 3

Core Temperature Estimation During Exercise by Leveraging Wearable Vital Sensors

3.1 Introduction

For preventing heat stroke, core temperature is important to evaluate the effect of heat environment and/or exercise on human body. Under heat environment and hard exercise, core temperature gradually rises, then our body transfers the heat to the air for decreasing the temperature by thermal regulation systems such as sweating and increase of skin blood flow. These thermal regulation systems usually work efficiently, therefore we can keep core temperature safe. However, a failure of the system can occur in severe environments. Especially in early summer, people are not accustomed to heat stress leading to heat stroke.

Generally, core temperature is measured as rectal or tympanic temperature, although their measurement is difficult due to its invasiveness. Inserting a measurement probe into those parts is dangerous during exercise and burdensome in daily life. Instead, axillary temperature or oral temperature is generally used as a reference of core temperature since they are easy to measure. However, we cannot use these temperature during activities because of the following reasons. First, we need to keep resting and closing an armpit and wait about 4 minutes in order to measure axillary temperature [64]. Second, it is dangerous to keep a measurement probe in a mouth during exercise.

Alternatively, a wireless ingestible thermometer called CorTemp [6] can be used to directly measure core temperature during exercise. However, the receiver is expensive and the thermometer is disposable leading to a considerable cost. An alternative is H.O.T. system which measures skin temperature [65] by a heat sensor inside a helmet used for military, public safety (firefighters, SWAT), HAZMAT, industrial, and racing. It displays alert if skin temperature rises over a threshold because extreme skin temperature is caused by abnormalities. Nevertheless, skin temperature does not always quickly

reflect changes in a human body including core temperature because it is affected by environmental factors such as wind and solar radiation. In addition, there is a 3M SpotOn System [8] which indirectly measures core temperature by a probe on a forehead, which is not appropriate to exercise since it needs a wired control unit assuming clinical use.

In this chapter, we propose a method to estimate core temperature by using Gagge's two-node model [26]. The two-node model is one of human thermal models [26, 27, 29] that represents a human body as two nodes: a core node and a skin node. It simulates changes of core and skin temperature by calculating heat production in a human body and heat exchange between the core node, the skin node and the ambient air. For this calculation, we leverage a wrist-worn gadget to obtain heart rate and skin temperature, and environmental sensors to measure the air temperature, humidity and the solar radiation. Interestingly, a previous study [32] proposes six individual parameters of the two-node model and optimizes them based on measurement of rectal and seven-point skin temperatures. Their parameters are related to individual thermoregulation function and physical conditions, such as a perspiration rate and a skin blood flow rate. The study confirms that appropriately setting these individual parameters is essential for accurate estimation of core temperature. Nevertheless, their method is not capable to assess core temperature in the exercise scenario since their method requires intrusive calibration in terms of measurement effort. To address this challenge, we propose an unobtrusive method for calibrating these parameters in the wild scenario (e.g., daily lives and exercise). Our method optimizes the number of parameters and ranges, and compares the change of measured skin temperature and 3,200 instances of skin temperature simulated by all the possible combinations of four parameters. We also introduced an extensional equations to the model in order to consider the effect of solar radiation.

We have evaluated our method through totally 52 hours of walking in heat environment by seven subjects. The results show that our method improves the accuracy by 12% at most compared to the default individual parameter set.

3.2 Core Temperature Estimation Method

This section presents our proposal of a method to estimate core body temperature during exercise. We note that our method leverages Gagge's two-node model to estimate the temperature, however we abstractly describe the model in this section to focus on the method of model calibration, and the detail will be described in the next section.

3.2.1 Overview

Figure 3.1 illustrates the overview of Gagge's two-node model. The model approximates a human body as a sphere composed of the core node and the surrounding skin node. It simulates the temperatures of these nodes by calculating the heat energy transfer between the core node, the skin node and

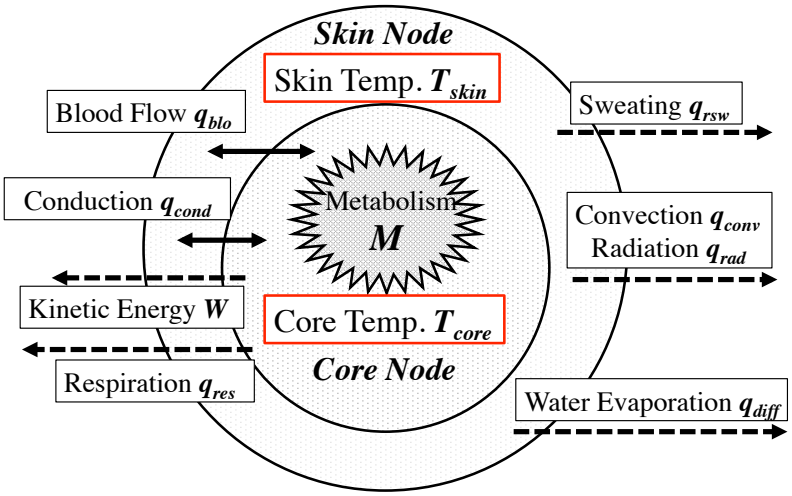


Figure 3.1: Gagge's two-node model overview

the ambient air. Accordingly, it is the most simple model due to its the minimum number of nodes compared to the other multi-node models [29]. Nevertheless, the study [32] proved its high accuracy when the optimal parameters were appropriately set. Therefore, we employ their method with a small extension to suit our scenario where we use the two-node model for outdoor exercise.

Figure 3.2 depicts the overall process of the method, where the essential inputs are given to the two-node model, and the model generates exhaustive simulation of skin temperatures, and then the best parameter set is obtained from comparing actually measured skin temperature and the simulated temperatures. The method basically requires three types of input as follows. (1) Static information to assess the physical parameter of the two-node model (i.e., body characteristics). This information is manually given to the model. (2) Environmental information consists of temperature, humidity, and solar radiation to evaluate the environmental effect on the human body. These values are measured by commodity devices deployed in the exercise environment, and given to the model in each unit time. (3) Biological information, namely the skin temperature for our calibration, and continuous heart rate measurements for assessing the heat production by metabolism. The heart rate is finally translated into the metabolic heat production rate in each unit time. We also measure the initial core temperature as tympanic temperature before exercise. The tympanic temperature can be easily measured by using some infrared thermometers on the market.

3.2.2 Core Temperature Estimation and Parameter Calibration

Suppose t unit time elapses since a user starts exercise. Firstly, the static information and the initial temperature of the skin node (${T_{skin}}^0$) and the core node (${T_{core}}^0$) are given to the model. Then, environmental and biological sensor readings from time 0 to $t-1$ are input to the model. After that, given a set θ_i of individual parameter values, the two-node model calculates the following sequences

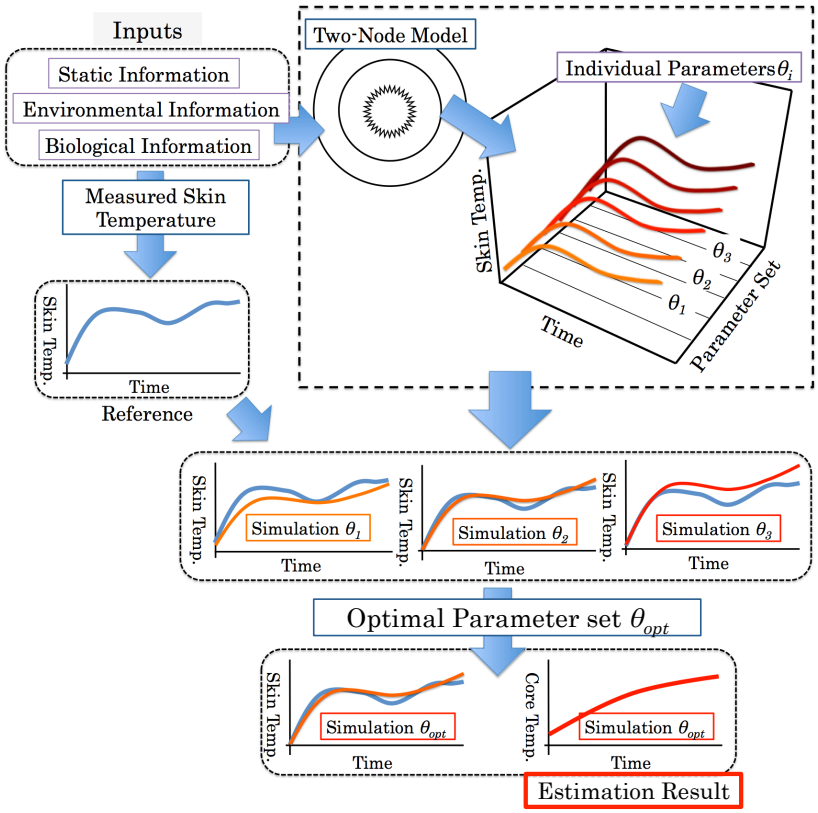


Figure 3.2: Exhaustive simulation for parameter optimization

$\boldsymbol{T_{skin}}^{\boldsymbol{t}}(\theta_i)$ and $\boldsymbol{T_{core}}^{\boldsymbol{t}}(\theta_i)$ of the skin and the core temperature at each unit time.

$$\begin{aligned} \boldsymbol{T_{skin}}^{\boldsymbol{t}}(\theta_i) &= \{{T_{skin}}^0, {T_{skin}}^1(\theta_i), {T_{skin}}^2(\theta_i), \ldots, {T_{skin}}^t(\theta_i)\}, \\ \boldsymbol{T_{core}}^{\boldsymbol{t}}(\theta_i) &= \{{T_{core}}^0, {T_{core}}^1(\theta_i), {T_{core}}^2(\theta_i), \ldots, {T_{core}}^t(\theta_i)\}, \end{aligned}$$

where ${T_{skin}}^t(\theta_i)$ and ${T_{core}}^t(\theta_i)$ denote the skin and the core temperature at time t for a given individual parameter set θ_i, respectively. The details of the parameters are described in the next section.

Figure 3.2 depicts the determination process of the optimal individual parameter set. We determine the optimal parameter set θ_{opt} by selecting the parameter set which minimizes the squared error between the simulated skin temperature and the measured skin temperature as below.

$$\theta_{opt} = \arg\min_{\theta_i} \sum_{j=1}^{t} |{T_{skin}}^j(\theta_i) - \hat{{T_{skin}}}^j|^2.$$

In the above equation, $\hat{{T_{skin}}}^j$ indicates the measured skin temperature at time j. Finally, the corresponding $\boldsymbol{T_{core}}^t(\theta_{opt})$ is our estimation result at time t. We note that θ_{opt} at time t is determined independently of the previous θ_{opt} determined at time $t-1$ in order to fully utilize all the given information.

Table 3.1: Range of individual parameters

α_1	5.04	5.67	**6.3**	6.93	7.56	(5 factors)
α_2	22.5	$\cdots$	45	$\cdots$	**75**	(8 factors)
α_3	75	$\cdots$	150	$\cdots$	**250**	(8 factors)
α_4	30	$\cdots$	**100**	110	120	(10 factors)

3.2.3 Two-node Model Parameters of Different Thermal Regulation

In hot environment, the human body usually widens blood vessels to emit heat from skin surface (vasodilation). It also increases sweat production to cool skin surface by swear evaporation. These responses are general reactions which occur when the core body temperature rises, but their capabilities (e.g., speed of response, amount of sweating) are widely different from person to person. In the study [32], Takada et al. replace some constants in the Gagge's two-node with the variables to explain the difference of human thermal responses. We further modify their proposal to suit our scenario and define two equations related to vasodilation and sweating responses during exercise in hot environment as below.

$$ {V_{blo}}^t = \alpha_1 + \alpha_2 \cdot ({T_{core}}^t - {T_{core}}^0), \tag{3.1} $$

$$ {m_{rsw}}^t = \{\alpha_3 \cdot ({T_{core}}^t - {T_{core}}^0) + \alpha_4 \cdot ({T_{core}}^t - {T_{core}}^0) \cdot ({T_{skin}}^t - {T_{skin}}^0)\} \cdot 10^{-3}. \tag{3.2} $$

${V_{blo}}^t$ and ${m_{rsw}}^t$ mean the blood flow rate and the amount of sweat productiom at time t. In these equations, $\alpha_1, \alpha_2, \alpha_3$ and α_4 are the variable parameters. α_1 and α_2 are the initial value of skin blood flow and the vasodilation coefficient, respectively. α_3 represents the sweat coefficient for exercise and α_4 means the coefficient of basic sweating response. These equations indicate that the skin blood flow and sweating amount increase according to the rise of core and skin temperatures represented as the gradients against their initial values $({T_{core}}^0, {T_{skin}}^0)$. Note that these gradient values are regarded as zero if they are negative.

We change the four parameters within the ranges shown in Table 4.2, where the total number of parameter combination is 3,200. The range is slightly modified from that proposed in the previous study [32] for the purpose of reducing the number of candidates. This aims to eliminate unnecessary parameters for representing the response in exercise under hot environment.

3.2.4 Consideration of Solar Radiation

Since Gagge's two-node model does not have an equation to represent the heat energy received via solar radiation despite its importance for assessing human body temperature in hot environment, we combine an extensional equation to the heat balance calculation in the skin node . We designed the received heat energy q_{dn} based on the reference [66]. Figure 3.3 illustrates the overview of heat absorption in the two-node model by direct solar radiation. In our case, the energy q_{dn} is defined as:

$$ q_{dn} = a \cdot A_p \cdot J_{dn}, \tag{3.3} $$

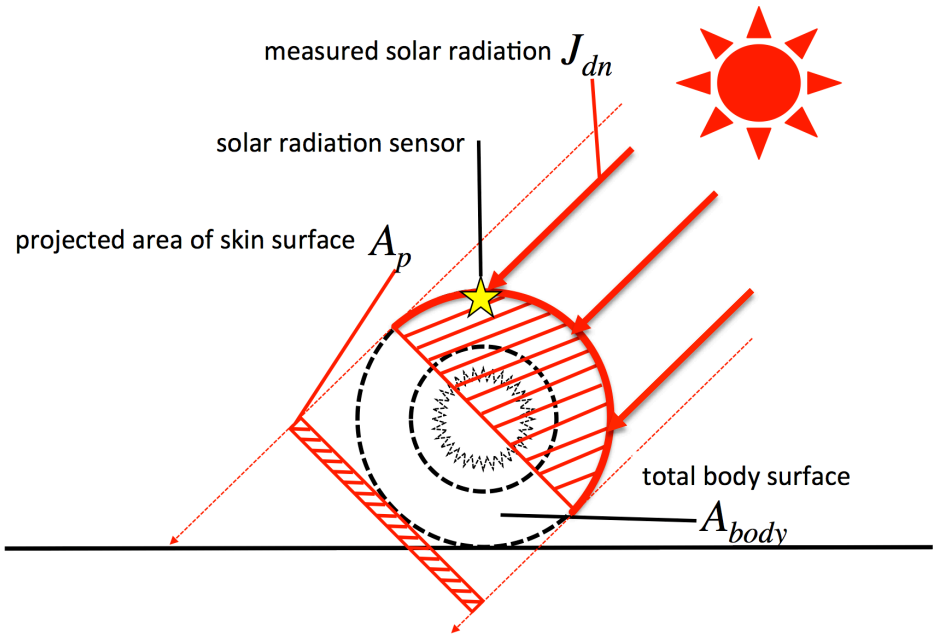


Figure 3.3: Consideration of solar radiation

where a is a coefficient of solar absorption in the skin surface and A_p is the projected area of the human body to a plane perpendicular to the direct solar radiation. We used $a = 0.4$ based on Reference [67]. J_{dn} denotes the measured solar radiation. Since A_p is difficult to be assessed on the real human body, we approximate the human body as a sphere by respecting Gagge's proposal. Thus, the projected area A_p is regarded as a quarter of the total area of human body surface as A_{body} (illustrated as Figure 3.3). Therefore, given measured solar radiation J_{dn} in each unit time, we consider the effect of solar radiation by combining the heat energy q_{dn} into the equation (3.7) in the next section, which represents the heat balance in the skin node.

3.3 Gagge's Two-node Model

This section elaborates on the detail of Gagge's two-node model, the method to give the sensor value, and its parameters representing difference of thermal response.

3.3.1 Detail of Two-node Model

Our method uses the two-node model originally proposed by Gagge et al. [26]. As indicated by its name, the model regards a human body as a combination of two nodes: the core and the skin node. The human body generally produces heat inside body (i.e., the core node), and then it emits heat from its surface (i.e., the skin node) to the ambient air by increase of skin blood flow and sweat evaporation.

General Equations in the Two-node Model

The model denotes the temperature of the core node as T_{core}, and that of the skin node as T_{skin}. With two fundamental equations as below, it simulates these body temperatures which chronologically vary by the environmental factors and/or exercise workload.

$$T_{core}{}^{t} = T_{core}{}^{t-1} + \Delta T_{core}{}^{t-1}, \tag{3.4}$$

$$T_{skin}{}^{t} = T_{skin}{}^{t-1} + \Delta T_{skin}{}^{t-1}. \tag{3.5}$$

Here, ΔT_{core} and ΔT_{skin} indicate changes of core and skin temperatures per unit time (we use 1 minute). The model repeats these calculation of ΔT_{core} and ΔT_{skin} for each time unit, and estimates the continuous changes of both temperatures. We note that we set the measured tympanic and skin temperatures to $T_{core}{}^{0}$ and $T_{skin}{}^{0}$ as the initial values instead of fixed values ($36.6°C$, $34.1°C$) used in the Gagge's proposal. Temperature changes of each node are defined as below.

$$\Delta T_{core} = \frac{M - W - q_{res} - (q_{cond} + q_{blo})A_{body}}{m_{core} \cdot c_{core}}, \tag{3.6}$$

$$\Delta T_{skin} = \frac{(q_{cond} + q_{blo} - q_{rsw} - q_{diff} - q_{conv} - q_{rad})A_{body}}{m_{skin} \cdot c_{skin}}. \tag{3.7}$$

In these equations, the change of temperature is calculated by dividing the sum of received and released heat energy by the product of the node weight and its specific heat. In short, the heat is produced by metabolism, then some of the heat is transferred to the skin node via blood flow and direct conduction, thereafter the remaining heat energy is used for rise of temperature in the core node (equation (3.6)). Similarly, the received heat in the skin node is emitted to the ambient air by the convection, radiation, sweat evaporation and water diffusion from skin surface, and remained heat energy is used for rise of temperature in the skin node (equation (3.7)). Hereafter, we further describe each component in these equations. We note that we highlight the given values and we omit the superscript t indicating time for simplicity, whenever it is unnecessary .

Heat Energy Calculation in the Core Node

The heat energy in the core node consists of (1) heat gain by metabolic heat production M, (2) energy used for physical activity W, (3) heat emission to the ambient air by respiration q_{res}, (4) heat transfer to the skin node by direct conduction q_{cond}, (5) heat transfer to the skin node via blood flow q_{blo}. The heat production M and the energy used for exercise W are the values to be given manually, accordingly we describe them in the next section. The other factors are defined as follows.

The heat loss by respiration q_{res} :

$$q_{res} = 0.0023 \cdot M(44 - \boldsymbol{\phi_{air}} \cdot P_{air}), \tag{3.8}$$

where the humidity ϕ_{air} is given by the sensor measurement and the saturated vapor pressure of the ambient air P_{air} is calculated with use of the air temperature (see the next section).

The heat loss by direct conduction q_{cond} :

$$q_{cond} = K_{min} \cdot (T_{core} - T_{skin}). \tag{3.9}$$

Here K_{min} means the minimum thermal conductance of the skin node. This equation indicates the heat transfer increases according to large gradient of the temperatures for the both nodes.

The heat loss by blood flow q_{blo} :

$$q_{blo} = c_{blo} \cdot V_{blo} \cdot (T_{core} - T_{skin}), \tag{3.10}$$

$$V_{blo} = \frac{6.3 + 75 \cdot (T_{core} - \boldsymbol{T_{core}}^{\mathbf{0}})}{1 + 0.5 \cdot (\boldsymbol{T_{skin}}^{\mathbf{0}} - T_{skin})}. \tag{3.11}$$

where the heat loss is represented as the product of the specific heat of blood, the volume of skin blood flow, and the gradient of the temperatures of the two nodes. The amount of skin blood flow V_{blo} increases according to the rise of core temperature. Note that increase of T_{skin} does not affect V_{blo} by regarding $(\boldsymbol{T_{skin}}^{\mathbf{0}} - T_{skin})$ as zero if $(T_{skin} \geq \boldsymbol{T_{skin}}^{\mathbf{0}})$. These equations mean the blood flow rate increases according to rise of core temperature, then the increase of the skin blood flow makes large heat transfer from the core node to the skin node. In summary, the heat energy increases by metabolism (e.g., intense exercise), then the heat energy is transferred to the skin node due to the gradient of the temperatures of the both nodes. At the same time, the human body increases its blood flow on the skin surface to efficiently emit the heat energy from the core node to the skin node.

Heat Energy Calculation in the Skin Node

The heat energy calculation in the skin node is composed of (1) heat transfer from the core node by direct conduction q_{cond}, (2) heat transfer from the core node via blood flow q_{blo}, (3) heat emission by sweat evaporation q_{rsw}, (4) heat emission by water diffusion on the skin surface q_{diff}, (5) heat exchange with the ambient air q_{conv}, (6) heat emission from the skin surface to the air by radiation q_{rad}. The components (1) and (2) were already described in the above, we describe the rest as below.

The heat loss by sweating q_{rsw} :

$$q_{rsw} = \min(q'_{rsw},\ 0.94 \cdot E_{max}), \tag{3.12}$$

$$q'_{rsw} = 0.7 \cdot m_{rsw} \cdot 2^{(T_{skin} - \boldsymbol{T_{skin}}^{\mathbf{0}})/3}, \tag{3.13}$$

$$m_{rsw} = \{250(T_{core} - \boldsymbol{T_{core}}^{\mathbf{0}}) + 100(T_{core} - \boldsymbol{T_{core}}^{\mathbf{0}})(T_{skin} - \boldsymbol{T_{skin}}^{\mathbf{0}})\} \cdot \frac{1}{1000}, \tag{3.14}$$

$$E_{max} = 2.2 \cdot h_{conv} \cdot (P_{skin} - \boldsymbol{\phi_{air}} \cdot P_{air}) \cdot F_{pcl}. \tag{3.15}$$

Where q'_{rsw} and E_{max} are ideal heat loss by sweating and theoretical maximum heat loss by water evaporation on the skin surface, respectively. This means the sweating is capable to make much heat lost from skin surface, meanwhile its capability is limited if the skin surface becomes entirely wet. Whenever q'_{rsw} does not exceed $0.94 \cdot E_{max}$, the actual heat loss q_{rsw} by sweat is equal to q'_{rsw}, otherwise the sweat evaporation rate saturates and the heat loss never exceeds $0.94 \cdot E_{max}$. q'_{rsw} is calculated by

Table 3.2: Constants

description	variable	value
skin specific heat capacity	c_{skin}	0.97
core specific heat capacity	c_{core}	0.97
blood specific heat capacity	c_{blo}	1.163
minimum skin thermal conductance	K_{min}	5.28
convective heat transfer coefficient	h_{conv}	4.3
radiation heat exchange coefficient	h_{rad}	5.23
efficiency for the passage of dry heat	F_{cl}	0.53
permeation efficiency factor	F_{pcl}	0.73

observing the amount of sweat m_{rsw} and the balance of the skin temperature against its initial value. The amount of sweat m_{rsw} becomes larger when the core temperature and the skin temperature make large gradient against their initial values. In equation (3.14), the first term indicates sweat increase by exercise and the second means basic sweat increase. E_{max} is assessed by the environmental contexts, where h_{conv} means the efficiency of heat transfer by convection between the skin and the air, F_{pcl} is a permeation efficiency factor, and P_{skin} denotes saturated vapor pressure on the skin surface.

The heat loss by insensible water-diffusion q_{diff} :

$$q_{diff} = 0.06 \cdot E_{max}. \tag{3.16}$$

This is the heat loss due to moisture always lost from the skin surface.

The combined heat transfer by convection and radiation q_{conv}, q_{rad} :

$$q_{conv} + q_{rad} = h_{total} \cdot (T_{skin} - \boldsymbol{T_{air}}) \cdot F_{cl}, \tag{3.17}$$

$$h_{total} = h_{conv} + h_{rad}. \tag{3.18}$$

The first equation means the heat is transferred from the skin to the air by observing the combined efficiency of convection and radiation, and the gradient of the tempeature between the skin surface and the air. F_{cl} indicates efficiency for the passage of dry heat from the skin surface through clothing to the air. We note that both of F_{pcl} and F_{cl} depend on the clothing insulation, but these values are fixed in our experiments in which we used similar types of clothing.

In these equations, we used the constant values defined in Table 3.2.

3.3.2 Manual Inputs to the Model

Here we elaborate on the essential inputs for the model described above. All the essential input values are summarize in Table 3.3 to 3.5.

Firstly, we give user's physical characteristics: the total area of the skin surface A_{body}, the body

mass of the core node m_{core}, that of the skin node m_{skin} as follows.

$$A_{body} = \boldsymbol{weight}^{0.425} \cdot \boldsymbol{height}^{0.715} \cdot \frac{71.84}{10000}, \tag{3.19}$$

$$m_{core} = \boldsymbol{weight} \cdot 0.95, \tag{3.20}$$

$$m_{skin} = \boldsymbol{weight} \cdot 0.05. \tag{3.21}$$

The total surface area A_{body} is assessed based on the subject's height and weight manually given [68]. The mass of each node is calculated from the total body mass [32].

Secondly, using measured heart rate (***heartrate***), the metabolic heat production is defined as following. Metabolic heat production M in equation (3.6) is estimated based on oxygen consumption $VO2$ derived from heart rate.

$$M = VO2 \cdot \boldsymbol{weight} \cdot \frac{5}{1000} \cdot \frac{1000 \cdot 4.186}{60}, \tag{3.22}$$

$$VO2 = VO2max \cdot \frac{\%VO2R}{100}, \tag{3.23}$$

$$VO2max = 15 \cdot \frac{maxHR}{\boldsymbol{restHR}}, \tag{3.24}$$

$$\%VO2R = 0.95 \cdot \%HRR + 6.8, \tag{3.25}$$

$$maxHR = 208.7 - 0.7 \cdot \boldsymbol{age}, \tag{3.26}$$

$$\%HRR = \frac{\boldsymbol{heartrate} - \boldsymbol{restHR}}{maxHR - \boldsymbol{restHR}} \cdot 100. \tag{3.27}$$

M is basically a product of mass-specific oxygen consumption $VO2$ and body weight (***weight***). The rest of equation (3.22) indicates unit conversion: firstly transferred from milliliter/minute to kilocalorie/minute by multiplying $\frac{5}{1000}$ according to reference [69], then converted kilocalorie/minute to Watt by multiplying $\frac{1000 \cdot 4.186}{60}$. The mass-specific oxygen consumption $VO2$ is calculated by equations (3.23) and (3.24) where the maximum oxygen consumption $VO2max$ derived from [70] is multiplied by $\%VO2R$. $\%VO2R$ indicates the oxygen consumption rate against its maximum value and is approximated by using the heart rate reserve $\%HRR$ as (3.27). $\%HRR$ is derived from the current heart rate (***heartrate***), the maximum heart rate $maxHR$ (3.26), and resting heart rate ***restHR*** according to [71, 72, 73].

Thirdly, the kinetic energy used for exercise W in equation (3.6) is calculated as the production of the metabolic heat energy M and kinetic energy efficiency Δ_{eff}:

$$W = M \cdot \boldsymbol{\Delta_{eff}}. \tag{3.28}$$

Δ_{eff} is a specific value for each exercise, for instance, 0.40 in walking, 0.44∼0.54 in running, and 0.23 in biking [74, 75].

Finally, the saturated vapor pressure due to skin temperature P_{skin} and ambient air temperature P_{air} are calculated by reference [76] as below.

$$P_{skin} = 6.11 \cdot 10^{(7.5 \cdot T_{skin}/(T_{skin}+237.3))}, \tag{3.29}$$

$$P_{air} = 6.11 \cdot 10^{(7.5 \cdot \boldsymbol{T_{air}}/(\boldsymbol{T_{air}}+237.3))}. \tag{3.30}$$

Table 3.3: Manual input

variable	description
$\boldsymbol{weight}$	Subject's Body Weight
$\boldsymbol{height}$	Subject's Body Height
$\boldsymbol{restHR}$	Measured Resting Heart rate
$\boldsymbol{age}$	Subject's Age
$\boldsymbol{T_{core}}^{\mathbf{0}}$	Initial Core Temperature (measured by [7] or [77])
$\boldsymbol{\Delta_{eff}}$	Kinetic Energy Efficiency of Exercise

Table 3.4: Input by wearable sensors [2]

variable	description
$\boldsymbol{T_{skin}}^{\mathbf{0}}$	Initial Skin Temperature
$\boldsymbol{heartrate}$	Measured Heart rate

Table 3.5: Input by environmental sensors [3]

variable	description
$\boldsymbol{T_{air}}$	Measured Air Temperature
$\boldsymbol{\phi_{air}}$	Measured Air Humidity

3.3.3 Thermal Response Parameters

For more precise simulation of human thermoregulation, Takada et al. have introduced individual parameters to some equations of Gagge's model [32]. They replaced constant values defined in equations (3.11) and (3.14) to variable parameters in order to represent the difference of the thermal regulation system (i.e., sweating and blood flow increase) for each person. Given the parameters $pr_1, pr_2, pr_3, pr_4, pr_5, pr_6$, equations (3.11) and (3.14) are redefined as below.

$$m_{rsw} = pr_3 \cdot (T_{core} - pr_1) \cdot (T_{skin} - pr_2) \cdot \frac{1}{1000}, \tag{3.31}$$

$$V_{blo} = \frac{pr_4 + pr_5 \cdot (T_{core} - pr_1)}{1 + pr_6 \cdot (pr_2 - T_{skin})}. \tag{3.32}$$

However, these equations are not suitable for our scenario (i.e. exercise in hot environment) because of the following reasons: (1) the term in equation (3.31) describes only the sweat production during rest; (2) the denominator of equation (3.32) represents vasoconstriction response in cold environment, which never happens in hot environment; (3) the parameters pr_1, pr_2 mean the initial values of core and skin temperature, however they are given by measuring these temperatures before exercise. Therefore, we do not use their original proposal, rather refine their equations to suit our scenario by using parameters

Table 3.6: Details of outdoor walking

Date	August 13-21, September 1-5 (10 Days)
Time	1 hour between 13:00-15:00
Subject	6 males and 1 female
Location	Sidewalk around Expo'70 Commemorative Park
Exercise	Walking at 5km/hour
Measured data	Wrist skin temperature, Heart rate, Core temperature (eardrum)
	Ambient temperature, Ambient humidity, Solar radiation
Weather	5 Days of Sunny and 5 Days of Cloudy
Temperature	31.2 ± 3.9 [℃] (Mean±SD)
Humidity	61.0 ± 13.1 [%] (Mean±SD)

Table 3.7: Subject information

ID	**Age**	**Height**[cm]	**Weight**[kg]	**Gender**
A	23	178	78	male
B	22	172	80	male
C	24	163	63	male
D	22	179	80	male
E	24	160	48	female
F	23	177	80	male
G	23	174	98	male

$\alpha_1, \alpha_2, \alpha_3, \alpha_4$ and measured initial temperatures ${T_{core}}^0, {T_{skin}}^0$ as below.

$$V_{blo} = \alpha_1 + \alpha_2 \cdot (T_{core} - {T_{core}}^0), \tag{3.33}$$

$$m_{rsw} = \{\alpha_3 \cdot (T_{core} - {T_{core}}^0) + \alpha_4 \cdot (T_{core} - {T_{core}}^0) \cdot (T_{skin} - {T_{skin}}^0)\} \cdot \frac{1}{1000}. \tag{3.34}$$

We here added α_3 in the same manner as α_4 to reproduce the sweating response during exercise as described in Gagge's model [26]. Also we simplified equation (4.5) by omitting vasoconstriction response. As described in Section 3.2, these four parameters will be optimized by observing a change of the skin temperature during exercise.

3.4 Performance Evaluation

3.4.1 Environment and Settings

We have evaluated the proposed method through real experiments where totally 52 exercise instances were collected in heat environment. Seven participants collected 1-hour walking exercise under the condition described in Table 3.6. The subjects consist of six males and one female as summarized in

Table 3.7. They wore Basis [2] smartwatch on their wrists to measure both skin temperature and heart rate and performed exercise for 60 minutes. Also they wore a tympanic thermometer DBTL-2 [7] to measure their tympanic temperature as the ground truth of core temperature. Note that we can safely measure the tympanic temperature by an infrared thermometer during walking since the exercise does not involve body contacts, despite the sensor requires to insert a measurement probe into an ear. They wore the similar clothes to set clothing-related parameters in the two-node model, and one of them bring an environmental sensor WBGT-203B[3] to measure temperature and humidity of the ambient air and a solar radiation sensor ML-01[78] to measure actual solar radiation during walking.

We used the absolute error averaged over time as a metric of performance to represent error at time t as:

$$\frac{1}{t} \cdot \sum_{i=1}^{t} |{T_{core}}^{i}(\theta_{opt}) - \hat{{T_{core}}^{i}}|,$$

where $\hat{{T_{core}}^{i}}$ indicates the core temperature actually measured at time i. This metric means the similarity of the core temperature curve over time to the ground truth. In other words, it represents how well the estimation reproduces the real human response.

Hereafter, we name our method as PROP for simplicity. For comparing PROP with other methods of parameter calibration, we introduce these two methods: (1) DEF: using standard parameters of the model (i.e., using default values as proposed by Gagge et al.), (2) OPT: using the optimal parameters based on actual core temperature (i.e., this shows the minimum error against ground truth), We note that OPT is the case of determining individual parameter set which achieves the minimum error against actual core temperature, despite the core temperature cannot be measured in the real scenario. Therefore, OPT indicates the theoretical minimum error achievable with the two-node model and its parameters. We introduce DEF to show the effectiveness of our method (i.e., the effect of calibrated parameters against using the their original values).

Finally, we measured the computation time for parameter calibration throughout 3,200 candidates of parameter set for 60-minute exercise using a workstation with Intel Xeon 2.66 GHz and 23.6 GB memory. It proves the average computation time for parameter calibration is about 1 second, highlighting our method is enough for the usage in real-time.

3.4.2 Result

We firstly evaluate the final performance of the method, namely the average error after 60-minute exercise finished. Figure 3.4 illustrates mean absolute error of PROP, OPT and DEF for subject A. This indicates PROP can reduce overall error by detecting better parameter set than DEF. It also shows an instance (2 Sept.) in which PROP gets worse than DEF. This infers our method may fail to find the better parameter set when the optimal parameter set is almost equivalent to default one. Nonetheless, the effect of wrongly determined parameter set is limited. We confirmed that PROP have worked better than DEF through almost all instances of seven subjects.

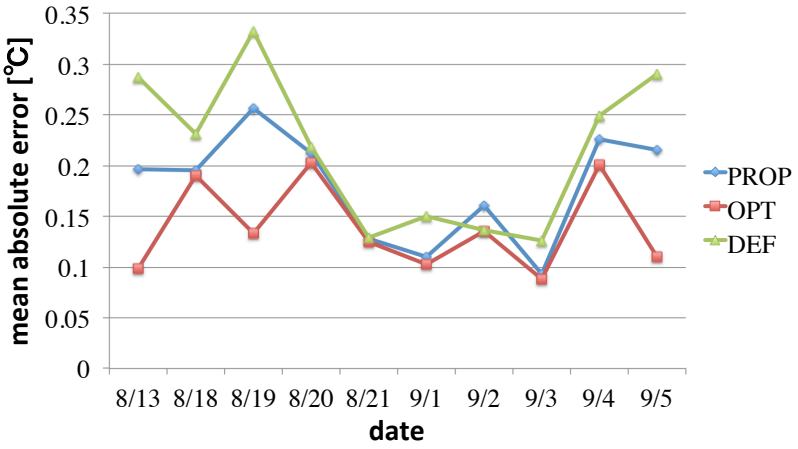


Figure 3.4: Overall errors for each day (Subject A)

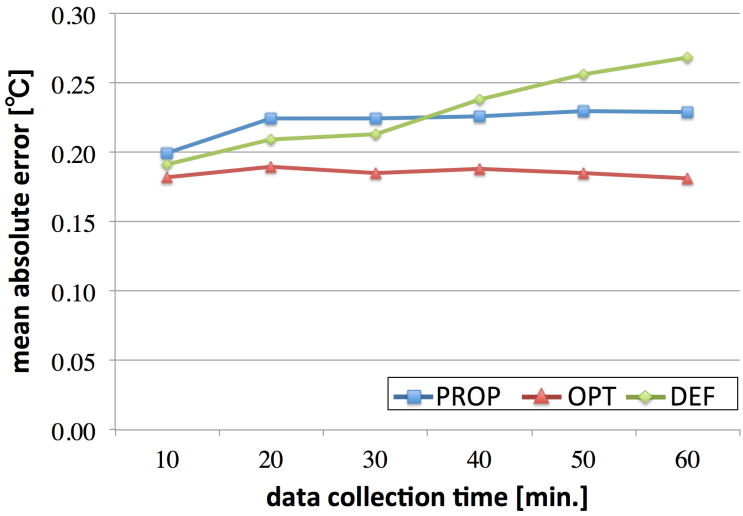


Figure 3.5: Average estimation error vs. calibration time

Next, we discuss the relationship between the time-course for calibration and the estimation error while varying the length of dataset used for calibration. Figure 3.5 illustrates the relationship between the averaged absolute error and 10- to 60- minute data length. We find the average error of DEF increases over time, despite the parameter set is fixed. It is also expectable that the error of DEF will be further greater in over 60-minute exercise. This is due to the two-node model itself which could not reproduce the actual transition of core temperature well. Evidently, OPT also shows inevitable error (0.18 $^{\circ}C$) between the two-node model and ground truth. Nevertheless, each method shows different error according to time course due to different capability of parameter calibration. The error of PROP

is larger than DEF immediately after exercise start due to wrongly calibrated parameters by small instances. In contrast, PROP shows that the error converges to 0.24 $^{\circ}C$ after 40 minutes, suggesting PROP could choose better parameter set than using default values. When 60 minutes elapsed, PROP shows error as 0.24 $^{\circ}C$ which is 12% error mitigation against that of DEF. Consequently, PROP achieves the smaller error compared to DEF, when the longer dataset is given for calibration. We also confirmed that the overall performance (i.e., average error after 60 minutes elapsed) of PROP, OPT and DEF were $0.04 \pm 0.38^{\circ}C$, $0.03 \pm 0.25^{\circ}C$ and $-0.22 \pm 0.36^{\circ}C$. From these results, we confirmed the advantage of PROP for online parameter calibration by observing the change of skin temperature.

3.5 Conclusion

This chapter described our proposal to estimate human core body temperature based on the two-node model. Our method employs wrist-worn devices to measure the skin temperature and heart rate, and then observes the change of skin temperature to determine the optimal parameter set of the two-node model. The optimal parameter set is determined by comparing the measured skin temperature and each of exhaustively generated 3,200 skin temperatures. The evaluation results show our method is capable to improve the performance by 12% at most compared with using the default parameter set in 60-minute outdoor walking exercise. It also proved our method could find the best combination the parameter within 1 minute, highlighting capability for usage in real-time. We note that some contents in this chapter refer our previous publications [79, 80, 81].

Chapter 4

Core Temperature Estimation in Complex Environment and Exercise

4.1 Introduction

In recent decades, the threat of global warming and catastrophic heat waves has triggered immense interest in heat stroke. Heat stroke causes severe health disorders that sometimes lead to death. To prevent heatstroke, health manuals are provided by World Health Organization (WHO) [4]. Environmental heat indices such as the wet bulb globe temperature (WBGT) and standard effective temperature (SET) can also evaluate the heat stress and provide guidelines for outdoor work and exercise. However, these approaches remain insufficient, thus the number of heat stroke patients has not decreased in recent years.

These methods fail mainly because environmental indicators cannot account for the rising core temperature inside human bodies. High core temperature is a fundamental precursor of heat stroke [5], but cannot be measured during exercise because core temperature is usually measured as the rectal, tympanic or esophageal temperature. In other words, it requires invasive thermometers and clinical authorization. Sensors that measure core temperature without clinical authorization have been recently released. For instance, CorTemp [6], a wireless ingestible thermometer, can continuously measure the core temperature, and the 3M SpotOn System [8] measures the core temperature indirectly by a probe placed on the forehead. However, CorTemp uses disposable sensors that are costly to replace, and 3M SpotOn System requires a wired control unit. Therefore, neither of these sensors are suitable for daily core-temperature monitoring. DBTL-2 [7] measures the tympanic temperature by an infrared thermometer probe, but continuous insertion of a measurement probe is disallowed in situations such as contact sports, in which players collide with each other.

Therefore, the core temperature during exercise should be estimated only from the available information. Several recent methods have estimated the core temperature using non-invasive sensors. Combining heart rate observations with a Kalman filter, Buller et al. estimated core temperatures

as hidden states [9]. The root mean squared error of their algorithm (over multiple datasets) was 0.21 $°C$. However, as acknowledged by the authors, this algorithm must overcome several challenges. First it must measure the actual core temperature by ingestible pills to determine the optimal bias (a model parameter). The bias must then be calibrated to suit the subject's age and fitness level. Yabuki et al. proposed a heat stroke alert system for outdoor workers that integrates computational fluid dynamics (CFD), sun and shade simulations and a core temperature estimation model [10]. Nevertheless, the three-dimensional CFD, sun and shade model must be constructed and simulated beforehand. Although these approaches estimate the core temperature with non-invasive sensors, they inevitably require preliminary investigation of the bias or the environment.

To solve the above problems, we have proposed a core temperature estimation method that integrates a Gagge's two-node model with wearable and environmental sensor measurements [79, 80, 81]. The method does not calibrate four parameters in advance. Instead, it instantly calibrates the four parameters by comparing the actual and simulated skin temperatures. We found the method could reasonably estimate the core body temperature during simple exercise, but the skin temperature for the calibration had to be measured for over 40 minutes and the parameters were difficult to estimate during complex exercise with variable intensity (e.g., ergometer exercise with variable intensity).

To address these challenges, we propose a novel method for core temperature estimation that satisfies the following requirements: (1) the core temperature is estimated by non-invasive sensors and minimal measurements; (2) on-site and lightweight calibration; (3) applicability to complex exercises involving warming up, practice, games, breaks, and other activities. Our method is based on Gagge's model and computes the core and skin temperature from the measurements of wearable and environmental sensors. For reproducing the human thermal response to variable exercise intensity, we represent the different thermal responses (i.e., the speed of sweating and blood flow response, and the delay of heat transfer between core and skin nodes) by two parameters. The effort for calibration is reduced by an on-site calibration method that samples the actual core temperature during breaks.

Our method was evaluated in real experiments conducted over more than 120 hours. The experimental activities included walking, running, biking, and tennis. Throughout rigorous evaluation using these datasets, we confirmed our method successfully estimated the core temperature with an average absolute error between 0.20 and 0.30 $°C$. We also developed a prototype of heat stroke caution system which notifies high core temperature in a particular threshold. It is evaluated through walking, running and biking datasets, and the performance of its notification (allowing 5-minute error against the actual timing) is proven as 0.514 precision and 1.0 recall in biking, 0.965 precision and 0.899 recall in running and walking.

Our contributions are summarized as below.

(1) The difference between the actual and simulated core temperatures is represented by two parameters. Our core temperature estimation method is applicable to complex exercise (interval running and walking, biking with variable exercise intensities and tennis practice).

(2) Our parameter calibration method is suitable for real exercise, which requires quick measurement of core temperature during breaks. The measurement is carried out within 15 seconds by a commercial-off-the-shelf infrared thermometer.

(3) To adapt Gagge's model to our scenario, we carefully integrate the effects of solar radiation, wind, and water ingestion into the original Gagge's model.

(4) The prototype of a heat stroke alert system shows moderate precision and recall even for complicated exercise data sets including walking, running and biking with interval breaks.

The rest of this chapter is structured as follows. We describe the method for instant calibration of six parameters in section 4.2. Various extensions to Gagge's two-node model are presented in section 4.3. Section 4.4 evaluates our proposed method in real experiments extending over 120 hours. A prototype of heat stroke alert system will be demonstrated in Section 4.5. Finally, section 4.6 concludes the chapter.

4.2 Improved Method for Core Temperature Estimation

In this section, we firstly describe the core temperature estimation and parameter calibration in our scenario. Then we briefly describe the fundamentals of Gagge's two-node model used in the core temperature simulation. Thereafter, we describe the parameters representing the thermoregulation differences. Finally, we describe the calibration method for the model parameters.

4.2.1 Use Scenario

Figure 4.1 illustrates the overall flow of our core temperature estimation during exercise, based on Gagge's two-node model with parameter calibration. In our scenario, subjects exercise with wearable sensors that measure their heart rates and skin temperatures. The ambient temperature, humidity and solar radiation are measured by environmental sensors deployed in the exercise field. To start the temperature simulation, we input the static information of the subjects (age, height, weight, and resting heart rate) and other information (clothing insulation, initial core temperature, and kinetic energy efficiency of the exercise) as described in Section 3.3.

After launching the simulation, we continuously input the sensor measurements (i.e., air temperature, humidity, solar radiation and heart rate) to Gagge's model for sequential prediction of the core temperature. These steps constitute the whole procedure of core temperature estimation. We also propose several extensions to Gagge's model. Specifically, we add three important heat transfer factors: solar radiation which increases the skin temperature, wind which assists heat transfer from the skin surface to the ambient air, and cold water intake which reduces the core temperature. We note that consideration of solar radiation is more precisely than that proposed in Chapter 3, namely we here consider the effect of human posture and clothing.

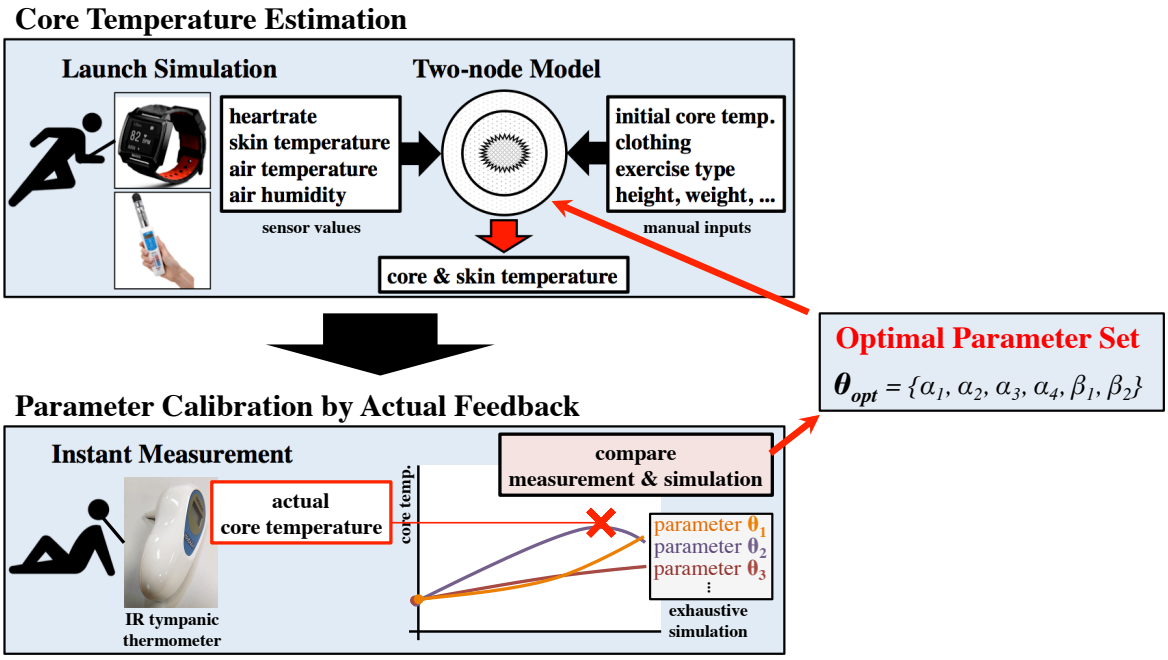


Figure 4.1: Scenario

4.2.2 Temperature Simulation by Gagge's Two-node Model

We here briefly summarize the essential equations in Gagge's two-node model. As previously described in Figure 3.1, the model approximates the human body as a sphere composed of the core node and the skin node. The skin and core node temperatures are simulated by calculating the heat transferred between two nodes and the ambient air. The two-node model is simpler than other multi-node models, as it contains fewer nodes [29]. The model updates the initial core and skin temperatures (which must be given in advance) based on the heat transfer between two nodes and the ambient air per unit time. The details of Gagge's model and the application procedure of the sensor data are provided in Section 3.3. We here show the fundamental equations as below.

$$\Delta T_{core} = \frac{M - W - q_{res} - (q_{cond} + q_{blo})A_{body}}{m_{core} \cdot c_{core}}, \tag{4.1}$$

$$\Delta T_{skin} = \frac{(q_{cond} + q_{blo} - q_{rsw} - q_{diff} - q_{conv} - q_{rad})A_{body}}{m_{skin} \cdot c_{skin}}. \tag{4.2}$$

Equations (4.1) and (4.2) respectively calculate the change of core temperature ΔT_{core} and skin temperature ΔT_{skin} per unit time. The core and skin temperatures at time t are computed as follows:

$$T_{core}{}^{t} = T_{core}{}^{t-1} + \Delta T_{core}{}^{t-1}, \tag{4.3}$$

$$T_{skin}{}^{t} = T_{skin}{}^{t-1} + \Delta T_{skin}{}^{t-1}. \tag{4.4}$$

By repeating these calculations, the model simulates the continuous changes of both temperatures.

4.2.3 Parameters Representing Difference of Thermoregulation

Despite its simplicity, Gagge's two-node model performs accurately when the individual parameters are appropriately given. Takada et al. introduced six parameters in two equations that account for individual differences in thermoregulation function [32]. In our study [80, 81], we modified their equations to suit our scenario: the initial core and skin temperatures are measured and the sweating response to exercise is represented by a new parameter. The refined equations are follows.

$$
\begin{aligned}
{V_{blo}}^t &= \alpha_1 + \alpha_2 \cdot ({T_{core}}^t - {T_{core}}^0), && (4.5)\\
{m_{rsw}}^t &= \{\alpha_3 \cdot ({T_{core}}^t - {T_{core}}^0) + \alpha_4 \cdot ({T_{core}}^t - {T_{core}}^0) \cdot ({T_{skin}}^t - {T_{skin}}^0)\} \cdot 10^{-3}. && (4.6)
\end{aligned}
$$

In these equations, the volume of skin blood flow and the amount of sweat increase when the core and the skin temperatures rise due to exercise and/or hot environment. The four parameters $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ represent the difference in the increasing speed of blood flow and sweat. Although this approach achieved performance improve on the core temperature estimation in the exercise with constant load, it did not showed performance improve in complex exercises.

To address this challenge, we propose two additional parameters that reproduce the actual core temperature response during complex exercise with variable intensity. With new parameters β_1 and β_2, we define the equations of skin blood flow ${V_{blo}}^t$, sweat amount ${m_{rsw}}^t$, and metabolic heat production M^t at time t as below.

$$
\begin{aligned}
{V_{blo}}^t &= \alpha_1 + \alpha_2 \cdot ({T_{core}}^{t-\beta_2} - {T_{core}}^0), && (4.7)\\
{m_{rsw}}^t &= \alpha_3 \cdot ({T_{core}}^{t-\beta_2} - {T_{core}}^0) + \alpha_4 \cdot ({T_{core}}^{t-\beta_2} - {T_{core}}^0) \cdot ({T_{skin}}^{t-\beta_2} - {T_{skin}}^0), && (4.8)\\
M^t &= \frac{1}{\beta_1} \cdot \sum_{j=t-\beta_1+1}^{t} M^j && (4.9)
\end{aligned}
$$

The detail of these equations and parameters β_1, β_2 is described in section 4.3.1. These parameters are calibrated using occasional measurements of actual core temperature, as described in section 4.2.4.

4.2.4 Parameter Calibration

The parameter calibration method improves the accuracy of the estimation. In our previous studies [79, 80], we proposed a calibration method that is suitable for exercise. The calibration criterion is the actual skin temperature. However, we found two shortcomings in the method: (1) the calibration requires over 40-minute measurements of skin temperature and (2) an effective parameter set is not always found, because the skin and core temperatures respond differently to the various phases of complex exercise.

To address these issues, this chapter proposes a new calibration method requiring only occasional measurements of core temperature. During warm-ups and breaks, the core temperature can be measured by an infrared tympanic thermometer such as [77]. The infrared thermometer provides quick and

easy measurements. To minimize the burden to users, our approach calibrates all parameters using a single measured core temperature. The parameters are calibrated as follows. Suppose that unit time t has elapsed since the start of an exercise session. The static information and the sensor measurements at times 0 to $t-1$ are input to the model. Then, given a set θ_i of individual parameter values, the two-node model estimates the following sequence $\boldsymbol{T_{core}}(\theta_i)$ of core temperatures at each unit time.

$$\boldsymbol{T_{core}}(\theta_i) = \{{T_{core}}^0, \ldots, {T_{core}}^t(\theta_i)\}, \tag{4.10}$$

where ${T_{core}}^j(\theta_i)$ denotes the estimated core temperature at time j for a given individual parameter set θ_i . The optimal parameter set θ_{opt} minimizes the squared error between the simulated and measured core temperatures:

$$\theta_{opt} = \arg\min_{\theta_i} \sum_{j \in \boldsymbol{t_{measure}}} ({T_{core}}^j(\theta_i) - \hat{T_{core}}^j)^2. \tag{4.11}$$

In this equation, $\hat{T_{core}}^j$ means the actual core temperature measured at time j, and $\boldsymbol{t_{measure}}$ is a set of times at which the core temperature is measured. Note that the number $\|\boldsymbol{t_{measure}}\|$ of measured core temperatures can vary in different situations. We used only one measurement through our experiments for minimizing user's effort. ${\boldsymbol{T_{core}}}^t(\theta_{opt})$ is the estimation result.

4.3 Extended Two-node Model

This section proposes various modifications to Gagge's two-node model. These modifications correct for several factors that deviate the simulated core temperature from the real one. The extended model includes (1) the delay of thermal response and heat transfer, (2) solar radiation, (3) wind, and (4) water ingestion. These extensions, along with the modifications to equations (4.1), (4.2), (4.5), and (4.6) of Gagge's two-node model, are summarized in Table 4.1. Equations (A-1), (B-1), (B-2), (C-1), and (D-1) in Table 4.1 are newly included in our extended model.

4.3.1 Representing Delay of Thermal Response

Figure 4.2 depicts the core temperatures (1) measured by the infrared sensor, (2) simulated by the original Gagge's model, and (3) simulated by our extended Gagge's model. The core temperatures in cases (1) and (2) deviate markedly, especially when subjects were resting. In Gagge's model, the subject's metabolism is indicated by heart rate; therefore, it immediately decreases at the start of a rest period. The core temperature also immediately decreases because the heat production declines. However, the human body does not respond in this way. Instead, it maintains a high temperature and sweats for a few minutes after the exercise is finished. Figure 4.2 also shows the initial fall of the actual core temperature, which was earlier reported in [82]. The authors of [83] argued that the core temperature initially falls by reflux blood from muscle, whose temperature is regularly lower than that of central blood. However, the initial fall was absent in some of our samples. Therefore, we conclude

Table 4.1: Equations of extended two-node model

Delay of Response (Sec.4.3.1)

(4.1′) $\Delta T_{core} = \dfrac{1}{m_{core} c_{core}} \cdot \{M' - W - q_{res} - (q_{cond} + q_{blo}) \cdot A_{body} - q_{water}\}$

(A-1) $M' = \dfrac{1}{\beta_1} \cdot \sum_{j=t-\beta_1+1}^{t} M^j$

(4.5′) ${V_{blo}}^t = \alpha_1 + \alpha_2 \cdot ({T_{core}}^{t-\beta_2} - {T_{core}}^0)$

(4.6′) ${m_{rsw}}^t = \{\alpha_3 \cdot ({T_{core}}^{t-\beta_2} - {T_{core}}^0) + \alpha_4 \cdot ({T_{core}}^{t-\beta_2} - {T_{core}}^0) \cdot ({T_{skin}}^{t-\beta_2} - {T_{skin}}^0)\} \cdot \dfrac{1}{1000}$

Solar Radiation (Sec.4.3.2)

(4.2′) $\Delta T_{skin} = \dfrac{1}{m_{skin} c_{skin}} \cdot \{(q_{cond} + q_{blo} - q_{rsw} - q_{diff} - q_{conv} - q_{rad}) \cdot A_{body} + q_{solar}\}$

(B-1) $q_{solar} = a \cdot A_p \cdot J_{solar}$

(B-2) $A_p = f_{eff} \cdot f_p \cdot f_{cl} \cdot A_{body}$

Wind (Sec.4.3.3)

(C-1) $h_{conv} = 13.36 \cdot v^{0.6}$

Water Ingestion (Sec.4.3.4)

(D-1) $q_{water} = m_{water} \cdot c_{water} \cdot ({T_{core}}^t - T_{water})$

that our model is not easily modified by a biological approach. Instead, we introduce a flexible and simple delay model that accounts for the initial fall and the discrepancy between the real and modeled data as described above. The delay model includes two delay parameters, β_1 and β_2. β_1 is the window size that delays the heat production M of the body. This parameter ensures the correct response when a subject starts to rest.

The heat production is delayed by a sliding window of size β_1. The larger the parameter, the greater the delay of thermal response of metabolism (see Eqs. (4.1′) and (A-1) in Table 4.1). The parameter β_2 considers the delay of the subject's thermal response (sweating and blood flow increase; see Eqs. (4.5′) and (4.6′)). In Gagge's two-node model, the subject's sweating and blood flow rate are computed from the increments in core and skin temperatures from their initial values. This model responds quickly to a temperature rise, but the response is much faster than the actual response. Hence, we simply delay the calculation of the sweating ${m_{rsw}}^t$ and blood flow rate ${V_{blo}}^t$ parameters at time t using the core and skin temperatures β_2 minutes ago. Namely, we calculate ${V_{blo}}^t$, ${m_{rsw}}^t$ based on the core temperature ${T_{core}}^{t-\beta_2}$ and the skin temperature ${T_{skin}}^{t-\beta_2}$ as described in Eqs. (4.5′) and (4.6′). Note that ${T_{core}}^j$ $(j < 0)$ equals ${T_{core}}^0$.

These parameters, and the other parameters (α_1, α_2, α_3, α_4 in Eqs. (4.5′) and (4.6′)), depend on the physical conditions, which vary on a daily basis. Hence, we optimize these parameters in

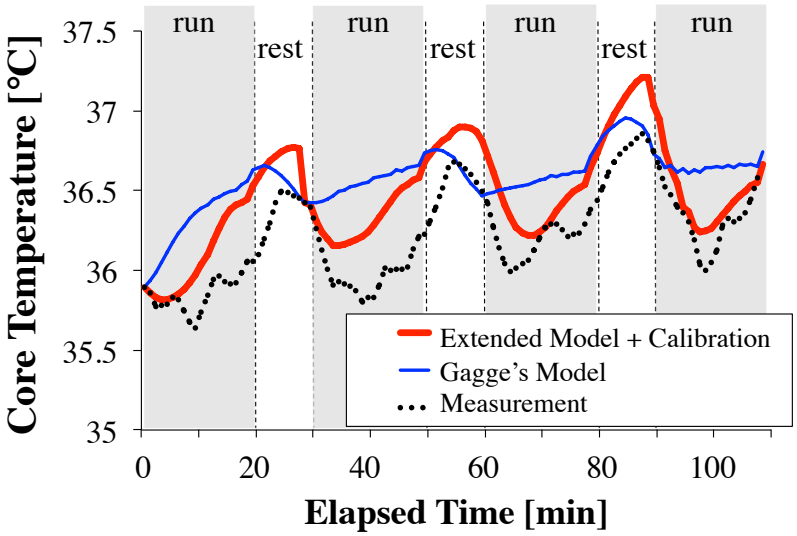


Figure 4.2: Discrepancy between simulated and measured core temperatures during a running exercise with breaks

our calibration. To determine the appropriate parameter ranges, we experimentally investigate the distributions of the calibrated β_1 and β_2 based on the dataset described in section V-B. The initial ranges of β_1 and β_2 were 1–15 and 0–10, respectively. We conducted exhaustive thermal simulations of 165 parameter combinations. The other individual parameters (α_1 to α_4) were fixed at their default values. The cumulative distributions of the optimal delay parameters β_1 and β_2 were calculated by Eq. (4.11) and are plotted in Figure 4.3. To balance accuracy and complexity (number of parameter candidates), we adjusted the range to 1–12 for β_1 and 0–5 for β_2. Both ranges cover 85% of the optimal parameters, but reduce the parameter combination size to 72. These delay parameters, and the four individual parameters described in section 4.2.4, were calibrated in the evaluation.

4.3.2 Consideration of Solar Radiation

Gagge's two-node model excludes the heat input from solar radiation, which significantly raises the skin temperature. To ensure that our model is applicable to both indoor and outdoor exercise, we integrate the amount of heat q_{solar} absorbed by the skin nodes under direct solar radiation into the two-node model. We note that we tried to combine the energy received by solar radiation to Gagge's model in our previous study [80], however the method could not consider the human attitude and worn clothing covering the skin surface. In contrast, we here consider the effect of solar absorption with original human shape wearing clothing. The heat energy q_{solar} of solar radiation was defined in [66], and is given by Eqs. (B-1) and (B-2) in Table 4.1. The variables a, A_p, and J_{solar} in Eq. (B-1) represent the solar absorption coefficient of skin, the projected area of the original human surface,

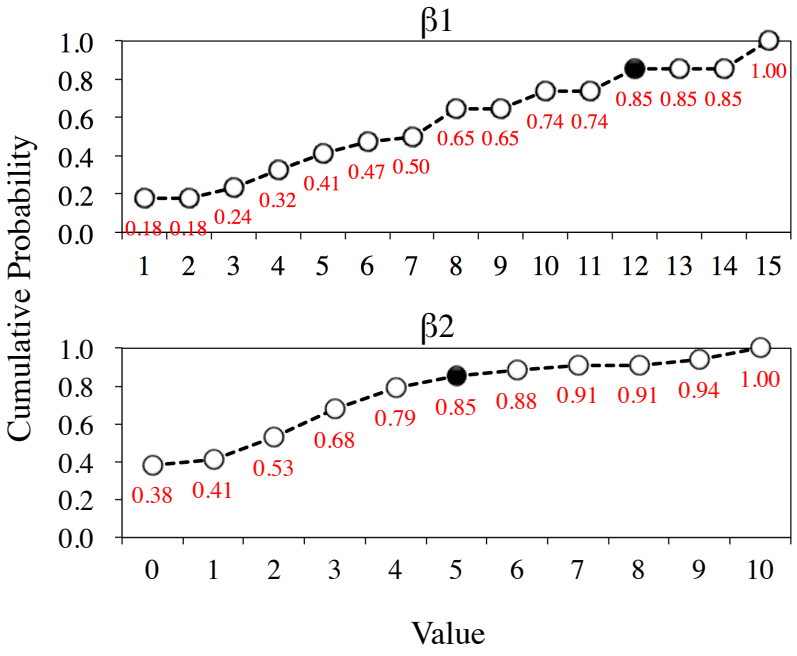


Figure 4.3: Distribution of optimal delay parameters

and the solar radiation power (measured by solar sensors), respectively. a is defined as in [84]. In Eq. (B-2), the projected area A_p is derived from the total surface area A_{body} of the human body, and the coefficients f_{eff}, f_p, and f_{cl}. During standing, the effective radiation area factor f_{eff} has been inferred as 0.725 [66], and the projected area factor f_p has been reported as 0.85 [85]. Considering the clothing worn by participants in our experiments, the ratio f_{cl} of unclothed skin surface (relative to total skin surface area) was set to 0.4.

4.3.3 Wind Effect

Although Gagge's model includes a sweating function, our experiments confirmed that it insufficiently models the heat lost by sweating during exercise. In the actual dataset described in section 4.4.2, the simulated heat loss by sweating saturated in over 48% of our outdoor experimental samples. However, this result is unnatural because it ignores the wind in real environments. Wind enhances the heat exchange between the skin surface and ambient air. Gagge's model represents the heat exchange by a fixed convection coefficient h_{conv}, although it depends on the wind velocity v [m/sec] (see [86] and Eq. (C-1) in Table 4.1). However, the speed of natural wind is difficult to measure as it largely depends on location and the surrounding buildings. Therefore, we consider the relative wind during the subject's movement (e.g., walking and running). The subject's speed is easily estimated from the distance traveled in a given time. Based on our experimental settings (section 4.4.2), we set v to 1.4 m/sec

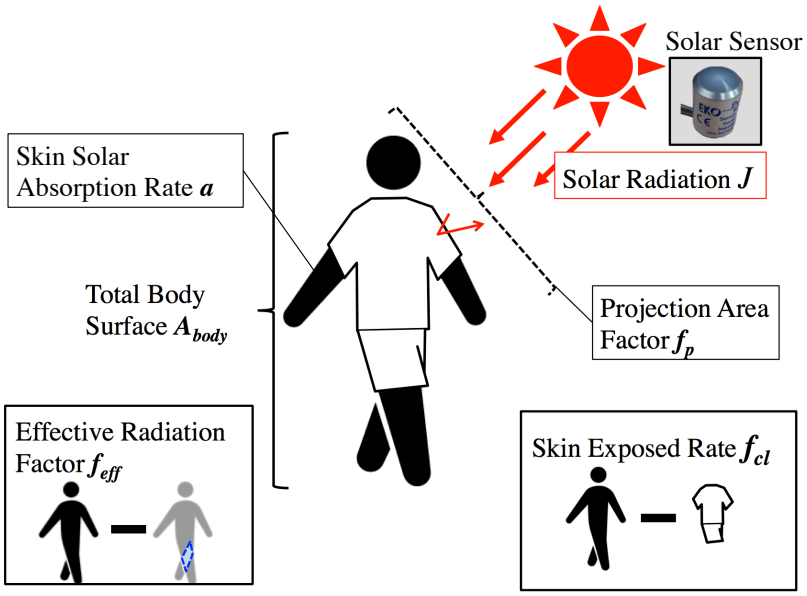


Figure 4.4: Detailed consideration of solar radiation

for walking and 2.5 m/s for running. After this modification, h_{conv} during walking and running was improved by 386% and 547%, respectively, over the default value. The relative wind velocity increases the maximum heat loss of sweating.

4.3.4 Decreasing Temperature by Drinking Water

In hot environments, we usually reduce our skin and core temperatures by drinking water and taking a rest. The authors of [87] compared the core temperatures of water-drinking and non-water-drinking subjects during exercise. By considering the temperature difference between water and the core nodes, they inferred that water intake encourages heat emission by sweat production, and that cold water ingestion reduces the core body heat. As Gagge's model assumes a continuously active sweating function, we consider only the heat lost by cold water ingestion. This heat transfer is described by a thermal energy formula as Eq. (D-1), in which c_{water}, m_{water} and T_{water} denote the specific heat of water, the amount of water ingestion and the water temperature, respectively. To alter the core temperature by this mechanism, we added the heat loss by cold water ingestion q_{water} as Eq. (4.1′).

4.4 Evaluation

4.4.1 Settings

Sensors

The proposed method was evaluated in walking, running, biking, and tennis practice experiments. In all exercises, the initial skin temperatures and heart rates of the subjects were measured by a Basis Peak [2] sensor worn on the wrist. The core temperature was assumed as the tympanic temperature measured by a DBTL-2 [7] or MC-510 [77]. During walking, running, and biking (which involve no body contacts), the tympanic temperature was continuously measured by a DBTL-2 measurement probe with an infrared thermometer. In tennis practice, subjects were requested to occasionally measure their tympanic temperature by an MC-510, because the continuous insertion of measurement probes is not appropriate in this exercise. The environmental information was measured by an ambient sensor WBGT-203B [3]. We also deployed a solar radiation sensor ML-01 [78] for measuring the solar radiation intensity.

Methods

As described in section 4.2.4, our method of six-parameter calibration requires at least one measurement of the subject's core temperature. Therefore, each sample was divided into two durations; one for calibration, the other for evaluation. The former and latter durations are denoted as $[0, t_1]$ and $[t_1, t_2]$, respectively, where t_1 is the time of measuring the core temperature by an infrared tympanic thermometer. Using Eq. (4.11) and the parameter candidates in Table 4.2, the calibrated parameter set θ_{PROP} in the proposed method is calculated as

$$\theta_{PROP} = \arg\min_{\theta_i} (T_{core}{}^{t_1}(\theta_i) - \hat{T_{core}}{}^{t_1})^2. \tag{4.12}$$

where $T_{core}{}^{j}$ and $\hat{T_{core}}{}^{j}$ respectively denote the estimated core temperature at time j and the actual core temperature measured at time j. Since we assume a single core temperature measurement at time t_1 in $[0, t_1]$, the actual core temperature is single-valued ($\hat{T_{core}}{}^{t_1}$). Hereafter, the proposed method is named **PROP**. We also introduce two additional methods. First is the **OPT** method, which uses the best parameter set determined by all core temperature measurements. This parameter set is given by

$$\theta_{OPT} = \arg\min_{\theta_i} \sum_{j=0}^{t_2} (T_{core}{}^{j}(\theta_i) - \hat{T_{core}}{}^{j})^2. \tag{4.13}$$

θ_{OPT} performs best because it uses all measured core temperatures in the calibration. Therefore, it indicates the theoretical limit of the core temperatures estimated by the extended two-node model with six parameters and calibration techniques. The other method, **DEF**, uses the default parameter set proposed by Gagge et al. [26], which assesses the baseline performance of the core temperature estimated by Gagge's two-node model. Note that our Gagge's two-node model integrates the extensions proposed in section 4.3.

Table 4.2: Ranges of individual parameters

Parameter	Range					# of candidates
α_1	5.04	5.67	**6.3**	6.93	7.56	5
α_2	22.5	$\cdots$	45	$\cdots$	**75**	8
α_3	75	$\cdots$	150	$\cdots$	**250**	8
α_4	30	$\cdots$	**100**	110	120	10
β_1	**1**	$\cdots$	6	$\cdots$	12	12
β_2	**0**	$\cdots$	3	$\cdots$	5	6

Table 4.3: Details of walking and running experiments

Date	July 28 - September 2 (10 Days)
Subject	6 males
Time	110 minutes between 12:00 and 15:00
Exercise	7.5 km Walking at 5km/hour & 10 km Running at 9km/hour
Measured data	Skin temperature, Heart rate, Core temperature (tympanic) Ambient temperature, Ambient humidity, Solar radiation
Temperature	33.3 ± 2.9 [$^\circ C$] (Mean±SD)
Humidity	45.6 ± 15.7 [%] (Mean±SD)

Error

As the performance metric, we employ the mean absolute error (MAE) of the core temperature in $[t_1, t_2]$. This metric, given by

$$\frac{1}{t_2 - t_1} \cdot \sum_{i=t_1+1}^{t_2} |{T_{core}}^i(\theta) - \hat{{T_{core}}^i}|, \tag{4.14}$$

computes the average distance between the estimated and measured core temperature curves.

4.4.2 Result in Walking and Running Experiments

We first evaluated the proposed method in real experiments conducted over 60 hours. In these experiments, 34 datasets were collected in a warm environment. The details of the experiment are shown in Table 4.3. The subjects were six males with average age, height, and weight of 22.8 ± 0.8 y, 173.5 ± 4.1 [cm] and 68.7 ± 8.1 [kg], respectively (where $\pm$ denotes standard deviation). Participants chose either running or walking according to their condition, as shown in Figure 4.5. The highlighted lines in this figure are the calibration durations. Figure 4.6 shows the route over which subjects were instructed to walk 7.5 km (3 rounds $\times$ 3 sets = 9 rounds) with 2 breaks or run 10 km (3 rounds $\times$ 4 sets = 12 rounds) with 3 breaks. During each 10-minute break, subjects were allowed to drink up to 250 ml of water at room temperature (= 30 $^\circ C$) if needed. Based on [74], the kinetic energy efficiency Δ_{eff} was set to 0.40 for walking and 0.44–0.54 for running. Δ_{eff} represents the ratio of metabolic energy

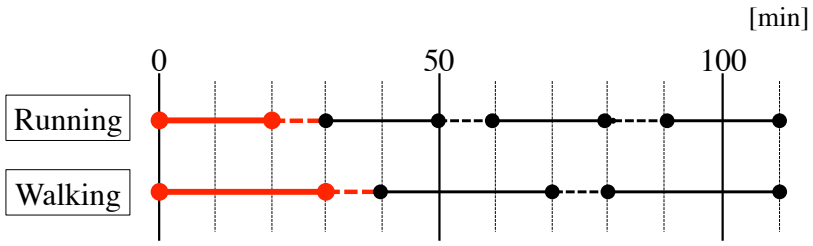


Figure 4.5: Walking and running schedules

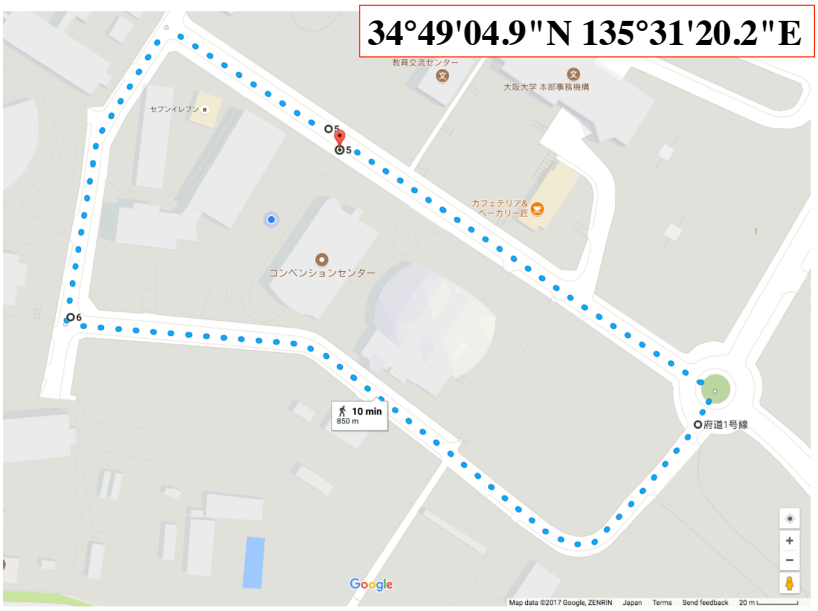


Figure 4.6: Exercise route

consumed in exercise, relative to the total metabolic energy (Eq. (3.28) in Section 3.3).

Table 4.4 shows the overall errors in the walking and running experiments, along with their standard deviations. The **PROP** method more successfully reduced the mean absolute error and standard deviation than **DEF**, meaning that the calibrated parameters better reproduced the actual response of the core temperature. The error of **OPT** in running was already satisfactory (0.20 $°C$), but the accuracy of the proposed method is expected to improve with more measurements of core temperatures. The error is larger in running than in walking because running raises the core temperature earlier and higher than walking. Relative to **DEF**, the MAE of **PROP** was 13% (0.23–0.20 $°C$) lower in walking and 30% (0.43–0.30 $°C$) lower in running. The errors in **DEF** confirm that Gagge's model and the default parameters are reasonable during walking (i.e., moderate exercise), but incur large error ($> 0.4°C$) during running (i.e., heavy exercise).

To reveal the details, we present the MAEs of each session in Tables 4.5 and 4.6. For example,

Table 4.4: Mean absolute errors in walking and running datasets

Exercise	**PROP**[$^\circ C$]	**OPT**[$^\circ C$]	**DEF**[$^\circ C$]	**# of dataset**
Walk	0.20±0.16	0.14±0.12	0.23±0.16	9
Run	0.30±0.24	0.20±0.18	0.43±0.36	25
Overall	0.28±0.23	0.19±0.17	0.39±0.34	34

Table 4.5: Mean absolute error changes in walking dataset [$^\circ C$]

Method	walk2	break2	walk3
PROP	0.19±0.15	0.16±0.12	0.21±0.18
OPT	0.14±0.12	0.12±0.08	0.14±0.13
DEF	0.19±0.14	0.31±0.14	0.25±0.16

Table 4.6: Mean absolute error changes in running dataset [$^\circ C$]

Method	run2	break2	run3	break3	run4
PROP	0.33±0.24	0.20±0.17	0.33±0.26	0.25±0.21	0.33±0.25
OPT	0.21±0.18	0.20±0.17	0.19±0.17	0.21±0.18	0.21±0.18
DEF	0.30±0.24	0.56±0.43	0.47±0.38	0.53±0.38	0.41±0.34

walk3 in Table 4.5 means the third walking period, and break2 in Table 4.6 means the second break. Note that walk1, rum1, and break1 are excluded because these durations are used in the calibration. Tables Tables 4.5 and 4.6 confirm that the error is larger in **DEF** than in **PROP** and **OPT**, especially during breaks. These results highlight the limitation of Gagge's two-node model. The large **DEF** errors were successfully reduced by **PROP** and **OPT**, because the delay parameters reproduce the actual core temperature response. However, the errors are larger in **PROP** than in **OPT**, especially during exercise. This indicates the difficulty of estimating the core temperature rise from single measurements of the actual core temperature.

4.4.3 Result in Biking

Next, the proposed method was evaluated on the dataset of seven subjects performing an indoor ergometer exercise. The experimental details are presented in Table 4.7. The subjects were six males and one female with average age, height and weight of 23.0 ± 0.8 y, 171.9 ± 7.5 [cm] and 75.3 ± 15.7 [kg], respectively (where ± denotes standard deviation). All participants performed the one-hour ergometer exercise on six days, yielding 42 datasets. As shown in Figure 4.7, the exercise intensity varied between 2.4 and 7.2 [W] in each five-minute interval. Note that no breaks were taken throughout this experiment. Therefore, the first 25 minutes were used in calibration (highlighted lines in Figure 4.7), and the following 35 minutes were dedicated to evaluation. As this experiment was conducted indoors, the solar radiation and wind were not considered. Water ingestion was also ignored as the subjects took no drinks throughout the experiment. Based on [75], the kinetic energy efficiency Δ_{eff} of biking was set to 0.23.

Table 4.7: Details of biking experiment

Date	Jan 14-30 (15 Days)
Subject	6 males and 1 female
Time	1 hour
Location	Air-conditioned room
Exercise	Cycling with an ergometer as shown in Figure 4.7
Measured data	Initial skin temperature, Heart rate, Core temperature (tympanic) Ambient temperature, Ambient humidity
Temperature	26.9 ± 0.9 [$^\circ C$] (Mean±SD)
Humidity	25.5 ± 4.8 [%] (Mean±SD)

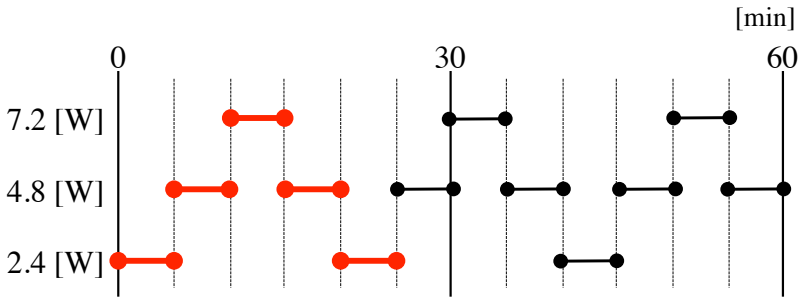


Figure 4.7: Schedule of biking experiment

Table 4.8: Mean absolute errors in biking dataset

Exercise	**PROP**[$^\circ C$]	**OPT**[$^\circ C$]	**DEF**[$^\circ C$]	**# of dataset**
Bike	0.20±0.21	0.14±0.21	0.27±0.25	42

Table 4.8 shows the mean absolute errors of the biking experiment and their standard deviations. **PROP** estimated the core temperature to within 0.20 $^\circ C$ of the measured value, and outperformed **DEF**. This improvement is attributable to the calibration, which observes the core temperature response to variable exercise intensity even when no breaks are taken, and hence selects a superior parameter set. Comparing Tables 4.4 and 4.8, we find that the overall performance was higher in walking and biking than in running. Therefore, Gagge's model is appropriate for moderate exercises such as walking and biking.

4.4.4 Result in Tennis Practice

In this experiment, we collected a 20-hour dataset of two days' tennis practice. The details are presented in Table 4.9 and Figure 4.8. The six subjects were amateur players with at least five years of

Table 4.9: Details of tennis experiments

	Day 1 (Aug. 1)	Day 2 (Aug. 2)
Subject	1 male and 1 female	4 females
Time	13:30 - 16:30	13:00 - 16:30
Temperature	39.1 ± 2.3 [$^{\circ}C$] (Mean±SD)	38.3 ± 4.0 [$^{\circ}C$] (Mean±SD)
Humidity	34.6 ± 4.0 [%] (Mean±SD)	38.4 ± 9.0 [%] (Mean±SD)
Location	An Outdoor Tennis court in Osaka University, Suita, Japan	
Measured data	Skin temperature, Heart rate, Core temperature (tympanic) Ambient temperature, Ambient humidity, Solar radiation	

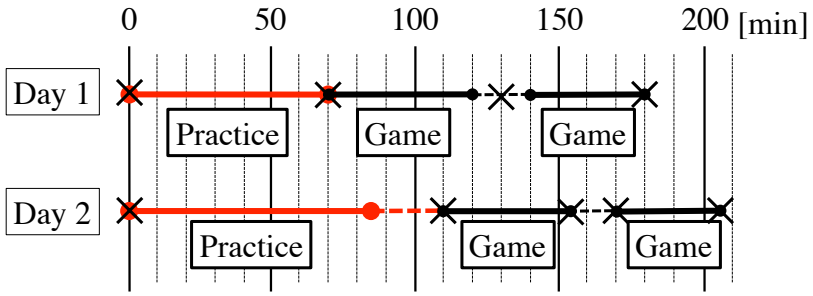


Figure 4.8: Schedule of tennis experiment

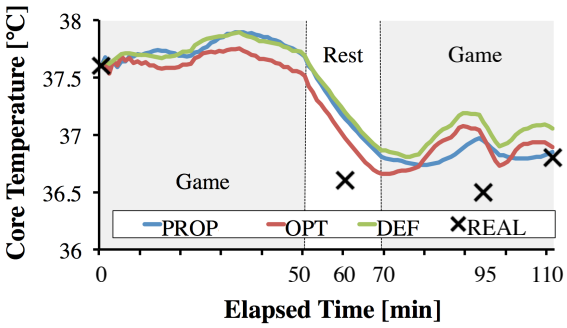
tennis experience. On Day 1, the average age, height, and weight of the competitors (2 participants) were 21.5 ± 0.7 y, 167.5 ± 12.0 cm, and 56.0 ± 17.0 kg, respectively. On Day 2, the age, height, and weight of the competitors (4 participants) were 51.3 ± 1.0 y, 160.5 ± 4.2 [cm], and 52.8 ± 3.9 [kg], respectively. Figure 4.8 illustrates the timetable of the both days, where participants freely organized their schedule of practice, game, breaks and beverage intake. On average, participants drank 1475 ml of water per day.

Although Δ_{eff} is required in simulations of Gagge's model, a standard value of Δ_{eff} is unavailable for tennis because this sport involves various motions. Hence, the motions of tennis were approximated as combined running and standing motions, which were switched as the subject's speed altered. If the player's speed exceeded 5 km/h, the motion was assumed as running; otherwise, it was assumed as standing. The speed was obtained by Adidas Speed Cell [88], which measures the average speed every five seconds. Following [74], the average speed over five seconds was translated into kinetic energy efficiency ($\Delta_{eff} = 0.44$–0.54), which is zero during standing. Finally, we estimated the average kinetic energy efficiency Δ_{eff} in each unit time.

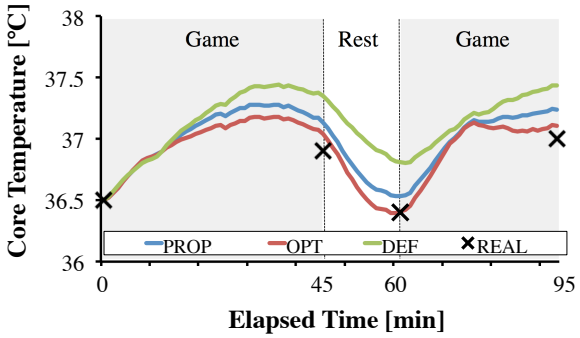
Figure 4.8 depicts the schedule on both days. The practice periods (highlighted) were assigned as calibration durations, and the games were used for evaluation. During their breaks in the evaluation

Table 4.10: Mean absolute error in tennis dataset

Day	**PROP**[$^\circ C$]	**OPT**[$^\circ C$]	**DEF**[$^\circ C$]	# of participants
1	0.28±0.25	0.26±0.22	0.67±0.30	2
2	0.30±0.30	0.24±0.29	0.52±0.34	4



(a) Core temperature estimation result (Day 1, Subject 2). REAL denotes the measured data



(b) Core temperature estimation result (Day 2, Subject 3). REAL denotes the measured data

Figure 4.9: Examples of core temperature estimation in tennis practice

periods, participants were requested to occasionally measure their own core temperatures using the MC-510 sensor [77]. Cross marks in the figure means timings of core temperature measurement. The first and second measurements were used in model initialization and calibration, respectively. The remainder were reserved for evaluation. Table 4.10 depicts the mean absolute error on each day. On both days, **PROP** estimated the measured core temperature to within 0.30 $^\circ C$. According to these results, our method determines a reasonable parameter set even in real sports environments. Meanwhile, the errors in **DEF** were the largest among our experiments. This indicates the difficulty of estimating core body temperature by **DEF** during real sports with complicated movements.

Figure 4.9 presents two examples of core temperature estimation. All methods simulated the actual trend of the players' core temperature, but **PROP** and **OPT** yielded more accurate values than **DEF**.

We infer that **DEF** can usefully determine the rough trend, but parameter optimization ensures a more accurate estimation.

4.5 Application Prototype of Heat Stroke Caution

This section demonstrates the prototype of our system for heat stroke alert.

4.5.1 Motivation

There are some studies of heat stroke sensing. Reference [89] proposes a sensor-equipped headgear to monitor in-hardhat temperature. It shows relationship between temperature inside headgear and core temperature. The major limitation of the method is that users have to wear headgears during sports. In the other study, a heat stroke alert application is developed on smartphones [90]. Although the method suits almost all situation since it relies on the air temperature and humidity, the heat risk assessment with environmental contexts is not capable to individually notify the heat stroke risk. Another study [91] proposes a wearable shirt with integrated e-textiles to prevent heat stroke for firefighters. It integrates heart rate with the air temperature and humidity.

Despite these efforts, there has not been a system which satisfies following criteria. (1) Minimum interference: the system has to be easily installed and does not interfere in user's activity. (2) Scalability: the system has to seamlessly expand in terms of device, environment and exercise. (3) Accuracy: the system has to precisely assess the body heat context and to individually detect heat stroke risk. To meet these criteria, our system leverages commercial-off-the-shelf smartwatches to measure heart rate and body surface temperature, then it estimates core body temperature based on the individually different heat context. The method fairly considers the effect of different environments and exercises as proved in Section 4.4, leading to precise assessment of heat stroke risk.

4.5.2 Heat Stroke Warning Algorithm

Rise of human core body temperature is known as a fundamental precursor of heat-related illness [5]. A study reveals that over $39.5°C$ core temperature is commonly observed among heat stroke patients [19]. Meanwhile, our goal is not detecting heat stroke patients but preventing users from heat stroke, accordingly we set a threshold for heat stroke warning to $38.0°C$ based on WMO (World Meteorological Organization) guideline [92]. We note that this threshold should be lower that the level of hyperthermia ($39.5\ °C$) for safer notification, and optimized based on the feedback from users (e.g., a higher threshold may be suitable for well-trained athletes). In the following evaluation, we evaluate precision and recall of heat stroke warning that arises when the core temperature exceeds $38.0°C$.

Table 4.11: Confusion matrix in ergometer exercise (**PROP**)

Notification	True Warning	
	Yes	No
Yes	**55**	120
No	3	**2425**

Table 4.12: Confusion matrix in ergometer exercise (**OPT**)

Notification	True Warning	
	Yes	No
Yes	**57**	100
No	1	**2445**

Table 4.13: Confusion matrix in ergometer exercise (**DEF**)

Notification	True Warning	
	Yes	No
Yes	**55**	149
No	3	**2396**

Table 4.14: Confusion matrix in running and walking (**PROP**)

Notification	True Warning	
	Yes	No
Yes	**311**	88
No	121	**3085**

Table 4.15: Confusion matrix in running and walking (**OPT**)

Notification	True Warning	
	Yes	No
Yes	**286**	62
No	146	**3111**

Table 4.16: Confusion matrix in running and walking (**DEF**)

Notification	True Warning	
	Yes	No
Yes	**159**	40
No	273	**3133**

4.5.3 Result

We here use the dataset presented in Section 4.4.2 and 4.4.3 for evaluating the performance of our prototype system. We note that we do not use the dataset collected in tennis practice (Section 4.4.4) due to its lack of continuous measurement of ground truth. In order to evaluate precision and recall, we initially classified all the estimated core temperature and measured core temperature into two categories: (1) $38.0°C$ or greater and (2) less than $38.0°C$. To clarify the effect of core temperature estimation error on the performance of heat stroke alert system prototype, we compare the confusion matrixes of heat stroke cautions generated by **PROP**, **OPT** and **DEF**. The results are summarized in Tables 4.11 to 4.16, where the rows show the number of warnings produced by our algorithm and the columns show the number of true warnings. We note that the numbers on the tables are the total time duration in minutes when core temperature is $38.0°C$ or grater / less than $38.0°C$. In ergometer exercise, only 58 out of 2,603 minutes are over the threshold (i.e., $38.0°C$). This is because the room temperature was not so high ($< 30°C$) and there was no solar radiation which constantly heats up human body in outdoor. Tables 4.11 to 4.13 indicate that $\{precision, recall\}$ of **PROP**, **OPT** and **DEF** in ergometer exercise are {0.314, 0.948}, {0.363, 0.983}, and {0.270, 0.948}. This suggests each algorithm of core temperature estimation can alert with very high recall, however the precision is still low. The major reason of low precision is overestimation, namely we find our method tends to estimate core temperature higher than groundtruth. Nevertheless, **PROP** can reduce false positives compared to **DEF**, accordingly our method of parameter calibration slightly improves the performance of heat stroke alert.

Tables 4.14 to 4.16 denote that $\{precision, recall\}$ of **PROP**, **OPT** and **DEF** through running and walking dataset are {0.779, 0.720}, {0.822, 0.662}, and {0.799, 0.368}. These results suggest **PROP** is most likely to estimate core temperature higher than the groundtruth, leading to the highest recall measure. **DEF** also shows moderate performance on precision, however its recall is too small due to

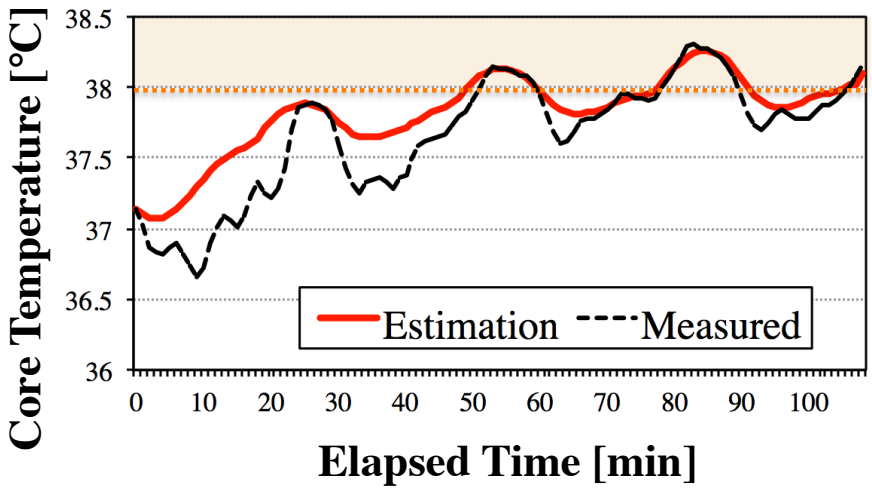


Figure 4.10: An example of core temperature estimation

underestimation of core temperature. This severely drops reliability of heat stroke alert. Although **PROP** and **OPT** show moderate performance in severely hot environments, our application requires higher recall for human safety. We consider the average error up to 0.30 $^{\circ}C$ proved in Section 4.4 is the major reason to deteriorate both precision and recall. Another reason is the fact that there are some subjects whose actual core temperature exceeds 38.0 ℃ for only a few minutes. In that case, correct notification becomes more difficult due to estimation error.

Figure 4.10 illustrates an example of core temperature estimation and heat stroke alert. It shows core temperature estimation is still challenging, namely reproducing actual response is a difficult task. On the other hand, the estimation result is useful for detecting whether the temperature reaches 38.0 $^{\circ}C$ or not. In this case, our algorithm can correctly alert it with a few false positives.

4.5.4 Result with acceptable error ΔT

We consider some error of timing should be acceptable because core temperature estimation is still difficult by limited sensors and we assume the threshold must be much lower than 39.5 ℃ (dangerous level) for safety. Therefore, we introduce acceptable error ΔT and evaluate precision and recall while changing the range of ΔT. For example, $\Delta T = 3$ means that we regard warnings which error of warning timing within three minutes as correct. Also, we introduce an offset value to improve reliability by incorporating the error caused by core temperature estimation to heat stroke caution. We note that the error for each trial is unexpectable because of individual difference, and daily difference in even the same subject, thus we mechanically give the offset as from 0.0 to 0.3 $^{\circ}C$. Figure 4.11 depicts the relationship between precision and recall, and acceptable error ΔT and the offset value in the ergometer data set. The results basically illustrates the trade-off between precision and recall which

varies according to offset value. In almost all cases, **OPT** shows the best performance on precision and recall due to its minimum error compared to **PROP** and **DEF**. Nevertheless, each algorithm can achieve around 0.5 precision and 1.0 recall if 5-minute error is accepted ($\Delta T = 5$) with zero offset (Figure (4.11(a))). Accordingly, we confirm adding offset does not dramatically improve reliability of the system in ergometer exercise. Precision is still challenging, however we believe that we can improve accuracy of notification by considering continuous alerts (i.e., raise warning in the case that the algorithm continuously detects high temperature) since our current algorithm detects high-core temperature at that moment.

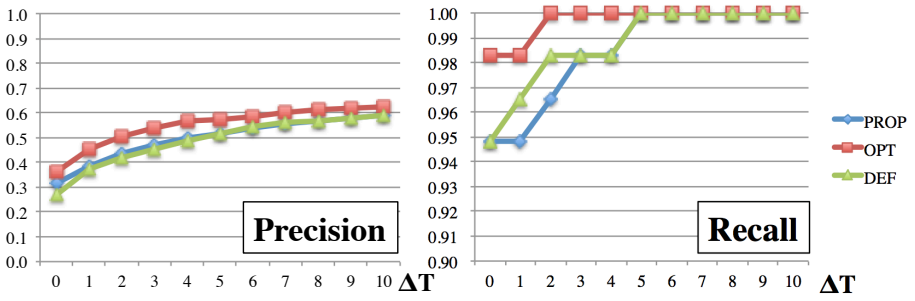
Figures 4.12 to 4.15 illustrate precision and recall with ΔT in running and walking while changing the offset value. The upper row on each figure depicts precision and the lower shows recall. In these results, we find recall of each method in walking is almost 1.0 even if we disallow any delay of detection timing. This result is as same as the result proved in ergometer exercise (Figure 4.11), meaning core temperature estimation usually gives overestimation in walking exercise. In contrast, precision and recall of **PROP** and **OPT** in running is moderate when $\Delta T = 0$ and gradually rise according to evolution of ΔT. This is because reproducing human body response in complicated exercise is difficult. Nevertheless, precision and recall of **PROP** in running improve with the increase of ΔT and exceeds 0.9 with $\Delta T = 8$ and no offset (Figure 4.12). Further, **DEF** shows lower recall in running, meaning default parameter set usually makes underestimation on core temperature estimation. In other words, parameter calibration is essential for heat stroke caution in running exercise. However, Figure 4.12 shows there are some false negatives even if we applied parameter calibration in core temperature estimation.

Figures 4.13 to 4.15 demonstrate the number of false negatives decreases according to higher offset. Also, Figure 4.14 shows recall of **PROP** and **OPT** in both exercises reach 1.0 by expecting 0.2 $^\circ C$ error as the offset. Namely, 0.2 $^\circ C$ offset successfully makes heat stroke caution for human safety in our dataset. It is true that precision decreases when 0.2 $^\circ C$ offset is given, however it is still over 0.7 when we allow 30-minute error of notification timing. We note that 0.2 $^\circ C$ offset can be defined as the best preference for safer caution, while 0.1 $^\circ C$ offset shows the best performance in terms of balance between precision and recall. Consequently, we recommend 0.2 $^\circ C$ or larger offset for heat stroke caution in hot environments and consider 0.1 $^\circ C$ also works well in the other environments.

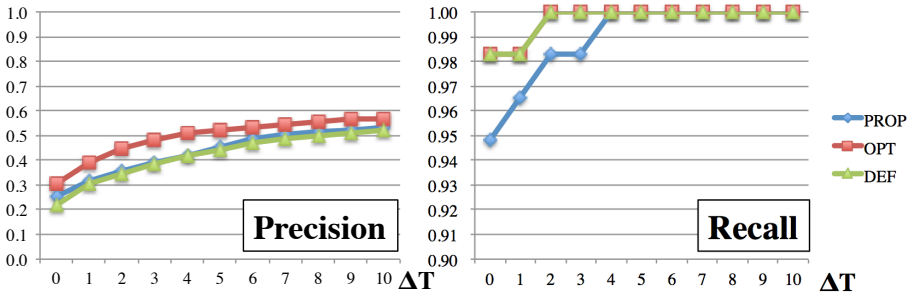
4.6 Conclusion

We proposed a method that estimates the core body temperature during exercise. The method extends Gagge's two-node model, and improves the accuracy of the estimation by optimizing the parameter set representing the different thermoregulation responses. To this end, it compares the occasionally measured core temperatures with the simulation output of the extended Gagge's model. A single tympanic temperature measured during an exercise break is sufficient for optimizing the parameter

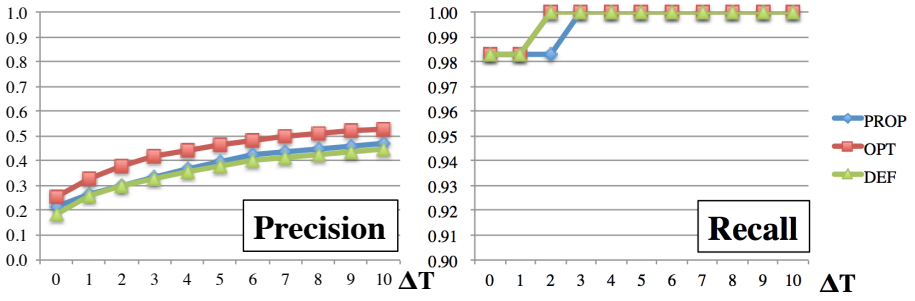
set. Moreover, our method successfully estimated the core temperature (within 0.3 $^{\circ}C$ error) during walking, running, biking, and tennis practice. Finally, we developed a prototype of heat stroke alert system, revealed our method can instantly detect high core temperature exceeding $38.0^{\circ}C$ with over 0.7 recall in running, walking and ergometer exercise, and confirmed precision and recall dramatically increase when we accept some error of warning timing. The results suggest expecting 0.2 $^{\circ}C$ error successfully removes false negatives of heat stroke alert in our datasets. We note that some contents in this chapter refer our publication [93, 94].



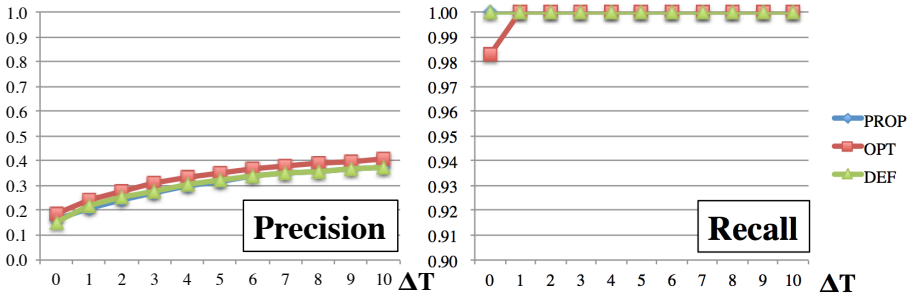
(a) no offset



(b) offset = $0.1°C$



(c) offset = $0.2°C$



(d) offset = $0.3°C$

Figure 4.11: Precision and recall in ergometer exercise with ΔT

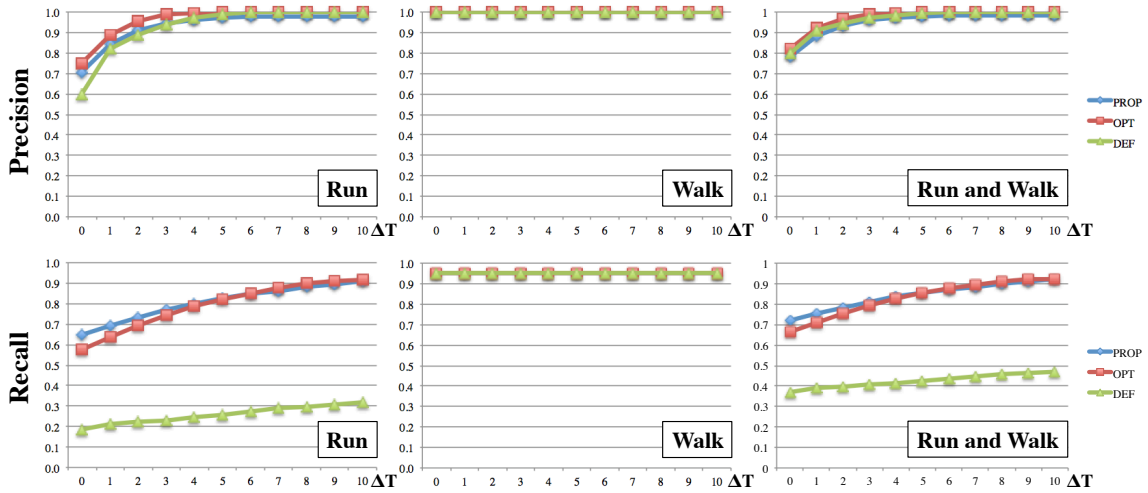


Figure 4.12: Precision and recall in running and walking with ΔT (no offset)

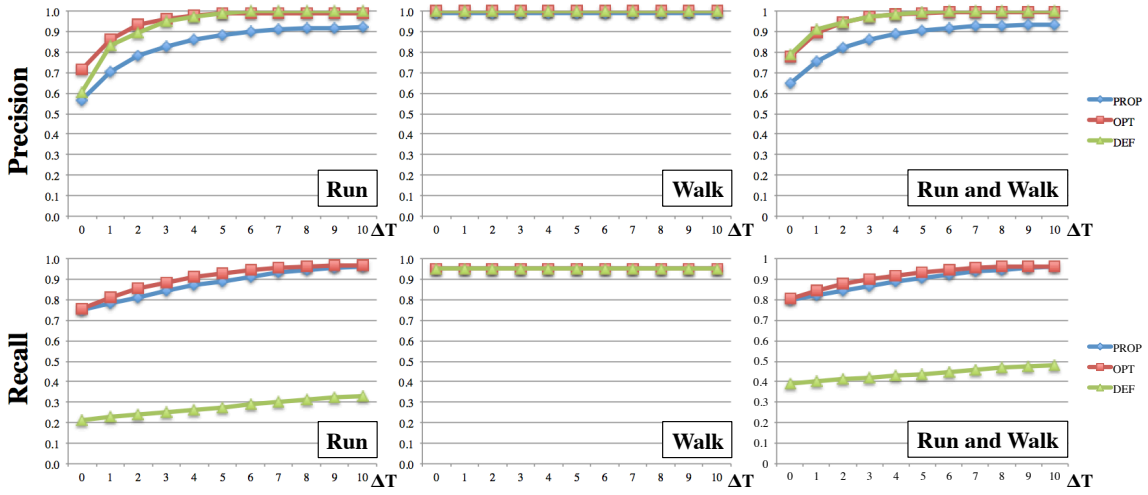


Figure 4.13: Precision and recall in running and walking with ΔT (offset = $0.1°C$)

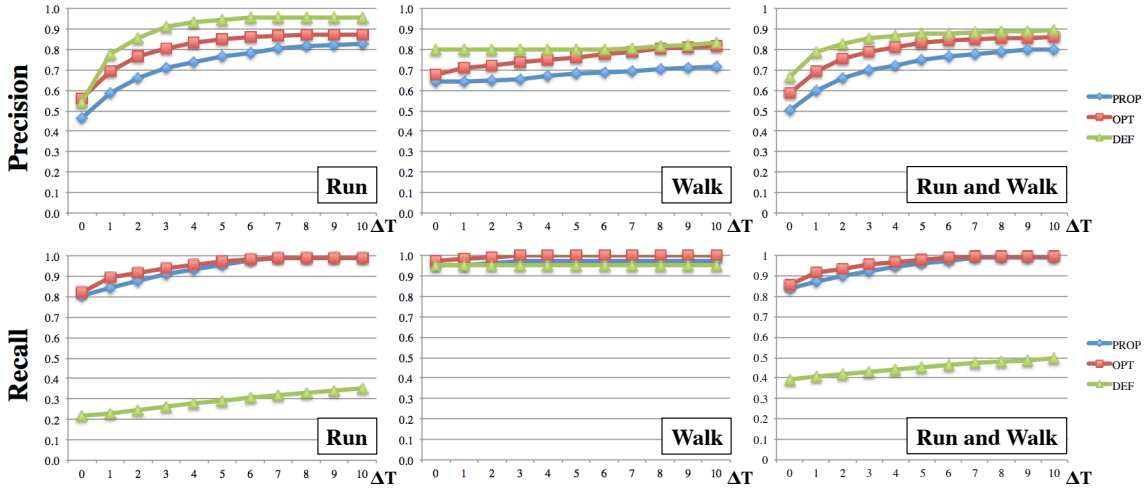


Figure 4.14: Precision and recall in running and walking with ΔT (offset = $0.2°C$)

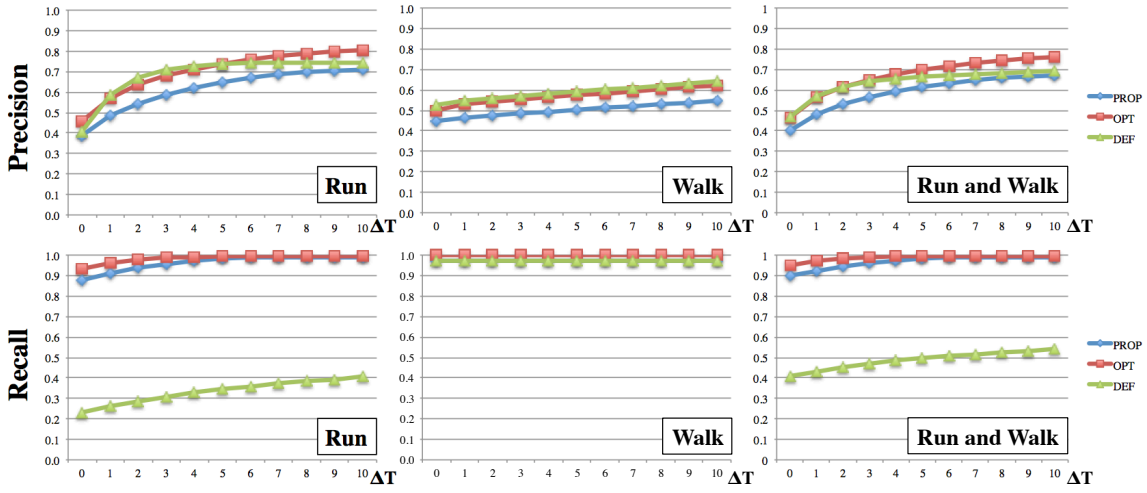


Figure 4.15: Precision and recall in running and walking with ΔT (offset = $0.3°C$)

Chapter 5

Fluid Intake Estimation by Wrist-worn Inertial Sensors for Dehydration Prevention

5.1 Introduction

Water accounts for about 60% of the human body, and when the body loses it in higher rate than its intake rate (through drinking), dehydration symptoms occur. The dehydration causes many severe health problems like organ and cognitive impairment, accordingly it is critical for the human to drink water in a sustained manner to avoid dehydration. Some people can sense the dehydration by feeling thirsty, dry or sticky mouth, then they can prevent it by correspondingly taking water immediately. However, many others (children, elderly people, athletes, busy people, etc.) are sometimes not aware of their dehydration, leading to many symptoms like anorexia, cramp, headache, nausea, and unconsciousness. Also the lack of in-body water causes rise of core body temperature, as the study reported that 1% water loss against body weight triggers 0.3 $^{\circ}C$ rise of core body temperature. These problems spark the need for a system that automatically tracks fluid intake as well as encourages users to appropriately drink water in order to prevent dehydration.

To approach this goal, many systems to track human fluid intake have started to emerge recently. Firstly smarter gadgets are developed for the purpose of keeping record of daily fluid intake such as Hidrate Spark [12], Trago [11], which always observe fluid level they own and encourage the user to drink more fluids to catch daily hydration goals. Also respective studies propose smart containers [59, 60, 61, 62] equipped with either accelerometer, capacitive sensor or RFID tags. These containers can capture deep context of user's fluid intake, namely both of fluid intake timing and each amount of intake. Nonetheless, these solution are neither scalable nor ubiquitous since the user has to buy as special container for tens of dollars or install a dedicated sensor in the container used for fluid intake. Further, they can track only the water consumed by the designated container, meaning the user has

to bring the container everywhere. This severely limits their wide applicability. To provide seamless estimation for multiple drinking scenarios, several studies propose methods of fluid intake estimation based on vision [13], audio [14] and wearable sensors [15, 16]. These approaches indirectly measure the amount of fluid intake by observing arm posture captured by camera, swallowing activity caught by audio analysis, and arm gestures grasped by inertial sensors embedded in smart watches. Nevertheless, these solutions still have some limitation: vision-based approach requires photos adequately captured with desirable angle and quality; audio analysis in the noisy environment is challenging and sometimes raises privacy concern; smart watches are not used alone, but combined with other sensing modalities. In summary, to our knowledge, there are no systems which meet following criteria for pervasive monitoring of fluid intake: (1) Accuracy: a fairly good estimation is essential; (2) Ease of distribution: leveraging existing sensor modalities or bootable on commodity sensors; (3) Scalability: applicable for recent commercial-off-the-shelf devices with minimal configuration or seamlessness; (4) Less interference: minimal intervention to the user while preserving accuracy, anonymity, and privacy.

To satisfy these requirements, we propose a ubiquitous and unobtrusive system to track the amount of human daily fluid intake by leveraging the inertial sensors embedded in commodity smart watches. The key idea is to firstly recognize the drinking moments and thereafter gauge the amount of fluid intake in every drinking episode. For precise recognition of drinking moments, our method combines two classifiers with different time-scale windows and extracts fine-grained drinking gesture following coarse-grained activity recognition. Inertial sensor readings are translated into motion features and then given to Conditional Random Field (CRF) to infer the sequence of high-level activities behind the sensor readings. Macro-scale CRF mainly segments drinking episode through daily human activities such as eating, walking and running. Further, micro-scale CRF infers sequential gestures carried out during each drinking episode: (1) lift the container, (2) sip and consume beverage, and (3) release the container. Finally, the system estimates the amount of fluid intake by observing the posture of user's arm during the sip period and the duration of sip gesture recognized by the previous step.

All the essential components of the system and the whole system have been evaluated over 80-hour dataset collected through five different situations and extra scenarios. Firstly, drinking moment could be segmented with 84% precision and 87% recall by macro-scale classifier, and sip gesture was detected with average precision and recall of 91% and 96% through 1615 drinking episodes. Further, our method estimated the amount of fluid intake for 1615 episode with 59% absolute error, but reduced it to 29.1% by considering individual difference of drinking manners with use of training dataset. The overall error (i.e., mean relative error which does not use absolute value) was proved limited to around 15% for 1615 fluid intakes. Finally, our system has been validated in the rigorous environments of either office or a whole day in weekend. The result showed the system could detect sip moments with more than moderate precision and recall, and the overall fluid intake (i.e., daily error of fluid intake amount) was assessed with the worst error -23%.

The rest of this chapter is organized as follows. Section 5.2 describes our dataset collected to

design and evaluate our method. Section 5.3 presents detailed design of the proposed method of fluid intake estimation. Sections 5.4 and 5.5 provide performance evaluation through our laboratory and wild experiments, respectively. Then we generalize our method with the extra datasets in Section 5.6, and discuss some important topics in Section 5.7. Finally, Section 5.8 concludes this chapter.

5.2 Datasets

For developing our algorithm to recognize drinking moments from noisy traces of wrist motion and to estimate fluid intake amount, a group of 22 subjects of different ages ad genders collected the necessary data. They were incentivized to collect datasets by either receiving money, given free beverages or voluntary contribution. We totally collected five datasets with different drinking environments in terms of used containers, consumed beverage types, location, and groundtruth collection method. The datasets are composed of over 80 hours traces of human activities and 1316 episodes of fluid intake through five different scenarios. We note that we also collected the other dataset for important remarks in Sections 5.6 and 5.7.

Through almost all experiments, the participants wore a smart wristband [95] on their dominant hands to capture the 6 axes motion of the gyroscope and accelerometer as illustrated in Figure 5.1, then they drank various types of beverages (e.g., water, green tea, etc.). We found a few lefty subjects wore the gadget on their left hand, accordingly we inverted the sensor readings of their y-axis and z-axis. We summarize the essentials of our five datasets in Table 5.1, where we introduce our datasets as Lab-macro, Lab-micro, Lab-micro+, Wild-office, Wild-day dataset, respectively. In this section, we elaborate on the methodology of collecting the data for each dataset.

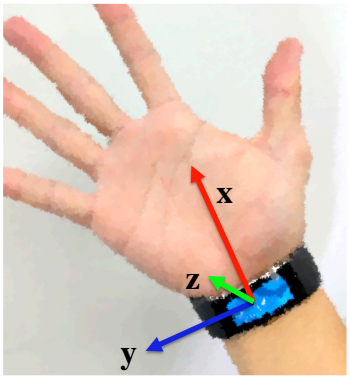


Figure 5.1: A smart band on the dominant hand (indicating three axes of a inertial sensor)

Table 5.1: The datasets collected for the evaluation at glance

Dataset	Scenario	Groundtruth	Amount
(1) Lab-macro	perform 8 different activities for 20 minutes/activity in the lab and outside the laboratory	hand-crafted by event buttons on Android App	22 hours
(2) Lab-micro	drink fixed amount of different beverages using bottles and cups in the laboratory	hand-crafted by event buttons on Android App	180 episodes
(3) Lab-micro+	drink free amount of different beverages using bottles in the laboratory	hand-crafted by event buttons on Android App and weight the bottle before and after the episode and submit to the same App	1069 episodes
(4) Wild-office	unobtrusively drink different beverages in the laboratory while participants are doing research work	wearable camera [96] and smart bottle [11]	27 hours & 40 episodes
(5) Wild-day	unobtrusively drink beverage in real-life while participants are in a weekend	wearable camera [96] and smart bottle [11]	31 hours & 27 episodes

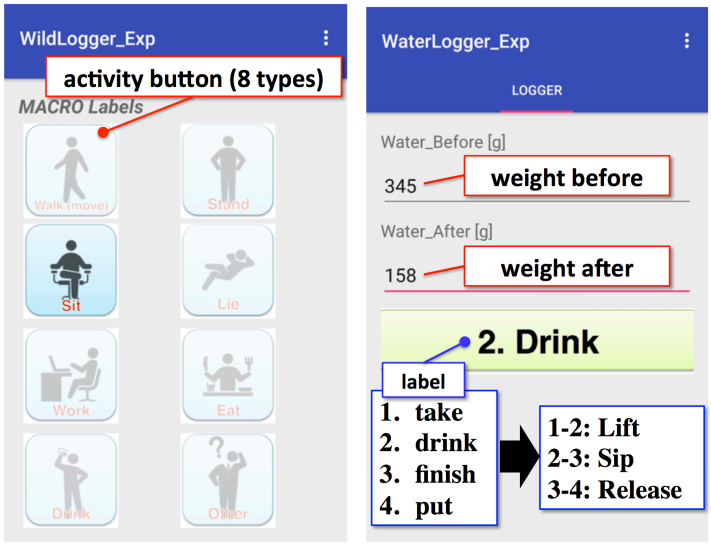
5.2.1 Datasets in the Lab-environment

Firstly we conducted three preliminary studies in our laboratory to develop the basic theory of our algorithm composed of the drinking activity recognition and the amount of fluid intake estimation algorithms. All the datasets were collected by predefined manners controlled by the subjects themselves. Groundtruth of each context (i.e., macro-activity class, micro-gesture or amount of fluid intake) was exactly recorded by using two Android applications that we have developed (Figures 5.2(a) and 5.2(b)). The details of each dataset follows below.

Lab-macro dataset

The nine subjects were hired by $30/subject to collect 20 minutes of each of the following activities: (*move, stand, sit, lie, work, eat, drink, other*). We asked subjects to collect at least 30 instances of fluid intake for *drink* class. The *other* activity class includes a mix of the following activities (*driving, cycling, being on a train, being in an elevator, etc.*). We ended up collecting totally 1325 minutes of these activities as shown in Table 5.2. Despite we asked them to collect 20 minutes for each activity, the collected dataset is not completely balanced since the real activity cannot be performed in a fixed time units.

Groundtruth collection: The subjects manually tagged their activity classes by clicking on of the eight buttons on the groundtruth app installed in their smartphones (Figure 5.2(a)).



(a) macro activities (eight types) App

(b) micro activities (four types) and fluid weight App

Figure 5.2: Android applications to tag the ground truth activities and amount of fluid intake

Table 5.2: Lab-macro dataset activities

Activity	Duration [min.]
drink	64
move	185
stand	207
sit	203
lie	165
work	199
eat	179
other	122
Σ	1325

Lab-micro dataset

This dataset is collected for the purpose of building the model to recognize the three micro-activities (lifting a container, sip, releasing a container) that constitute the drinking activity as well as the fluid intake amount of each drinking episode. 10 participants were offered free beverages to collect datasets and groundtruth labels. We here prepared 50, 100 and 150 grams of beverages in cups and plastic bottles to fairly compare the difference caused by individual subjects and container types. Each subject provided 18 instances of fluid intake (with 3 different fluid levels $\times$ 2 container types $\times$ 3 times), accordingly we collected totally 180 drinking episodes using plastic bottles and paper cups.

Groundtruth collection: The participants manually tagged four key events for each drinking

Table 5.3: Statistics of drinking activities of each subject in Lab-micro+ dataset

Subject	# of drinking	Duration (mean ± SD) [sec.]
1	76	6.90 ± 1.45
2	108	3.02 ± 0.84
3	79	4.01 ± 0.87
4	90	3.20 ± 0.96
5	77	5.61 ± 1.34
6	36	8.55 ± 2.31
7	62	5.22 ± 1.99
8	39	7.10 ± 1.93
9	86	3.74 ± 1.01
10	91	4.64 ± 1.22
11	67	5.47 ± 1.66
12	74	3.94 ± 1.14
13	32	5.87 ± 1.45
14	52	6.23 ± 1.32
15	43	7.32 ± 2.61
16	57	6.77 ± 1.72
Overall	1069	5.06 ± 2.09

activity to assign groundtruth labels to the three primary gestures (*Lift, Sip, Release*) that take place during drinking by operating buttons on the groundtruth app on their smartphones (Figure 5.2(b)). Note that the participants did not tag the amount of consumed beverage due to designated amount of each intake (50, 100 or 150 grams).

Lab-micro+ dataset

We uses this dataset for validating our method. We offered 16 participants free beverages sealed in plastic bottles, then they collected large and diverse dataset in our laboratory. They were instructed to freely drink something by bottles during their daily research work, namely they drank free amount of fluid whenever they wanted. This made widely distributed duration and amount of each drinking activity. Finally, the subjects have minimum and maximum numbers of drinking episodes of 32 and 108 as shown in Table 5.3, and totally generated 1069 drinking instances with groundtruth labels. Figure 5.3 also shows the box plot of the amount of fluid intake in drinking instances for each subject. These insights indicate the sip duration and fluid intake amount vary from person to person, highlighting the existence of personal differences regarding the drinking behavior.

Groundtruth collection: We similarly collected the groundtruth of drinking micro-activities by Android app in Figure 5.2(b). The subjects additionally registered the groundtruth of fluid intake for each drinking episode by weighting the container in grams unit before and after the intake using a digital weight scale (Figure 5.4(a)).

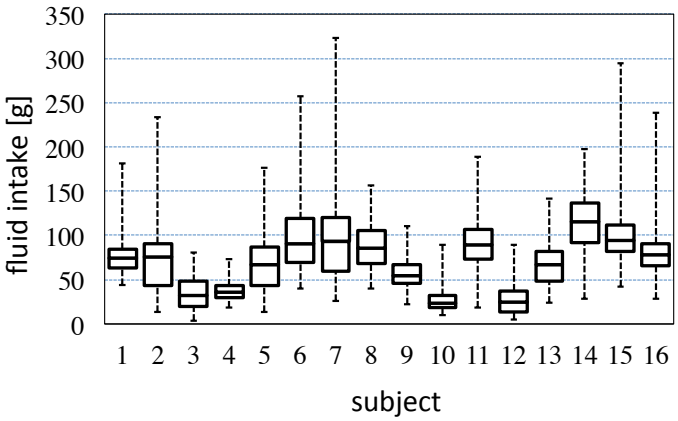


Figure 5.3: Statistics of the amount of fluid intake for each subject (Lab-micro+ dataset)

(a) Weight scale (laboratory experiments) (b) Trago smart bottle (wild) (c) Mofily YoCam (wild)

Figure 5.4: Gadgets used for the groundtruth collection

5.2.2 Datasets in the Wild-environment

To validate our method with rigorous scenarios, we carried out two experiments by uncontrolled manners (i.e., collected dataset without interfering the subject's life). Here subjects just wore a smart band to track arm motion and carried the dedicated gadgets to unobtrusively record groundtruth of their activity and fluid intake.

Wild-office dataset

This dataset is collected in our office by an unobtrusive manner. We collected this dataset through the whole day across three different days which include all office-related activities as well as drinking activities. The average durations of the stay hours, count of fluid intakes, and amount of each fluid intake were revealed as 9.1 hours, 13.3 times and 71.1 grams, respectively. The total amount of fluid intake in each dataset were 1242, 892 and 789 grams.

Groundtruth collection: We used a neck-worn camera (Mofily YoCam [96]) and a smart bottle (Trago smart bottle [11]) for groundtruth collection without any interference. The camera took a picture every 5 seconds to provide clues about the current user activity. The smart bottle could automatically keep track of amount of each fluid intake.

Wild-day dataset

This dataset composed of two full tracks of wrist motion during a whole day in the weekend (i.e., datasets from getting up to going to bed). Two inexperienced subjects (a male and a female) voluntarily collected datasets, in which they consumed beverages 18 and 9 times (totally 1250 and 759 grams), respectively.

Groundtruth collection: Similarly, we collected the groundtruth by using the neck camera and the smart bottle. We note that the subject used a portable weight scale and his neck camera automatically recorded the weight scale reading when the subject could not use the smart bottle (e.g., dinner in a restaurant).

5.3 Algorithm Design

Figure 5.5 illustrates the overview of our proposal for fluid intake estimation which is composed of the following three consecutive modules. (1) macro-scale activity recognition for separating drinking moments from other daily activities (by **Macro-scale Activity Classification Module**), (2) micro-scale activity recognition to capture the sip gesture through inferring the sequential micro-activities (by **Micro-scale Activity Classification Module**), and (3) estimation of fluid intake amount for each sip gesture period based on the arm posture observation (by **Amount of Fluid Intake Estimation Module**). We describe the detail of each module in the rest of this section.

5.3.1 Macro-scale Activity Classification Module

This module works as the initial filter to separate drinking activities from other activities (*move, stand, sit, lie, work, eat, other*). The human activities are intrinsically connected seamlessly and some activities have implicit rules of inter-activity transition (e.g., eating class follows to sitting class). To represent connection and transition rules of activities, we employ the Conditional Random Field (CRF) to recognize eight classes. CRF is a discriminative model for inferring a consistent sequence of

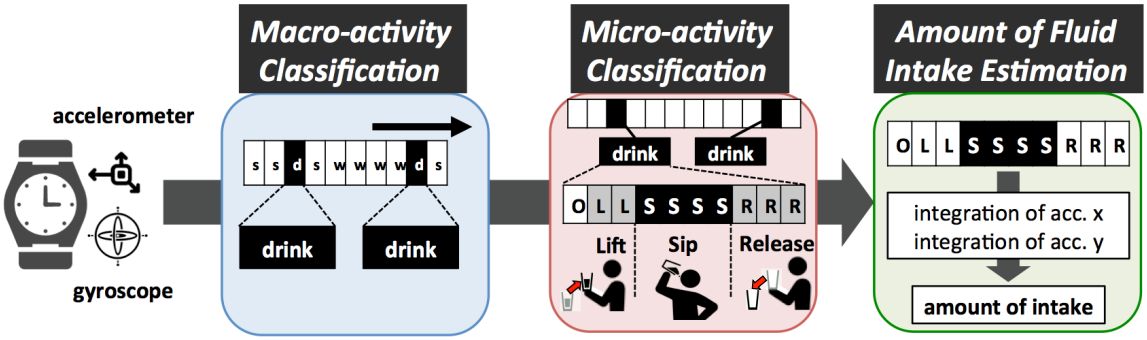


Figure 5.5: An overview of our algorithm consisting of three modules

labels (i.e., activities) by observing the whole sequence of given features. It represents the relationship between observables (i.e., features extracted from sensors) and the unobservable sequence of actual states by unary potential function ϕ_i and the relationship between hidden states by pairwise potential function ψ_{ij} to indicate chronological dependencies among hidden states. These assumption help the model to learn the natural connection of human activities and reject unexpected activity transitions. With these potential functions, the model defines the conditional probability distribution $p(\boldsymbol{y}|\boldsymbol{x})$ and standardization term $Z(\boldsymbol{x})$ as below.

$$p(\boldsymbol{y}|\boldsymbol{x}) = \frac{1}{Z(\boldsymbol{x})} \prod_{i \in S} \phi_i(y_i|\boldsymbol{x}) \prod_{i \in S} \prod_{j \in N_i} \psi_{ij}(y_i, y_j|\boldsymbol{x}), \tag{5.1}$$

$$Z(\boldsymbol{x}) = \sum_{\boldsymbol{y}} \{ \prod_{i \in S} \phi_i(y_i|\boldsymbol{x}) \prod_{i \in S} \prod_{j \in N_i} \psi_{ij}(y_i, y_j|\boldsymbol{x}) \}. \tag{5.2}$$

Here, $\boldsymbol{x}$, $\boldsymbol{y}$, y_i, S, N are the sequence of observable features, the transaction of actual activity classes, the state on the site i, the set of all sites, and the set of neighbor sites. Figure 5.6 illustrates the structure of CRF we use, where the activities in the sequence line up. In our case, $\boldsymbol{x}$ means the features extracted from the inertial sensor readings and $\boldsymbol{y}$ indicates the sequence of user activity y_i which belongs to one of the primary eight classes (*move, stand, sit, lie, work, eat, drink, other*). Neighbor N represents the sites y_{i-1} and y_{i+1} in the case of Chain-CRF (Figure 5.6). We generated this CRF model with PyStruct library [97] and the model was trained by the block-coordinate Frank-Wolfe algorithm [98].

Model parameter selection: To leverage CRF, the inertial sensor readings and groundtruth labels are segmented into sliding windows with fixed length, then features are extracted in each window and the representative label of the window is determined by finding the most frequently appeared labels among expected classes (eight classes in macro-activity classification phase). The window length and overlap rate are variable and affect the performance of activity classification. Thus, we observed the average duration of drinking instances, and then carefully defined candidates of the length and overlap rate of sliding windows as presented in Table 5.4. The best combinations of these parameters is empirically chosen by the grid-search algorithm [99] over all possible combinations so as to get the

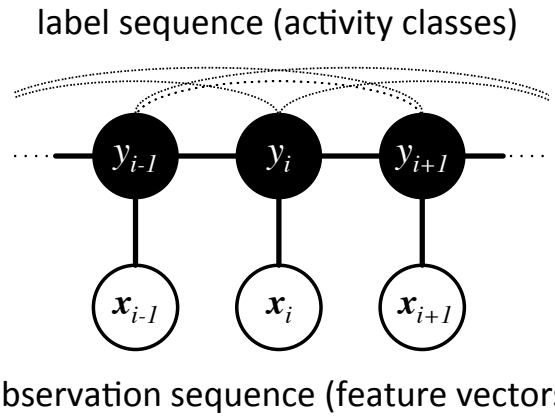


Figure 5.6: The Conditional Random Field (CRF) model

Table 5.4: Candidates of feature-extraction window parameters

classifier	**parameter**	**value**
macro activities	window length [sec.]	4.0, 5.0, 6.0, 7.0, 8.0, 9.0, 10.0
	window overlap [%]	0, 25, 50
micro activities	window length [sec.]	0.5, 1.0, 1.5, 2.0
	window overlap [%]	0, 25, 50

highest performance for recognizing *drink* class. We finally found 8 seconds and 0% overlap rate were the best parameters for sliding window used to recognize macro-scale drinking episodes.

Feature extraction and selection: We defined the following 28 different statistical features using 3-axis acceleration values indicated as a_x, a_y, a_z and 3-axis gyroscope values indicated as g_x, g_y, g_z: The average and standard deviation for each of the 3 axes of the accelerometer and gyroscope values, correlation coefficients for each pair of the 6 axes, and manipulation (the ratio of magnitude 3-axis gyroscope to that of acceleration which means complexity of hand movement [49]). Since the gyroscope and accelerometer values vary to different ranges, accordingly we scale each feature value before applying the classifier for the fair comparison among features. We applied the standardization to scale each feature to follow a Gaussian distribution with zero mean and unit variance. Despite the performance of the classifier using all the extracted features (i.e., 28) is good, it is known that high dimensional features often cause over-fitting and increase the computational cost. To reduce the dimension of the feature space, we apply the backward-feature selection method [100] to empirically remove any unnecessary features. Further, to balance the performance and computational cost, we pick the feature combination that satisfies both the minimal dimension and the reduction of the classification performance by 1% at most compared to the best case. We finally got **6 features** as the most effective combination of feature for the macro-activity classifier as represented in Table 5.6.

Table 5.5: The candidate features for the macro- and micro- activity classifiers

feature	description	number
avg	average value of readings within the window	6
dev	standard deviation of value within the window	6
corr	correlation coefficient for each pair of value in the window	15
manipulation	average ratio of gyro norm to acceleration norm: $\frac{1}{W}\sum^{W}\frac{\lvert g_x\rvert+\lvert g_y\rvert+\lvert g_z\rvert}{\lvert a_x\rvert+\lvert a_y\rvert+\lvert a_z\rvert}$	1

Table 5.6: The optimized features for the macro- and micro- activity classifiers

Classifier	Selected features
Macro-activity	$avg(a_x), avg(a_y), corr(a_x, a_y), corr(g_x, g_y), corr(g_y, g_z), manipulation$
Micro-activity	$avg(a_x), avg(a_y), avg(g_y), avg(g_z), corr(a_x, a_y), corr(g_y, g_z), manipulation$

5.3.2 Micro-scale Activity Classification Module

The goal of this module is to recognize the three consecutive micro-activities: (lift the container, sip, release the container) that constitute the drinking activity. Figure 5.7 illustrates an instance of drinking activity, where three micro-activities are sequentially conducted and generate the unique pattern of the waveforms. Based on this property, we define the three consecutive gestures of drinking as {*Lift, Sip, Release*} and additionally introduce the *Other* gesture to represent gestures associated with drinking sequence (e.g., opening a cap of a bottle, fetching the container). To represent the flow of these activities, we use CRF as a classifier and infer the sequential transitions of gestures by given input features extracted from inertial sensor readings. The parameters of sliding window are defined and optimized by the same way as macro-scale activity classification module, then we got 0.5 second and 50% as the best combination for this module. The input are also chosen from 28 candidates in Table 5.5 and scaled by the standardization algorithm. Then, the best feature set was determined by balancing the classification performance and the lowest dimension of feature vector. Finally, we ended up using only **7 features** for the micro-activity classifier in Table 5.6.

5.3.3 Amount of Fluid Intake Estimation Module

This module aims to estimate the amount of fluid intake for each drinking episode. It uses the *Sip* moments from the result in the previous module, then it extracts some features from the moment in order to estimate the amount of fluid intake. We firstly attempted to leverage the length of the *Sip* duration for the estimation since the volume of fluid intake is intuitively proportional to the *Sip* duration. It worked well in some cases, but gave bad results in many other cases. Figure 5.8 shows the box plot of the *Sip* duration at different amounts of fluid intake using a bottle or a cup (i.e., Lab-micro dataset), in which we find drinking the same amount of fluid takes a different duration that spans a large range of values (i.e., large variance), especially when using a bottle. This is mainly due to the

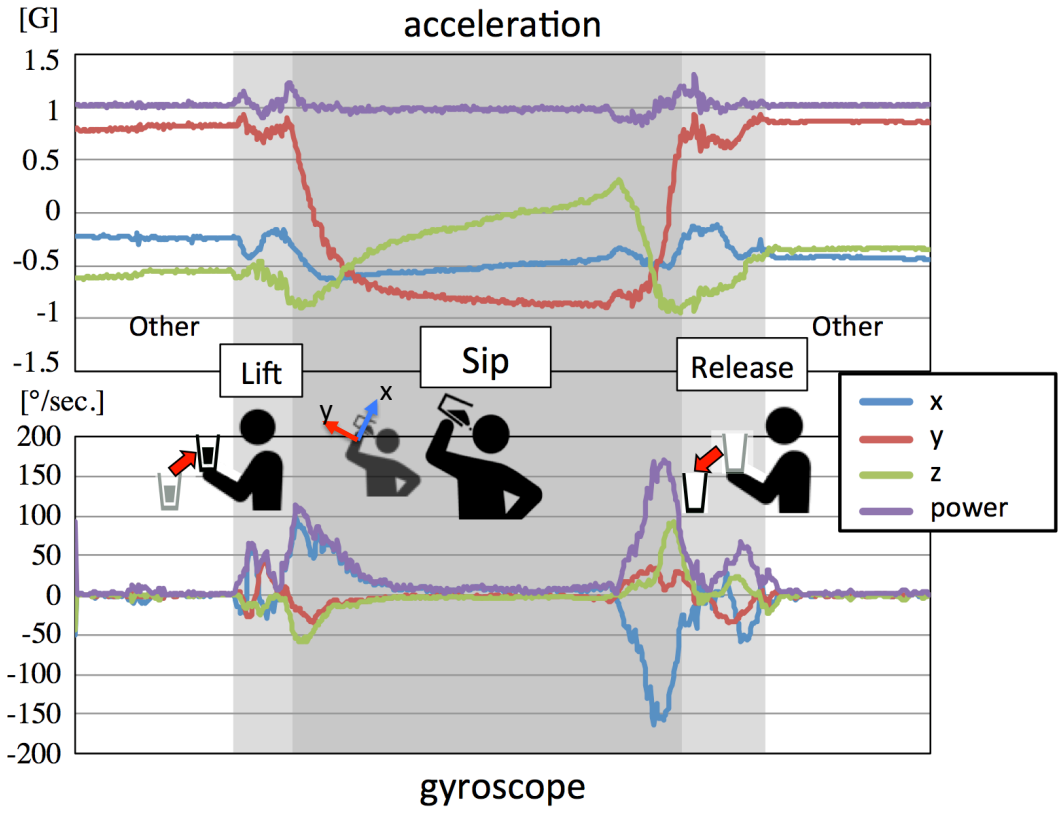
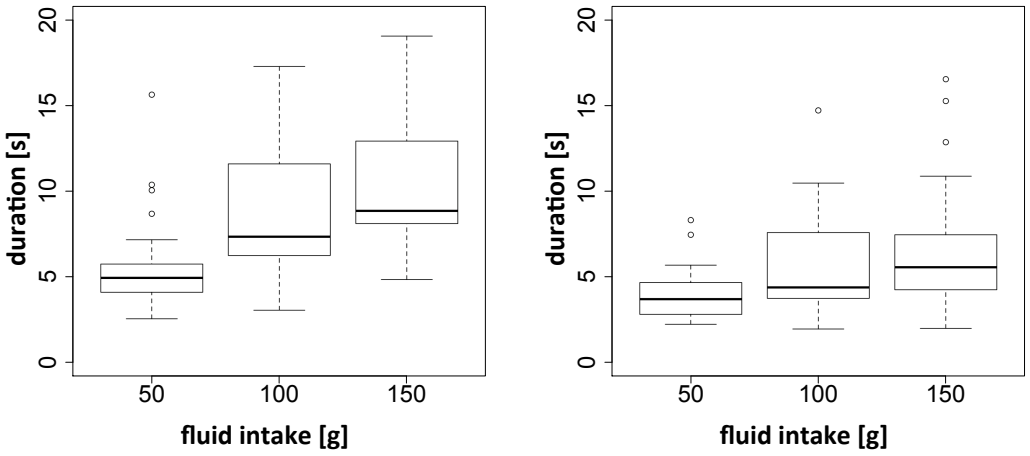


Figure 5.7: Drinking micro-activities sequence and their inertial sensor patterns

difference in the arm posture during drinking (i.e., tilt or the inclination of the container) which affects the speed of flow of the fluid. Additionally, the arm posture during the *Sip* period depends on the amount of fluid in the container. The container inclination increases when it has a small amount of fluid and vice-versa.

Key insights: These problems suggest a duration of the *Sip* gesture is insufficient to explain the amount of fluid intake. To address this challenge, we use the acceleration values. Our intuition is that the 3 axes of the acceleration can represent 3-D arm posture since the accelerations have gravity components which represent the 3-D orientation of arm when the user drinks at stand-still. In other words, the continuous change of 3-D orientation of the arm decides the shape of waveform and peak values of X- and Y- axis accelerations as illustrated in Figure 5.7. As supporting this hypothesis, Figure 5.9 shows the different shapes of waveforms of X- and Y- accelerations at five different instances with different amount of remained fluid in a bottle, suggesting the acceleration values have potential to finely estimate the amount of fluid intake.

Further, we integrate the container inclination (i.e., acceleration values) along with the *Sip* duration for the estimation of the amount of fluid intake. We used the integration of primary axes of acceleration



(a) duration (bottle)

(b) duration (cup)

Figure 5.8: The average *Sip* duration against the fluid intake amount of 180 drinking episodes using a bottle or a cup

(i.e., X and Y) to represent both the width (*Sip* duration) and magnitude of the peak area (i.e., container inclination) as the feature for the estimation of fluid intake amount. Note that the integration values also capture the change of posture during the *Sip* gesture and Z-axis is not used due to its perpendicular to the orientation of the drinking gesture.

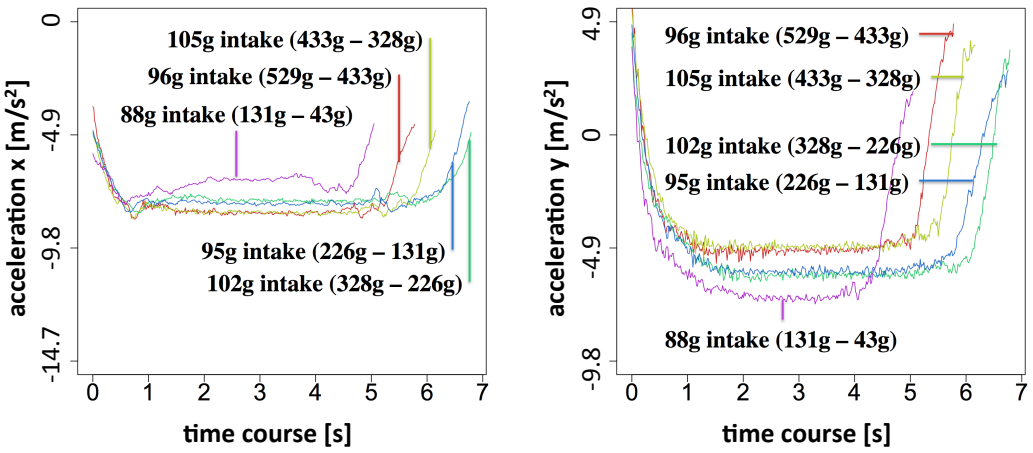
The estimation model: Figure 5.10 illustrates the relationship between the *Sip* duration, the integration value of x and y accelerations and fluid intake through 1069 instances in Lab-micro+ dataset. We find the Pearson correlation coefficients among these features and the fluid intake amount are $R = 0.69, -0.60, -0.55$, which highlight that the *Sip* duration, integration values of x and y accelerations have a good positive or negative correlation with fluid intake amount. Therefore, we build a linear regression model to predict the fluid intake amount based on the integration of the x and y acceleration. With predictor variables $\{I_{a_x}, I_{a_y}\}$, we propose a novel model for the estimation of the fluid intake amount M [g] as:

$$M = p_1 * I_{a_x} + p_2 * I_{a_y} + p_3, \tag{5.3}$$

where p_1, p_2, p_3 represent linear regression parameters, namely two coefficients for each predictor variable and the intercept. We also define a simple model that employs only one predictor variable d_{Sip} (i.e., duration of the *Sip* gesture):

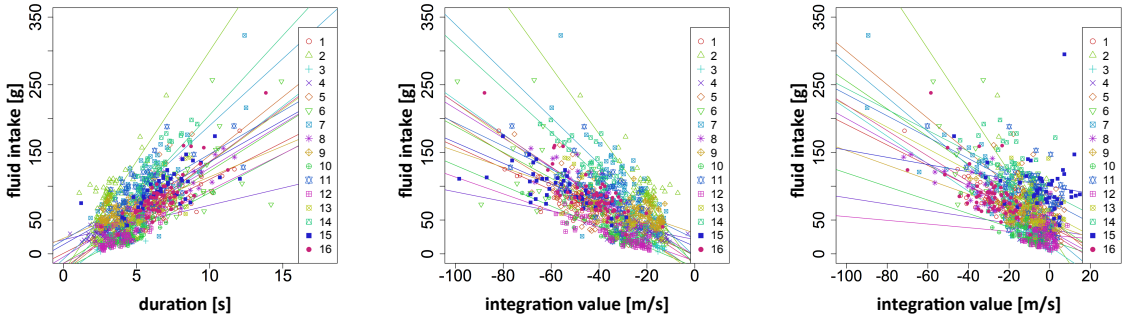
$$M = p_4 * d_{Sip} + p_5, \tag{5.4}$$

This is referred as the benchmark model later. The two models performance are quantified in Section 5.4 .



(a) X-acceleration (b) Y-acceleration

Figure 5.9: Waveforms of X and Y accelerations for different five amounts of fluid intake



(a) *Sip* duration (b) Integration of X-acceleration (c) Integration of Y-acceleration

Figure 5.10: Relationship between the *Sip* duration and integration value of X and Y accelerations against the fluid intake using 16 subjects data

5.4 Performance Evaluation

In this section, we evaluate the performance of the different system modules using the three controlled datasets collected in the laboratory. We use the leave-one-subject-out (LOSO) cross-validation method which validates an unseen subject using the trained model through other known subjects.

Table 5.7: Confusion matrix of the macro-scale activity classification

	move	stand	sit	lie	work	eat	**drink**	other	*Precision*	*Recall*
move	989	96	18	3	30	244	14	12	57.0%	70.3%
stand	94	1246	39	14	47	101	7	11	63.7%	79.9%
sit	158	290	392	17	236	281	55	97	26.4%	25.7%
lie	69	168	450	14	212	287	29	18	12.3%	1.1%
work	138	100	407	61	505	251	11	24	40.6%	33.7%
eat	27	6	28	0	77	1153	52	8	43.0%	85.3%
drink	10	1	3	0	0	85	1328	94	**83.6%**	**87.3%**
other	249	50	148	5	137	282	93	209	44.2%	17.8%

5.4.1 The Macro-scale Activity Classification

Table 5.7 shows the precision and recall of macro-activity classifier generated by applying 9-fold LOSO cross-validation on the Lab-macro dataset. The rows mean true labels and the columns mean predicted labels. It shows that our system can accurately detect the drinking activity achieving 83.6% precision and 87.3% recall. Many of the drinking false negative samples are classified as eating activities given that both have a bit similar arm movement patterns. Since we are targeting the detection of the drinking activity, the extracted features are picked to mainly capture the arm movement during these activities. Therefore, some activities that do not impact the arm posture greatly like (sit, lie, work) are usually misclassified. On the other hand, the eating activity which has a bit similar arm movement to the drinking activity has a high recall rate, however, its precision is low as many activities are misclassified as eating. Consequently, macro-scale classification results of our system (precision =83.6% and recall=87.3%) are comparable to the results achieved by [55] (86.5% precision and 84.3% recall) which fuses dedicated inertial sensors and a throat microphone (i.e., not scalable).

5.4.2 The Micro-scale Activity Classification

Table 5.8 shows the precision and recall of the drinking micro-activities recognition generated by applying 16-fold LOSO cross-validation on the Lab-micro+ dataset. The table indicates that *Sip* gesture can be recognized with higher precision and recall rates (90.7% precision and 96.3% recall) than other gestures due to the unique arm posture (i.e., the arm is inclined and suspended) during the *Sip* period. The high recall rate of the *Sip* gesture guarantees reliable tracking of fluid intake. However, the 9.3% of false positives may cause overestimation of the intake amount. Nevertheless, the effect of false negatives detection of *Sip* may partially balance the overestimation caused by the false positives as discussed in details in the next section. Our *Sip* gesture classification accuracy is comparable to that achieved by the method in [16] which reports 84% precision and 94% recall. However, they fused dedicated wearable inertial sensors worn on the wrist with a couple of magnetic distance sensors that are worn on the shoulder and the wrist.

Table 5.8: Confusion matrix of the micro-scale activity recognition

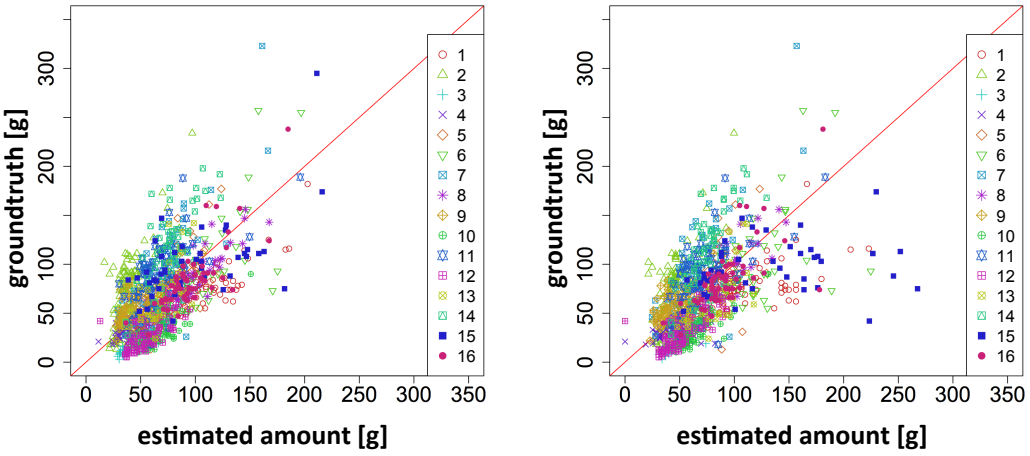
	Other	Lift	**Sip**	Release	*Precision*	*Recall*
Other	8635	1708	622	2680	68.9%	63.3%
Lift	1526	13480	786	395	84.9%	83.3%
Sip	3	462	21534	369	**90.7%**	**96.3%**
Release	2372	221	791	17241	83.4%	83.6%

5.4.3 The Amount of the Fluid Intake Estimation

We evaluated the performance of the fluid intake amount estimation by training the parameters of estimation model in Equations (5.3) and (5.4), and validating them through 16-fold LOSO cross-validation on the Lab-micro+ dataset. During training, the data of all unknown subjects are aggregated (i.e., aggregated training). Figure 5.11 depicts scatter plots of the estimated fluid intake amount and ground-truth amount using the proposed model (using the *Sip* duration and the acceleration integration) and the benchmark model (using the *Sip* duration only). We use the mean absolute percentage error (MAPE) of the estimated amount of fluid intake against the groundtruth amount to evaluate the performance. The resultant MAPE of the proposed and benchmark models are 58.9% and 59.2%, respectively. This large deviation between the estimated and groundtruth amounts was expected given the fewer sensors used by our system (Mirtchouk et al. used dedicated four sensors and showed 47.2% MAPE).

It should be noted that the MAPE measures the error in the amount of fluid intake of each episode. On the other hand, the overall amount of fluid intake error reflects the average deviation between the totally estimated amount of fluid intake of a subject and the groundtruth amount of the total fluid intake of the same subject. Based on the current training models, the average overall fluid intake errors for a subject using the proposed and benchmark models are 34.6% and 34.9% respectively. The overall error is smaller than MAPE since overestimation samples cancel out the underestimation.

To enhance the amount of fluid intake estimation, we constructed a tailored training of model parameters for each subject (i.e., individualized training). Particularly, 10-fold cross validation is used for training and validating each subject own data. Figure 5.12 plots the scatter of estimated fluid intake amount against the groundtruth amount by the individually-trained model. The MAPE value of the proposed model dropped to 29.1% and increased to 64.8% using the benchmark model. This result shows that individual training based on our proposed model of the fluid intake amount estimation gives better result compared to that obtained using the aggregated training as it captures the personal differences in the drinking habits (e.g., arm posture during the *Sip* period). However, the benchmark model result gets worse for two reasons. First, the *sip* duration is not a personal-based quantity as it mainly depends on the human's thirst level. Second, generally the number of samples used in the individual training is lesser than those in the aggregated training. The overall amount of fluid intake error for a subject using the proposed and the benchmark models are 15.1% and 42.5% respectively.



(a) 2-D acceleration model (Eq.(5.3))

(b) benchmark model (Eq.(5.4))

Figure 5.11: Scatter plot of the estimated amount and groundtruth amount of fluid intake (generalized model)

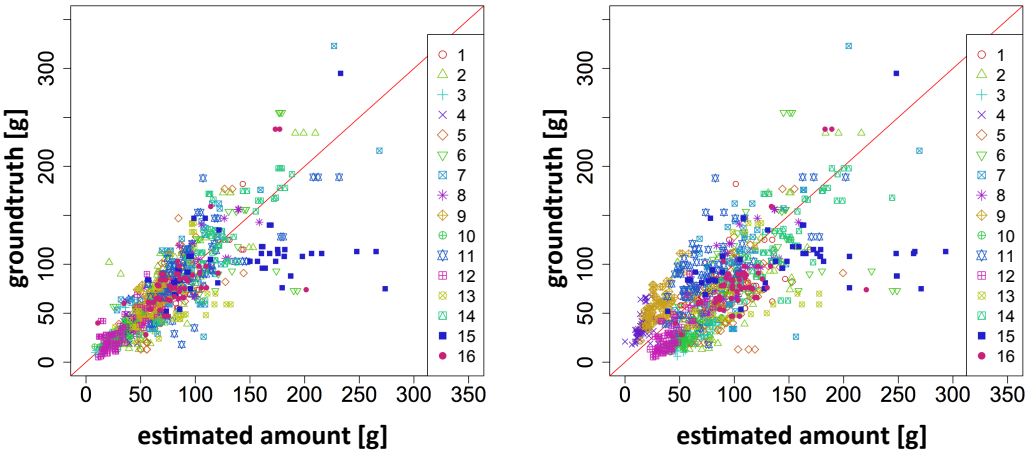
This means that our method can consistently estimate the amount of fluid intake of a subject with an average overall error of 15%. Later, we discuss how many training instances are need to construct an individualized model for a subject.

5.5 In-situ Evaluation

This section aims to validate that our system can consistently estimate the amount of fluid intake in real-life scenarios. We use the data collected in the wild either in the office work (Wild-office dataset) or through daily life outside the office (Wild-day dataset). In this section, we used the pre-trained models for macro and micro activity classifiers, and fluid intake estimation constructed by LOSO cross validation in Section 5.4.

5.5.1 Validation using Data Collected During Office Work

To validate our system, we estimate the accumulative amount of fluid intake of 3 participants while they are doing daily research work for 3 different days. The {*precision*, *recall*} of the drinking recognition for each subject are {100%, 94.1%}, {100%, 92.3%} and {100%, 100%}. This high accuracy (achieved even when they are behaving naturally) is due to the introduction of the *Other* gesture as one of the micro-activities of the drinking activity which helps to remove outlier segments of the extracted drinking traces. Note that we find that some drinking activities in this challenging environment have the following sequence of micro-activities: *Other* $\rightarrow$ *Lift* $\rightarrow$ *Sip* $\rightarrow$ *Release* $\rightarrow$ *Other* due to the error



(a) 2-D acceleration model (Eq.(5.3))

(b) benchmark model (Eq.(5.4))

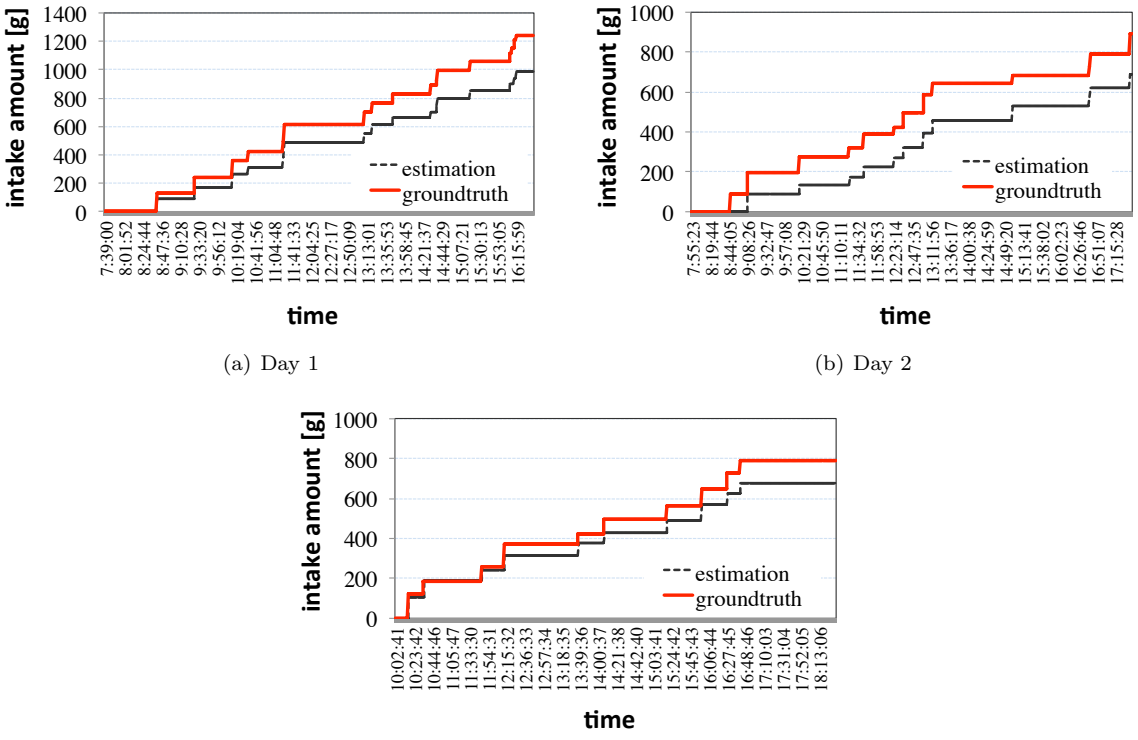
Figure 5.12: Scatter plot of estimated amount and groundtruth amount of fluid intake (individually optimized model)

in extracting the exact drinking moments, presenting some noise in the start and end of the extracted drinking traces. Given that subjects almost used bottles for drinking and given the restricted subject activities in the laboratory, these factors further contributed to the good accuracy.

Figure 5.13 visualizes the discrete time instants during the day working hours on the x-axis and the accumulation of the fluid intake on the y-axis. The figures illustrate the deviation between the estimated and ground truth amounts in 3 different days. The overall errors of daily fluid intake amount estimations are -20.7%, -22.9% and -14.2%. This underestimation of the groundtruth fluid amounts is mainly owing to the false negatives in the drinking recognition, thus missing some fluid intake opportunities. These results suggest that the amount of fluid intake can be captured in real-life scenarios with small underestimations using smartwatches. Generally, underestimation is preferable than overestimation in health support applications.

5.5.2 Validation Using Data Collected in Real-life in a Weekend

To validate that our system archives its goal of estimating the amount of fluid intake in very challenging scenarios, two subjects (male and a female) were asked to collect data during a weekend from waking up to bedding time. During the day, they performed many indoor activities (cooking, walking, taking a nap, etc.) and outdoor activities (shopping, eating at a restaurant, cycling, etc.). The precision and recall of the drinking activity recognition for both subjects are {100%, 66.6%} and {53.8%, 77.8%}. The male subject has a low recall rate (i.e., many false negatives), because he had dinner in a restaurant where he used cups for drinking despite the micro-activity classifier has been trained by a dataset



(a) Day 1

(b) Day 2

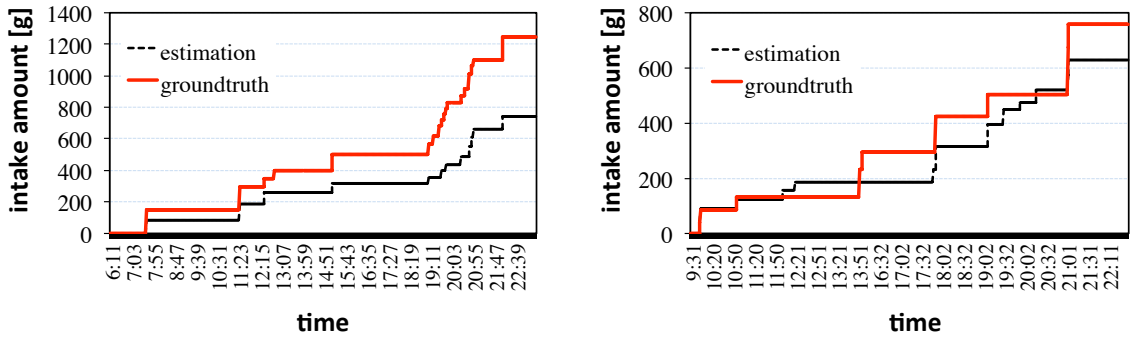
(c) Day 3

Figure 5.13: Accumulative amounts of daily fluid intake for 3 different working days (Wild-office dataset)

collected through fluid intake with bottles. On the other hand, the female subject has low precision rate (many false positives) because she used her hand for eating at lunch time, thus many eating moments are misclassified as drinking moments.

Figure 5.14 visualizes the discrete time instants over a weekend day on the x-axis and accumulation of the fluid intake on the y-axis using a generalized training model. The overall errors of the daily fluid intake amount for the two subjects are -40.9% and -17.1%, respectively. This underestimation is mainly due to the high false negatives in the *Sip* gesture recognition of the male subject. Figure 5.14(a) shows that the gap between the estimated and groundtruth amounts starts to widen from 7:00pm when the male subject was in a restaurant and used cups to drink water and other alcoholic beverages. However, the gap between the estimated and groundtruth amount is small for the female subjects as she almost used a bottle to drink fluids (Figure 5.14(b)). It should be noted that no individualized training models are applied to get an insight of the system performance for first-hand users.

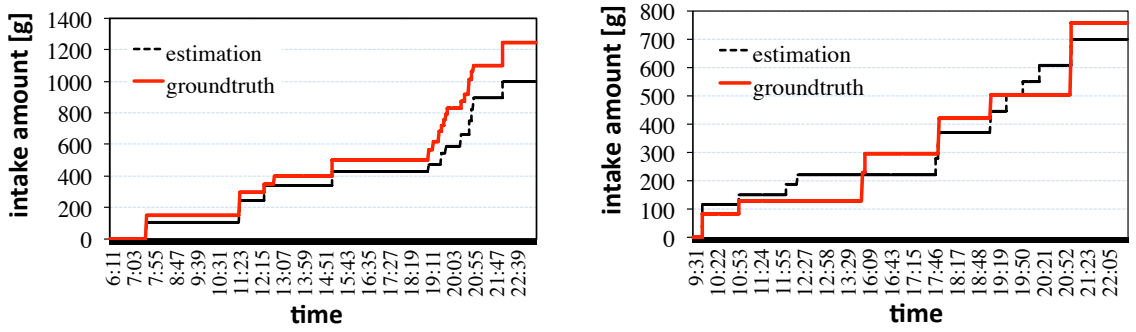
By applying an individualized training models where each subject estimation model is trained by



(a) the male subject

(b) the female subject

Figure 5.14: The accumulative fluid intake amount over a weekend of two subjects (Wild-day dataset- generalized training)



(a) the male subject

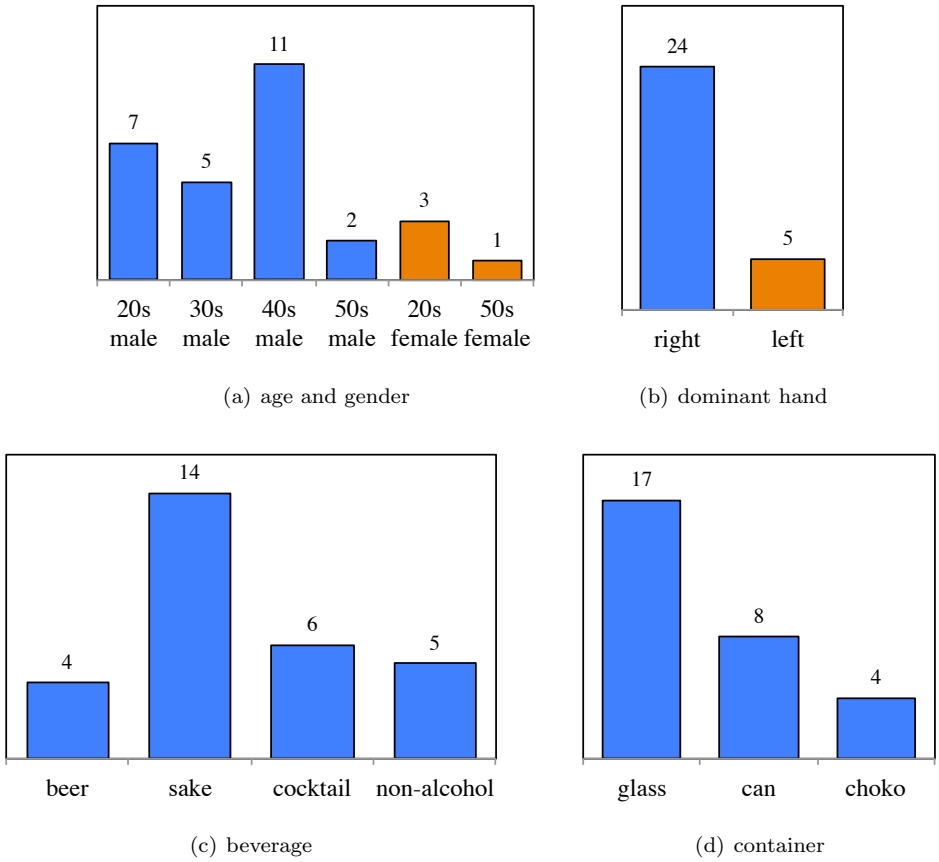
(b) the female subject

Figure 5.15: The accumulative fluid intake amount over a weekend of two different subjects (Wild-day dataset- individual training)

his/her own data collected after this experiment, the overall errors of the daily fluid intake amount estimations fall to -20% and -7.6% for the male and female subjects respectively. Figure 5.15 clarifies that the error of fluid intake amount estimation drops for both subjects after using only self data for training.

5.6 Generalization of the system

This section aims to demonstrate that our method could generalize over various user groups, beverages/container types, and smartwatch models.



(a) age and gender

(b) dominant hand

(c) beverage

(d) container

Figure 5.16: Statistics of the "Nombe Challenge" experiment

5.6.1 Users Generalization

To validate that our system could generalize over users of different ages and genders with different drinking manners, we carried out "Nombe Challenge" ("nombe" means a drunk person in Japanese) in a demonstration session held in an academic conference in Japan. Totally, 29 subjects participated by drinking different beverages using multiple container types. Figure 5.16 shows the statistics of participants gender, the used hand of drinking, and the beverage and the fluid container types. We used leave-one-sample-out cross-validation due to lack of enough samples for each class. The precision and recall of the *Sip* gesture recognition are 87.3% and 94.9%, and the average overall error of the amount of fluid intake is 4.7%. This highlights the ability of our system to track the fluid intake amount consistently for different user groups.

Table 5.9: 240 intake episodes of 17 beverages using 10 container types

17 beverages		10 containers	
type	# of episodes	type	# of episodes
apple juice	6	beer glass	16
beer	2	plastic bottle	42
cider	15	can	55
coffee latte	2	steel bottle	2
coke	10	glass	25
drinkable jerry	5	mug	37
green smoothie	4	paper bottle	6
green tea	20	paper cup	31
hot cocoa	5	straw	23
hot coffee	26	Trago bottle [11]	3
iced coffee	25		
lemonade	15		
milk	7		
energy drink	29		
still water	10		
wheat tea	49		
drinkable yogurt	10		
Σ	240	Σ	240

5.6.2 Beverage and Container Generalization

To validate the generalization nature of our system over different beverage and container types, we collected an extra dataset of the authors' daily life totaling 240 drinking episodes comprising 17 beverage types and 10 container types (Table 5.9). Due to the limited size of data, we trained the classifier for 6 major container categories using leave-one-sample-out cross-validation. Table 5.10 summarizes the performance of the *Sip* micro-activity recognition and the error of fluid intake amount estimations for each container type. Evident from the table, the *Sip* gesture recognition precision and recall values are over 94% for five major container types except using the straw because it does not need to lift the fluid container. Overall, drinking using bottles achieved the worst accuracy (not worse than straw) as many types of bottles are used. Finally, the system can consistently achieve an overall fluid intake amount estimation error less than 30% for all container/beverage types except drinking with straw.

5.6.3 Smartwatches

Nowadays, there are many commodity smartwatches with different limitations on the maximum frequency of sensors sampling. To demonstrate that our system could generalize to different smartwatch models having different sampling rates, we asked a subject to wear both Microsoft Band 2 and Sony Smartwatch 3 at the same time. She is instructed to perform each of the 8 macro-activity classes for 10 minutes/class and 30 episodes of fluid intake. We used the pre-trained CRF classifier (based on the controlled dataset) for the macro and micro activity recognition, and the fluid intake estimation

Table 5.10: The performance of *Sip* gesture recognition and the fluid intake amount estimation using 6 different container types

container type	glass	bottle	can	mug	paper cup	straw
						
precision	94.8%	94.5%	97.7%	94.2%	95.9%	77.0%
recall	95.4%	97.3%	99.1%	97.5%	98.9%	88.0%
overall error (proposed)	-11.7%	-29.7%	8.6%	16.8%	6.8%	35.3%
overall error (benchmark)	21.3%	-46.8%	10.7%	18.8%	6.2%	36.3%
# of instances	41	53	55	37	31	23

Table 5.11: Comparing the system performance using two smartwatches at different sensor sampling rates

Device	Frequency	*drink* activity recognition		*Sip* gesture recognition		Overall error of intake amount	
		Precision	Recall	Precision	Recall	Proposed model	Benchmark Model
Microsoft Band 2	62 Hz	92.5%	80.4%	91.0%	93.9%	24.6%	25.2%
	31 Hz	92.5%	80.4%	91.0%	94.1%	24.5%	25.3%
	16 Hz	93.0%	80.6%	90.6%	93.7%	24.0%	25.9%
	8 Hz	92.9%	80.6%	88.4%	93.8%	23.2%	31.4%
Sony SmartWatch 3	62 Hz	92.1%	87.0%	71.4%	89.7%	33.3%	64.3%
	31 Hz	92.1%	87.2%	73.5%	90.5%	33.5%	59.5%
	16 Hz	91.4%	87.6%	72.2%	87.6%	27.6%	56.7%
	8 Hz	92.1%	87.2%	74.2%	86.8%	26.9%	45.3%

models. Table 5.11 compares the performance of our system using the two smartwatches at different sensor sampling rates. It shows that the performance does not deviate much either between different smartwatches or different sampling rates. This suggests that our system can work well on different smartwatches, even those who have restrictions on sensors sampling rates for the power saving preference.

5.7 Discussion

In this section, we discuss the limitations, some practical considerations, and potential applications of our proposed method.

Limitations: Our system builds on two hypotheses: The user wears the smartwatch on the dominant hand that she usually uses to drink and she usually drinks while being at standstill. The first hypothesis cannot be relaxed but the user could be asked during the app setup to wear the watch in her dominant hand. Although the second hypothesis is applicable in many cases, we are currently

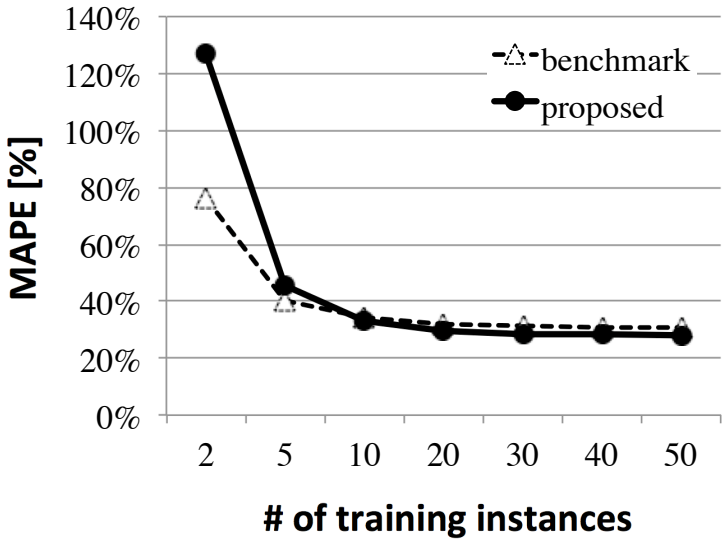


Figure 5.17: Performance of individual's fluid intake amount estimation at different number of her training instances

working to relax it to allow for the estimation of the fluid intake amount while the user is moving which may happen in some cases (e.g., during exercise). We note that these limitations are also common in the previous studies [16, 15, 55].

Energy consumption: Smartwatch inertial sensors are naturally sampled all the time to track the user daily exercises and to be aware of her arm movement to automatically turn the display on or off. Therefore, the proposed system virtually does not consume any excess energy to operate. The only energy footprint of our system is the real-time processing of sampled sensors to estimate the fluid intake amount. This footprint, however, is similar to that of other smartwatch health-related apps (e.g., pedometer).

Adopting individual models: We have shown that adopting individual models for training the regression models can refine the user's amount of fluid intake estimation. However, it is critical to know how much data is enough for our algorithm to build an individualized model that can enhance the system performance. Figure 5.17 quantifies the effect of the number of individual's training instances on the error of the amount of one's fluid intake estimation. We used Lab-micro+ dataset for this and picked up subjects whose counts of fluid intake were over 60. Then, we trained and validated individual models through 10-fold cross-validation for each subject. The figure shows that initially the error drops significantly when the number of training instances increases. Later on, it saturates quickly after tens of instances. The key takeaway is that small number of individual training instances are enough to calibrate the models of one's fluid intake estimation.

Container type: During the previous evaluations (except in wild scenarios), we have assumed

		trained model		Confusion error
		bottle	cup	
validation data	bottle	40.9%	43.3%	+2.4%
	cup	42.8%	31.0%	+11.8%

Table 5.12: The effect of container type confusion on the estimation error of fluid intake amount

that the system is aware of container type and accordingly use the corresponding model to estimate the fluid intake. To quantify the effect of wrongly applied model on the accuracy of fluid intake amount estimation, we used the trained bottle model to estimate the amount of fluid intake using cup and vice-versa. To remove the personal effect, we picked some drinking episodes of the same subject to form two datasets: one contains drinking episodes using bottles while the other contains drinking episodes using cups. Thereafter, we applied 10-fold cross validation where the fluid intake estimation model is trained with 90% of samples of a given container (e.g., bottle) dataset and validated by 10% of sample in both containers (bottle and cup) dataset.

Table 5.12 shows the performance of fluid intake amount estimation model using different mixes of trained and validation data. The diagonal values shows the MAPE of the fluid intake using the same container dataset for both training and validation. The off diagonal values shows the MAPE of the fluid intake amount estimation using different container datasets for training and validation. The last column shows the error in the fluid intake estimation that emanated when the container type is confused (i.e., the system used a trained model of a given container type to estimate the amount of fluid intake using other container type). It confirms that confusing the container type does not have a large impact on the fluid intake amount estimation. Finally, this experiment verifies that our system can maintain a good accuracy of the estimation of the amount of fluid intake using different containers, when the system confuses the container type or when the estimation model is not yet build for a given container.

Potential applications: Our method has many potential applications like hydration notification, automated generation of fluid intake log, and alcoholic avoidance system. It can be integrated with an exercise tracking app to estimate the amount of fluid intake and outtake respectively, thus notifying the user of potential dehydration risks when the accumulative outtake amount is much higher than the intake amount. Additionally, tracking the amount of the daily fluid intake can be integrated with the daily food intake estimated by other apps to generate health monitoring logs (e.g., how much sugar was ingested, how much caffeine the user took). Finally, It can be used with a minor modification to track the individual alcohol intake rate to avoid alcohol overdose which may cause severe health problems like cancer.

5.8 Conclusion

We presented a novel, ubiquitous and unobtrusive system for fluid intake estimation that leverages only the commodity smartwatches embedded sensors. It first recognizes drinking moments by the macro-scale activity classification. Thereafter, it extracts the *Sip* micro-activity that takes place during the drinking activity. Finally, it estimates the amount of fluid intake based on the *Sip* duration and the arm posture during the *Sip* period. We evaluated the system by collecting large and diverse datasets in different environments (indoor and outdoor, controlled and uncontrolled) that spans for more than 80 hours by 51 participants of different genders and experiences using different container and beverage types. Our results show that the system can detect the drinking activity accurately (with 84% precision and 87% recall) as well as identify the *Sip* micro-activity accurately (with 91% precision and 96% recall). Based on this, it can estimate the amount of daily fluid intake in grams with an error limited to 15%. Finally, the system can be generalized over different user groups and is robust to different smartwatch models and container types. We note that we have submitted the journal version of this chapter to IPSJ journal and ACM IMWUT and they are currently under review [101, 102].

Chapter 6

Conclusion

This dissertation has presented a novel approach to heat stroke prevention using core temperature estimation and fluid intake monitoring by smart wrist-worn devices. The primary goal of this dissertation is to realize a pervasive method of core temperature estimation with wide applicability and high accuracy. We have designed a method which relies on the well-known human thermal model and common sensors on the market, and addressed the trade-off between accuracy and its cost in terms of calibration effort and interference in user activity. Also, we have coped with a challenge of the human thermal model in complex exercise and environment by introducing both of a novel parameter set to enhance representation capability of the model and extensional equations leading to wide applicability to a variety of environments. The other goal is to actualize dehydration prevention to maintain the reliability of the human thermal model. The major challenge is to precisely estimate drinking moments and amount of fluid intake by using sensing modalities embedded in the state-of-the-art smart watches. We have addressed this by using two activity classifiers with different time scales and observing the duration as well as the posture in drinking episodes. Through this dissertation, we have elaborated on three primary contributions to catch these goals as follows.

Firstly, we design and evaluate a method to estimate core body temperature which can be concurrently optimized for individuals with little effort. The method employs Gagge's two-node model for core temperature simulation as well as the optimization of four parameters representing differences of human thermoregulation capability. This realizes the estimation of core temperature which can be seamlessly deployed in any environment with minimal interference in user activity as well as wide scalability owing to requiring only common devices on the market. Despite limited information was available during exercise, we confirmed the method successfully estimated core temperature and it could choose an effective set of four parameters in moderate exercise.

Secondly, we propose essential extensions for Gagge's two-node model to realize core temperature estimation in the wild. To suit the two-node model to exercises with variable workload, we propose two delay parameters which approximate potential delays of human responses. The model is also extended to consider the effect of solar radiation, wind, and water ingestion. The method is evaluated through

over 120 hours of walking, running, biking, and tennis and achieves up to 0.30 degree Celsius error in each exercise. Further, a prototype of heat stroke caution has been developed and showed effectiveness in a real application.

Thirdly, we propose automated estimation of fluid intake to prevent dehydration by leveraging inertial sensors in wrist-worn gadgets. We collected data in both of laboratory and wild experiments and designed a novel algorithm to capture drinking episode from noisy motion traces of the wrist. The algorithm recognizes human activities with macro and micro scale classifiers so as to firstly separate drinking activity from the other activities and then recognize small gestures constituting a series of fluid intake. We also propose a robust model for estimating amount of fluid intake which uses not only the duration of sipping gesture, but also the posture of the arm. The method has been validated through rigorous datasets and shown 15% error for overall fluid intake amount.

Through these contributions, commercial-off-the-shelf smart devices on the wrist have been utilized to assess rise of core temperature and amount of fluid intake. Our study leaves potentials of further studies for reinforcing performance of core temperature estimation and fluid intake assessment. For instance, we expect the estimation of core temperature to be enhanced by either precisely considering energy production and expenditure in sports using unique information for each sport (e.g., measuring the kicking power in soccer), or employing additional sensing modalities such as a thermography to finely measure skin temperature. The performance of gauging fluid intake is also foreseen to be improved by considering deep circumstance regarding each drinking episode (e.g., container type used, fluid type consumed, location and time of drinking). Fusing smart wrist-worn devices with smartphones, infrastructures, and other sensorized gadgets will help these extensions. Consequently, this dissertation has established the foundation of pervasive method for preventing heat stroke for all mankind.

Acknowledgement

I would like to express my sincerely appreciation to continued support of my supervisor Professor Teruo Higashino of Osaka University through trials and tribulations of this Ph.D thesis. Again I express my heartiest gratitude to him for his encouragement and invaluable comments in preparing this thesis.

I am very grateful to Professor Masayuki Murata, Professor Takashi Watanabe, Professor Toru Hasegawa and Professor Morito Matsuoka of Osaka University for their invaluable comments and helpful suggestions concerning this thesis.

I am heartily grateful to Assistant Professor Akira Uchiyama for the precious advices and technical discussions provided through out the research.

I would like to thank Associate Professor Hirozumi Yamaguchi and Akihito Hiromori of Osaka University, and Associate Professor Takaaki Umedu of Shiga University for their valuable comments.

I also thank Assistant Professor Moustafa Elhamshary of Tanta University for his friendly cooperation.

Thanks go to everyone of Higashino laboratory for their feedback, encouragement and support.

Finally, I would like to thank my family and my friends for their help and understanding.

Bibliography

[1] Japanese Ministry of the Environment, "Precautions for exercise," http://www.wbgt.env.go.jp/pdf/envman/3-3.pdf.

[2] Basis, "Basis peak — the ultimate fitness and sleep tracker," 2014. [Online]. Available: http://www.mybasis.com/

[3] Kyoto Electronics Manufacturing Co., Ltd., "Heat stroke checker [wbgt-203a/213a]," 2012. [Online]. Available: http://www.kyoto-kem.com/en/product/wbgt2xx/

[4] C. Koppe, S. Kovats, G. Jendritzky, B. Menne, D. J. Breuer, and D. Wetterdienst, *Heat waves: risks and responses*. Regional Office for Europe, World Health Organization, 2004.

[5] L. R. Leon and B. G. Helwig, "Heat stroke: role of the systemic inflammatory response," *Journal of Applied Physiology*, vol. 109, no. 6, pp. 1980–1988, 2010.

[6] H. Inc., "Cortemp sensor," 2013. [Online]. Available: http://www.hqinc.net/cortemp-sensor-2/

[7] T. S. Inc., "Tympanic temperture logger dbtl-2," 2013. [Online]. Available: http://www.t-science.jp/doc/dbtl-2.html

[8] 3M, "3m spoton system - product brochure," 2013. [Online]. Available: http://multimedia.3m.com/mws/media/878163O/spoton-system-brochure.pdf

[9] M. J. Buller, W. J. Tharion, C. M. Duhamel, and M. Yokota, "Real-time core body temperature estimation from heart rate for first responders wearing different levels of personal protective equipment," *Ergonomics*, vol. 58, no. 11, pp. 1830–1841, 2015.

[10] N. Yabuki, T. Onoue, T. Fukuda, and S. Yoshida, "A heatstroke prediction and prevention system for outdoor construction workers," *Visualization in Engineering*, vol. 1, no. 1, p. 1, 2013.

[11] TRAGO, "Trago cap," 2015, https://trago.co/pages/trago-cap.

[12] Hidrate, "Hidrate spark," 2015, http://hidratespark.com/.

[13] J.-L. Chua, Y. C. Chang, M. H. Jaward, J. Parkkinen, and K.-S. Wong, "Vision-based hand grasping posture recognition in drinking activity," in *Proceedings of 2014 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*. IEEE, 2014, pp. 185–190.

[14] H. Kalantarian and M. Sarrafzadeh, "Audio-based detection and evaluation of eating behavior using the smartwatch platform," *Computers in biology and medicine*, vol. 65, pp. 1–9, 2015.

[15] M. Mirtchouk, C. Merck, and S. Kleinberg, "Automated estimation of food type and amount consumed from body-worn audio and motion sensors," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 2016, pp. 451–462.

[16] O. Amft, D. Bannach, G. Pirkl, M. Kreil, and P. Lukowicz, "Towards wearable sensing-based assessment of fluid intake," in *Proceedings of 2010 8th IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops)*. IEEE, 2010, pp. 298–303.

[17] U.S. Environmental Protection Agency, "Climate change indicators in the United States, 2016. Fourth edition." 2016, http://www.epa.gov/climate-indicators.

[18] P. A. Mackowiak, S. S. Wasserman, and M. M. Levine, "A critical appraisal of 98.6 f, the upper limit of the normal body temperature, and other legacies of carl reinhold august wunderlich," *Jama*, vol. 268, no. 12, pp. 1578–1580, 1992.

[19] Y. Shapiro and D. Seidman, "Field and clinical observations of exertional heat stroke patients." *Medicine and science in sports and exercise*, vol. 22, no. 1, pp. 6–14, 1990.

[20] E. F. Adolph *et al.*, *Physiology of Man in the Desert*. New York & London: Interscience Publishers, Inc., 1947.

[21] R. C. Cork, R. W. Vaughan, and L. S. Humphrey, "Precision and accuracy of intraoperative temperature monitoring," *Anesthesia & Analgesia*, vol. 62, no. 2, pp. 211–214, 1983.

[22] D. J. Casa, S. M. Becker, M. S. Ganio, C. M. Brown, S. W. Yeargin, M. W. Roti, J. Siegler, J. A. Blowers, N. R. Glaviano, R. A. Huggins, *et al.*, "Validity of devices that assess body temperature during outdoor exercise in the heat," *Journal of athletic training*, vol. 42, no. 3, p. 333, 2007.

[23] R. Gonzalez, Y. Nishi, and A. Gagge, "Experimental evaluation of standard effective temperature a new biometeorological index of man's thermal discomfort," *International journal of biometeorology*, vol. 18, no. 1, pp. 1–15, 1974.

[24] Y. Epstein and D. S. Moran, "Thermal comfort and the heat stress indices," *Industrial health*, vol. 44, no. 3, pp. 388–398, 2006.

[25] IDC, "Worldwide wearables market to nearly double by 2021, according to IDC," 2017, https://www.idc.com/getdoc.jsp?containerId=prUS42818517.

[26] A. Gagge, "An effective temperature scale based on a simple model of human physiological regulatory response," *ASHRAE TRANSACTIONS*, vol. 77, no. 2192, pp. 247–262, 1971.

[27] J. A. Stolwijk, "A mathematical model of physiological temperature regulation in man," National Aeronautics and Space Administration (NASA), Tech. Rep. CR-1855, 1971.

[28] C. E. Smith, "A transient three-dimensional model of the human thermal system," *PhD. Thesis of Kansas State University*, pp. 1–177, 1991.

[29] S. Tanabe, K. Kobayashi, J. Nakano, Y. Ozeki, and M. Konishi, "Evaluation of thermal comfort using combined multi-node thermoregulation (65mn) and radiation models and computational fluid dynamics (cfd)," *Energy and Buildings*, vol. 34, no. 6, pp. 637–646, 2002.

[30] D. Fiala, K. J. Lomas, and M. Stohrer, "Computer prediction of human thermoregulatory and temperature responses to a wide range of environmental conditions," *International Journal of Biometeorology*, vol. 45, no. 3, pp. 143–159, 2001.

[31] Y. Kobayashi and S.-i. Tanabe, "Development of jos-2 human thermoregulation model with detailed vascular system," *Building and Environment*, vol. 66, pp. 1–10, 2013.

[32] S. Takada, H. Kobayashi, and T. Matsushita, "Thermal model of human body fitted with individual characteristics of body temperature regulation," *Building and Environment*, vol. 44, no. 3, pp. 463–470, 2009.

[33] J. D. Hardy, E. F. Du Bois, and G. Soderstrom, "The technic of measuring radiation and convection one figure," *The Journal of Nutrition*, vol. 15, no. 5, pp. 461–475, 1938.

[34] J. H. Choi, J. Lee, H. T. Hwang, J. P. Kim, J. C. Park, and K. Shin, "Estimation of activity energy expenditure: accelerometer approach," in *Proceedings of the 27th Annual International Conference of Engineering in Medicine and Biology Society (EMBS 2005)*. IEEE, 2006, pp. 3830–3833.

[35] J.-H. Choi, J.-W. Lee, and K.-S. Shin, "A generalized calorie estimation algorithm using 3-axis accelerometer," *Journal of Biomedical Engineering Research*, vol. 27, no. 6, pp. 301–309, 2006.

[36] J. Lester, C. Hartung, L. Pina, R. Libby, G. Borriello, and G. Duncan, "Validated caloric expenditure estimation using a single body-worn sensor," in *Proceedings of the 11th international conference on Ubiquitous computing*. ACM, 2009, pp. 225–234.

[37] A. Zhan, M. Chang, Y. Chen, and A. Terzis, "Accurate caloric expenditure of bicyclists using cellphones," in *Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems*, 2012, pp. 71–84.

[38] Microsoft, "Kinect for Windows," 2017, https://developer.microsoft.com/en-us/windows/kinect.

[39] P.-F. Tsou and C.-C. Wu, "Estimation of calories consumption for aerobics using kinect based skeleton tracking," in *Proceedings of 2015 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 2015, pp. 1221–1226.

[40] L. Rabiner and B. Juang, "An introduction to hidden markov models," *ieee assp magazine*, vol. 3, no. 1, pp. 4–16, 1986.

[41] J. Lafferty, A. McCallum, and F. C. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," 2001.

[42] D. L. Vail, M. M. Veloso, and J. D. Lafferty, "Conditional random fields for activity recognition," in *Proceedings of the 6th international joint conference on Autonomous agents and multiagent systems*. ACM, 2007, p. 235.

[43] F. Peng and A. McCallum, "Information extraction from research papers using conditional random fields," *Information processing and management*, vol. 42, no. 4, pp. 963–979, 2006.

[44] W. Gu, M. Jin, Z. Zhou, C. J. Spanos, and L. Zhang, "Metroeye: Smart tracking your metro trips underground," in *Proceedings of the 13th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*. ACM, 2016, pp. 84–93.

[45] O. Amft and G. Troster, "On-body sensing solutions for automatic dietary monitoring," *IEEE pervasive computing*, vol. 8, no. 2, pp. 62–70, 2009.

[46] C. Merck, C. Maher, M. Mirtchouk, M. Zheng, Y. Huang, and S. Kleinberg, "Multimodality sensing for eating recognition," in *Proceedings of the 10th EAI International Conference on Pervasive Computing Technologies for Healthcare*. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2016, pp. 130–137.

[47] T. Rahman, A. T. Adams, M. Zhang, E. Cherry, B. Zhou, H. Peng, and T. Choudhury, "Bodybeat: a mobile system for sensing non-speech body sounds." in *MobiSys*, vol. 14, 2014, pp. 2–13.

[48] E. Thomaz, I. Essa, and G. D. Abowd, "A practical approach for recognizing eating moments with wrist-mounted inertial sensing," in *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 2015, pp. 1029–1040.

[49] Y. Dong, J. Scisco, M. Wilson, E. Muth, and A. Hoover, "Detecting periods of eating during free-living by tracking wrist motion," *IEEE journal of biomedical and health informatics*, vol. 18, no. 4, pp. 1253–1260, 2014.

[50] E. Thomaz, A. Bedri, T. Prioleau, I. Essa, and G. D. Abowd, "Exploring symmetric and asymmetric bimanual eating detection with inertial sensors on the wrist," in *Proceedings of the 1st Workshop on Digital Biomarkers*. ACM, 2017, pp. 21–26.

[51] Y. Shen, J. Salley, E. Muth, and A. Hoover, "Assessing the accuracy of a wrist motion tracking method for counting bites across demographic and food variables," *IEEE journal of biomedical and health informatics*, vol. 21, no. 3, pp. 599–606, 2017.

[52] H. Kalantarian, N. Alshurafa, and M. Sarrafzadeh, "Detection of gestures associated with medication adherence using smartwatch-based inertial sensors," *IEEE Sensors Journal*, vol. 16, no. 4, pp. 1054–1061, 2016.

[53] A. Bedri, R. Li, M. Haynes, R. P. Kosaraju, I. Grover, T. Prioleau, M. Y. Beh, M. Goel, T. Starner, and G. Abowd, "Earbit: Using wearable sensors to detect eating episodes in unconstrained environments," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 3, pp. 37:1–37:20, Sept. 2017. [Online]. Available: http://doi.acm.org/10.1145/3130902

[54] C. Bi, G. Xing, T. Hao, J. Huh, W. Peng, and M. Ma, "Familylog: A mobile system for monitoring family mealtime activities," in *Proceedings of 2017 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. IEEE, 2017, pp. 21–30.

[55] Y. Mengistu, M. Pham, H. M. Do, and W. Sheng, "Autohydrate: A wearable hydration monitoring system," in *Proceedings of 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2016, pp. 1857–1862.

[56] M. Mirtchouk, D. Lustig, A. Smith, I. Ching, M. Zheng, and S. Kleinberg, "Recognizing eating from body-worn sensors: Combining free-living and laboratory data," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 3, pp. 85:1–85:20, Sept. 2017. [Online]. Available: http://doi.acm.org/10.1145/3131894

[57] F. Calise, "Hydrate," 2013, https://play.google.com/store/apps/details?id=com.frankcalise.h2droid.

[58] C. LLC., "iHydrate," 2016, http://www.ihydrateapp.com/.

[59] M.-C. Chiu, S.-P. Chang, Y.-C. Chang, H.-H. Chu, C. C.-H. Chen, F.-H. Hsiao, and J.-C. Ko, "Playful bottle: a mobile social persuasion system to motivate healthy water intake," in

Proceedings of the 11th international conference on Ubiquitous computing. ACM, 2009, pp. 185–194.

[60] B. Dong, R. Gallant, and S. Biswas, "A self-monitoring water bottle for tracking liquid intake," in *Proceedings of the 2014 IEEE Healthcare Innovation Conference (HIC).* IEEE, 2014, pp. 311–314.

[61] P. H. Dietz, D. Leigh, and W. S. Yerazunis, "Wireless liquid level sensing for restaurant applications," in *Sensors, 2002. Proceedings of IEEE*, vol. 1. IEEE, 2002, pp. 715–720.

[62] A. Jayatilaka and D. C. Ranasinghe, "Real-time fluid intake gesture recognition based on batteryless UHF RFID technology," *Pervasive and Mobile Computing*, vol. 34, pp. 146–156, 2017.

[63] HydraCoach, "Hydracoach," 2015, https://hydracoach.com/.

[64] "Temperature measurement in paediatrics," *Paediatrics & Child Health.*, vol. 5(5), pp. 273–276, 2000.

[65] hothead technologies, "H.O.T. system," 2013. [Online]. Available: http://www.hotheadtechnologies.com/portfolio/product-2/

[66] N. Kunioka, M. Tokumoto, and M. Shinohara, "Study on the heating value of human body by solar radiation," *Summaries of Technical Papers of Annual Meeting Architectural Institute of Japan*, vol. 1998, pp. 387–388, 1998, (in Japanese).

[67] K. Kuwabara et al., "P-26 solar reflectance, transmittance and absorptance of different clothing materials," *Proceedings of Symposium on HES*, vol. 34, pp. 209–212, 2010.

[68] D. Du Bois and E. F. Du Bois, "Clinical calorimetry: tenth paper a formula to estimate the approximate surface area if height and weight be known," *Archives of internal medicine*, vol. 17, no. 6_2, pp. 863–871, 1916.

[69] J. M. Janot, "Calculating caloric expenditure," *IDEA Fitness Journal*, vol. 2, no. 6, pp. 32–33, 2005.

[70] N. Uth, H. Sørensen, K. Overgaard, and P. K. Pedersen, "Estimation of vo2max from the ratio between hrmax and hrrest–the heart rate ratio method," *European journal of applied physiology*, vol. 91, no. 1, pp. 111–115, 2004.

[71] S. R. Colberg, D. P. Swain, and A. I. Vinik, "Use of heart rate reserve and rating of perceived exertion to prescribe exercise intensity in diabetic autonomic neuropathy," *Diabetes care*, vol. 26, no. 4, pp. 986–990, 2003.

[72] J. Karvonen and T. Vuorimaa, "Heart rate and exercise intensity during sports activities," *Sports Medicine*, vol. 5, no. 5, pp. 303–311, 1988.

[73] H. Tanaka, K. D. Monahan, and D. R. Seals, "Age-predicted maximal heart rate revisited," *Journal of the American College of Cardiology*, vol. 37, no. 1, pp. 153–156, 2001.

[74] G. Cavagna and M. Kaneko, "Mechanical work and efficiency in level walking and running," *The Journal of Physiology*, vol. 268, no. 2, pp. 467–481, 1977.

[75] J. H. van Beek, F. Supandi, A. K. Gavai, A. A. de Graaf, T. W. Binsl, and H. Hettling, "Simulating the physiology of athletes during endurance sports events: modelling human energy conversion and metabolism," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 369, no. 1954, pp. 4295–4315, 2011.

[76] O. Tetens, "Uber einige meteorologische begriffe," *Z. Geophys.*, vol. 6, pp. 297–309, 1930.

[77] OMRON, "Omron gentle temp 510 (in japanese)," 2002. [Online]. Available: http://www.healthcare.omron.co.jp/product/mc/mc-510.html

[78] E. Instruments, "Solar Sensor ML-01," 2014. [Online]. Available: http://eko.co.jp/meteorology/met_products/0018.html/

[79] T. Hamatani, A. Uchiyama, and T. Higashino, "Estimating core body temperature based on human thermal model using wearable sensors," in *Proceedings of the 30th Annual ACM Symposium on Applied Computing*. ACM, 2015, pp. 521–526.

[80] ——, "Real-time calibration of a human thermal model with solar radiation using wearable sensors," in *Proceedings of the 2015 workshop on Wearable Systems and Applications*. ACM, 2015, pp. 45–50.

[81] ——, "Estimating core body temperature under hot environment based on human thermal model using wearable sensors," *IPSJ Journal*, vol. 56, no. 10, pp. 2033–2043, 2015.

[82] T. Oka and S. Obara, "Changes in temperature of external auditory canal and sweat rate during 60-min pedaling exercise," *Journal of human sciences, Faculty of Integrated Arts and Sciences, the University of Tokushima*, vol. 7, pp. 1–9, 1999, (in Japanese).

[83] N. Ohnishi and T. Ogawa, "Initial fall in core temperature following the start of exercise," *Japanese Journal of Physical Fitness and Sports Medicine*, vol. 39, no. 6, p. 626, dec 1990. [Online]. Available: http://ci.nii.ac.jp/naid/110001924047/en/

[84] J. Fohr, *Heat and Moisture Transfer between Human Body and Environment*. John Wiley & Sons, 2015.

[85] M. Oguro, E. Arens, H. Zhang, K. Tsuzuki, and T. Katayama, "Measurement of projected area factors for thermal radiation analysis on each part of the human body," *Journal of Archit. Plann. Environ.*, 2001.

[86] S. Yoshida, "Cfd analysis of outdoor thermal environment incorporated multifractional human thermoregulation model," *NAGARE: Journal of Japan Society of Fluid Mechanics*, vol. 30, no. 2, pp. 87–96, 2011, (in Japanese).

[87] A. Yorimoto, S. Nakai, T. Yoshida, and T. Morimoto, "Relationship between drinking behavior and body temperature during exercise in heat," *Japanese Journal of Physical Fitness and Sports Medicine*, vol. 44, no. 3, pp. 357–363, 1995.

[88] adidas, "adidas bluetooth smart compatible speed_cell," 2011. [Online]. Available: http://www.adidas.com/us/bluetooth-smart-compatible-speed_cell/G75090.html/

[89] J. B. Mitchell, K. R. Goldston, A. N. Adams, K. M. Crisp, B. B. Franklin, A. Kreutzer, D. X. Montalvo, M. G. Turner, and M. D. Phillips, "Temperature measurement inside protective headgear: comparison with core temperatures and indicators of physiological strain during exercise in a hot environment," *Journal of occupational and environmental hygiene*, vol. 12, no. 12, pp. 866–874, 2015.

[90] Y. Xue, S. Hosokawa, S. Murata, S. Kouchi, and K. Fujinami, "An on-body placement-aware heatstroke alert on a smartphone," in *Computer Applications for Communication, Networking, and Digital Contents*. Springer, 2012, pp. 226–234.

[91] G. Florea, R. Dobrescu, D. Popescu, and M. Dobrescu, "Wearable system for heat stress monitoring in firefighting applications," in *Proceedings 2nd Int. Conf. Inf. Technol. Comput. Networks (ITCN '13)*, 2013, pp. 129–134.

[92] G. McGregor, P. Bessemoulin, K. Ebi, and B. Menne, "Heat waves and health: Guidance on 7 warning system development 8," *WORLD METEOROLOGICAL ORGANIZATION*, vol. 14, p. 15, 2010.

[93] T. Hamatani, A. Uchiyama, and T. Higashino, "On-line core temperature estimation using wearable sensors during sport with variable exercise intensity," *IPSJ Journal*, vol. 58, no. 11, pp. 1818–1831, 2017.

[94] ——, "Heatwatch: Preventing heatstroke using a smart watch," in *Proceedings of 2017 IEEE Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*. IEEE, 2017, pp. 661–666.

[95] Microsoft, "Microsoft Band — Official Site," 2014. [Online]. Available: https://www.microsoft.com/microsoft-band/en-us

[96] MOFILY, "Yocam: The world's most versatile waterproof life camera," 2015, http://www.getyocam.com/.

[97] A. C. Müller and S. Behnke, "pystruct - learning structured prediction in python," *Journal of Machine Learning Research*, vol. 15, pp. 2055–2060, 2014.

[98] S. Lacoste-Julien, M. Jaggi, M. Schmidt, and P. Pletscher, "Block-coordinate frank-wolfe optimization for structural svms," *arXiv preprint arXiv:1207.4747*, 2012.

[99] C.-W. Hsu, C.-C. Chang, C.-J. Lin, *et al.*, "A practical guide to support vector classification," 2003.

[100] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *Journal of machine learning research*, vol. 3, no. Mar, pp. 1157–1182, 2003.

[101] T. Hamatani, M. Elhamshary, A. Uchiyama, and T. Higashino, "Fluidmeter: Gauging the human daily fluid intake using smartwatches," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, (submitted).

[102] ——, "Proposal of fluid intake estimation using inertial sensors in a smartwatch," *IPSJ Journal*, (submitted).