

CARNEGIE MELLON UNIVERSITY

DOCTORAL THESIS

**A Bio-sensing and Reinforcement Learning
Control System for Personalized Thermal Comfort
and Energy Efficiency**

by

Chenlu ZHANG

September 26, 2019

A thesis submitted in partial fulfillment of the requirements

for the degree of Doctor of Philosophy

in

Building Performance and Diagnostics

School of Architecture

Doctoral Committee:

Vivian Loftness (Chair), University professor, School of Architecture

Sivaraman Balakrishnan, Ph.D., Assistant Professor, Department of Statistics

Ömer Tugrul Karagüzel, Ph.D., Assistant Teaching Professor, School of Architecture

Azizan Abdul Aziz, Data Analytics Professor, School of Architecture

“Do the difficult things while they are easy and do the great things while they are small. A journey of a thousand miles must begin with a single step.”

Lao zi

Abstract

A comfortable indoor thermal environment plays a crucial role in preserving occupant health and productivity. In most office building today, the indoor thermal environment is regulated by heating, cooling, and air-conditioning (HVAC) systems with static schedule-based rules. While prevalent, this control strategy has resulted in low thermal satisfaction rates and energy waste. A growing number of researchers are focusing on occupant-centric building controls and applying various advanced control methods to improve thermal comfort and energy efficiency. However, it is still challenging to integrate occupants' personalized requirements into a control system with a capability of learning from the environment. This thesis has developed a bio-sensing and reinforcement learning control system for continuously integrating occupants' bio-signals into the operation of different heating, cooling, and ventilation systems, learning through interaction to achieve personalized thermal comfort and energy savings.

A bio-sensing and reinforcement learning control (Bio-REAL) system is comprised of a bio-sensing network, multiple Bio-REAL agents, and a negotiator. The bio-sensing network uses smart wristbands to measure occupants

wrist temperature in real-time. The Bio-REAL agent initiates the best control actions on behalf of each occupant in response to the wrist temperature, subjective feedback, and environmental conditions. The negotiator resolves conflicts in the control actions initiated by different Bio-REAL agents to maximize collective thermal comfort and minimize energy consumption. A state-of-art reinforcement learning algorithm, double Q learning with experiment replay and neural network approximation, is applied to train the Bio-REAL agents.

This thesis evaluates the Bio-REAL systems using three types of experimental techniques: simulation experiments, preliminary field and simulation experiments, and field experiments. The simulation experiment trains a Bio-REAL system with three virtual occupants and an office room with a variable air volume (VAV) system in a heating season. The three virtual occupants are simulated using classic thermal comfort models. The room of a small-sized office building is modeled by the EnergyPlus simulation tool. The preliminary heating season field and simulation experiments gather data from six occupants, providing inputs to create the personalized occupant models. The experimental test space is a room with water-sourced radiators for heating and modeled by the EnergyPlus tool. The co-simulation with personalized occupant models and EnergyPlus model assesses the performance of the Bio-REAL system. The cooling season field experiment evaluates the real-world performance of the Bio-REAL systems with fourteen occupants in a tropical climate, occupying a studio with ambient temperature controls and shared controls of ceiling fans.

The three types of experiments each demonstrated that the Bio-REAL system has more advantages for improving thermal comfort and energy efficiency compared to the conventional control systems based on thermal comfort models and static schedules. With the combinations of bio-sensing and learning capability, the Bio-REAL system was able to derive dynamic and adaptive control policies, mapping occupants' personalized requests and the changes of indoor and outdoor environmental conditions to optimum control actions.

The Bio-REAL system contributes an innovative approach for controlling building conditioning systems, to deliver thermal comfort for each individual at the lowest energy possible, with benefits for occupant health and productivity, as well as sustainability. The Bio-REAL research addresses individual differences in thermal comfort for multi-occupant spaces with limited individual controls. It also addresses a range of heating and cooling choices from ambient to task systems. The structure and learning process of the Bio-REAL system, the strategies for the simulations, and the real-world implementation offer creative solutions for building control systems, contributing to the application of the Internet of Things and artificial intelligence in buildings.

Acknowledgements

I would like to express my sincere gratitude to my advisor Prof. Vivian Loftness for her continuous support of my Ph.D. study and related research, for her patience, vision, and immense knowledge. I would also like to thank Prof. Ömer Tugrul Karagüze, Prof. Sivaraman Balakrishnan, and Prof. Aziz Abdul-Aziz for their insightful feedback and comments, as well as encouragement. My sincere thanks also go to Prof. Khee Poh Lam and Mr. Bertrand Lasternas, who provided me with the valuable opportunity to conduct an in-depth field research at the National University of Singapore.

I thank my labmates Zhiang Zhang, Wannan Zhang, Siliang Lu, Yuqi Pan, Linhao Li, Youngjoo Son, Yulu Wang, and James Katungyi for their contributions to my research experiments, for the stimulating discussions, for the day and nights we were working together, and for all the fun we have had in the last five years. Sincerely thanks to my friends and classmates at Carnegie Mellon University and the National University of Singapore who participated in the experiments voluntarily. Without your precious support, it would not be possible for me to complete the research.

I am deeply grateful to my family. I want to thank my father Youpeng Zhang, my mother Xiaoe Liu, my brother Yi Zhang, my sisters-in-law Yuan Xiong, and my niece Ruiqi Zhang for their unconditional love, trust, and

support throughout my Ph.D. life. I also wish to thank my best friend Jie Yang for her long-term accompany and help on my writing, and supporting me in my bad days.

Last but not least, my heartfelt thanks go to my partner Ruiji Sun for his care, his generosity, his help on the graphs of my thesis, sharing my burden of daily chores, encouraging me when I felt less confident and making me happy when I felt down. Thank you for being there in all the good and bad times.

Contents

Abstract	ii
Acknowledgements	vii
1 Introduction	1
2 Literature Review of Sensing, Learning, and Controls for Thermal Comfort and Energy Efficiency	5
2.1 Variables Underlying Thermal Comfort and Energy Efficiency	5
2.2 Sensing Technologies for Thermal Comfort and Energy Efficiency	8
2.3 Controllable Systems for Thermal Comfort and Energy Efficiency	11
2.4 Building Controls for Thermal Comfort and Energy Efficiency	14
2.4.1 Classic Building Controls	14
2.4.2 Building Controls based on Thermal Comfort Models .	16
2.4.3 Learning Methods for Building Controls	23
2.4.4 Reinforcement Learning and Controls	26
2.4.5 Control with Decomposition	30
2.5 Summary of Literature Review	31

3	Bio-sensing and Reinforcement Learning Control System with Multiple Agents and A Negotiator	33
3.1	Wireless Bio-sensing Network	33
3.2	Reinforcement Learning Control System	35
3.2.1	Bio-REAL Agents	35
3.2.2	Negotiator	38
4	Research Hypotheses	41
4.1	Main Hypothesis	41
4.1.1	Hypothesis 1	42
4.1.2	Hypothesis 2	42
5	Simulation Experiment: VAV with Electric Reheat for Heating	43
5.1	Objective	43
5.2	Simulation Setup	44
5.3	Occupant Models: Fanger PMV and Pierce Tow-Node	45
5.4	Building Model	48
5.5	The Bio-REAL Control System	48
5.5.1	Action, Reward, and State Design for the Three Bio-REAL Agents	49
5.5.2	Negotiation	50
5.5.3	Architecture and Hyper-parameters of the Bio-REAL system	51
5.6	Simulation Run	51
5.7	Experiment results	52

5.8	Conclusion and Discussion	55
6	Preliminary Field and Simulation Experiment at CMU: Water Sourced Radiators for Heating	57
6.1	Objective	57
6.2	Simulation Setup	58
6.3	Human Subject Experiment at CMU	60
6.3.1	Experiment Room	60
6.3.2	Experiment Subjects	60
6.3.3	Experiment Design	62
6.3.4	Measurement and Equipment	62
6.3.5	Data Analysis: Environmental Conditions	63
6.3.6	Data Analysis: Biological Responses	65
6.3.7	Data Analysis: Thermal Satisfaction Variation	65
6.4	Occupant Models	68
6.4.1	Personalized Thermal Comfort Models	68
6.4.2	Personalized Skin Response Models	74
6.5	Building Model	74
6.6	Bio-REAL Agents	77
6.6.1	Action, Reward, and State Design for the Six Bio-REAL Agents	77
6.7	Simulation Run	80
6.8	Experiment results	83
6.8.1	Result Analysis of Case 1	84

6.8.2	Result Analysis of Case 2	86
6.8.3	Result Analysis of Case 3	87
6.8.4	Result Analysis of Case 4	91
6.9	Conclusion and Discussion	92
7	Field Experiments at NUS: Smart Ceiling Fans for Cooling	97
7.1	Objective	97
7.2	Experiment Setup	98
7.2.1	The Net Zero Energy Building at NUS	98
7.2.2	Experiment Room	99
7.2.3	Facilities of the Experiment Room	100
7.2.4	Experiment Subjects	104
7.2.5	Measurement and Equipment	106
7.3	The Bio-REAL Control System	109
7.3.1	Bio-REAL agents	109
7.3.2	Learning Process	111
7.4	Training Experiment	113
7.4.1	Experiment Process	114
7.4.2	Experimental Conditions	114
7.4.3	Experiment Data Analysis	115
7.4.4	The Results for Agent Training	126
7.4.5	Multi-agent Negotiation	135
7.5	Evaluation Experiment	136
7.5.1	Experiment Design	136

7.5.2	Negotiation Example	138
7.5.3	Evaluation Result	140
7.6	Conclusion and Discussion	144
8	Conclusion	147
8.1	Summary of Experimental Findings	148
8.1.1	Simulation Experiment:VAV with Electric Reheat for Heat- ing	148
8.1.2	Preliminary Field and Simulation Experiment at CMU: Water Sourced Radiators for Heating	148
8.1.3	Field Experiment at NUS : Smart Ceiling Fans for Cooling	150
8.2	Contribution	151
8.2.1	Comfort and Energy Benefit	151
8.2.2	The Application Of Internet of Things (IoT) in Buildings	151
8.2.3	The Application of Artificial Intelligence in Buildings .	153
8.2.4	The application Of Digital Twin in Buildings	154
8.3	Limitation and Future Work	155
8.3.1	Large-scale Application in Buildings	155
8.3.2	Multi-agent Communication Network	156
8.3.3	The Ensemble of Multiple Learning Algorithms	157
	Bibliography	159

List of Figures

2.1	Variables for thermal comfort	7
2.2	Aircuity sensor suite (left) and Aclima indoor environmental sensors (right)	9
2.3	Nest(left) and Comfy (right) mobile application	9
2.4	Microsoft smart band 2TM (left) and glasses designed by Ghahramani et al. (right)	10
2.5	Controllable system for thermal comfort and energy efficiency	12
2.6	Heating, ventilation, and air conditioning system (HVAC) Example	13
2.7	Closed-loop temperature control	15
2.8	Thermal comfort metrics (Kim, Schiavon, and Brager, 2018; 10551, 2019)	16
2.9	Closed-loop control to maintain thermal comfort	17
2.10	The agent and environment interaction interface for reinforcement learning	27
2.11	The role of learning and controls	31
3.1	Bio-sensing and reinforcement learning control system	34
3.2	The bio-sensing Network	35

5.1	The integration of the classic occupant models, the building model, and the Bio-REAL control system.	44
5.2	The relationship of mean skin temperature and negative PPD in the specific thermal condition	47
5.3	The relationship of air temperature and mean skin temperature in the specific thermal condition	47
5.4	Building model for the simulation experiment	48
5.5	The learning speed and the control performance of the Bio-REAL control system	52
5.6	The group reward \overline{PPD} , indoor air temperature, and mean radiant temperature using the optimum dynamic control policy	54
6.1	The integration of the personalized occupant models, the building model, and the Bio-REAL control system.	59
6.2	The room for the human subject experiment	60
6.3	Locations of the six experiment subjects	61
6.4	PPD (Predicted Percent Dissatisfied) heatmap created by 1000 data samples from the PMV/PPD model	71
6.5	Prediction results of the PPD based neural network	72
6.6	EnergyPlus model for the experiment room at CMU	76
6.7	The comparison of the TMY3 and RMY2017 Pittsburgh weather data	78
6.8	Case 1 and Case 2: Static Control Schedule with Fixed Setpoint	82

6.9	The comparison of the percentage of dissatisfaction for the six subjects and the five subjects in Case1 (yellow area indicates occupied time)	86
6.10	The comparison of the percentage of dissatisfaction for the six subjects and the five subjects in Case 1 (yellow area indicates occupied time)	87
6.11	Indoor conditions in Case 2 (yellow area indicates occupied time)	87
6.12	(A) The dynamic temperature setpoint schedule learned and executed by the Bio-REAL control system in Case 3 for the first scenario, (B) indoor conditions, and (C) dissatisfaction percentage created by the learned schedule.(yellow area indicates occupied time)	90
6.13	Comparison of the setpoint schedule in Scenario 1 and Scenario 2	91
6.14	(A) The dynamic temperature setpoint schedule learned and executed by the Bio-REAL control system in case 4 for the first scenario, (B) energy consumption, and (C) dissatisfaction percentage achieved by the learned schedule.(yellow area indicates occupied time)	93
7.1	The Net Zero Energy building at NUS (Image from School of Design and Environment)	98
7.2	The location of the experiment room at NUS	99
7.3	The experiment room at NUS	99

7.4	The table layout of the experiment room at NUS	100
7.5	The VAV system of the experiment room at NUS	100
7.6	The capacity of the VAV system in the experiment room at NUS	101
7.7	Four vertical locations for air velocity measurement	102
7.8	Unoccupied temporal pattern of the air velocity in the experi- ment room	103
7.9	Experiment subjects at NUS	105
7.10	Thermal comfort questionnaire for the experiment at NUS . . .	108
7.11	The Bio-REAL control system at NUS learning from the real experiences	109
7.12	The training process in practice	113
7.13	air movement acceptability and wrist temperature variation of neutral-preferred subjects in the four action sequences. Dotted line = wrist temperature at 27 °C and variations in fan speeds. Colors = acceptance to the air movement.	120
7.14	air movement acceptability and wrist temperature variation of warm-preferred subjects in the four action sequences.	121
7.15	air movement acceptability and wrist temperature variation of cool-preferred subjects in the four action sequences.	122
7.16	air movement acceptability and wrist temperature variation of no preference subjects in the four action sequences.	123
7.17	air movement acceptability and wrist temperature variation of irregular-preference subjects (e.g., subject 5) in the four action sequences.	124

7.18 Wrist temperature distribution when voted acceptable to the air movement among the 14 subjects across different (A) age and (B) BMI	125
7.19 Q-value (expected air movement acceptability) of fan speeds at different wrist temperature for neutral-preferred subjects .	127
7.20 Predicted best fan speed (action) at different wrist temperature for neutral-preferred subjects	128
7.21 Q-value (expected air movement acceptability) of fan speeds at different wrist temperature for the warm-preferred subjects	129
7.22 Predicted best fan speed at different wrist temperature for the warm-preferred subjects	130
7.23 Q-value (expected air movement acceptability) of fan speeds at different wrist temperature for the cool-preferred subjects .	131
7.24 Predicted best fan speed at different wrist temperature for the cool-preferred subjects	132
7.25 Q-value (expected air movement acceptability) of fan speeds at different wrist temperature for the no-preferences subjects	132
7.26 Q-value (expected air movement acceptability) of different fan speeds at different wrist temperature for the irregular-preference subjects	134
7.27 Predicted best fan speed at different wrist temperature for the irregular-preference subjects	135

7.28	Acceptable range of operative temperature and air speeds for the comfort zone at humidity ratio 0.010 in ASHRAE standard 55	138
7.29	Fan speed schedules for the evaluation experiment	141
7.30	Comparison in thermal comfort rate between baseline and Bio-REAL control	143
7.31	Comparison in air movement acceptability rate between baseline and Bio-REAL control	143
8.1	The Bio-REAL system contributes to the application of IoT in buildings	152
8.2	The Bio-REAL control system for large-scale application	156
8.3	Multi-agent communication network	157
8.4	The ensemble of multiple learning algorithms	158

List of Tables

2.1	Summary of literature relevant to personalized thermal comfort models	20
2.2	Literature summary of building controls with artificial intelligence	24
3.1	Algorithm: Double Q learning with experience replay and neural network approximation, and negotiation	39
5.1	Clothing insulation, metabolic rate, weight and height of the three occupants	46
5.2	Comparison of the optimal control policy and static schedule .	53
6.1	Demographic information of the experiment subjects at CMU	61
6.2	Variable and equipment of the CMU experiment	64
6.3	Experimental condition during the 22 experiment sessions . .	64
6.4	Biological responses of the six subjects during the 22 experiment sessions	66
6.5	Thermal satisfaction distribution of the six subject	67

6.6	4-level thermal satisfaction in different operative temperature and relative humidity conditions for the six subjects (1.0 Clo, 1.0-1.1 Met, 0.1 m/s).	69
6.7	Prediction result of the personalized thermal comfort mode . . .	73
6.8	Prediction result of the personalized skin response mode . . .	75
6.9	Thermal satisfaction reward/penalty designed for the Bio-REAL agent	80
6.10	The evaluation results of the four cases with two scenarios . . .	85
7.1	Air velocities at different vertical locations and fan speeds . . .	103
7.2	Demographic information of the experiment subjects at NUS . . .	105
7.3	Variable and equipment of the NUS experiment	107
7.4	air movement acceptability reward/penalty designed for the Bio-REAL agent	111
7.5	Action sequences of the experience collection experiment at NUS	115
7.6	Experimental condition during the experience collection experiment	116
7.7	air movement acceptability and wrist temperature collected from the experiment	117
7.8	The evaluation experiment at NUS	137
7.9	Wrist temperature and Q-value snapshot for the five subjects in Session5	139
7.10	Weighted sum Q-Value	139

Nomenclature

Agent A software learner and decision-maker in reinforcement learning

Bio-sensing A sensing technology measuring human's biological responses to the environment

Co-simulation Different subsystems which form a coupled problem modeled and simulated in a distributed manner.

Control Policy A mapping from situations to control signals

Control Schedule A mapping from time to control signals

Control System A software program manages, commands, directs, or regulates the behavior of other devices or systems using control loops.

Digital Twin A digital replica of a living or non-living physical entity.

Internet of Thing A system of interrelated computing devices, mechanical and digital machines, objects, or people

Learning Environment The thing that an agent interacts with, comprising everything outside the agent

Markov Decision Process A discrete time stochastic control process

Participatory-sensing A process using mobile devices and cloud services to collect systematic data and form interactive sensor networks

Reinforcement Learning A goal-directed learning from interaction. It is one of three basic machine learning paradigms, alongside supervised learning and unsupervised learning.

Thermal Comfort Model A mathematical model predicting occupant's condition of mind that expresses satisfaction with the thermal environment

List of Abbreviations

ANN	Artificial neural network
ASHRAE	The American society of heating, refrigerating and air-Conditioning engineers
BMI	Body mass index
Bio-REAL	Bio-sensing and reinforcement learning control
DRL	Deep Reinforcement learning
EER	Estimated energy requirement
GA	Genetic algorithm
HVAC	Heating, ventilation, and air conditioning
IoT	Internet of things
KNN	K-nearest neighbors algorithm
MPC	Model predictive control
MS	Microsoft
PMV	Predicted mean vote
PPD	Predicted percentage dissatisfied
RMSE	Root mean square error
RL	Reinforcement learning
SDK	software development kit
SVM	Support vector machine
VAV	Variable air volume

List of Symbols

Clo	clothing insulation	$Clo (0.155 m^2 \cdot ^\circ C/W)$
Met	metabolic rate	$Met (58.2 W/m^2)$
MRT	mean radiant temperature	$^\circ C (^\circ F)$
MST	mean skin temperature	$^\circ C (^\circ F)$
RH	relative Humidity	%
T_a	air temperature	$^\circ C (^\circ F)$
T_{in}	indoor temperature	$^\circ C (^\circ F)$
T_{out}	outdoor temperature	$^\circ C (^\circ F)$
T_r	radiant temperature	$^\circ C (^\circ F)$
T_{skin}	skin temperature	$^\circ C (^\circ F)$
T_{wrist}	wrist temperature	$^\circ C (^\circ F)$
V_a	air velocity	m/s
s, s'	states	
a	an action	
r	a reward	
S	set of all states	
A_t	action at time t	
S_t	state at time t, due to S_{t-1} and A_{t-1}	

R_t reward at time t , due to S_{t-1} and A_{t-1}

π control policy

$q(s, a)$ value of taking action a in state s

Q, Q_t array estimates of action-value function

Chapter 1

Introduction

The importance of indoor comfort to human health, productivity, and well-being have been emphasized in many studies (Samet and Spengler, 2003; Parsons, 2014). Indoor comfort includes four primary aspects: thermal comfort, visual comfort, acoustic comfort, and air quality. Thermal comfort was ranked by building occupants as the most important aspect of indoor comfort (Frontczak and Wargocki, 2011). ASHRAE (2013) defined thermal comfort as “the condition of mind that expresses satisfaction with the thermal environment”. Indoor thermal environment is regulated by building heating, cooling, and ventilation systems, which are big energy consumers of buildings.

Buildings accounted for about 40% of total U.S. energy consumption in 2018 (EIA, 2019). About 48 % energy in office buildings was used by mechanical systems for space heating, cooling, and ventilation (Pérez-Lombard, Ortiz, and Pout, 2008). Despite such a large amount of energy use for regulating the thermal environment, the thermal satisfaction rate was low according to the survey results from office occupants. (Choi, Loftness, and Aziz, 2012; Loftness et al., 2009 Huizenga et al., 2006;). One primary causes for the

low thermal satisfaction rate and the high energy consumption are improper building controls.

Today, in most office buildings, the thermal environment is regulated only by indoor temperature with static schedule-based rules. This control strategy allows no variation of temperature setpoint and has no consideration of occupants' thermal requirements, resulting in energy waste and thermal discomfort. (Hoyt et al., 2005). A growing number of researchers are focusing on occupant-centric building control that integrates human's requirement into building control loop to improve thermal comfort and energy efficiency. However, there are lots of practical challenges:

First, thermal comfort is a complex subjective response, requiring a broader set of environmental, physiological, and psychological variables to be considered (Fanger, 1970). These variables are neither static nor consistent, therefore, infrequent data is not representative. The analysis of building energy consumption also needs lots of information, such as external load, internal load, and the conditions of building mechanical systems. Current information only is not adequate. Future and past information are also necessary.

Second, the data for most of these variables is impractical to be obtained. In most traditional office buildings, the only available information is room temperature. In modern office buildings, environmental data and the conditions of mechanical systems can be available in the building automation system (BAS). However, the data from occupants is hardly accessible. Advanced sensing technology can be leveraged. However, the cost, intrusiveness, accuracy are the obstacles for these technologies to be implemented in

practice.

Third, HVAC systems are the only building systems that support automatic control in most office buildings. There are many building systems, such as shading devices and windows, can be used for space heating, cooling, and ventilation, yet often are ignored. Integrative control of these systems could effectively improve thermal comfort and energy efficiency. However, these systems have diverse working mechanisms and need to be controlled in different ways made the coordination of these systems non-trivial.

Last, mapping the data of thermal comfort and energy consumption to the control actions for the building systems is another non-trivial task, especially for multi-objective controls. The control theory in engineering and artificial intelligence has provided plenty of algorithms to develop control models or systems for the mapping from the data to optimum control actions. However, in shared multi-occupant spaces with limited individual controls, building controls need to resolve the conflicts in different occupants' thermal requirements and balance the thermal comfort and energy efficiency. These require a delicate design in control systems and proper selection in control algorithms.

In the following chapters, the author will explain these challenges in detail, propose a solution to resolve these challenges, and demonstrated the feasibility and effectiveness of the solutions through three different experiments.

Chapter 2

Literature Review of Sensing, Learning, and Controls for Thermal Comfort and Energy Efficiency

This chapter first summarizes the significant variables for thermal comfort and energy efficiency, then reviews conventional and state-of-art sensing technologies for measuring the variables. After that, the controllable system for thermal environment regulation is summarized. Methods and algorithms for building controls are compared. Last, the remaining gaps in building control for thermal comfort and energy efficiency are concluded.

2.1 Variables Underlying Thermal Comfort and Energy Efficiency

Thermal comfort can be estimated from a broad set of variables, including environmental factors, personal factors, physiological responses, subjective

responses, and behavior response, as shown in 2.1. The six primary thermal comfort factors are air temperature, radiant temperature, humidity, airspeed, clothing insulation, and activity level (Fanger, 1970). Outdoor weather stimulation changes occupants' preference by behavioral adjustment (e.g., changing clothing and opening windows), physiological adaptation (e.g., acclimatization) and psychological adaptation (e.g., temperature expectation). (De Dear and Brager, 1998). Occupants' physiological responses, including skin temperature of different body parts, skin wetness or sweat rate, peripheral blood flow, and heart rate, would be more effective estimator for thermal comfort if they can be precisely detected (Kurz, 2008; Fanger, 1970; Gagge, 1986, Zhang et al., 2010b). Besides, occupants' behavior, such as changing clothes, operating windows, adjusting thermostats, and so forth indirectly reflect their feeling to the environment. Occupants are often encouraged to directly report their feedback, which is often regarded as the ground truth of thermal comfort.

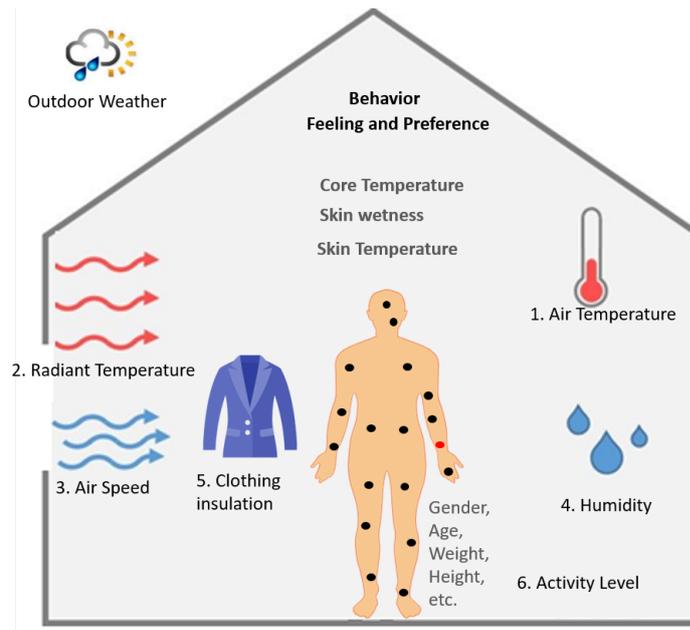


FIGURE 2.1: Variables for thermal comfort

The variables affecting energy performance for space heating, cooling, and ventilation in buildings are numerous and always intricately interrelated (Zhao and Magoulès, 2012). The building mechanical systems consume energy to deliver a proper thermal environment by adding or removing heat from space. The space heat gain depends on outdoor weather, internal heat gain, and building structure and characteristics. For example, Because of the transparent surfaces of the building envelope, the solar radiation enters space and contributes to heat gain. Due to indoor and outdoor weather differences, heat also transfers through exterior walls, roofs, and floors through conduction, convection, or radiations (ASHRAE., 2017). Building occupants, lights, appliances, and equipment generate heat, the density of which determines the internal heat gain. Some of these heat sources contribute to load only after a time lag. Moreover, the behavior and efficiency of the building mechanical systems also determines their energy performance.

2.2 Sensing Technologies for Thermal Comfort and Energy Efficiency

Today's sensing technologies have created opportunities to measure or detect all above-mentioned variables, although some of them are still immature. In traditional office buildings, thermal comfort-related data is barely available. The sensors or meters on thermostats are the only tools that provide indoor environment information. It only measures the indoor temperature at zone level and assumes that assuming that one thermal zone has a similar thermal environment. Fortunately, there are usually abundant sensors deployed in the building mechanical systems to monitor their working statuses, including supply airflow, refrigerant temperature, and pressure sensors. Besides, energy meters or sub-metering can provide building energy consumption (Ahmad et al., 2016). The machinery and energy data are valuable sources for energy-efficient controls.

With the availability of low-cost and high-performance sensors and network infrastructure, wired and wireless indoor environmental sensor networks have been deployed in more and more modern office buildings. The network has dense sensor nodes. One sensor node fuses a variety of sensors, such as air temperature sensors, relative humidity sensors, and carbon dioxide sensors, as shown in Figure 2.2. It enables more accurate and comprehensive evaluation of indoor thermal environment (Kojima, 2011).



FIGURE 2.2: Aircuity sensor suite (left) and Aclima indoor environmental sensors (right)

Due to the pervasive and ubiquitous mobile devices nowadays, the industry tends to develop mobile applications for thermostats, as shown in Figure. Occupants can report their thermal preferences through the applications, which communicate to the building automation systems. These subject feedback from occupants will be stored and processed for better controls (Nouvel and Alessi, 2012; Yang and Newman, 2012; Jazizadeh et al., 2011). This sensing strategy is named as participatory sensing by some researchers (Jazizadeh et al., 2012; Jazizadeh, Marin, and Becerik-Gerber, 2013; Lam, Yuan, and Wang, 2014). Sparse and none-continuous participation is the limitation of participatory sensing.



FIGURE 2.3: Nest(left) and Comfy (right) mobile application

To achieve continuous assessment of thermal comfort, researchers are investigating the less-intrusive technologies, bio-sensing. Wearable devices, such as smart wristbands and glasses, are getting popular for bio-sensing.

The manufactured wrist bands, such as Fitbit Charge HRTM and Microsoft Smart Band (Figure 2.4 left), can be applied in offices to monitor occupants' physiological responses (Hasan, Alsaleem, and Rafaie, 2016; Chaudhuri et al., 2018; Liu, 2018). There were also specially designed wristbands for thermal comfort monitoring. For example, researchers designed wristbands to measure the skin temperature of the radial artery, ulnar artery, and upper wrist ((Sim et al., 2016)) and wrist sweat rate (Sim, Yoon, and Cho, 2018). Besides wristbands, researchers (Ghahramani et al., 2016; Li, Menassa, and Kamat, 2018) also developed glasses with infrared sensors to measure face skin temperature (Figure 2.4 right). In addition to measure physiological response, the presence or absence of an occupant can be detected by tracking the location of the wearable devices (Zhao et al., 2014). However, not all occupants are willing to wear these devices. The low acceptability is potentially the limitation of the wearable devices.



FIGURE 2.4: Microsoft smart band 2TM (left) and glasses designed by Ghahramani et al. (right)

Researchers are then seeking for non-contact and non-intrusive approaches, such as ambient intelligence and computer vision (CV). Ambient intelligence aims to make our environment responsive and sensitive to our presences based on data from ambient sensors, smart devices, appliances, and so forth (Cook, Augusto, and Jakkula, 2009). Although still in its infancy, it could estimate occupant behaviors and predict occupant's preferences or desires

(Stavropoulos et al., 2012). Computer vision aims to gain a high-level understanding from digital images or videos (Szeliski, 2010). Recently, researchers used RGB cameras and Photoplethysmography (PPG) techniques to capture occupants' thermoregulation state Jazizadeh and Jung, 2018. However, the accuracy for CV to detect physiological conditions is not yet promising. The cost-effective sensing technologies that provide accurate measurement will have a higher priority to be selected.

2.3 Controllable Systems for Thermal Comfort and Energy Efficiency

The controllable systems for thermal comfort and energy efficiency are building-specific. In general, the author categorized them as centralized ambient systems, decentralized ambient systems, group task systems, and individual task systems, as shown in Figure 2.5. These systems may control the thermal environment at zone levels, group levels, or individual levels. Zone-level controls affect all occupants in a thermal zone, while group-level controls influence a small portion of occupants in a thermal zone. Individual controls change the thermal comfort of one occupant or local comfort of specific body parts of an occupant.

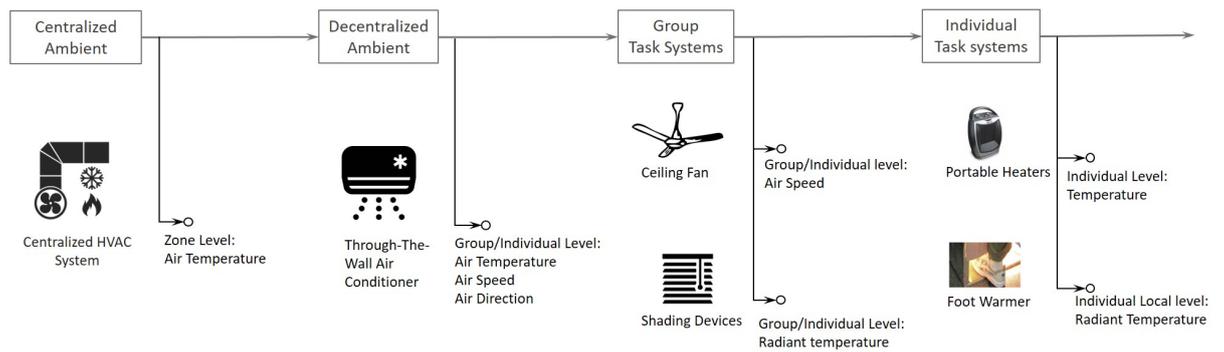


FIGURE 2.5: Controllable system for thermal comfort and energy efficiency

Among the four types of controllable systems, the centralized ambient system is the most energy-intensive one and plays the most crucial role in thermal environment regulation in office buildings. It serves multiple or large spaces and can create a uniform and stable thermal environment. A typical centralized ambient system is a centralized HVAC system. HVAC is a complex system and has various components, including chiller, boiler, cooling tower, air handling units (AHU), fan coil units (FCUs), variable air volume boxes (VAVs) and so forth as shown in Figure 2.6. These components are controlled to maintain temperature setpoints of thermal zones. In most office buildings, the zone setpoint is static schedule-based. Some office buildings may allow manually manipulating the setpoints through a thermostat or automatically adjust the setpoints by a virtual supervisor. The automatic control of HVAC setpoints is supervisory controls, while the automatic control of HVAC components is local controls (Wang and Ma, 2008).

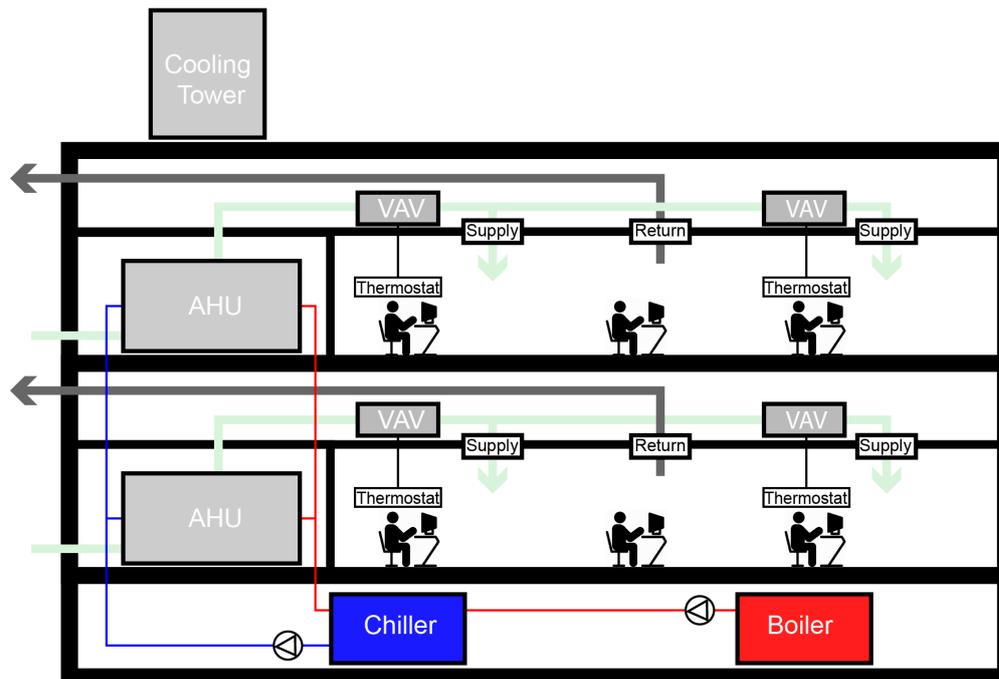


FIGURE 2.6: Heating, ventilation, and air conditioning system (HVAC) Example

The decentralized ambient systems typically serve a single or small space. The terminal units of them can be individually controlled by occupants. They usually create a less-uniform thermal environment. They are mostly the direct expansion or DX types, such as packaged through-the-wall air conditioner and commercial outdoor packaged systems Bhatia, 2011.

The group task systems refer to the task devices shared by a group of people, such as ceiling fans, shading devices, and windows. They serve single or small spaces and change the thermal environment at individual levels or group levels that require negotiation among the group. The individual task systems serve individuals in the task area, such as portable heaters, table fan, foot warmer Zhang et al., 2010a, heated office chair.

The task systems and ambient systems can be integrated to task ambient systems, which maintain an acceptable thermal environment in the ambient

space, while the task components can be individually controlled by occupants in their localized zones Bauman and Arens, 1996. The task ambient systems can save energy consumption by allowing higher ambient setpoint in summer and lower ambient setpoint in winter.

2.4 Building Controls for Thermal Comfort and Energy Efficiency

In this section, the author reviews classic building controls and learning methods of building controls, including the methods for the HVAC supervisory control, the HVAC local control, and the controls of other controllable systems.

2.4.1 Classic Building Controls

Conventionally, the HVAC supervisory control schema is static schedule-based with no consideration of occupants' requirements. There are usually separate control schemes in heating seasons, cooling seasons, and swing seasons. The occupied setpoint and unoccupied setpoint are also different. For example, a weekday schema in a heating season can be that temperature setpoint is 21 °C from 7 AM to 5 PM and 16 °C other time of the weekday. The schedule-based controls have little flexibility that often leads to thermal discomfort and energy waste (Hoyt et al., 2005).

The HVAC components work together with closed-loop control to maintain the scheduled temperature setpoint, as shown in Figure 2.7. The most

traditional controller for local HVAC controls is on/off controller or bang-bang controller. With on/off controllers, the components of HVAC system is turned off as soon as the process temperature rises above the setpoint and turned on when the process temperature drops below the setpoint minus a hysteresis (Johnson, 1999). The bang-bang control with dead-band can avoid frequent changes caused by on/off controls because no change is made in the dead zone (Dounis and Caraiscos, 2009). However, overshoots of the process temperature are still exists and cause over-heating or over-cooling.

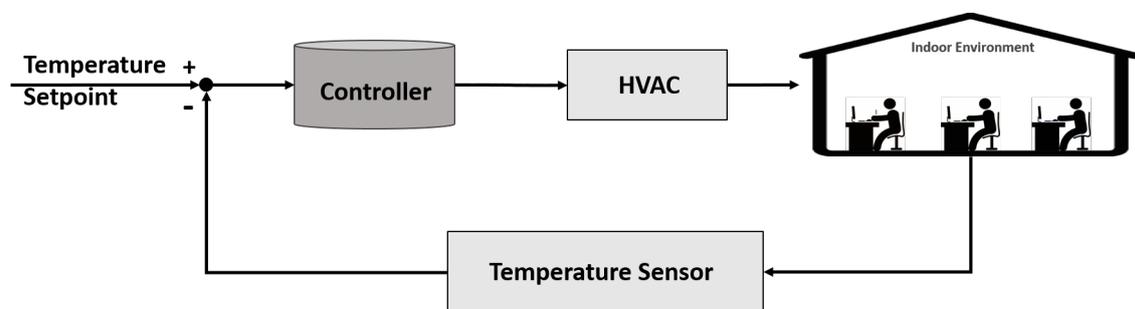


FIGURE 2.7: Closed-loop temperature control

The classic controller, PID (Proportional, integral, derivative), could reduce overshoots because its derivative term can lower the rate of error (the difference between the process temperature and setpoint) and flatten the action trajectory. Its integral term can force the PID controller to reach the setpoint timely by summing instantaneous errors over time. However, if the gains of P, I, and D terms are improperly selected, the PID controllers alone could make the entire system unstable (Zhong, 2006). They also have poor control performance for non-linear processes having responses delays (Shaikh et al., 2014).

2.4.2 Building Controls based on Thermal Comfort Models

Being aware of the weaknesses of the schedule-based controls, a growing number of researchers are focusing on occupant-centric building controls that place occupants' requirements into building control loop. Occupants' requirements for the thermal environment can be quantified by different metrics, such as thermal sensation, satisfaction, acceptability, and preferences (Kim, Schiavon, and Brager, 2018; 10551, 2019), as shown in Figure 2.8. Thermal sensation is the most frequently used one. Thermal satisfaction is often for post-occupancy evaluation. Thermal acceptability is for evaluating occupants' tolerance to the thermal environment. Thermal preferences is a good metric if thermal comfort models will be used for controls because it directly suggests a direction for control. This thesis uses sensation, satisfaction, and preferences as thermal comfort metrics.



FIGURE 2.8: Thermal comfort metrics (Kim, Schiavon, and Brager, 2018; 10551, 2019)

Since occupants' direct requirements are often not available, research tends to integrate thermal comfort models into building control loop, as shown in Figure 2.9. A thermal comfort model is a mathematical model that can be a regression, classification, or probability distribution. It predicts occupant's

thermal comfort requirement (e.g., sensation or preferences) based on different kinds of inputs mentioned in section 2.1. The controllers will act on the heating, cooling, and ventilation systems if their current thermal comfort levels predicted by the thermal comfort models don't meet the comfort objective. Researchers have developed a variety of thermal comfort models to ensure their prediction close to the occupants' real thermal comfort feedback.

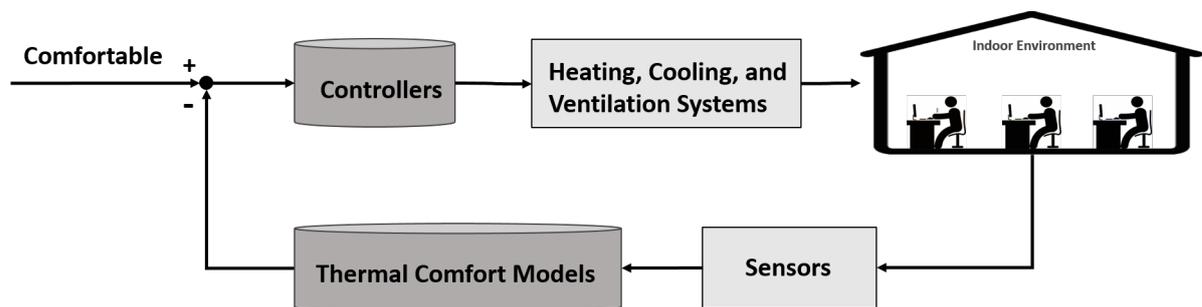


FIGURE 2.9: Closed-loop control to maintain thermal comfort

The most well know thermal comfort model is Fanger's PMV (predicted mean vote) model (Fanger, 1970). It relies on two theories: a necessary condition for thermal comfort is heat balance, for which the internal heat production should be equal to the heat loss from a body; a sufficient condition for thermal comfort is that the mean skin temperature and sweat secretion inside narrow limits. It was developed using data from extensive chamber studies with more than one thousand participants and links the PMV index for thermal sensation with six factors, which are air temperature, mean radiant temperature, relative humidity, air velocity, clothing insulation, and activity level (i.e., metabolic rate). This model was proven to be accurate for building occupants in near-sedentary activity and steady-state conditions (Doherty and Arens, 1988; De Dear and Brager, 1998). It was adapted by the ASHRAE

(The American Society of Heating, Refrigerating and Air-Conditioning Engineers) thermal comfort standards (ASHRAE 55, 2013) to build standard thermal comfort zones when air velocity is lower than 0.2 m/s. This thesis considers it as a valuable benchmark or initial model for determining a comfort zone for the sealed, mechanically controlled office buildings.

Another widely used comfort model is the Pierce two-node model Gagge, 1986. It models the human body as two concentric cylinders to represent body core and skin shell. Although both the PMV model and two-node model use the heat balance equation as the means of deriving the physiological parameters underlying thermal comfort, the PMV model doesn't explicitly estimate the actual value of the physiological parameters, while the two-node model estimates the skin temperature, core temperature, and skin wetness. Although an evaluation study has shown that the two-node model tended to underestimate skin wetness and core temperature and overestimate skin temperature (Doherty and Arens, 1988), this thesis considers it as a valuable initial model for bio-responses estimation.

To build a model that is more applicable to the actual workplace, especially for the naturally ventilated buildings, De Dear et al. (De Dear and Brager, 1998) developed the adaptive model from 160 buildings worldwide. The adaptive model for naturally conditioned spaces was adopted by ASHRAE-55. It is given as a linear regression model describing the correlation between optimum indoor temperature and outdoor temperature. The range of optimum indoor temperature derived from this model was about twice larger than that estimated by the PMV model (De Dear and Brager, 1998). The

larger acceptable temperature range grants more flexibility for building controls with the goal of energy saving. However, the biggest limitation of it is ignoring the six important factors of thermal comfort (Fanger and Toftum, 2002).

A broad set of field studies have demonstrated individual differences in thermal comfort due to the differences in gender, age, weight, height, historical thermal experience, economic level, and so forth (Karjalainen, 2012; Indraganti and Rao, 2010; Indraganti, Ooka, and Rijal, 2015). However, the above-mentioned thermal comfort models are aggregated models and not accurate if applying to individuals or a small group of occupants. Hence, researchers have emphasized on personalized thermal comfort models predicting individuals' thermal comfort, which have diversities in input variables, outputs, modeling algorithms, evaluation metrics, and continuous learning methods, as summarized in Table 2.1.

There are mainly two approaches to personalized thermal comfort models. One approach is incorporating personal identifiers, such as age, gender, weight, height, and race, into the thermal comfort model. This type of model needs to be trained by data from many different occupants (Chaudhuri et al., 2017; Hasan, Alsalem, and Rifaie, 2016; Lam, Yuan, and Wang, 2014). The personalization is credited to different input values. In contrast, another approach is creating a thermal comfort model for each occupant and training the model using individual data. In this approach, the values of model parameters and even the model structure are different for different occupants,

TABLE 2.1: Summary of literature relevant to personalized thermal comfort models

Personal	Environment	Physiological	Output	Data	Modeling	Evaluation	Application		
Author, Year	Factors	Responses	Subject	Data/Sensing	Algorithm	Accuracy	Convergence(C)	Online	
Chaudhuri, 2017	Clo, Gender, Activity Level	T_{air} , MRT, V_{air} , RH, T_{out}	N/A	3	RP-884, Singapore	SVM, KNN, logistic regression	SVM = 73.14% Confusion matrix	N/A	N/A
Lam, 2014	AL, A, G, W, H	T_a , T_{out}	N/A	7	Filed, 5 days, 13 subjects	Piece-wise linear regressions	Accuracy = 75%, Energy saving = 18%	N/A	Group T_a setpoints, 87 subjects, 1 hour
Auffenberg, 2015	N/A	T_{air} , RH, T_{out}	N/A	Varied Scale	RP-884, 10 cities	User-specific Bayesian network	RMSE = 1, RMSE = 1.3 for PMV	Relearning, C = 10 samples	N/A
Ghahramani, 2015	N/A	T_a	N/A	3	Field, 1-2 mon, 33 subjects	Bayesian network, three Gaussian	Accuracy = 70.14%	Relearning, KS remove samples	N/A
Damm, 2011	Different Clo set	T_a	N/A	3	Field, 6851 votes, 28 subjects	Logistic regressions	N/A	Congruency of starting and real	Model predictive control for blind
Liu, 2007	Different Clo set	T_a , V_a , RH, MRT	N/A	3	Field, 5 months, 113 subjects	ANN as regression	Veracity = 80%	Relearning, keep 20 new samples	Personal air conditioner
Dai, 2017	N/A	N/A	13 body locations	7	Field, 969 votes, 11 subjects	SVM Gaussian or linear kernel	88.7%/(forehand, cheek, forearm, hand)	Relearning, C = 28 samples	N/A
Hasan, 2016	N/A	T_a , RH	$T_{tourist}$ heart rate	7	Field, MS Band, 5 subjects	EER, ANN	Over-fitting	N/A	N/A
Zhang Hui, 2010	N/A	N/A	19 body locations	9	Chamber Tests	T_{skin} - T_{skin} set point, logistic regression	R-square >0.9, residue SD <0.95	N/A	Task-ambient conditioning
This Research	N/A	T_{air} , RH, V_{air} , MRT, T_{out}	$T_{tourist}$	4	MS Band, Field and Simulation	ANN combined PMV	Personalized Pattern	Updated learning	Train control agents

so the personalization is credited to the model itself. Since a small individual model requires fewer data to train, this approach is preferred by most researchers.

The inputs used by the individual models are either environmental variables or biological variables. The prediction methods in existing research include neural network, Gaussian distribution, and logistic regressions, support vector machine (SVM), K-nearest neighbors(KNN), linear discriminant analysis (LDA), linear regression, and decision trees. Liu, Lian, and Zhao (2007) built a neural network to classify thermal comfort levels based on indoor air temperature, radiant temperature, air velocity, and relative humidity. The neural network model was then imbedded in an air conditioner. Ghahramani, Jazizadeh, and Becerik-Gerber (2014) created personalized thermal comfort profiles to correlate the thermal discomfort and room temperature. The comfort profiles and the zone-level energy models profiles determined the HVAC temperature setpoint. Ghahramani, Tang, and Becerik-Gerber (2015) and Daum, Haldi, and Morel (2011) built a model for each comfort levels. Ghahramani, Tang, and Becerik-Gerber (2015) created three Gaussian distributions to predict the probabilities of being uncomfortably cool, comfortable, and uncomfortable warm. Daum, Haldi, and Morel (2011) built three logistical regressions separately for the three comfort levels. Among the biological inputs, skin temperature is the most extensively studied one. Different features of the skin temperature have been investigated, including skin temperature of different body parts (Zhang et al., 2010b), mean skin temperature (MST)

(Höppe, 1999), the gradient of MST over time interval (Choi, 2010), the gradient of skin temperature over different body locations (Sim et al., 2016; Dai et al., 2017), the deviation of skin temperature and neutral skin temperature (Zhang et al., 2010b), and heat loss estimated from MST (Liu et al., 2014). Besides skin temperature, Höppe (1999) pointed out that airspeed is an important factor for thermal comfort at a high sweat rate. Heart rate and the change in rates can be an effective indicator of activity level in warm conditions (Choi and Loftness, 2012).

For personalized thermal comfort model development, the domain knowledge, such as heat balance theory and the principle of thermoregulation in humans, can be employed to determine not only the input variables but also model structures. For instance, in Lam, Yuan, and Wang (2014)'s model, thermal comfort is equal to heat generation plus heat loss. The heat generation has a linear relationship with the Estimated Energy Requirement (EER), which is calculated based on the personal factors of age, gender, weight, and height. The heat loss is a function of indoor and outdoor temperature. Aufenberg, Stein, and Rogers (2015) designed a Bayesian network based on the correlation among the input variables. The structure of the model explicitly describes the relationship between inputs and outputs.

Although domain knowledge has been employed, all the thermal comfort models are data-driven. A large amount of data is required to train the model so that the personalized thermal comfort models can have a good performance. If long-term data is not available, an initial model can be valuable

complementation. Online learning (Bauer, Koller, and Singer, 1997) techniques should be applied to gradually adapt the initial model to a personalized model with more and more incoming data. Daum, Haldi, and Morel (2011) built a group comfort profile using data from a group of occupants, then adapt it to each occupant. Liu, Lian, and Zhao (2007) demonstrated that a neural network trained by sample data from an occupant with one kind of clothing and activity level can effectively adapt to the same occupant with other kinds of clothing and activity level, and other occupants. Nouvel and Alessi (2012) used the deviation between personal feedback and PMV calculated feedback to update the effect of metabolic rate and built personalized PMV model.

2.4.3 Learning Methods for Building Controls

In addition to integrating comfort models in building controls, many researchers implemented learning methods to build intelligent controllers and overcome the drawback of the classic controllers. These researches have diversities in control objectives, controlled variables, and algorithms, as summarized in Table 2.2.

Fuzzy control was popular in comfort controls because it has the strength in handling the fuzzy inputs and comfort (e.g., warm) is a fuzzy input or partial truth to control systems (Dounis and Manolakis, 2001). For example, Jazizadeh et al. (2012) and Ghahramani, Jazizadeh, and Becerik-Gerber (2014) applied the fuzzy logic to map thermal preferences to ambient temperature. One major weakness of the fuzzy systems is that prior expert knowledge is

TABLE 2.2: Literature summary of building controls with artificial intelligence

Personal	Objectives	Supervisory control (Set-point)	Control of local components	Control level	Human in Loop	Algorithm/method	Evaluation
Author, year	Energy Thermal Comfort Other		Zone	Local Indivi		Comfort	Energy
Erickson, 2012	Y	Y	Y	Y	Y	AMV correct PMV	Satisfaction vs Static $C(T_{in}-T_{out})=10.1\%$
Jazizadeh, 2012	Y	Y	Y	Y	Y	Fuzzy, Minimize error	N/A
Kolokotsa, D, 2002	Y	Y	Shading, Lighting, etc.	Y	Y	GA, Fuzzy system	Range of PMV index vs on/off control 35% saving
Liu, 2007	Y	Y	Compressor, Fan, Valve	Y	Y	ANN, P	N/A
Lu, Lu, 2005	Y	Y	Fan and pump pressure	Y	Y	Neuro-fuzzy system	N/A vs fixed setpoint
Lang, 2015	Y	Y	Flow rate of an AHU	Y	Y	MPC, Autoregressive moving average	N/A 27.8% saving vs PID control
Liu and Henze, 2006	Y	Y	Thermal and storage	Y	Y	Tabular Q learning	Cost saving vs 7 cases
Dalamagkidis, 2007	Y	Y	Heat pump, window, etc.	Y	Y	TD learning + linear function	PPD and Energy vs on/off and fuzzy-PD
Vazquez-Canteli, 2017	Y	Y	Y	Y	Y	Q learning + ANN	N/A
Boman, 1998	Y	Y	Personal desire	Y	Y	Multi-agent + simple rules	N/A
Hagras, 2003	Y	Y	Visual, Safety	Y	Y	Multi-agent + fuzzy system + GA	Compare the rules with Mendel-Wang method
This Research	Y	Y	Ceiling Fan	Y	Y	DDQN + Experience replay	Learning speed, comfort rate, energy saving

required to build the fuzzy rule sets. If the prior knowledge is incomplete or incorrect, the fuzzy system must be tuned. The tuning processes are time-consuming and error-prone (Kruse, 2008). There were also Fuzzy P, PI, and PID controller developed to employ fuzzy logic in closed-loop control. The hybridization can offer the advantages of both controllers.

In complex HVAC controls, Genetic algorithm (GA) was extensively used to tune the fuzzy system since it can take various HVAC constraints into account (Alcalá et al., 2003). Besides, the fuzzy system and GA together can also be used to perform optimal control where the GA calculates optimum set points and the fuzzy system maps the setpoint to controlled variables (Kolokotsa et al., 2002).

Like fuzzy systems, artificial neural networks (ANN) can also represent a complex non-linear relationship. Most of the time, ANN are preferred to fuzzy systems because ANN can be trained by observed data instead of prior expert knowledge. Kanarachos and Geramanis (1998) developed an adaptive controller using neural networks to handle the nonlinearities of the hydronic heating system. However, ANN has two recognized limitations for building controls. One is that a lot of observed data are required to make ANN come into play. Another is that it is not straightforward to extract the comprehensive rules in the neural network (Kruse, 2008).

The neural-fuzzy systems can combine the strength of both neural network and fuzzy systems. In the case of cooperative neural-fuzzy systems, ANN learns parameters from the fuzzy system either off-line or on-line. In the case of a hybrid neural-fuzzy system, ANN and fuzzy system are fully

fused. For example, Lu et al. (2005) implemented a hybrid neural-fuzzy system to find variable pressure setpoints for fan and pump by giving the mass flow rate of chilled water. In their study, the controller system was a neural network and the fuzzy rules were the neurons.

Model predictive control (MPC) has been quite popular among researchers and industry during the last decade because it provides valuable prospects for dealing with system dynamics, time delays, and so forth. MPC uses dynamic models of process and prediction on disturbance (e.g., variation in occupancy and outdoor weather) to determine the optimal open-loop control sequence over a short-time future horizon by minimizing an objective function. A variety of simulation-based research has shown that MPC has outperformed other control methods for energy saving because it considers future disturbances and exploits system dynamics Shaikh et al., 2014. Its major drawback is that the computational burden may prevent it from a real-time implementation (Lamoudi, Alamir, and Béguery, 2011).

2.4.4 Reinforcement Learning and Controls

The reinforcement learning (RL) has drawn growing attention among building control researchers in recent years due to its successful in artificial intelligent fields. RL can be either model-based or model free. In a typical RL problem, an agent aims to learn the best action in different situations through interaction with an environment. As shown in Figure 2.10, the interaction is that, at each time step t , an agent receives a state (S_t) that represents the situation of the environment, and on that basis selects an action (A_t) from an

action set. In response to the action, the environment sends a reward (R_{t+1}) to the agent and changes to a new state (S_{t+1}). One interaction is one learning episode. The agents usually needs to be trained by hundreds of episodes. The mapping from states to actions is called the agent's policy (π_t). The goal of the agent is to learn an optimal policy that maximizes the accumulated rewards it receives in the long run. The detail description of RL learning can be found in Sutton, Barto, and Bach (2018).

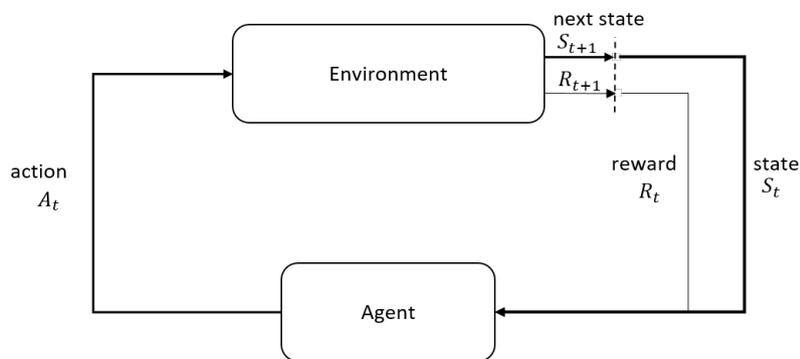


FIGURE 2.10: The agent and environment interaction interface for reinforcement learning

RL can be either model-based or model free. Model-based RL relies on a model of the environment, which is usually a Markov decision process. The goal of model-based RL is to solve the Markov decision process. “Model-free” reinforcement learning (RL) needs no model and trains learning agents through interaction with a simulated or real environment. Q-learning is one of the most popular “model-free” RL algorithm (Watkins and Dayan, 1992). It estimates the expected reward, also called the Q-value, of a state-action pair. Formally, the Q-value under a given policy π is shown in Equation 2.1, where γ is a discount rate that trades off the immediate reward and the

k -time-steps later reward.

$$Q_{\pi}(s, a) = E_{\pi} \left[\sum_{k=0}^{\infty} \gamma^k R_{t+k+1} \mid S_t = s, A_t = a \right] \quad (2.1)$$

Ultimately, an optimal Q-value, $Q^*(s, a) = \max(Q(s, a))$, will be learned. The Q-value quantifies the value of different actions in different states. The optimal policy is the one that can achieve the maximum Q-value. Q-learning is often combined with function approximation to save the time and data needed for learning. With approximation, the Q-value in a tabular form is approximated by a characterized function with parameter θ (i.e., $Q(s, a, \theta) \approx Q(s, a)$) so that Q-value can be generalized from the examples of them. Most methods in supervised learning, including linear combinations of features, decision trees, and neural network, can be used as function approximators.

Deep reinforcement learning combines deep learning architecture with reinforcement learning algorithm (François-Lavet et al., 2018). Model-free DRL can be value-based or policy-based. The value-based DRL estimates the optimal value function, such as Q-value. The policy-based DRL directly search for the optimal policy. DRL uses deep neural networks to represent the value function or policy. It has been successful in complicated tasks and achieved super-human-level performance in playing video games. The most well-known application of deep RL is AlphaGo from Google DeepMind Group (Silver et al., 2016). This research group has proposed a series of deep RL algorithms. Some state-of-art algorithms are Double Q learning (Mnih et al., 2013; Mnih et al., 2015), Double Deep Q learning (Van Hasselt, Guez, and

Silver, 2016), dueling deep Q learning (Wang et al., 2015), and Asynchronous advantage actor-critic (A3C) (Mnih et al., 2016). Deep RL had successful real-world applications, such as robotics (Gandhi, Pinto, and Gupta, 2017) and self-driving car (Pan et al., 2017)).

There were some applications of RL in building controls. Liu and Henze (2006) implemented the tabular Q-learning to control zone air temperature setpoint and thermal storage discharge rate for cost-saving. In which, the state-action value function is a massive lookup table. The performance of the tabular Q-learning will diminish with the increase of the state-action space's dimensionality. Dalamagkidis et al. (2007) used a linear function to approximate action-value function to control heat pump, ventilation subsystems, and windows. Their reward function synergized the objectives of energy efficiency, thermal comfort, and air quality. Vázquez-Canteli, Kämpf, and Nagy (2017) applied fitted Q-learning that approximate the Q-value with a neural network to minimize energy consumption while maintaining a target room temperature. Zhang et al. (2019) implemented A3C to adjust HVAC supply water temperature. The A3C agents were first trained by a calibrated building simulator then deployed in the real-world building. Zhang, Zhang, and Loftness (2019) developed double deep Q learning agents for each occupant to achieve personalized thermal comfort controls while saving energy consumption.

2.4.5 Control with Decomposition

Building controls have decomposability in nature because they require simultaneously complete many tasks and achieve multiple objectives (Lamoudi, Alamir, and Béguery, 2011). With increased requirement on comfort and energy efficiency, it seems unwise to leave the burden of building controls on a single central controller. Break a complex problem into several sub-problems can split the burden of problem solvers. To split the computation load, Lamoudi, Alamir, and Béguery (2011) designed a distributed MPC system, which has a zone layer and a coordination layer. In the zone layer, one model predictive controller is responsible for temperature adjustment in one zone. The coordination layer's job is optimally dispatching resources between zones. Other distributed MPC examples can be found in Moroşan et al. (2010) and Ma, Anderson, and Borrelli (2011). Besides, the multi-agent system (MAS) also divides one problem into several sub-problems, which are solved by their representative agents. Boman et al. (1998) designed a MAS, where agents represented different entities of a building, such as occupants, rooms, and environmental parameters. In the MAS of Hagraş et al. (2003), one agent is tied to one objective. The objectives were the comfort, cost, and safety. The researches have shown that decomposition can improve computational efficiency, system reliability, reconfigurability, and responsiveness, to name a few.

2.5 Summary of Literature Review

Building controls should make the best of the resources available in the building to improve thermal comfort and energy efficiency. In other words, the control system designer should take full advantage of the data mentioned in section 2.1, the sensing systems mentioned in section 2.2, and the heating, cooling, ventilation systems mentioned in section 2.3. The learning and controls systems play the role to bridge these systems, as shown in Figure 2.11. A learning and control method mentioned in section 2.4 should be correctly selected to mobilize the resources or systems and intelligently control them.

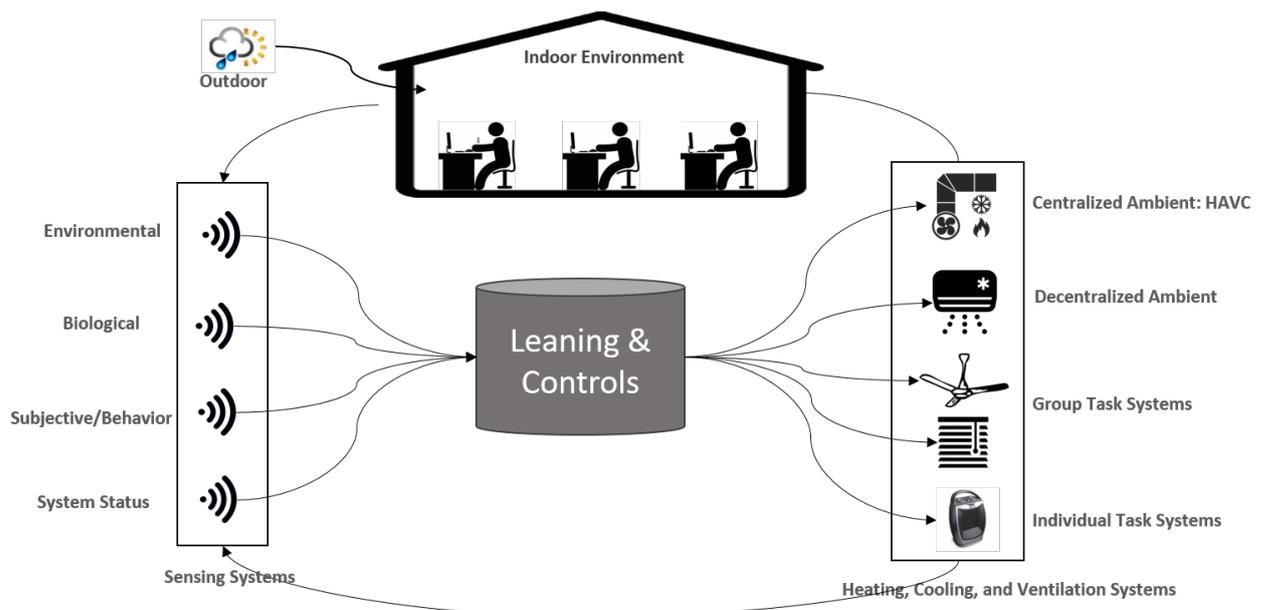


FIGURE 2.11: The role of learning and controls

Because of the practical constraints, previous researchers barely integrate bio-responses into building control loop. They relied on environment data

and subjective feedback to gain thermal comfort information. However, environment data can not explain individual differences. Their methods of obtaining feedback are intrusive. Therefore, subjective feedback is usually non-continuous. Since thermal comfort varies time by time, sparse feedback is not representative and can be misleading. This thesis proposed a bio-sensing approach to continuously integrate personalized thermal comfort requirement into building control loop.

Every building is different due to its location, geometry, building characteristics, the types of heating, cooling, and ventilation systems, and so on. For each building, the energy model mapping from energy consumption to control actions is a complex non-linear function. It is non-trivial to build an energy model that describes the behavior of a building accurately, not mention for many buildings. Moreover, solving complex non-linear models in the process of building controls is computationally expensive. Therefore, the model-based building controls have less priority than the model-free ones. The rule-based control, no matter its fuzzy rule or other static rules, can not deal with the internal and external disturbance to buildings, such as weather and occupancy variations. An adaptive approach is preferable. Model-free deep reinforcement learning, a method of learning from interaction, can generate control actions adaptively.

This thesis proposed a bio-sensing and reinforcement learning control (Bio-REAL) system to improve thermal comfort and energy efficiency. The control framework has multiple control agents and a negotiator to take advantage of decomposition.

Chapter 3

Bio-sensing and Reinforcement

Learning Control System with

Multiple Agents and A Negotiator

Figure 3.1 shows the proposed bio-sensing and reinforcement learning control (Bio-REAL) system that has a wireless bio-sensing network and a reinforcement learning control system.

3.1 Wireless Bio-sensing Network

The responsibility of bio-sensing is continuously integrating occupants' biological responses into building control loop. After comparing the intrusiveness, cost, detection accuracy of different bio-sensing technologies, this thesis selected Microsoft Band 2™ for four reasons. First, although remote bio-sensing, such as infrared cameras, can measure biological responses and less intrusive, the wearable devices with contact to the skin can provide a more accurate measurement. Second, wristbands are getting more pervasive and

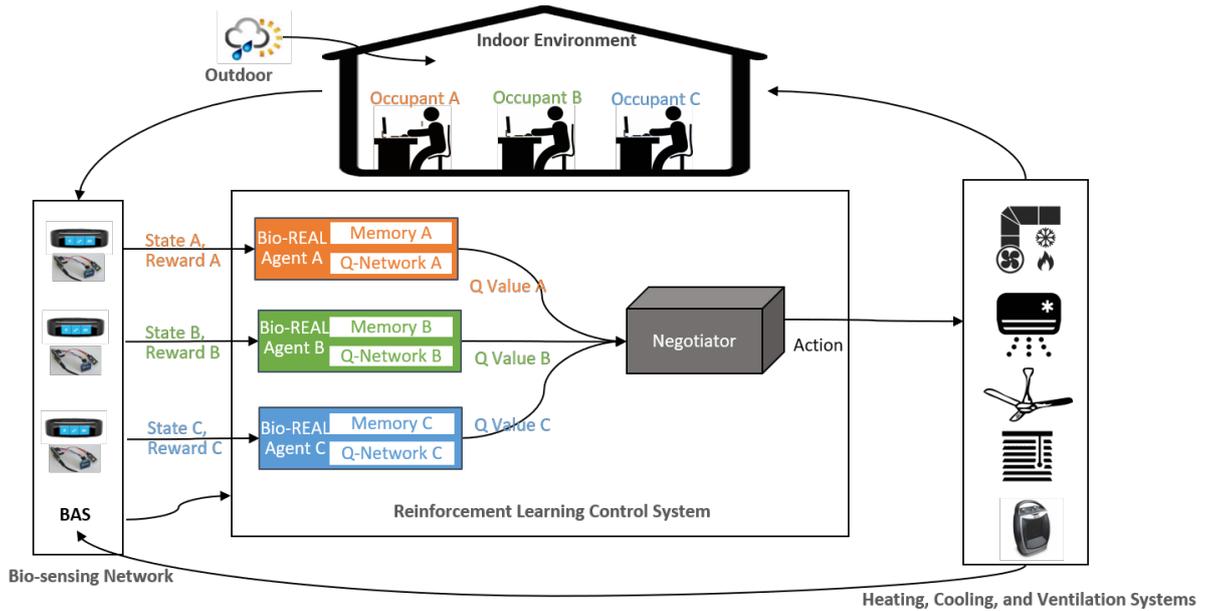


FIGURE 3.1: Bio-sensing and reinforcement learning control system

are less intrusive compared to other wearable devices, such as smart clothes or glasses. Third, researchers found that wrist temperature is closely correlated with thermal sensation (Choi and Loftness, 2012). Last, the Microsoft Band is one of the cheapest smart wristbands in the market that can measure wrist temperature.

The Microsoft Band 2™ has abundant sensors. Besides the skin temperature sensor, there are heart rate, RR interval, galvanic skin response, ultraviolet radiation exposure intensity, light intensity, air pressure sensors, and so forth. It provides SDK (software development kit) so that the third-party application developers can access the sensors available on the band. Based on the SDK, the thesis developed a bio-sensing network, as shown in Figure 3.2. A mobile application was developed and installed on each occupants' cell phone. During the sensing process, the mobile app accesses the sensors on the wristband through Bluetooth then sends the sensor data to the server

through Wi-Fi or cellular data. On the server, the Flask, a web framework, receives the data and saves the data into the InfluxDB, an open-source time-series database.

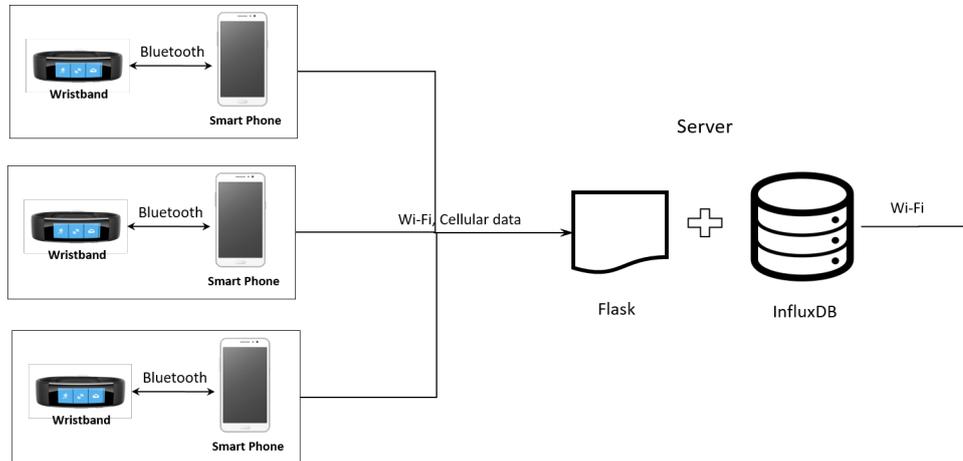


FIGURE 3.2: The bio-sensing Network

3.2 Reinforcement Learning Control System

The reinforcement learning control (REAL) system comprises multiple Bio-REAL agents and a negotiator. The REAL system decomposed the task of satisfying all occupants in a shared thermal zone into the subtasks of pleasing each occupant, energy-saving, and negotiating the conflict in occupants. A Bio-REAL agent acts on behalf of an occupant. The negotiator is responsible for conflict negotiation. The multi-Agents structure is to ensure computational efficiency, flexibility, and reliability of the REAL system.

3.2.1 Bio-REAL Agents

The objective of a Bio-REAL agent is to optimize thermal comfort of an occupant it represents and energy efficiency. The design of state, action, and

reward is the core to achieve the objectives. The algorithm training the agent is also significant to the performance of the control system.

State. The state can be any information that are useful for the agent to learn. It describes the current, historical, and future situation related to the objectives. It can be any variables related to thermal comfort and energy consumption, as mentioned in section 2.1. The more variables included in the state, the more representative the state is. However, the state-space increases with the number of variables. A high dimensional state-space requires more data to train the agent and results in computational cost. Therefore, the most relevant variables should be selected for the state to balance the representativeness and space dimension. In this thesis, occupant's skin temperature is the essential variables for the state. Indoor and outdoor environmental variables are chosen as supplements.

Action. The action is building specific and depends on the controlled variables of the controllable heating, cooling, and ventilation systems available in the building, as mentioned in section 2.3. For thermal controls, actions can be any activities that change the thermal environment, including supervisory controls, such as adjusting HVAC setpoint, and local-level controls, such as turning off fan coil units. It even can be the changes to occupants' behaviors. In this thesis, the actions varied in the three types of experiments because of the different heating and cooling systems.

Reward. Since the technical objective of an RL agent is to maximize the accumulated reward in the long run, the reward in this thesis was quantified personalized thermal comfort of each occupants and energy efficiency.

The thermal comfort is quantified by thermal satisfaction levels, air acceptability levels, etc. The weighting on thermal comfort and energy efficiency determines the priority in making control decisions.

Algorithm. Among the model-based, value-based, and policy-based reinforcement learning algorithms mentioned in section 2.3, the author selected double Q learning with experience replay and neural network approximation (Van Hasselt, Guez, and Silver, 2016) to train the Bio-REAL agents. This value-based RL algorithm was preferred to policy-based one because it estimates Q-value that is valuable for the negotiation step. Moreover, this double Q learning algorithm maintains two Q-networks (neural network approximating Q-value): online-network and target-network. The two Q-networks have the same architecture but different parameters. For each update step, the online-network is used to determine the action with the highest Q-value and the target network is for Q-value determination. The online-network is updated every step. The update equation is shown in Table 3.1 line 13. Every τ step, the target-network copies the weight from the online-network (Table 3.1 line 14). The decoupling of prediction and evaluation solves the over-optimism in Q learning. The samples used in training, such as occupants' wrist temperature, is time-series data. Hence, the data samples are highly correlated with each other, which leads to instability of RL (Mnih et al., 2015). The thesis used the experience replay (Lin, 1993) to reduce the correlation. For the experience replay, the experience (state, action, next state, and reward) are stored in the agents' memory bank for some time and sampled uniformly to updated the Q-network, as shown in Table 3.1 line 10 and

11. The algorithm is wrapped by Gym, an open-source interface for the RL tasks (OpenAI, 2019).

3.2.2 Negotiator

A Negotiator is to resolve the conflict in the decisions made by each Bio-REAL agent. A Bio-REAL agent selects an action or makes decisions according to its Q-value. The action with the highest Q-value given the current state is the best for the agent. The best actions for different agents can be disparate. Therefore, the negotiator selects the action maximizing the weighted sum of all agents' Q-value, as shown in Table 3.1 line 7. This action is the best one for the group in the shared thermal comfort zones. During the training process, the negotiator usually selects ϵ -greedy action rather than the best action. The ϵ -greedy method is a strategy of balancing exploration and exploitation. With ϵ -greedy, the agent has $1-\epsilon$ probability to select the best action and ϵ probability to select a random action. The exploration with random action enables the agents to try new actions and visit new states, which could be better than the already visited ones.

TABLE 3.1: **Algorithm:** Double Q learning with experience replay and neural network approximation, and negotiation

1. **Initialize** action set A and the weight w for each agent
 2. **Initialize** Memory D , and two Neutral Network Q_θ and Q_θ^- for each agent
 3. **Initialize** discount rate γ and learning rate α
 4. Repeat forever:
 5. For each agent i :
 6. Observe state S_t^i and predict the Q value $Q_\theta^i(S_t^i, A^i)$
 7. **Negotiation:** select ϵ - greedy actions a_t^{i*} that can maximize $\sum_i w^i Q_\theta^i(S_t^i, A^i)$
 8. For each agent i :
 9. Observe resultant reward R_{t+1}^i and next state S_{t+1}^i
 10. **Experience Replay:** Append the transition $(S_t^i, A_t^i, S_{t+1}^i, R_{t+1}^i)$ to D
 11. **Experience Replay:** Uniformly sample a mini-batch from D
 12. for each sample in the mini-batch:
 13. **Updating:** update θ of the Q-network:

$$\theta \leftarrow \theta + \alpha \left(R_{t+1}^i + \gamma Q_{\theta^-}^i(S_{t+1}^i, \operatorname{argmax}_{a^i \in A^i} Q_\theta^i(S_{t+1}^i, a^i)) - Q_\theta^i(S_t^i, A_t^i) \right) \nabla_\theta Q_\theta^i(S_{t+1}^i, A_t^i)$$
 14. Copy θ to θ^- every τ step
-

Table 3.1 shows the general procedure of the algorithm. The specific learning procedure is different for the three types of experiments.

Chapter 4

Research Hypotheses

The Bio-REAL control system will be evaluated by three experiments. The first experiment is a simulation experiment, which is a proof of concept to demonstrate that the Bio-REAL control system is feasible. The second is a field and simulation experiment to evaluate the performance of the Bio-REAL control system in a heating season. The third is a field experiment conducted in a cooling season. The three experiments will address the following research hypotheses.

4.1 Main Hypothesis

The Bio-REAL system with a bio-sensing network, multiple personalized Bio-REAL agents, and a negotiator, taking the inputs of occupant wrist temperature and controlling different heating and cooling systems, will effectively improve thermal comfort and save energy consumption.

4.1.1 Hypothesis 1

An indoor environment regulated by a Bio-REAL system will guarantee a higher thermal satisfaction rate as compared to static schedule-based systems.

4.1.2 Hypothesis 2

An indoor environment regulated by a Bio-REAL system can guarantee lower energy consumption, without sacrificing thermal satisfaction, as compared to static schedule-based systems.

Chapter 5

Simulation Experiment: VAV with Electric Reheat for Heating

5.1 Objective

This simulation experiment was designed to demonstrate the feasibility of the Bio-REAL system. Moreover, the architecture and the hyper-parameters of the Bio-REAL system were tuned to ensure reliable performance. More specifically, the learning environment was simulated by three classic occupant models and a building model. Three control agents and a negotiator of the Bio-REAL control system were created. For the Q-network, the hidden-layers, the nodes of each layer, and the activation function were determined. The performance in thermal satisfaction and energy saving of the Bio-REAL control system were quantified and compared to the standard and comfort-oriented static schedule-based controls. The experiment also quantified the learning speed of the Bio-REAL system.

5.2 Simulation Setup

In the simulation experiments, three classic occupant models, a building model, and the Bio-REAL system worked in an integrative way, as shown in Figure 5.1. The occupant models and the building model formed the learning environment for the Bio-REAL control system. The classic occupant models were to simulate occupants' responses. The building model was built using the EnergyPlus simulation tool (Crawley et al., 2001) to simulate indoor environmental conditions and energy consumption.

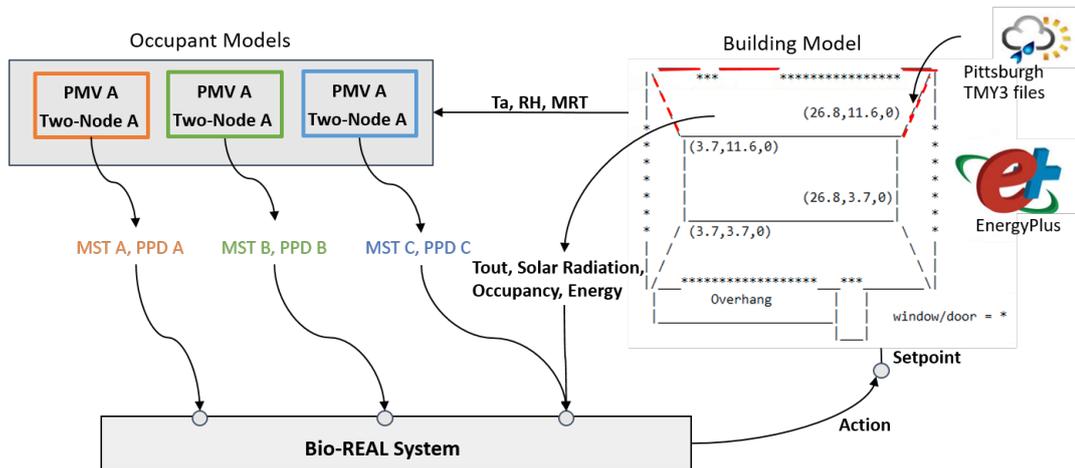


FIGURE 5.1: The integration of the classic occupant models, the building model, and the Bio-REAL control system.

Figure 5.1 also shows the data flow loop of the learning and control process: the Bio-REAL system selects a temperature setpoint, the building model simulates the air temperature (T_a), relative humidity (RH), and mean radiant temperature (MRT) of the room and the energy consumption of the HVAC system given the setpoint. T_a , RH , and MRT are the inputs for the occupant models, which output thermal satisfaction levels and mean skin temperature. The outputs from the occupant models, the indoor environmental

conditions, the outdoor weather including outdoor temperature (T_{out}) and solar radiation, the occupancy status, and the energy consumption from the building model are the inputs of the Bio-REAL control system.

5.3 Occupant Models: Fanger PMV and Pierce Two-Node

Three occupants are modeled using Fanger's PMV (predicted mean vote) model (Fanger, 1970) and Pierce two-node model (Doherty and Arens, 1988). The PMV model was used to simulate occupants' thermal satisfaction feedback. Although Fanger's PMV model was designed to predict the mean votes of a large group of occupants on thermal comfort, this experiment used the PPD (predicted percentage dissatisfied) converted from PMV to quantify the thermal satisfaction of individual occupants. Since the Pierce two-node model can predict the mean skin temperature (MST) of a sedentary occupant accurately (Doherty and Arens, 1988; Doherty and Arens, 1988), it is adopted to simulate the skin responses of each occupant. The PMV-PPD model has six inputs, as shown in Equation 5.3. The Pierce two-node model has eight inputs, as shown in Equation 5.3.

$$PPD = Fanger(T_a, RH, MRT, V_a, Clo, Met)$$

$$MST = PierceTwoNode(T_a, RH, MRT, V_a, Clo, Met, Weight, Height)$$

where T_a ($^{\circ}\text{C}$ or $^{\circ}\text{F}$) is air temperature, RH (%) is relative humidity, MRT ($^{\circ}\text{C}$ or $^{\circ}\text{F}$) is mean radiant temperature, V_a (m/s) is air velocity, Clo (clo) is clothing insulation level, Met (met) is metabolic rate, $Weight$ (kg or lb) is body weight, and $Height$ (m or ft and in) is body height.

The experiment aimed to design three occupants that have different thermal preferences. Hence, the individual differences are modeled by assuming that the three occupants have different dress preferences, metabolic rate, weight, and height, as summarized in Table 5.1.

TABLE 5.1: Clothing insulation, metabolic rate, weight and height of the three occupants

	Occupant A	Occupant B	Occupant C
	0.67	0.89	1.10
Clothing Insulation (clo)	(knee-length skirt + long-sleeve shirt + full slip)	(overalls trousers + long-sleeve shirt + T-shirt)	(ankle-length skirt + long-sleeve shirt + suit jacket)
Metabolic Rate (Met)	1.0 (reading, seated)	1.0 (writing, seated)	1.1 (typing, seated)
Weight	60 kg (132 lb)	100 kg (220 lb)	85 kg (187 lb)
Height	1.65 m (5 ft and 5 in)	1.81 m (5 ft and 11 in)	1.81 m (5 ft and 11 in)

Based on the settings in Table 5.1, the PPD , MST , and set-point preferences of the three occupants will be different even when they are in the same thermal environment. For example, Figure 5.2 and Figure 5.3 show the relationships of MST and negative PPD and the relationship of MST and T_a for the three occupants when the thermal environment is that RH is 30 %, MRT

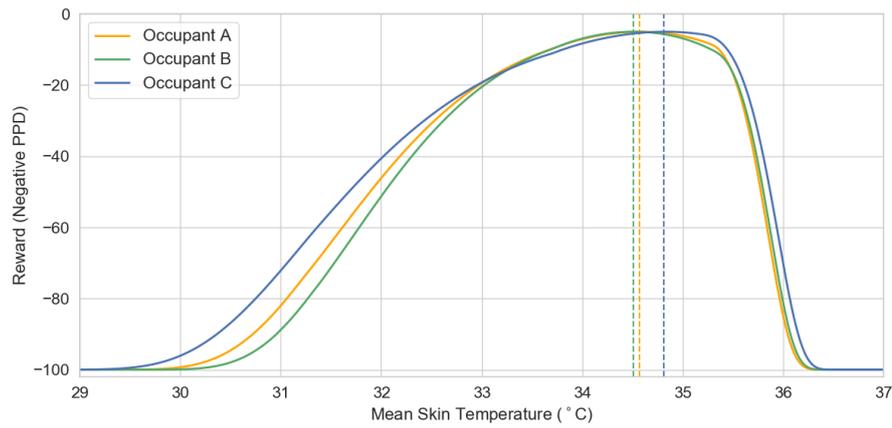


FIGURE 5.2: The relationship of mean skin temperature and negative PPD in the specific thermal condition

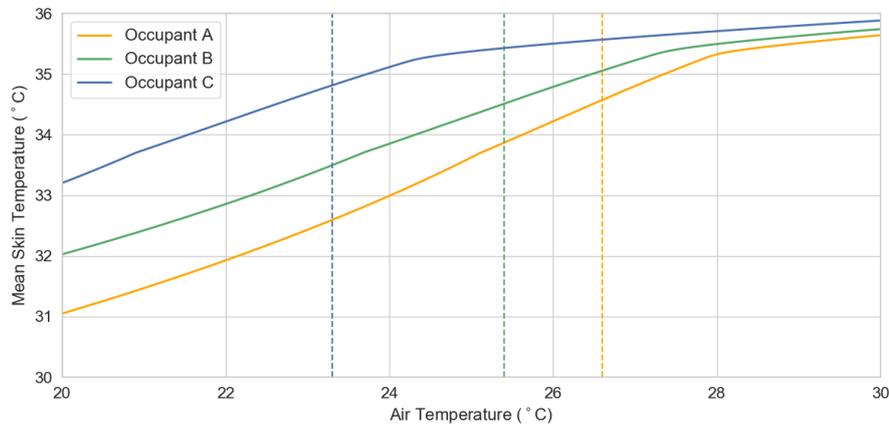


FIGURE 5.3: The relationship of air temperature and mean skin temperature in the specific thermal condition

is $2\text{ }^{\circ}\text{C}$ lower than T_a , and V_a is 0.1 m/s . As indicated by the Figures, to achieve the lowest PPD, the MST should be $34.57\text{ }^{\circ}\text{C}$ for occupant A, $34.51\text{ }^{\circ}\text{C}$ for occupant B, and $34.81\text{ }^{\circ}\text{C}$ for occupant C. The air temperature that can achieve the lowest PPD should be $26.62\text{ }^{\circ}\text{C}$ for occupant A, $25.40\text{ }^{\circ}\text{C}$ for occupant B, and $23.35\text{ }^{\circ}\text{C}$ for occupant C. Based on the occupant models, the air temperature that can achieve the lowest PPD averaged from three occupants is around $25\text{ }^{\circ}\text{C}$.

5.4 Building Model

The building model is modified from a DOE (department of energy) reference building. It is a one-story building with four exterior zones and one interior conditioned zones. The building is located in Pittsburgh, PA, USA, so the TMY3 (typical meteorological year) weather files of of Pittsburgh Airport are used to simulate outdoor weather. The HVAC systems of the building are single duct VAV (variable air volume) systems with electric reheat. The efficiency of the electric heating coils is 100 %. The three occupants are assumed in the Northern zone of the building from 8:00 AM to 6:00 PM. The size of the Northern zone and the location of windows/doors are shown in Figure 5.4.

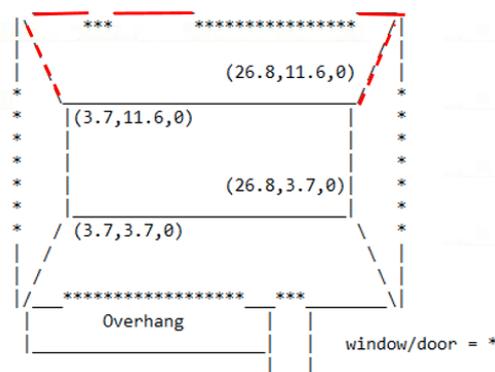


FIGURE 5.4: Building model for the simulation experiment

5.5 The Bio-REAL Control System

The Bio-REAL system in this experiment has three Bio-REAL agents. The Bio-REAL agent adjusts the thermal environment on behalf of the three occupants. The three elements of RL, state, action, and reward for an agent to learn is described below.

5.5.1 Action, Reward, and State Design for the Three Bio-REAL Agents

State. In this experiment, occupant's mean skin temperature, outdoor air temperature, and diffuse solar radiation are selected as the state. The mean skin temperature is for the Bio-REAL agents to understand occupants' current thermal comfort status. The diffuse solar radiation and outdoor air temperature can help the Bio-REAL agents to foresee the possible energy consumption of different actions. Since the experiment room faces North, the direct solar radiation was ignored.

Action. The control action of this experiment was adjusting temperature setpoint. The action set A is shown in Equation 5.1:

$$A \equiv \Delta T_{setpoint} \in \{-2, -1, 0, 1, 2\} \quad (5.1)$$

where $\Delta T_{setpoint}$ is the desired °C changes from the current temperature setpoint. Because the agents may take naive actions at the beginning of the learning process, the temperature setpoint is restricted to the range of 20 °C to 30 °C to avoid the risk of extreme over-cooling or over-heating. Moreover, since the agent takes actions every 15 minutes, the highest setpoint variation is no higher than 2 °C to avoid too heavy temperature fluctuation within a short period.

Reward. The objective of the Bio-REAL agent is to maximize the thermal satisfaction of the occupant it represents. The experiment quantified the individual thermal satisfaction as the PPD and designed the reward as the

negative PPD, as shown in Equation 5.3. In short, the agent will receive a penalty if the occupant feels thermally uncomfortable.

$$R_t = -PPD_t \quad (5.2)$$

5.5.2 Negotiation

This thesis applied the double Q-learning with experience replay and neural network approximation algorithm to train the Bio-REAL agents, as described in Chapter 3. Besides negotiating the conflicts in occupants, the negotiator in this experiment is also responsible for saving energy. This experiment assumed that the temperature setpoint T_{set} closest to outdoor air temperature T_{out} is the most energy-efficient. Therefore, the negotiation procedure in this experiment was:

1. Select two actions A_t^{i*} that can maximize $\sum_i w^i Q_\theta^i(S_t^i, A^i)$ using the ϵ -greedy method
2. Select a_t^{i*} that can minimize $|T_{set} - T_{out}|$ from A_t^{i*}
3. Execute the action a_t^{i*}

In this experiment, the weight w for each occupant is 1.

5.5.3 Architecture and Hyper-parameters of the Bio-REAL system

After trying different architecture and hyper-parameters for the Bio-REAL control system, the following Q-network and hyper-parameters were determined.

Q-network. The Q-network has two hidden-layers, each of them has 24 nodes. The activation function is rectifier for the input layer and the first hidden layer and linear regression for the second hidden layer. Keras (Keras, 2019), a neural network API, is used to program the Q-networks. The loss function is Huber loss. The optimizer is

Hyper-parameter. The discount rate is 0.99. The learning rate is 0.001. The size of memory D is 1,000. The intervals for the target-network copying from the online-network, τ , is 10,000. The exploration probability, ϵ , is 0.05.

5.6 Simulation Run

The run period of the simulation was one weekday, January 01 (Monday). One simulation run is equivalent to one learning episode for the Bio-REAL agents. The simulation timestep was 15 minutes, indicating that the control system adjusts the temperature setpoint every 15 minutes. The Bio-REAL system only worked from 8:00 AM to 6:00 PM when there are three occupants, so the number of steps for each episode was 40. The temperature setpoint for the unoccupied period (from 12:00 AM to 8:00 AM and from 6:00 PM to 11:45 PM) was 20 °C. The Bio-REAL system was evaluated by 200

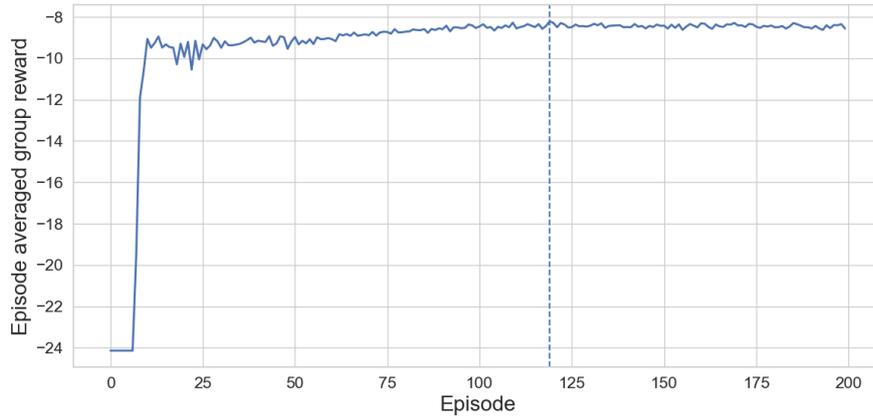


FIGURE 5.5: The learning speed and the control performance of the Bio-REAL control system

episodes. The group PPD (\overline{PPD}) averaged over the three occupants and 40 steps, as shown in Equation (5), was calculated every episode to analyze the learning speed and control performance of the Bio-REAL system.

$$\overline{PPD} = \frac{1}{3} \sum_{i=1}^3 \frac{1}{40} \sum_{t=1}^{40} PPD_t^i \quad (5.3)$$

5.7 Experiment results

The learning speed was quantified by the number of episodes needed for the Bio-REAL systems to converge to an optimum, where the reward (negative PPD) is maximized and sustained in the following episodes. The control performance was quantified by the \overline{PPD} and the HVAC daily electricity consumption after the convergence. As shown in Figure 5.5, the learning efficiency was 112 episode, indicating that the control system can perform optimally after interacting with the same occupants 112 days. The sub-optimum can be achieved after 25 episodes.

The control performance of the Bio-REAL system was compared with the comfort-oriented standard static schedule-based controls, as shown in Table 5.2. Since 25 °C was the air temperature setpoint that can achieve the lowest PPD for the three occupants, as mentioned in section 5.3, the comfort-oriented schedule control selected 25 °C as the static setpoint. The comfort-oriented schedule can also be considered as the optimum static schedule for thermal comfort. 22 °C is a typical indoor temperature setpoint in winter, so the standard static schedule-based controls selected 22 °C as the static setpoint.

TABLE 5.2: Comparison of the optimal control policy and static schedule

	Bio-REAL system	Comfort-oriented schedule	Standard schedule
		25 °C	22 °C
Group PPD	8.21	8.25	17.12
HVA Electricity Consumption	207.62kWh (708.40 kBTU)	207.31kWh (707.34kBTU)	196.80kWh (671.48kBTU)

Figure 5.6 shows the optimum dynamic control policy derived by the Bio-REAL system. As the Figure presented, the optimum dynamic control policy increased the setpoint by 2 ° every timestep from 8:00 AM to 9:15 AM so that the setpoint raised from 20 °C to 26 °C, then it maintained the setpoint as 25 °C from 9:15 AM to 4:15 PM. From 4:15 PM to 6:00 PM, the setpoint varied between 25 °C and 26 °C. The indoor air temperature varied the same as the temperature setpoint. The mean radiant temperature (MST) was affected by both temperature setpoint and solar radiation. The lower MST in the early

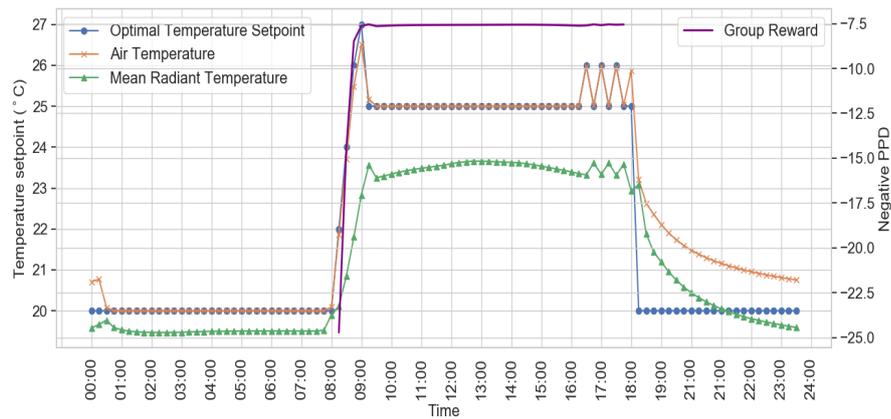


FIGURE 5.6: The group reward \overline{PPD} , indoor air temperature, and mean radiant temperature using the optimum dynamic control policy

morning and late afternoon was due to the decreases in solar radiation. Although there was variation in the optimum dynamic temperature setpoints, the group reward (\overline{PPD}) can be maintained at its maximum, around -7.5, from 9:00 AM to 6:00 PM. The lower reward from 8:00 AM to 9:00 AM was just because of the experiment constraint that the maximum increment in setpoint was 2 °C. Furthermore, the Bio-REAL system was able to learn that the setpoint should be higher (e.g., 26 °C) if there is a drop in solar radiation.

The results in Table 5.2 shows that the \overline{PPD} of the Bio-REAL system with the optimum dynamic control policy was 0.49% higher than that of occupant model-based controls using 25 °C and 52% higher than that of standard static schedule-based controls using 22 °C. However, the energy consumption of the controls with the Bio-REAL system was the highest.

5.8 Conclusion and Discussion

This experiment demonstrated the feasibility of the Bio-REAL system with a simple simulated learning environment comprised classic occupant models and the EnergyPlus model of a room with a VAV system. The classic occupant models were the PMV model and Pierce two-node model. The room was the Northern zone of a one-story DOE reference building. The simulation run was one weekdays (January 01). Three personalized Bio-REAL agents and a negotiator was created for the control system. Besides, this experiment tuned and determined the architecture and the hyper-parameters of the Bio-REAL systems.

The experiment results showed that the Bio-REAL control system converged to an optimum after learning with 112 episodes. The optimum dynamic control schedule created by the Bio-REAL system can achieve 0.49% thermal comfort improvement, compared to the optimal static control schedule (25 °C) generated based on the occupant models. The Bio-REAL system had 52% better performance than the standard static control schedule (22 °C).

However, the Bio-REAL system fails to reduce energy consumption because it aims to maximize thermal satisfaction and, on that basis, minimize energy consumption. Given the simulation setup in this experiment, the set-points optimizing thermal comfort are always the energy-intensive ones. The simulation setup in this experiment also has little randomness. Since the response of real occupants and the dynamic of a real indoor environment can be stochastic, the learning efficiency can be lower in practice.

Moreover, occupants' mean skin temperature is not readily accessible in practice. Only the skin temperature of specific body parts can be obtained non-intrusively using current bio-sensing technologies, such as smart wristbands. The following experiments investigated other approaches of balancing group thermal comfort and energy efficiency, integrate bio-sensing technology, and tested the Bio-REAL system with real occupants in real office buildings.

Chapter 6

Preliminary Field and Simulation

Experiment at CMU: Water Sourced

Radiators for Heating

6.1 Objective

This field and simulation experiment was designed to evaluate the performance of the Bio-REAL control system in a heating season with a more complex and realistic learning environment, compared to the simulation experiment in Chapter 5. More specifically, personalized occupant thermal comfort models were created for each occupant to build individual thermal comfort zones based on data collected from the preliminary field experiment. Six control agents and a negotiator of the Bio-REAL control system were created. The achievement in thermal satisfaction and the saving on energy consumption of the Bio-REAL control system were quantified and compared to the baseline control. Another objective of this experiment is to evaluate

the generalizability of the Bio-REAL control system by training and testing the system with different learning environments. Moreover, this experiment demonstrated the flexibility of the Bio-REAL control system with the change of occupancy and the objective function from comfort to energy.

6.2 Simulation Setup

In the simulation experiments, six personalized occupant models, a building model, and the Bio-REAL control system worked in an integrative way, as shown in Figure 6.1. The occupant models and the building models formed the learning environment for the Bio-REAL control system. To make the simulated learning environment as close as the real environment, the author conducted a human subject experiment to collect human's subjective and biological responses at indoor conditions with varied temperature and humidity. Then, these responses were used to build personalized occupant models. Besides, a building model was created by the EnergyPlus tool (Crawley et al., 2001) to simulate the room conducted the human subject experiment.

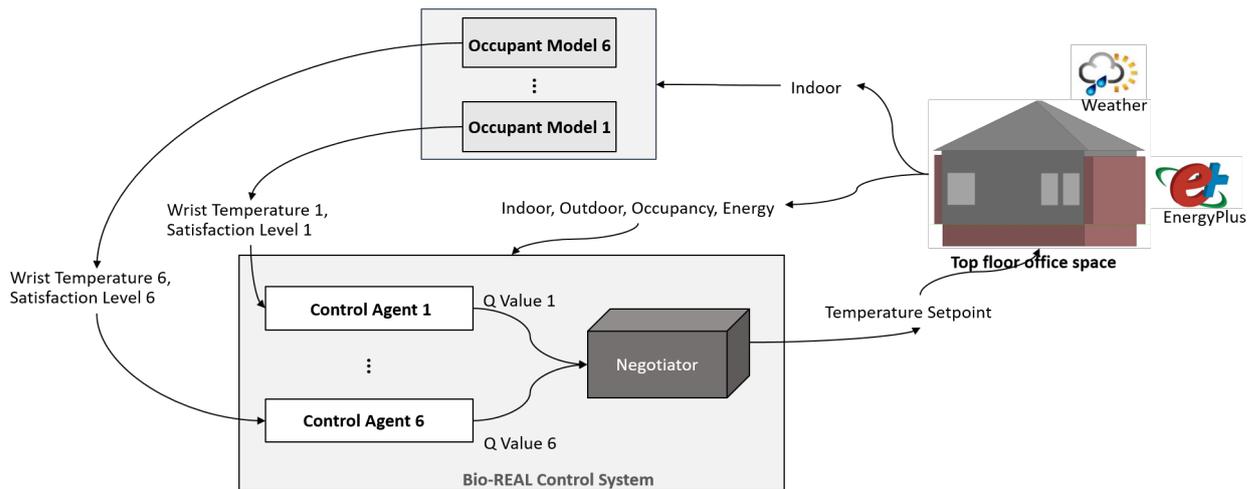


FIGURE 6.1: The integration of the personalized occupant models, the building model, and the Bio-REAL control system.

The data flow of the simulation process is also described in Figure 6.1. During the simulation, the Bio-REAL control system decides a temperature setpoint, based on which, the building model simulates the indoor conditions of the room and the energy consumption of the HVAC system. The indoor conditions, including air temperature and relative humidity, are the inputs for the personalized occupant models, which output individual thermal satisfaction levels and wrist skin temperature. These outputs are then the inputs of the Bio-REAL control system. Besides, the indoor and outdoor conditions (outdoor air temperature and solar radiation), the number of occupants, and the energy consumption from the building model are also the inputs of the Bio-REAL control system.

6.3 Human Subject Experiment at CMU

6.3.1 Experiment Room

The human subject experiment was undertaken from February 19th, 2018 to April 11th, 2018 in a room of Margaret Morrison Carnegie Hall ($40^{\circ}26'31.5''\text{N}$, $79^{\circ}56'29.5''\text{W}$) at Carnegie Mellon University (CMU), Pittsburgh, Pennsylvania, United States. The experiment room ($6.6\text{m} \times 7.5\text{m} \times 3.0\text{m}$) is on the fourth floor, which is the top floor of the hall, as shown in Figure 6.2. Above the room is a pitched roof. The experiment room has one exterior wall that is facing North and has two windows. There are two water radiators in the room for heating. Four 1500kw heaters were placed at the four corners of the room to provide additional heat if necessary.



FIGURE 6.2: The room for the human subject experiment

6.3.2 Experiment Subjects

Six healthy graduate students at CMU participated in the experiment. They are three female and three male. Their demographic information is listed in

Table 6.1. The body mass index (BMI) was calculated based on their height and weight.

TABLE 6.1: Demographic information of the experiment subjects at CMU

	Age	BMI
Mean	26.17	21.39
Standard Deviation	3.25	3.11

Each subject has a fixed workspace with one chair and one desk. The locations of the workspaces are illustrated in Figure 6.3. The six experiment subjects were required to wear long pants, shoes, and a sweater or long sleeve shirt (thick) during the experiment so that their clothing insulation is around 1.0 Clo. They were also asked to work on "office-type" work, such as reading, typing, and web surfing, to make sure their metabolic rate is 1.0-1.1 met.

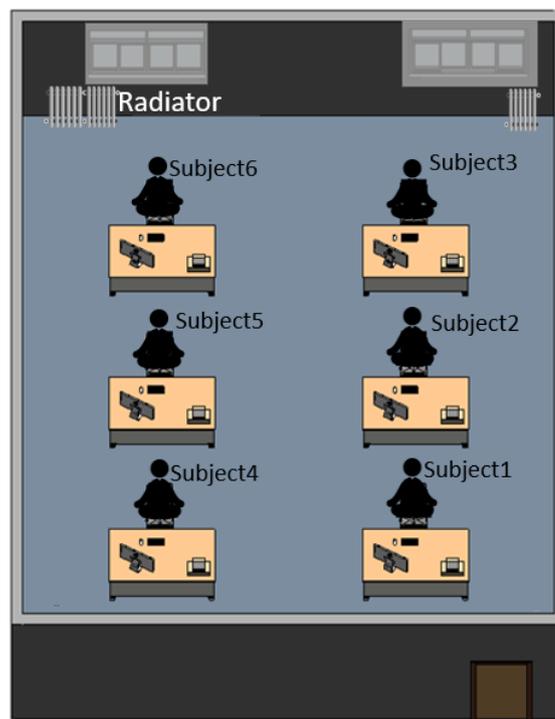


FIGURE 6.3: Locations of the six experiment subjects

6.3.3 Experiment Design

The experiment was single-blinded and included repeated measures of thermal comfort function and biological responses on the same subject. There were 22 experiment sessions, most of which were conducted after 7 PM to avoid the impact of solar radiation. Each session lasted 2-3 hours to eliminate the effect caused by long-time exposure. Each subject participated in the 6-7 experiment session to obtain enough individual samples.

The experimentally controlled variable was air temperature, which was varied slowly either from 18 °C to 30 °C or from 30 °C to 18 °C during each session by turning on/off the two radiators and the four heaters. On average, the temperature increased or decreased by 1 °C every 10 minutes. The change in air temperature caused the variation to mean radiant temperature and relative humidity.

6.3.4 Measurement and Equipment

The variables measured in the experiment can be divided into three groups: environment, biological, and subjective responses. The environment responses measured were air temperature, relative humidity, and radiant temperature. The air temperature and relative humidity were measured by self-constructed wireless sensor kits. Each of them was placed on the desk of each workplace during the experiment. The sensor kit is made up of a WiFi ESP8266 Microcontroller, a DHT22 Temperature/Humidity Sensor, and an OLED (organic light-emitting diode) Display. The radiant temperature was measured by

the REED instruments SD-2010 SD Series WBGT heat stress meter that has a black globe and a data-logger. The biological responses, including wrist temperature, heart rate, and RR interval, were measured by the bio-sensing network described in Section 3.1. Each subject wore a wrist band on the same hand during the experiment. Subjects were reminded to report their thermal sensation and thermal satisfaction through a web survey every 5 minutes. The variables measured, their measurement intervals, and equipment for the measurement were summarized in Table 6.2. Except for the radiant temperature, all the data collected were saved into a database through a wireless sensor network as described in Section 3.1. The radiant temperature was stored in the data-logger of the REED WBGT meter.

6.3.5 Data Analysis: Environmental Conditions

During the 22 experiment sessions, the air temperature was controlled no higher than 29.78 °C and no lower than 17.85 °C. The fluctuations of the radiant temperature were almost the same as that of the air temperature at different locations of the experiment room, hence the authors assumed that the operative temperature of the experiment was equal to the air temperature. Relative humidity was affected by both air temperature and indoor humidity ratio. The statistics of the operative temperature and relative humidity are summarized in Table 6.3.

TABLE 6.2: Variable and equipment of the CMU experiment

	Variable	Interval	Equipment
Environmental	Air Temperature (°C)	10 seconds	
	Relative Humidity (%)		
	Radiant Temperature (°C)	5 seconds	
Biological	Wrist Temperature (°C)	30 seconds	
	Heart Rate (bpm)	0.1 seconds	
	RR-interval (seconds)		
Subjective	Thermal Sensation	5 minutes	Web Survey
	Thermal Satisfaction		

TABLE 6.3: Experimental condition during the 22 experiment sessions

	Operative Temperature	Relative Humidity
Max	29.78 °C	40.67 %
Min	17.85 °C	9.17 %
Mean	24.38 °C	22.78 %

6.3.6 Data Analysis: Biological Responses

Three types of biological responses were recorded: wrist temperature, heart rate, and RR-interval. Their effectiveness was evaluated based on their statistics and their correlation with operative temperature (Cor_{T_o}) and thermal sensation (Cor_{sen}). Table 6.4 shows the Pearson correlation and some statistics of the bio-responses. As shown in the table, The mean of these bio-responses was all in a reasonable range. The mean wrist temperature of the six subjects was between 31 °C and 33 °C. The mean heart rate of them was around 72-74. The mean RR-Interval was around 0.8. Besides, the wrist temperature had a strong correlation with the operative temperature for all the six subjects. Its correlations with thermal sensation were all above 0.5. However, the Cor_{T_o} and Cor_{sen} for both heart rate and RR-interval were low and not consistent among subjects. For some subjects, they were positively correlated. For others, they were negatively correlated. Therefore, only wrist temperature was considered as the effective input for the Bio-REAL control system. Heart rate and RR-interval were not used in the later simulation experiment.

6.3.7 Data Analysis: Thermal Satisfaction Variation

There was 1,348 thermal sensation and satisfaction feedback collected from the six subjects. More than 200 for each of the them. Since thermal satisfaction was an important input for the Bio-REAL control system in later experiment, the author focused on analyzing satisfaction rather than sensation.

TABLE 6.4: Biological responses of the six subjects during the 22 experiment sessions

	S1	S2	S3	S4	S5	S6	
Wrist Temperature	Cor_{T_0}	0.83	0.85	0.90	0.89	0.83	0.91
	Cor_{sen}	0.50	0.60	0.77	0.46	0.58	0.67
	Mean	32.23	32.56	31.96	31.14	31.61	31.96
	Max	35.25	35.48	35.87	35.88	36.06	36.14
	Min	28.54	27.77	27.42	27.12	27.48	27.43
Heart Rate	Cor_{T_0}	0.42	0.37	0.26	0.15	-0.05	-0.02
	Cor_{sen}	0.19	0.43	0.27	0.02	-0.06	0.27
	Mean	72.34	74.03	74.62	74.20	72.88	74.25
	Max	83.80	90.28	85.96	92.85	103.67	105.00
	Min	62.73	60.89	60.12	60.01	62.69	61.47
RR-Interval	Cor_{T_0}	-0.18	-0.25	-0.32	0.15	0.06	0.10
	Cor_{sen}	-0.06	0.33	0.23	0.04	-0.04	-0.22
	Mean	0.80	0.79	0.78	0.83	0.80	0.78
	Max	0.95	1.04	0.94	0.97	0.95	0.98
	Min	0.57	0.61	0.54	0.56	0.63	0.54

TABLE 6.5: Thermal satisfaction distribution of the six subject

	S1	S2	S3	S4	S5	S6
Strongly Satisfied (3)	1	0	3	0	0	29
Satisfied (2)	55	51	55	40	27	92
Slightly Satisfied (1)	48	36	28	55	19	68
Neutral (0)	30	55	43	80	34	22
Slightly Dissatisfied (-1)	34	50	45	39	125	7
Dissatisfied (-2)	38	23	31	4	27	4
Strongly Dissatisfied (-3)	8	3	6	0	0	0
Total	214	218	211	218	232	222

The votes of the thermal satisfaction levels are not evenly distributed, as shown in Table 6.9. The number of votes on strongly satisfied and strongly dissatisfied are much less than others. Therefore, the 7-level thermal satisfaction was grouped into a 4-level thermal satisfaction. Slightly satisfied, satisfied, and strongly satisfied were all grouped as satisfied to treat satisfied levels the same. Neutral and slightly dissatisfied were in their original groups. Dissatisfied and strongly dissatisfied were classified as the dissatisfied.

The relationship between the six subjects' 4-level thermal satisfaction and indoor environment (operative temperature and relative humidity) were visualized in Table 6.6. The visualization of the data discovers some characteristics of the six subjects. For example, Subject 1 is more likely to be satisfied if the temperature is lower than 24 °C. Subject 6 has a relatively larger thermal comfort zone, while Subject 5 has a relatively smaller comfort zone. These experiment data demonstrated individual differences in thermal comfort. Clearer characteristics of the six subjects were identified by the personalized occupant models.

6.4 Occupant Models

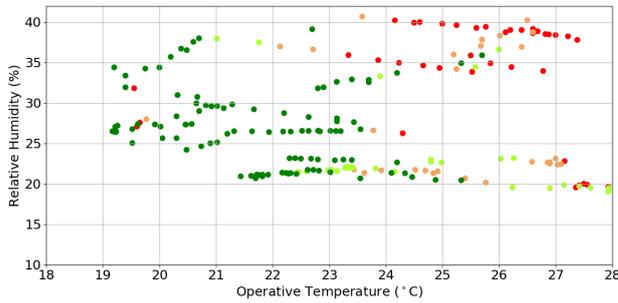
6.4.1 Personalized Thermal Comfort Models

The personalized thermal comfort models were developed for the six subjects to predict their thermal satisfaction levels at the operative temperature between 18 °C and 28 °C and relative humidity between 10% and 40%. The

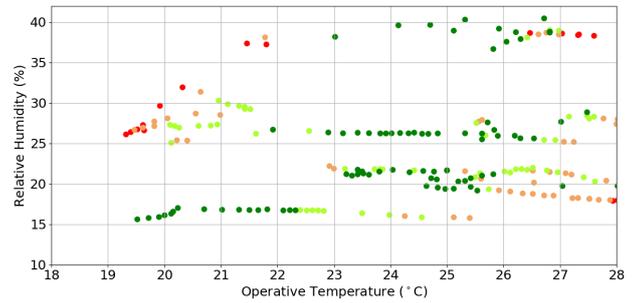
TABLE 6.6: 4-level thermal satisfaction in different operative temperature and relative humidity conditions for the six subjects (1.0 Clo, 1.0-1.1 Met, 0.1 m/s).

● Dissatisfied ● Slightly Dissatisfied ● Neutral ● Satisfied

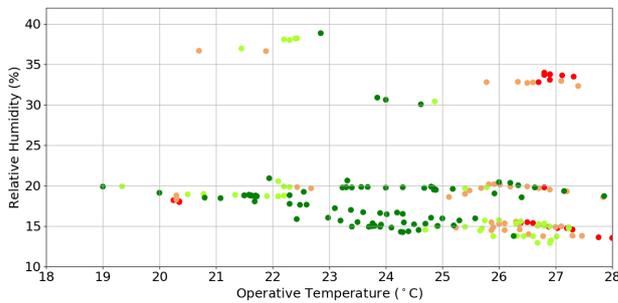
Subject1 (n=214)



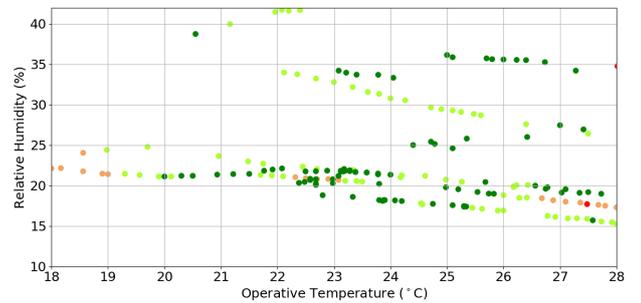
Subject2 (n=218)



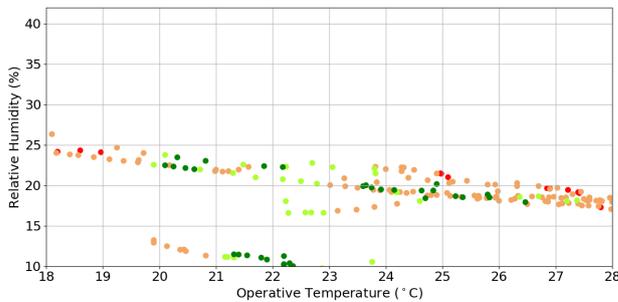
Subject3 (n=221)



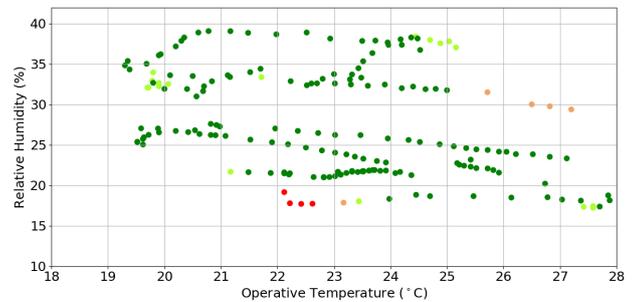
Subject4 (n=228)



Subject5 (n=232)



Subject6 (n=222)



existing methods for creating a personalized thermal comfort model were developing machine learning models based on real data collected from occupants in the field. These methods only focused on the prediction accuracy and not able to establish an individual thermal comfort zone for each occupant because the data sample from the field experiment was usually not big enough or representative. For example, as indicated by Table 6.6, from the data sample of subjects 5, it is hard to tell the thermal satisfaction of this subject for relative humidity above 30%. Therefore, to overcome the limitations, the author used a new approach to build the personalized thermal comfort model: (1) generate 1,000 data samples from the PMV model; (2) develop a initial neural network using these data samples; (3) update the neural network using the individual data collected from the human subject experiment to develop personalized thermal comfort models.

Among the six inputs of the PMV/PPD model, the author assumed that the clothing insulation and metabolic rate are 1.0, air velocity is 0.1m/s. 1000 air and mean radiant temperature data were sampled from 18 °C to 28 °C. 1000 relative humidity data were sampled from 10% to 40%. These input data was plugged into the PMV/PPD equation to output 1000 PPD data. The heatmap of the 1000 PPD data is shown in Figure 6.4.

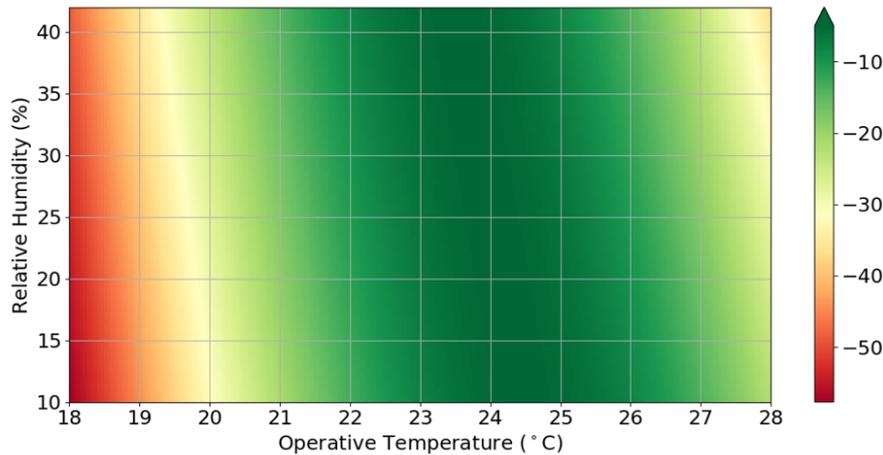


FIGURE 6.4: PPD (Predicted Percent Dissatisfied) heatmap created by 1000 data samples from the PMV/PPD model

Same as the real thermal satisfaction responses, the PPD value was also grouped into four satisfaction levels. PPD less than 10% was considered as satisfied because it is equal to PMV between -0.5 and 0.5. PPD between 10% and 25% ($0.5 < \text{PMV} < 1.0$ and $-1.0 < \text{PMV} < -0.5$) was regarded as neutral. PPD between 25% and 50% ($1.0 < \text{PMV} < 1.5$ and $-1.5 < \text{PMV} < -1.0$) was slightly dissatisfied. PPD that greater than 50% ($\text{PMV} > 1.5$ and $\text{PMV} < -1.5$) was dissatisfied.

The processed 1000 input and PPD data sample were used to train a neural network. The neural network has two hidden layers, each of which has eight nodes. The activation of the input layer and hidden layers are a rectified linear unit. The activation of the output layer is a softmax function. The loss function is categorical cross entropy that can compare the distribution of the predictions with the true distribution. Figure 6.5 visualizes the prediction result of the neural network. The tool Keras (Keras, 2019) and scikit-learn (learn, 2019) were used to build and train the neural network.

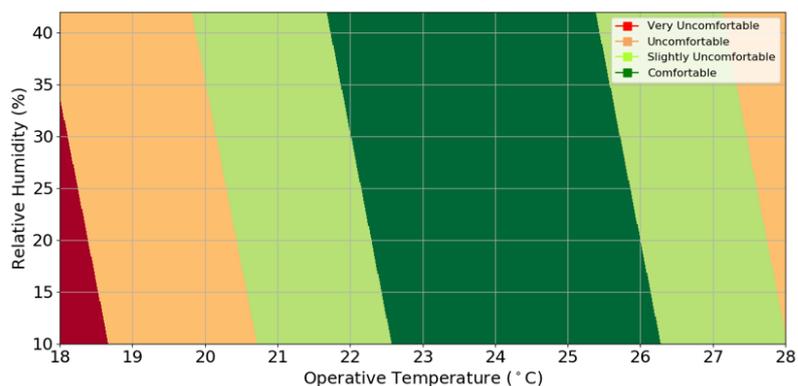


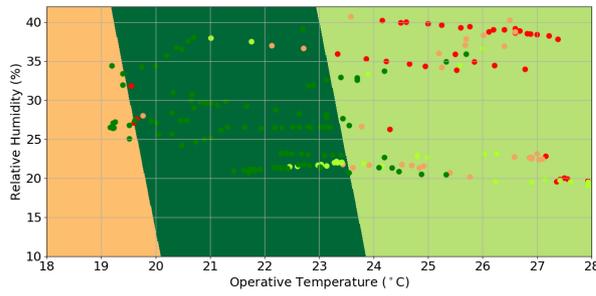
FIGURE 6.5: Prediction results of the PPD based neural network

The neural network trained by the PPD data was updated by the individual data collected from the six subjects to obtain six personalized thermal comfort models. Table 6.7 visualized the prediction results of the six personalized thermal comfort models. The results displayed the individual thermal comfort zones, which are different from the thermal comfort zones showed by the PMV/PPD model (Equation 6.5). The results also presented clearer characteristics of the six subjects than the visualization in Table 6.6. For example, subject 5 has the smallest thermal comfort zone, indicating that he/she requires a more delicately controlled thermal environment. Subject 6 and Subject 2 have the first and the second largest thermal comfort zones, suggesting that they have more flexibility in the thermal environment. The thermal comfort zone of Subject 4 is warmer than that of Subject 1. Subject 3 dislikes high relative humidity. Only Subject 4 and Subject 5 will be dissatisfied if the operative temperature is lower than 19 °C.

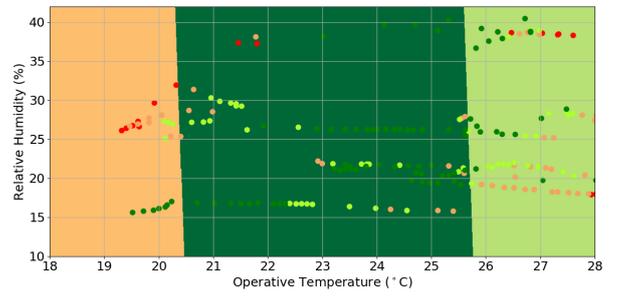
TABLE 6.7: Prediction result of the personalized thermal comfort mode

● Dissatisfied ● Slightly Dissatisfied ● Neutral ● Satisfied

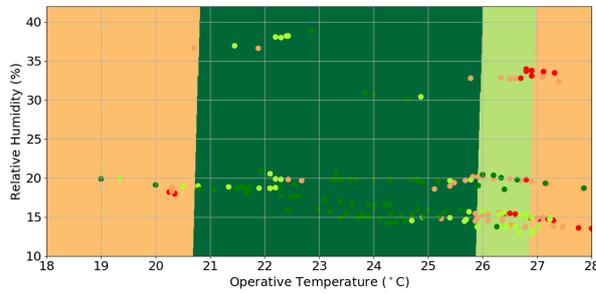
Subject1 (n=214)



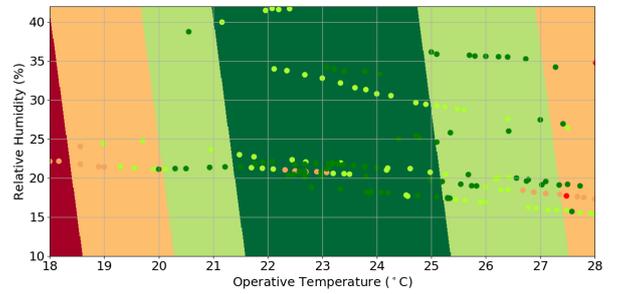
Subject2 (n=218)



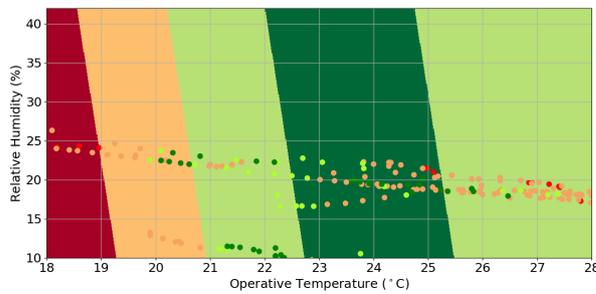
Subject3 (n=221)



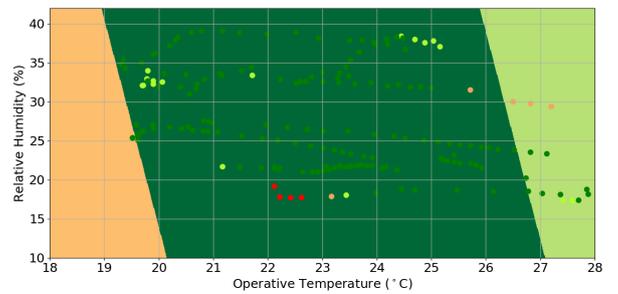
Subject4 (n=228)



Subject5 (n=232)



Subject6 (n=222)



6.4.2 Personalized Skin Response Models

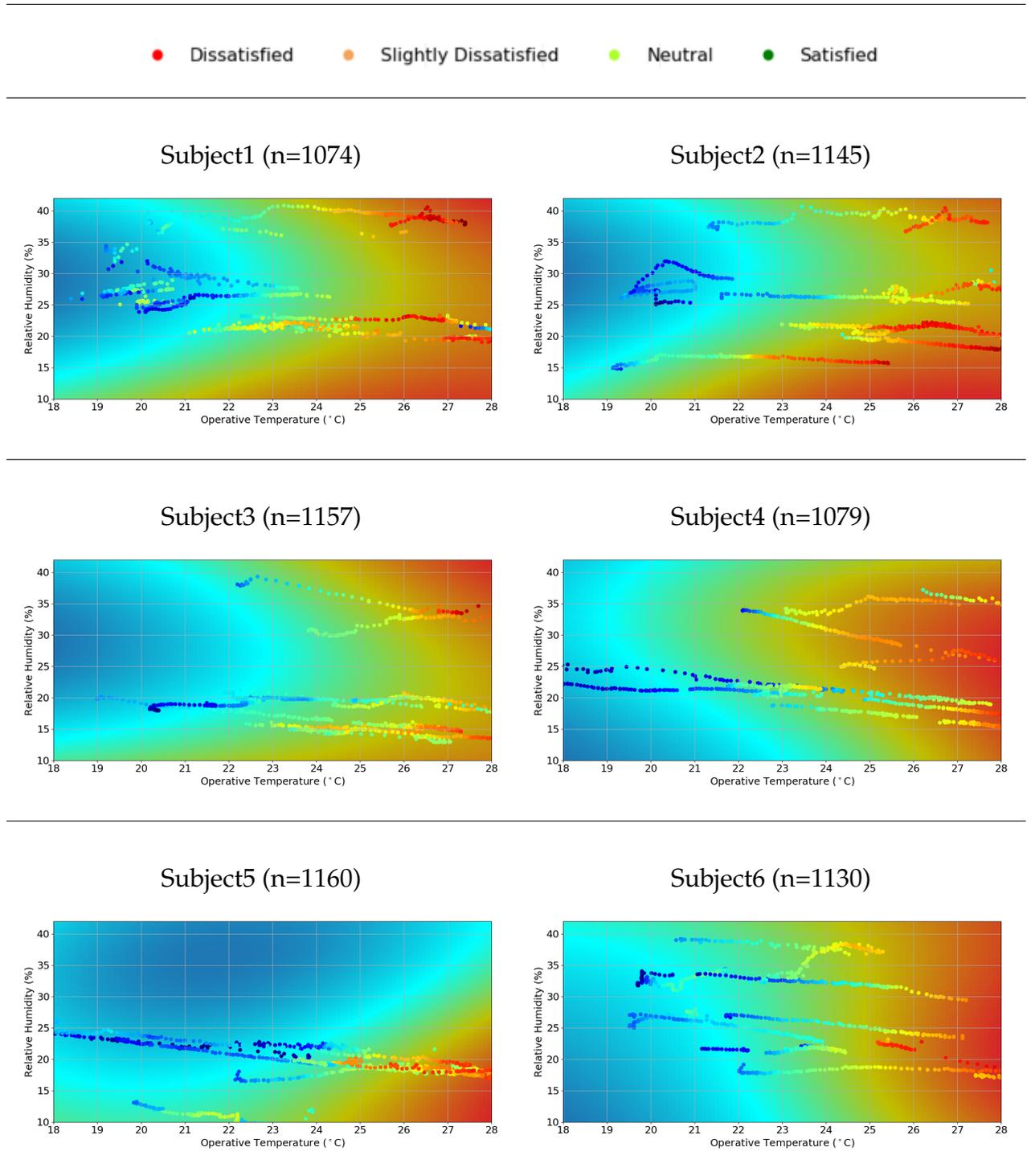
The personalized skin response models were developed for the six subjects to predict their wrist temperature responses at the operative temperature between 18 °C and 28 °C and relative humidity between 10% and 40%. The author applied a machine learning algorithm, support vector regression, to build the models. Six support vector regressions were trained separately by data samples (wrist temperature, operative temperature, relative humidity) collected from each subject (scatter plot in Table 6.8). The heatmap in Table 6.8 presents the prediction results of the personalized skin response models.

As the heatmap showed, the wrist temperature increases with the increase of operative temperature for all the six subjects. However, the relationship between wrist temperature and relative humidity is not monotonic. For example, with the rise of the relative humidity, Subject 1's wrist temperature decreased at the beginning but started to increase when relative humidity is above 30%. This non-monotony may be due to lacking data samples.

6.5 Building Model

The author used the EnergyPlus simulation tool to model the room conducted the human subject experiment. The room size, window location, and window size were modeled as same as the real room, as shown in Figure 6.6. Considering the room and its adjacent rooms have similar thermal conditions during winter, the author assumed that heat is not transferred across

TABLE 6.8: Prediction result of the personalized skin response mode



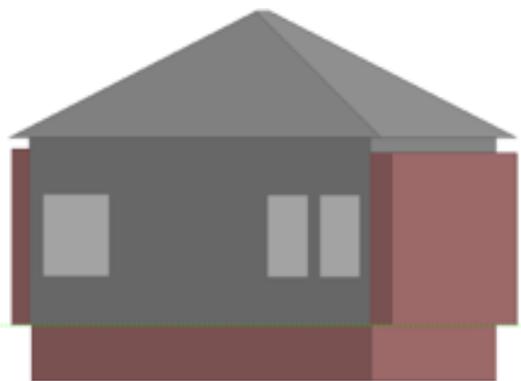


FIGURE 6.6: EnergyPlus model for the experiment room at CMU

in all the surfaces except for its wall facing North and roof. Therefore, adiabatic surfaces were used in thermal modeling to represent the surfaces that are between two zones at similar conditions. The HVAC system of the room is modeled as a water radiator of the EnergyPlus. The hot water for the water radiator is supplied by district heating.

For simplicity, the internal load of the model is the same every day during the run period because their occupancy, lighting, and equipment schedules are the same. These schedules assumed that every day from 8:30 AM to 5:30 PM, the room is fully occupied, all lights and equipment are on. The total occupied time was 9 hours per day.

The external load of the model is different every day during the run period, as shown in Figure 6.7. As mentioned before, to evaluate the generalization capacity of the Bio-REAL control system, the training and testing environment are different. The differences are the outdoor weather. The training environment used Pittsburgh Typical Meteorological Year 3 (TMY3) weather data, while the testing environment used the Pittsburgh 2017 Real Meteorological Year (RMY2017) weather data. As shown in Figure 6.7, outdoor air

dry bulb temperature and air humidity ratio of RMY2017 is slightly higher than those of TMY3 from Tuesday to Friday. Diffuse solar radiation per area of RMY2017 is twice higher than that of TMY3 on Thursday and Friday.

6.6 Bio-REAL Agents

The key to the Bio-REAL control system is to build control agents that can make optimum control decisions. In this simulation experiment, six Bio-REAL agents were built, one for each subject. The objective of an agent is to maximize thermal satisfaction while minimizing energy consumption. The three elements of DRL, state, action, and reward for an agent to learn is described below.

6.6.1 Action, Reward, and State Design for the Six Bio-REAL Agents

State. The state was designed to provide enough information for the agent to understand the current situation. In this experiment, the state involved five variables: wrist temperature to represent the biological status, air temperature, mean radiant temperature, and relative humidity to represent the indoor conditions, outdoor air temperature, and diffuse solar radiation to represent the outdoor conditions. The biological status and indoor conditions provided information for the agent to understand the thermal state of the subjects. The outdoor conditions were for understanding the HVAC system's energy consumption.

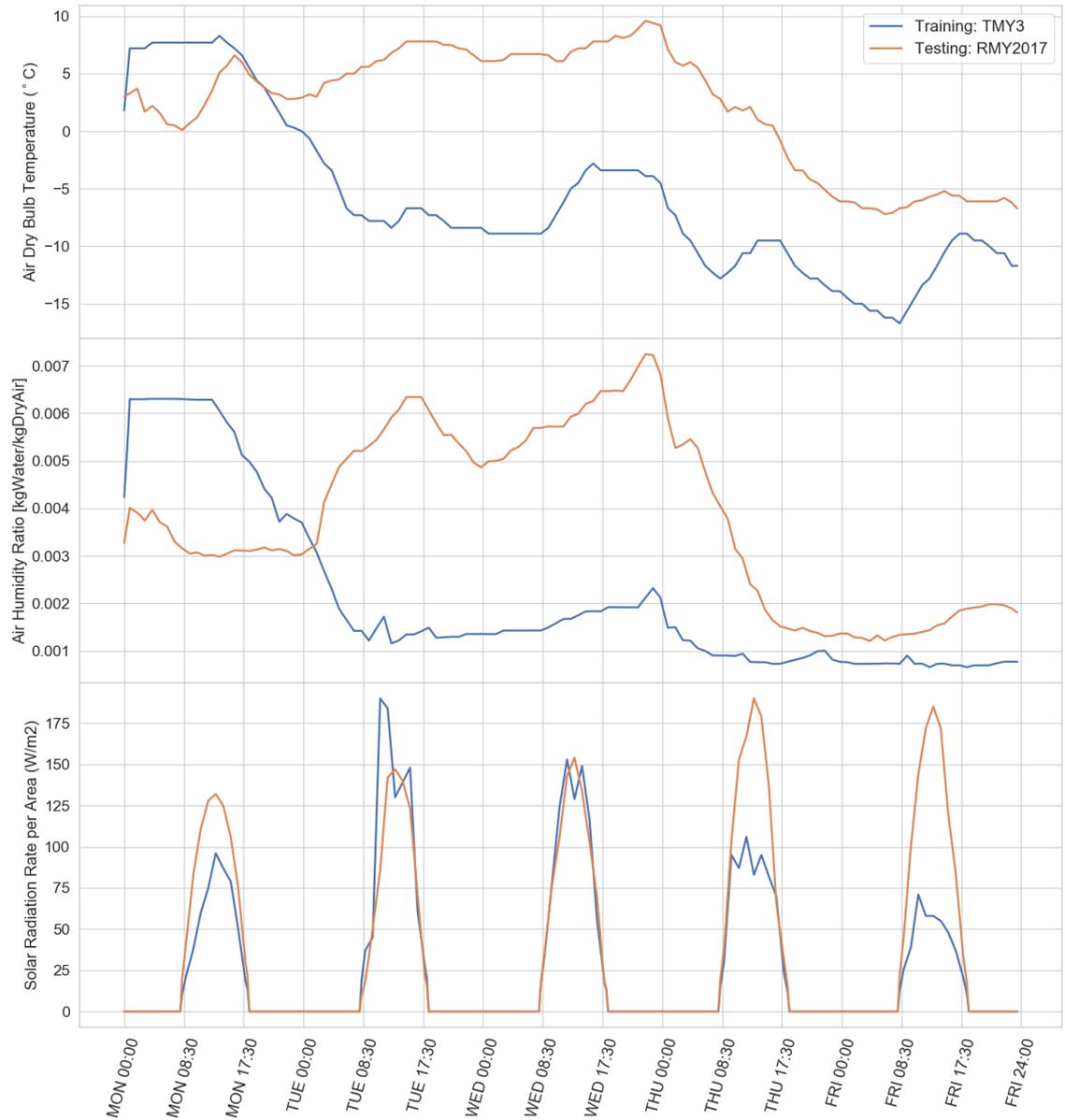


FIGURE 6.7: The comparison of the TMY3 and RMY2017 Pittsburgh weather data

Action. Same as Chapter 5, the control action of this experiment was also adjusting temperature setpoint, as shown in Equation 5.1. The setpoint is restricted to the range of 16 °C to 28 °C because it is not reasonable to set temperature outside this range in a heating season.

Reward. The control objective is to maximize thermal satisfaction and energy efficiency in the long run. Thus, the reward to the agent is designed as the penalty on thermal dissatisfaction and energy consumption, as shown in equation 6.1.

$$reward = dissatisfaction\ penalty \mathbb{1}(occupied) + \lambda * energy\ penalty \quad (6.1)$$

The symbol $\mathbb{1}$ is an indicator function, indicating that the thermal dissatisfaction penalty is only considered during the occupied time. The hyperparameter λ ($0 - \infty$) is designed to weight the importance of thermal satisfaction and energy efficiency. λ 's value should be selected based on building owner or facility manager's preferences. If energy efficiency is considered more important, λ should be higher. If occupant thermal satisfaction is more important, λ should be lower. The personalized thermal comfort models estimate the six subjects' thermal satisfaction as satisfied, neutral, slightly unsatisfied, unsatisfied, which were quantified as penalty/reward, as shown in Table 6.9.

TABLE 6.9: Thermal satisfaction reward/penalty designed for the Bio-REAL agent

Satisfied	Neutral	Slightly Unsatisfied	Unsatisfied
0	-1	-2	-3

The author designed two forms of energy penalty: (1) negative energy consumption and (2) negative temperature setpoint, to penalize energy consumption in different ways.

$$\text{energy penalty} = -\text{energy consumption} \quad (6.2)$$

$$\text{energy penalty} = -(\text{temperature setpoint} - 18) \quad (6.3)$$

Equation 6.2 indicates that the higher the energy consumption, the higher the penalty. This form of penalty is used by this experiment when energy efficiency has a priority. Equation 6.3 indicate that a higher temperature setpoint has a higher penalty. This form is used when the objective is to drive the temperature setpoint to a lower value for energy saving. The 18 °C is the lower limit of the temperature setpoint range.

6.7 Simulation Run

Two different scenarios were considered in this simulation experiment. The first scenario assumed that the six subjects always sit in the room and wear clothes with 0.1 Clo during the occupied time of the five weekdays. In the

second scenario, Subject 5, who has the narrowest individual comfort zone, is absent. In each scenario, four cases were studied to evaluate the performance of the Bio-REAL control system in different perspectives:

Case 1: Energy Efficient Oriented Static Baseline

Case 1 is a baseline that has a fixed setpoint used by DOE reference building for small offices in climate zone 5a in a heating season. The occupied temperature setpoint used is 21 °C, as shown in Figure 6.8. The setback or unoccupied temperature setpoint is 16 °C, that is rounded from the DOE setback 15.6 °C. This baseline was regarded as energy efficient because the occupied setpoint is relatively low for the six subjects.

Case 2: Thermal Satisfaction Oriented Static Baseline

Case 2 is also a baseline that has a fixed setpoint and setback (Figure 6.8), but the setpoint is selected based on the individual thermal comfort zones of the six subjects (Table 6.7) . These thermal comfort zones show that, when relative humidity is from 10% to 40%, the operative 23 °C can satisfy the six subjects most of the time. Therefore, 23 ° is the best static setpoint for the six subjects or the five subjects except subject 5. The setback of this case is also 16 °C.

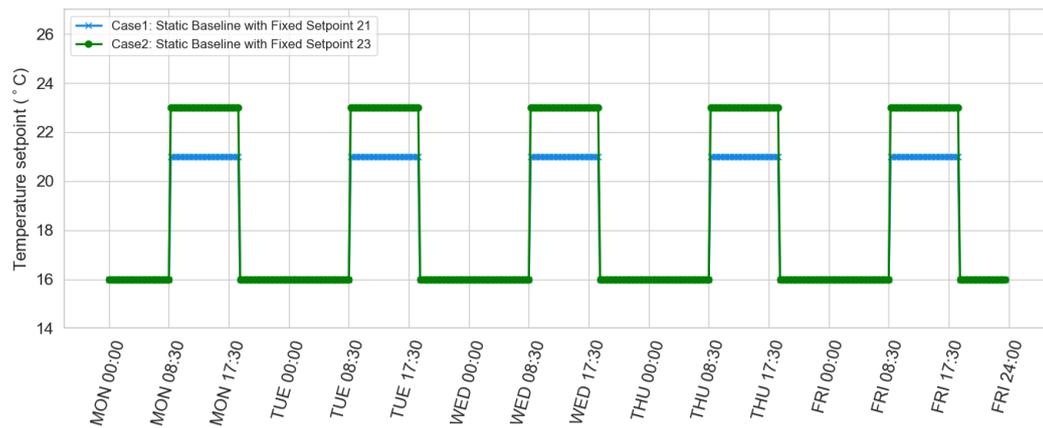


FIGURE 6.8: Case 1 and Case 2: Static Control Schedule with Fixed Setpoint

Case 3: Thermal Satisfaction Oriented Dynamic Bio-REAL Control

In Case 3, the Bio-REAL control system with six trained agents and a negotiator was implemented to adjust the setpoint. The control emphasized more on thermal satisfaction of the subjects. The energy penalty of the reward function was Equation 6.3. The highest energy penalty of the case was -1 , calculated from $-0.1 \times (28 - 18)$, equal to the lowest dissatisfaction penalty. Since the energy penalty was no higher than dissatisfaction penalty all the time, the agent sought the control strategy that can meet thermal satisfaction of the subject and, on that basis, lowered setpoint to save energy consumption.

Case 4: Satisfaction and Energy Efficiency Balanced Dynamic Bio-REAL Control

Case 4 also implemented the Bio-REAL control system, the objective of which is to minimize thermal dissatisfaction and energy consumption simultaneously. The energy penalty of the reward function was Equation 6.2. The

weighting factor λ was tuned to weight energy consumption and thermal dissatisfaction evenly. λ was $\frac{1}{150000}$ for the first scenario. It was $\frac{1}{130000}$ for the second scenario.

For all the four cases, the run period of the simulation was five weekdays, from January 01 (Monday) to January 05 (Friday). One simulation run was equivalent to one learning episode for the Bio-REAL agents. The simulation timestep is 15 minutes, meaning that the control system updated the temperature setpoint every 15 minutes. Each episode had 480 steps (5 days \times 24 hours \times 4 times per hour). Case 1 and Case 2 were static-schedule based control, so training was not needed. In Case 3, 4, and 5, the agents of the Bio-REAL control system was trained by a training environment for 1000 episodes. Every two training episodes, the Bio-REAL control system was evaluated by a testing environment. The performance of the control system was evaluated 500 times during the training process.

6.8 Experiment results

The performance of the control strategies was determined by the achievement in thermal satisfaction and energy saving. The thermal satisfaction performance was quantified by two metrics: percentage of dissatisfaction (PD) (Equation 6.4) and total satisfaction level (SL) averaged from the five weekdays (Equation 6.5). SL has more information than PD because it describes

not only the number of dissatisfied subjects but also the level of their dissatisfaction. The value 180 in the equations is the total number of satisfaction feedback collected during the occupied hours (5 days \times 9 occupied hours per day \times 4 times per hour). N of the equation is the number of subjects (6 for the first scenario and 5 for the second scenario)

$$PD = \frac{1}{180} \sum_{i=1}^{180} \frac{\text{number of dissatisfied subjects at time } i}{N} \times 100\% \quad (6.4)$$

$$SL = \frac{1}{180} \sum_{i=1}^{180} \sum_{j=1}^N \text{-thermal satisfaction level of subjects } j \text{ at time } i \quad (6.5)$$

The performance in energy saving (E) was quantified by daily district heating hot water energy averaged from the five weekdays (Equation 6.6). The value 480 was the number of steps in a simulation run.

$$E = \frac{1}{5} \sum_{i=1}^{480} \textit{ith 15minutes district heating hot water energy} \quad (6.6)$$

The PD , SL , and E of the five case studies obtained from the testing environment was listed in Table 6.10.

6.8.1 Result Analysis of Case 1

Case 1 with setpoint 21 °C during the occupied period resulted in a very high percentage of dissatisfaction and low satisfaction level. A detailed look at the 15 minutes percentage of dissatisfaction is shown in Figure 6.9. On Monday and Friday, there were 100% dissatisfaction existed. More than 60 % of subjects were not satisfied all the time for both Scenario 1 with six subjects

TABLE 6.10: The evaluation results of the four cases with two scenarios

		PD	SL	Daily E
	Case 1 Baseline 21 °C	74.81%	-8.583	12.49 kWh (42,626 BTU)
Six	Case 2 Baseline 23 °C	22.50%	-1.467	15.03 kWh (51,293 BTU)
Subjects	Case 3 Satisfaction	2.00%	-0.200	41.08 kWh (140,160 BTU)
	Case 4 Balanced	22.13%	-1.380	15.00 kWh (51,167 BTU)
	Case 1 Baseline 21 °C	69.78%	-6.617	12.49 kWh (42,626 BTU)
Five	Case 2 Baseline 23 °C	7.11%	-0.4167	15.03 kWh (51,293 BTU)
Subjects	Case 3 Satisfaction	2.44%	-0.210	20.67 kWh (70,515 BTU)
	Case 4 Balanced	6.56%	-0.372	14.72 kWh (50,221 BTU)

and Scenario 2 with five subjects. However, the energy consumption of this case is the lowest among the five cases.



FIGURE 6.9: The comparison of the percentage of dissatisfaction for the six subjects and the five subjects in Case1 (yellow area indicates occupied time)

6.8.2 Result Analysis of Case 2

Although much lower than Case 1, Case 2's percentage of dissatisfaction averaged over time was still more than 20 % for the scenario with six subjects. As shown in Figure 6.10, although the 15 minutes percentage of dissatisfaction was higher at the beginning of the occupied period, it was lower than 20 % most of the time.

A detailed look at the indoor conditions created by this case is shown in Figure 6.11 to further examine the causes of the high percentage of dissatisfaction. Figure 6.11 shows that the indoor air temperature was almost as same as the temperature setpoint, which is 23 °C during the occupied period. However, the mean radiant temperature (MRT) responded slowly to the change of the setpoint. It took almost 8 hours for the MRT to reach its

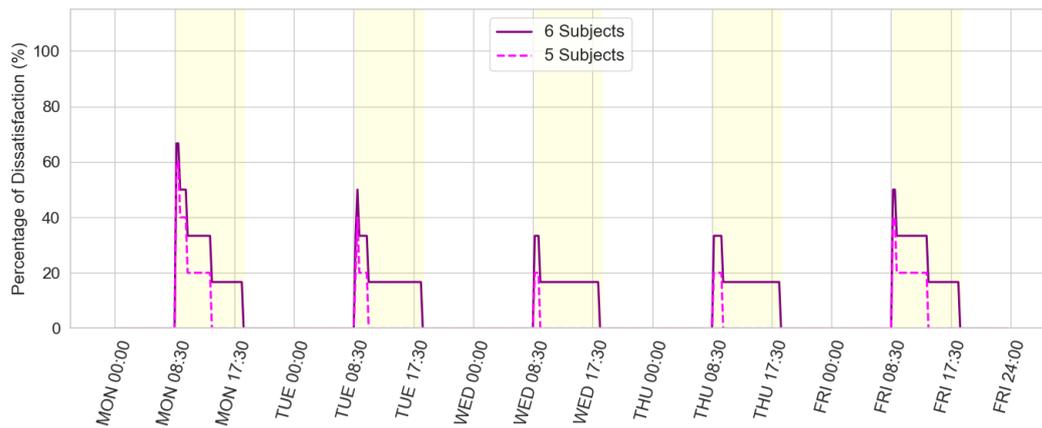


FIGURE 6.10: The comparison of the percentage of dissatisfaction for the six subjects and the five subjects in Case 1 (yellow area indicates occupied time)

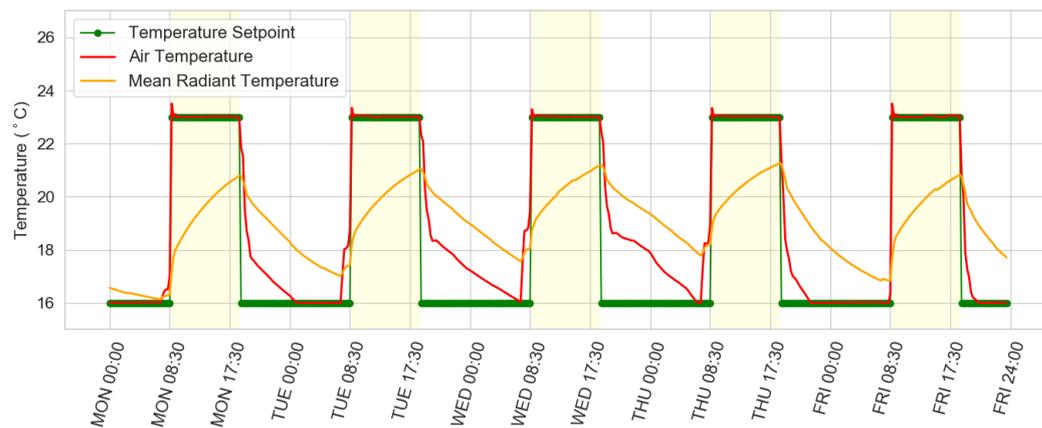


FIGURE 6.11: Indoor conditions in Case 2 (yellow area indicates occupied time)

highest point, 21 °C. The MRT also drops slowly after the setpoint reset to 16 °C. The slow rise of the MRT was the principal cause to the high percentage of dissatisfaction in Case 2. Besides, the energy consumption of Case 2 is 20% higher than that of Case 1 due to the higher setpoint.

6.8.3 Result Analysis of Case 3

The Bio-REAL control system in Case 3 had good achievement in thermal satisfaction. The percentage of dissatisfaction of case 3 is around 2% for both

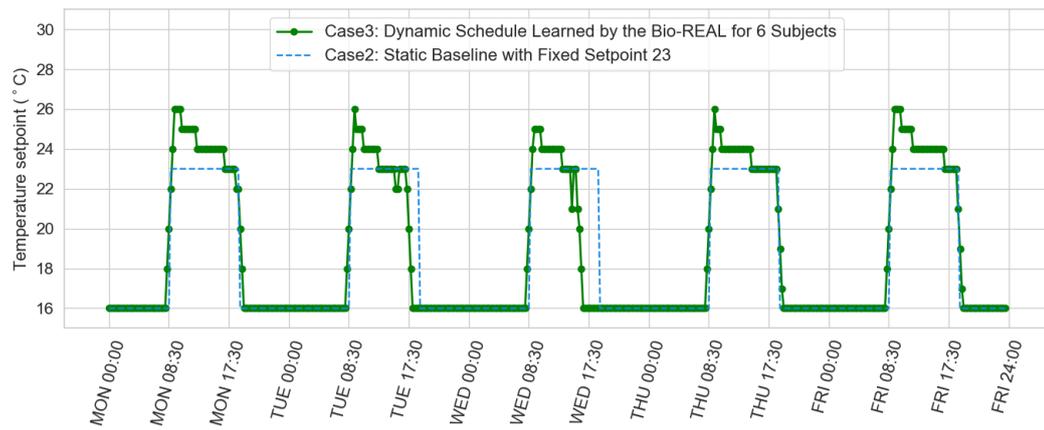
scenarios. As shown in Figure 6.12a and 6.12b, the Bio-REAL control system was able to set the unoccupied temperature setpoint as 16 °C to save energy consumption. It can also raise the temperature setpoint to 26 °C at the beginning of the occupied period to compensate for the poor MRT for thermal satisfaction. Furthermore, the Bio-REAL control system learned the advantage of the thermal mass to save energy consumption and maintain thermal satisfaction. For example, on Wednesday, the control system lowered the temperature setpoint to 16 °C and still maintained a satisfactory indoor condition between 3:00 PM and 5:30 PM, because the Bio-REAL system learned the delayed drop of the MRT.

In Case 3, the thermal dissatisfaction occurred only at the beginning of the occupied time, as shown in Figure 6.12c. One explanation of this is that the highest setpoint variation is 2 °C. The Bio-REAL control system cannot raise the setpoint more than 2 °C each step. Another reason is that the Bio-REAL control was not able to raise the temperature setpoint before the occupied time. Therefore, the occupied time was shifted to 7:00 AM in Case 4 to make pre-heating possible.

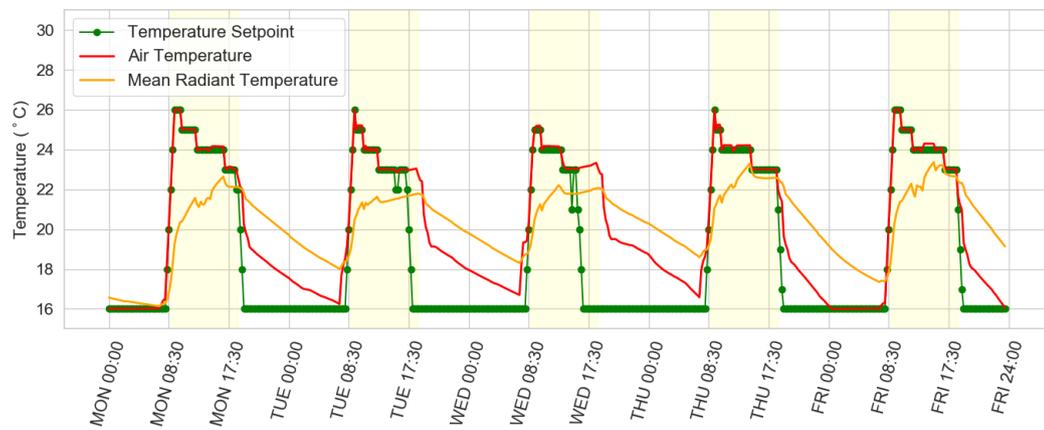
Although with low thermal dissatisfaction, the energy consumption of Case 3 is the highest one among the four cases. For Scenario 1 with six subjects, the energy consumption of Case 3 is 2.7 times more than that of Case 2. It can be concluded that, in this simulation setup, high thermal satisfaction is at the cost of high energy consumption.

However, the energy consumption of Scenario 2 was two times less than that of Scenario 1. Because Subject 5 who has narrow and water comfort zone

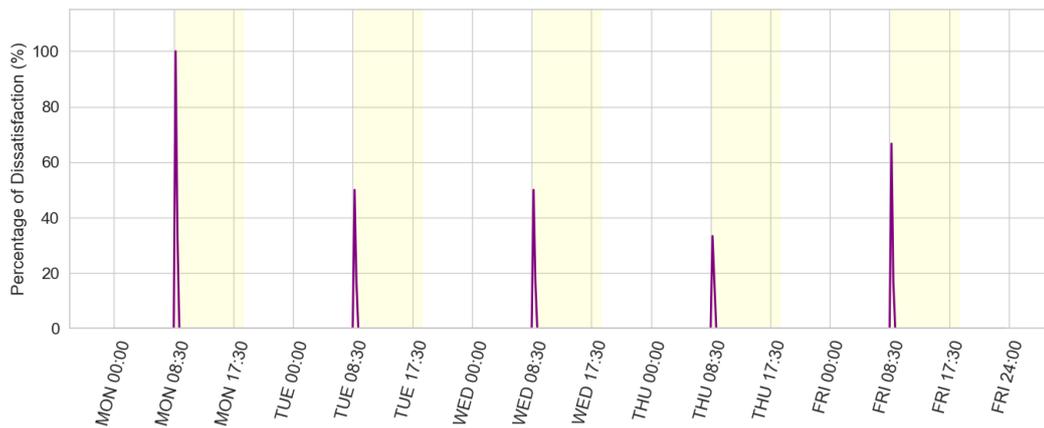
in winter was absent in Scenario 2, the dynamic setpoint in Scenario 2 was generally lower than that of Scenario 1, as shown in Figure 6.13.



(A)



(B)



(C)

FIGURE 6.12: (A) The dynamic temperature setpoint schedule learned and executed by the Bio-REAL control system in Case 3 for the first scenario, (B) indoor conditions, and (C) dissatisfaction percentage created by the learned schedule.(yellow area indicates occupied time)

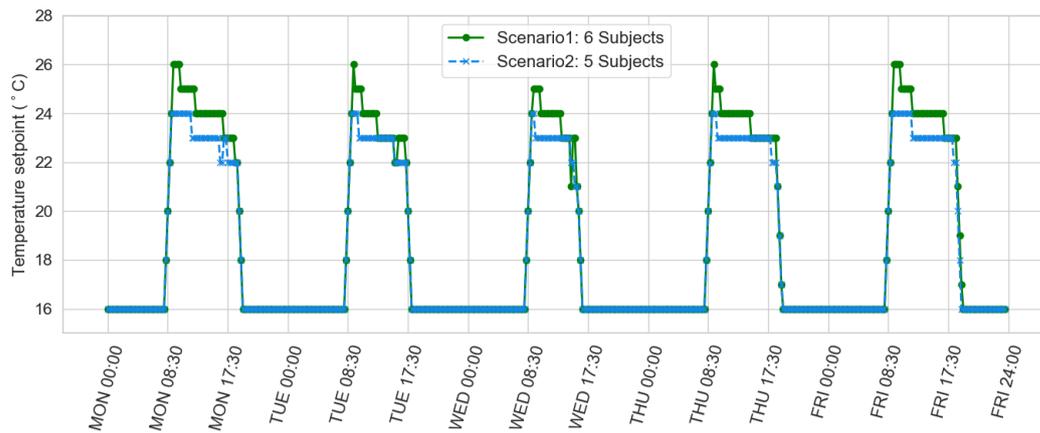


FIGURE 6.13: Comparison of the setpoint schedule in Scenario 1 and Scenario 2

6.8.4 Result Analysis of Case 4

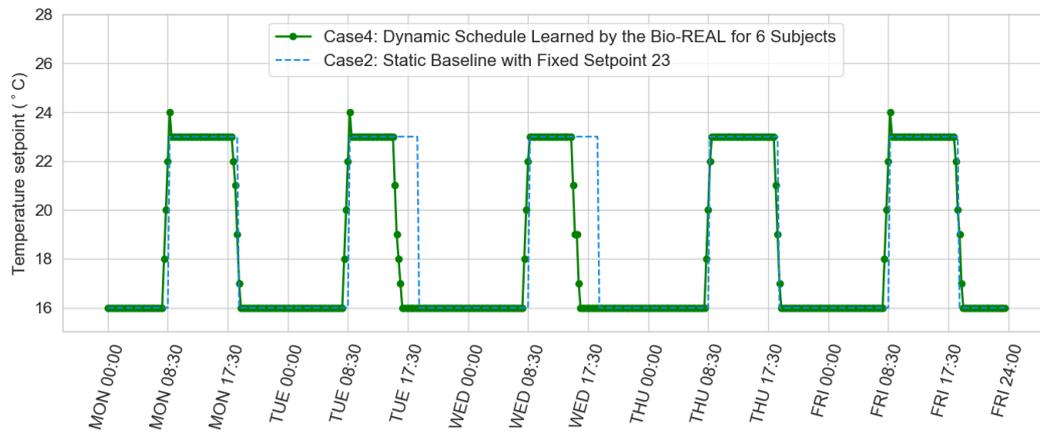
The dynamic schedule in Case 4 achieved delicate improvement on both thermal comfort and energy efficiency comparing to Case 2 baseline, even though the Case 2 with occupied setpoint 23 °C was already the optimum static schedule and has little potential to be improved. As shown in Figure 6.14a, the dynamic schedule had occupied setpoint as 23 °C most of the time similar to Case 2. However, there were two significant differences between Case 2 and Case 4 that contributed to the improvement. First, the Bio-REAL control system started to increase the setpoint before 8:30 AM and raised the setpoint to 24 °C or 23 °C so that the percentage of dissatisfaction can be lower at the beginnings of the occupied time. Moreover, similar to case 3, it lowered the setpoint to 16 °C during the occupied time, so that the energy consumption can drop to 0 (Figure 6.14b) and the thermal dissatisfaction percentage was still below 20% (Figure 6.14c).

The reason that both energy consumption and percentage of dissatisfaction were lower on Wednesday and Tuesday was the higher outdoor air temperature at these days (Figure 6.7)

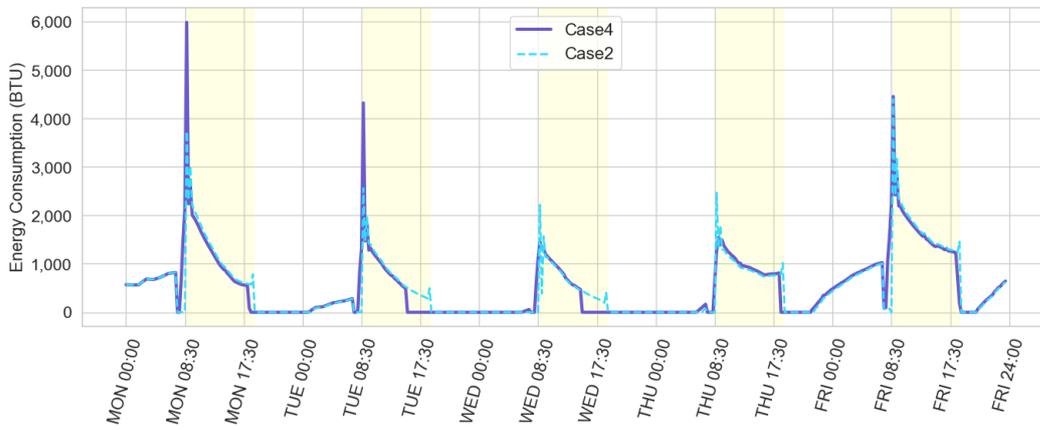
6.9 Conclusion and Discussion

This field and simulation experiment evaluated the performance of the Bio-REAL control system with learning environments simulated reality. The learning environments comprised occupant models developed using data collected from a human subject experiment and an EnergyPlus model of the experimental room with water radiators. The occupant models contained personalized thermal comfort models and skin response models. The personalized thermal comfort models created six individual thermal comfort zones, which were different from the PMV comfort zone. Six personalized Bio-REAL agents was created for the control system.

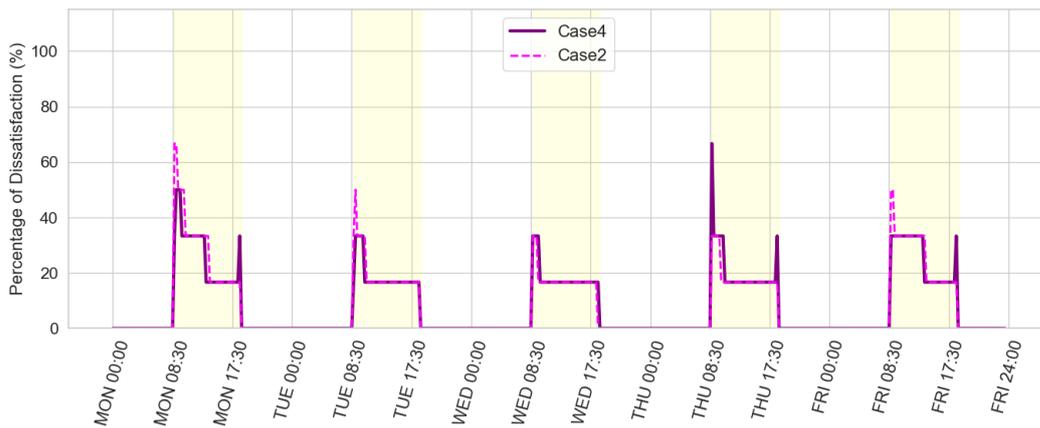
The learning environment for training and testing was different in term of their outdoor weather conditions. The experiment comprised four cases with two different scenarios. Case1 and Case2 were the baselines with static schedules. In Case3 and Case4, the Bio-REAL control system was trained and evaluated. The objective function of the Bio-REAL control system in Case3 weighted more on thermal satisfaction. The weighting on thermal satisfaction and energy efficiency in Case4 was balanced. There were six subjects in Scenario 1 and five subjects in Scenario 2. The simulation run was five weekdays from January 01 to 05. The Bio-REAL control system converged



(A)



(B)



(C)

FIGURE 6.14: (A) The dynamic temperature setpoint schedule learned and executed by the Bio-REAL control system in case 4 for the first scenario, (B) energy consumption, and (C) dissatisfaction percentage achieved by the learned schedule.(yellow area indicates occupied time)

less than 600 episodes (simulation run) for all cases. The time taken for one simulation run on Windows 10 system with random access memory 16GB and processor speed 3.4GHZ was around 1 minute. The total training time was about 10 hours.

The experimental data showed that wrist temperature was strongly correlated with thermal sensation. The testing results showed that the dynamic schedule generated by the Bio-REAL control system had more advantage in improving thermal satisfaction and energy efficiency than static schedules. In the case of balancing comfort and energy for five subjects, the Bio-REAL control can alleviate dissatisfaction by 8.4% and save energy consumption 2.1% compared to the optimum static schedule (23 °C) generated based on personalized thermal comfort models. Moreover, the Bio-REAL control system was able to exert the thermal mass to save energy consumption. Besides, with the change of control objectives and occupancy, the Bio-REAL control system had the flexibility to adjust the control schedules. Although the testing and training environment was different, the good testing results demonstrated the generalizability of the Bio-REAL control system.

However, the Bio-REAL control system in this experiment was not able to pre-heat the room without assistance. The pre-heating in Case4 was because the occupied time was moved up one and a half hours. This issue could be solved if historical information and weather prediction are in the state of the learning environments. Furthermore, the action designed in this experiment only allowed setpoint to be varied two at maximum. The performance of the Bio-REAL control system could be better if there is more flexibility in setpoint

variation.

Nevertheless, this field and simulation experiment not only demonstrated the advantage of the Bio-REAL control system but also provided guidance for applying Deep Reinforcement Learning in building control. The experiment showed a way of making multiple agents work together in practice by introducing a negotiator. Moreover, this experiment proved that reward functions significantly influenced the decision made by Bio-REAL agents. Reward functions should be carefully designed to guide the Bio-REAL agents. Straightforward reward signals can improve the performance of the agents.

Chapter 7

Field Experiments at NUS: Smart Ceiling Fans for Cooling

7.1 Objective

The previous experiments demonstrated the performance of the Bio-REAL control system by simulation. The objective of the field experiments was to present its real-world performance in a tropical climate with convective cooling. The field experiments comprised a training experiment and an evaluation experiment. The training experiment collected data from occupants and buildings. 14 Bio-REAL control agents were trained using the experimental data. The learning procedures were carefully designed to guarantee learning stability. The evaluation experiment assessed the performance of the Bio-REAL control system with well-trained Bio-REAL agents. The achievement in thermal comfort improvement and the saving on energy consumption of the Bio-REAL control system were quantified and compared to the baseline control.

7.2 Experiment Setup

7.2.1 The Net Zero Energy Building at NUS

The training and evaluation experiments were undertaken From December 17th, 2018 to January 11th, 2019 in the Net Zero Energy building (Figure 7.1) at the National University of Singapore (NUS). The building has six floors and houses a mix of research laboratories, design studios, and teaching and common learning spaces. In a tropical country like Singapore, mechanical cooling typically accounts for as much as 60 % of building energy consumption. An innovative hybrid cooling system was designed for the Net Zero Energy building to save energy. The hybrid cooling system incorporates typical variable air volume (VAV) systems and ceiling fans (SDE, 2018). With boosted airspeed from ceiling fans, a higher temperature setpoint of the VAV systems is possible without sacrificing thermal comfort. The higher temperature setpoints in Singapore can reduce building energy consumption.



FIGURE 7.1: The Net Zero Energy building at NUS (Image from School of Design and Environment)

7.2.2 Experiment Room

The experiment room (10.6m × 8.0m × 4.2m) is on the west side of the fourth floor of the Net Zero Energy building (the red box of Figure 7.2). Its eastern, northern, and western wall are all window walls. It has 16 LED ceiling light, four smart ceiling fans, and a VAV system, as shown in Figure 7.3.



FIGURE 7.2: The location of the experiment room at NUS

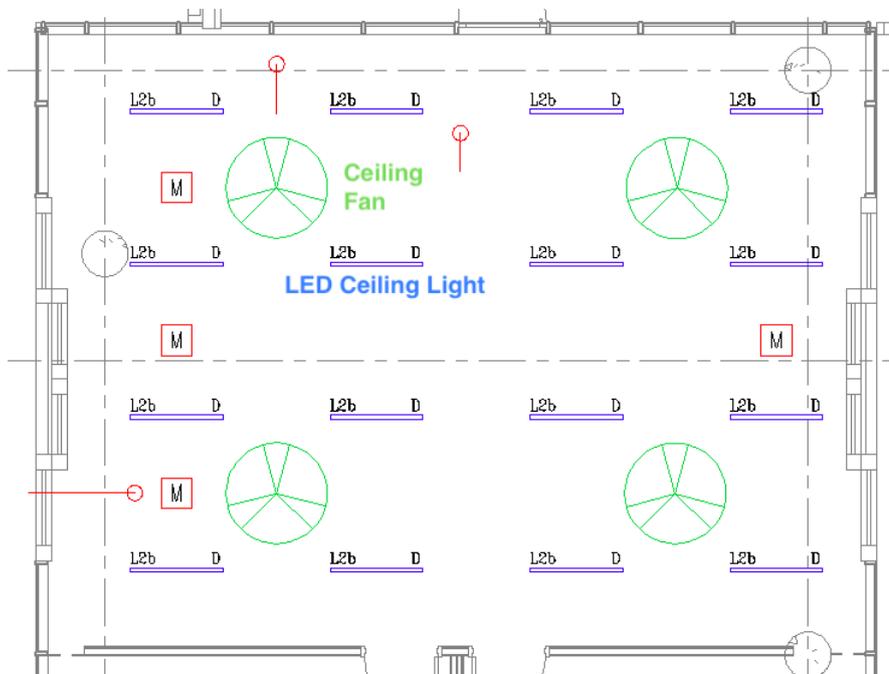


FIGURE 7.3: The experiment room at NUS

The layout of the testing room is shown in Figure 7.4. There were four tables under each fan. The horizontal distances from the four tables to the center of the ceiling fan are all 0.6m.

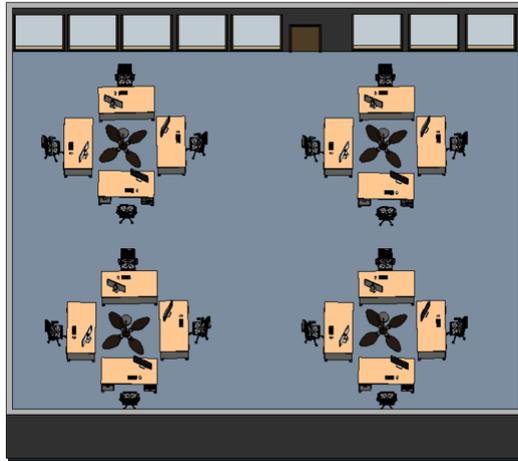


FIGURE 7.4: The table layout of the experiment room at NUS

7.2.3 Facilities of the Experiment Room

The VAV system is all outdoor air system that has no return vent. There are four supply vents installed on the ceiling of the experiment room. Figure 7.5 shows their locations. The supply vents are all 450 x 200 mm double deflection grille.

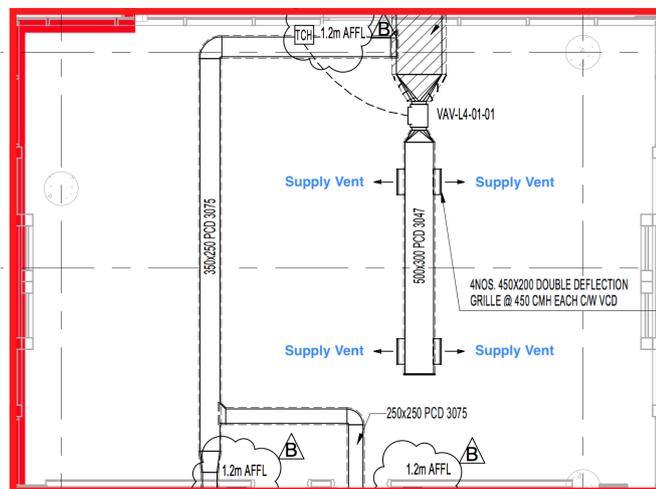


FIGURE 7.5: The VAV system of the experiment room at NUS

The VAV system's cooling capacity was examined to design the Bio-REAL control system. Since Singapore weather is very stable, without any mechanical cooling, the indoor air temperature is around 28 °C every day. Figure 7.6 shows the time taking for the VAV system to cool the air temperature from 28 °C to 25 °C. If the temperature setpoint was 27 °C, it took about 37 minutes to lower the indoor air temperature from 28 °C to 27 °C. If set the setpoint as 25 °C, the time consumed for the air temperature drops from 27 °C to 26 °C is 25 minutes and 50 minutes from 26 °C to 25.5 °C. This cooling capacity revealed a long delay for the indoor environment responding to the change of VAV setpoint, which is a big challenge for the Bio-REAL agents to learn.

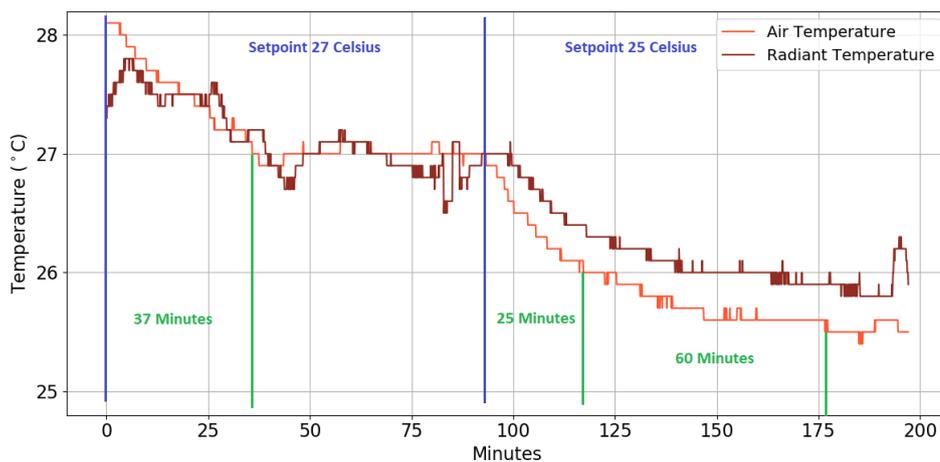


FIGURE 7.6: The capacity of the VAV system in the experiment room at NUS

Although ceiling fan is a relatively simple cooling system, it creates an asymmetric and dynamic indoor environment. The ceiling fan of the experiment room supports seven adjustable fan speeds. The air velocities at four different vertical locations (0.1m, 0.6m, table level, and 1.1m) and four

different horizontal locations (centers of the four tables) were measured under three different fan speeds (Speed2, Speed4, and Speed6) to understand the unoccupied spatial patterns created by the ceiling fan in the experiment room (Figure 7.7).

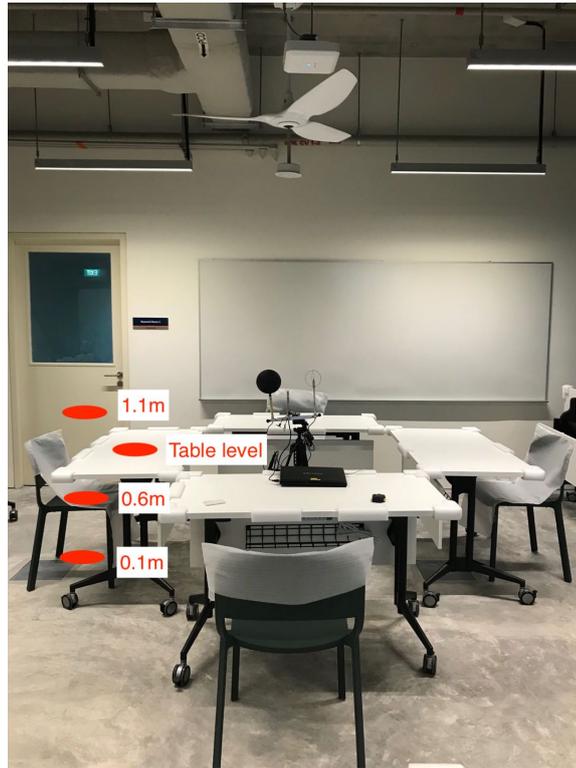


FIGURE 7.7: Four vertical locations for air velocity measurement

As shown in Table 7.1, the increases in fan speeds led to the rises in air velocity for all four locations. The air velocity was quite different at the four vertical locations. The air velocity at 0.1m was the highest due to both the rebounding air from the floor and Venturi effect. At table level, the air velocity was also very high because of the rebounding air from the table. The air velocity at 1.1m was the lowest one. Besides, since the air velocity at the center of the four tables was similar to each other, the occupants sitting at the four horizontal locations can experience similar air-movement.

TABLE 7.1: Air velocities at different vertical locations and fan speeds

	Fan Speed 2	Fan Speed 4	Fan Speed 6
1.1 m(Head level)	0.028 m/s	0.054 m/s	0.167 m/s
Table level	0.379 m/s	0.693 m/s	0.998 m/s
0.6 m (Waist level)	0.044 m/s	0.135 m/s	0.170 m/s
0.1 m (Foot level)	0.693 m/s	1.235 m/s	1.435 m/s

Figure 7.8 shows the unoccupied temporal pattern of the air velocity in the experiment room. For each fan speed, the air velocity had a cyclical fluctuation due to the turbulence. The higher the fan speed, the higher the turbulence intensity (calculated as $\frac{\sqrt{v_i - v_{mean}}}{v_{mean}}$). The figure also indicates that the airflow at the table level and 0.1m had higher turbulence intensities than that at other locations.

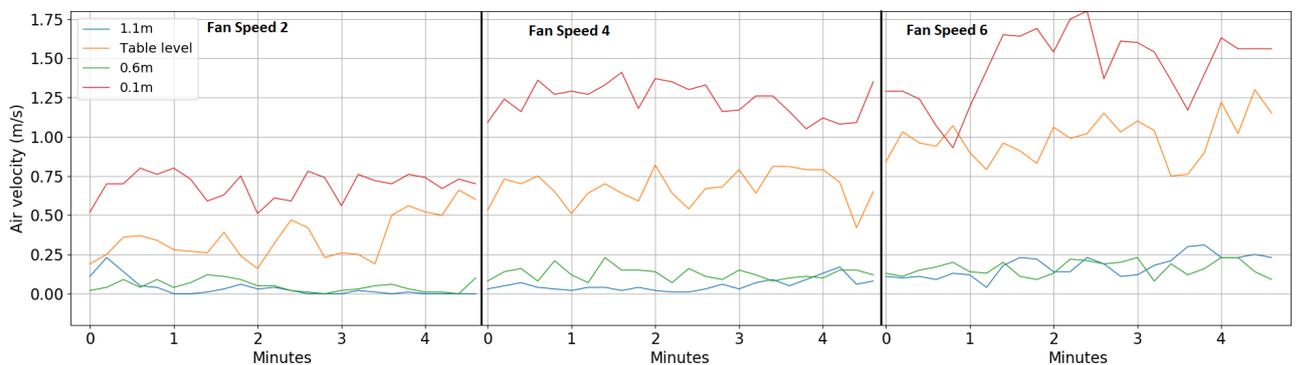


FIGURE 7.8: Unoccupied temporal pattern of the air velocity in the experiment room

7.2.4 Experiment Subjects

Fourteen healthy graduate students and staffs at NUS participated in the experiment. They are seven females and seven males. The statistics of their demographic data are listed in Table 7.2. Their preferences to Singapore weather and ceiling fans were also surveyed before the experiments. Six of the experiment subjects like the Singapore weather, while eight of them dislike it because it's too hot or humid. Almost all of them like the ceiling fan for cooling. Moreover, the time of their stay in Singapore is distinct. The shortest time is one month, while the longest one is more than nine years.

During the experiment, the experiment subjects were required to always sit at the same location with long pants, covered shoes, and short-sleeve T-shirts, which are the typical office wear (around 0.6 clo) in Singapore, as shown in Figure 7.9. They worked on "office-type" activities, such as reading, typing, and web surfing so that their metabolic rate is 1.0-1.1 met. They always put their hands on the table while working. Because of their wear, their uncovered areas are hands, wrist, forearm, and head. Also, since the airflow at table level was stronger than that at head level, the subjects' hands, wrist, and forearm would be more sensitive to the cooling effect of the fans.

TABLE 7.2: Demographic information of the experiment subjects at NUS

	Age	BMI
Mean	28.64	21.97
Standard Deviation	3.98	3.04



FIGURE 7.9: Experiment subjects at NUS

7.2.5 Measurement and Equipment

The variables measured in both experiments can be grouped into three categories: environment, biological, and subjective responses. The environment responses measured included air temperature, relative humidity, radiant temperature, and air velocity. A Delta Ohm thermal microclimate analysis tool was used to measure the environmental conditions. The microclimate tool comprised HP3217R combined relative humidity and air temperature probe and TP3275 globe temperature probe. The air velocity was measured by a omnidirectional hotwire probe (an accuracy of ± 0.001 m/s). The Microsoft Band 2TM measured the bio-responses. Subjects were reminded to report their subjective responses every 10 minutes through a web survey, as shown in Figure 7.10. The survey had 14 questions, including the overall and local thermal sensation, thermal comfort, thermal preference, air movement acceptability, and air movement preference. The variables measured and the measurement intervals and equipment were summarized in Table 7.3.

What is your overall thermal sensation?

Cold
 Cool
 Slightly Cool
 Neutral
 Slightly Warm
 Warm
 Hot

What is your Local thermal sensation?

Head:

Cold
 Cool
 Slightly Cool
 Neutral
 Slightly Warm
 Warm
 Hot

Chest

Cold
 Cool
 Slightly Cool
 Neutral
 Slightly Warm
 Warm
 Hot

Upper Arm

Cold
 Cool
 Slightly Cool
 Neutral
 Slightly Warm
 Warm
 Hot

Forearm:

Cold
 Cool
 Slightly Cool
 Neutral
 Slightly Warm
 Warm
 Hot

Wrist:

Cold
 Cool
 Slightly Cool
 Neutral
 Slightly Warm
 Warm
 Hot

Hand:

Cold
 Cool
 Slightly Cool
 Neutral
 Slightly Warm
 Warm
 Hot

Knee:

Cold
 Cool
 Slightly Cool
 Neutral
 Slightly Warm
 Warm
 Hot

Calf:

Cold
 Cool
 Slightly Cool
 Neutral
 Slightly Warm
 Warm
 Hot

Foot:

Cold
 Cool
 Slightly Cool
 Neutral
 Slightly Warm
 Warm
 Hot

Are you comfortable with the thermal environment?

Comfortable
 Slightly Uncomfortable
 Uncomfortable
 Very Uncomfortable

How do you want to change the thermal environment?

Warmer
 No Change
 Cooler

Is the air movement in your workspace acceptable?

Acceptable
 Slightly Unacceptable
 Unacceptable
 Very Unacceptable

How do you want to change the air movement?

Want Less
 No Change
 Want More

FIGURE 7.10: Thermal comfort questionnaire for the experiment at NUS

7.3 The Bio-REAL Control System

In this field experiment, the Bio-REAL control system directly interacted the real occupants and building, learning from the experiences in the real world, as shown in Figure 7.11.

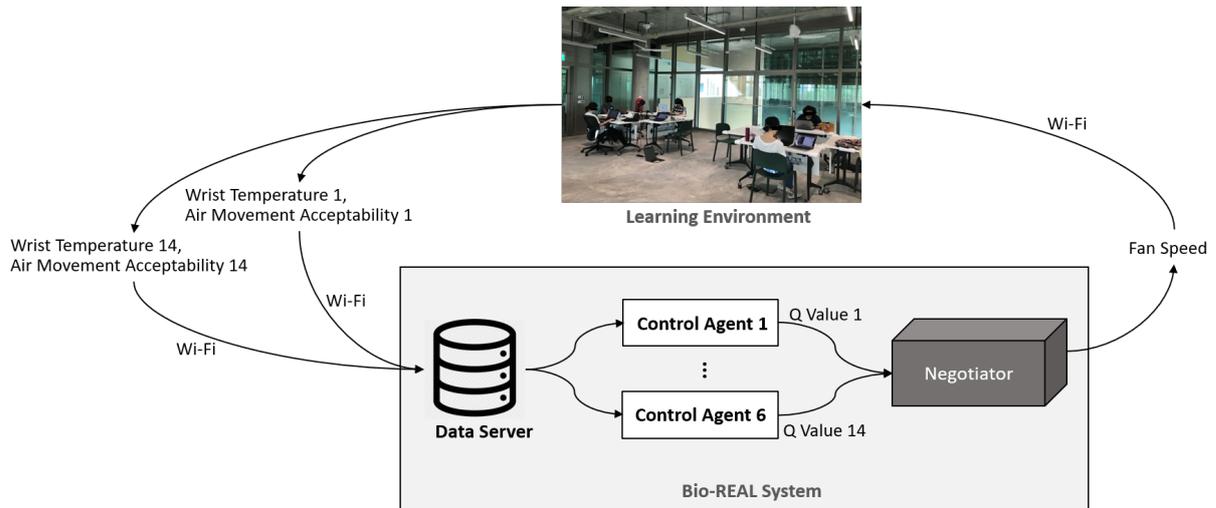


FIGURE 7.11: The Bio-REAL control system at NUS learning from the real experiences

During the interaction, the Bio-REAL control system decides a control action and sends the control decision to the cooling system through Wi-Fi. A data server collects the state and reward from the learning environment and dispatches them to the Bio-REAL agents for them to learn. .

7.3.1 Bio-REAL agents

Fourteen Bio-REAL agents were built, one for each subject. The action, state, and reward for an agent are described below.

Action. The analysis in Section 7.2.3 showed that there was a long delay for the indoor environment responding to the change of the VAV setpoints. Therefore, setpoint adjustment was not considered as the control action of

the Bio-REAL control system. The air temperature was controlled as 27 °C during the the two experiments because it is the expected room temperature for the real operation. The action in this experiment is adjusting fan speeds. The actions set is shown in Equation 7.1.

$$A \equiv \text{Fan Speed} \in \{0, 2, 4, 6\} \quad (7.1)$$

This thesis didn't incorporate the speed1, speed3, speed5, and speed7 into the action set because a smaller action-space can improve the learning efficiency. The Bio-REAL control system executes an action every 10 minutes. The responsibility of the Bio-REAL control system in this experiment is to learn the optimum dynamic fan speed schedules.

State. The options for the state can be indoor and outdoor environmental conditions, as well as occupants' biological responses. Because Singapore weather is very stable and the air temperature was maintained at 27 °C during the experiments, except for air velocity, the indoor and outdoor environment has little variation. Therefore, the state was only the biological response, wrist temperature. More specifically, the state was the latest wrist temperature 10 minutes after actuating an action. This one-dimensional state-space can also improve learning efficiency.

Reward. Since ceiling fans are not energy-intensive, the control objective was only thermal comfort optimization. Among the 14 subjective survey

TABLE 7.4: air movement acceptability reward/penalty designed for the Bio-REAL agent

	Acceptable	Slightly Unacceptable	Unacceptable	Very Unacceptable
Reward	0	-1	-2	-3

questions, air movement acceptability best represents the condition of minds to the air movement. Moreover, the air movement acceptability can represent thermal comfort feedback because the air movement is the only variable for the indoor condition, other conditions, such as air temperature, radiant temperature, and relative humidity are very stable. Hence, the reward was the quantified air movement acceptability 10 minutes after actuating an action. Table 7.4 shows the mapping from the air movement acceptability to the reward.

7.3.2 Learning Process

One challenge of implementing deep reinforcement learning (DRL) in real-world is overcoming the issues caused by limited learning samples. Especially at the beginning of the learning process, there are not enough data samples to learn so that the performance of the DRL agent can be quite poor. To overcome this challenge, the thesis divided the learning process of the Bio-REAL control system into two phases: training and testing/evaluation

phase. The two experiments were conducted respectively for training and evaluation. There were three main modifications made to the training process described in Section 3.2 for the real-world implementation.

First, the training process was split into data collection, Q-network updating, and agent examination to ensure learning stability, as shown in Figure 7.12. During the day, the agents collect data from the experiment subjects, process the raw data into interaction experiences (state, action, next state, reward), then save the data into agents' memory bank. At night, the agents replay the experiences and update their Q-networks. Fourteen agents were trained individually by following the day and night process.

Second, the action sequences during the training phase were designed by prior knowledge. The algorithm in section 3.2 encourages the agents to explore randomly at the beginning of the learning so that the agents can visit different state-action pairs enough times. Sufficient exploration is the key for the agents to find the optimum state-action. However, random exploration is time-consuming. To save training time, the author designed the sequences of the actions to ensure the agent visit different state-action pairs at least once in a shorter time, as shown in Table 7.5.

Last, the agents with updated Q-network won't play a part in building controls until they are examined. The examination was based on the Q-value predicted by the Q-network. The Q-value represented the expected air movement acceptability of fan speeds at a specific wrist temperature. The higher Q-value, the better. The best fan speed at a given wrist temperature for each subject is the one that has the highest Q-value. Therefore, a well-trained

agent should meet the conditions that the Q-value of higher fan speeds is higher at higher wist temperature, and vice versa. There could be some agents won't meet the conditions even with adequate training because the occupants they represent have irregular preferences. The Bio-REAL control agent only plays its role in the control system if it is well-trained and if the occupant it represents is in the room, as indicated by Figure 7.12.

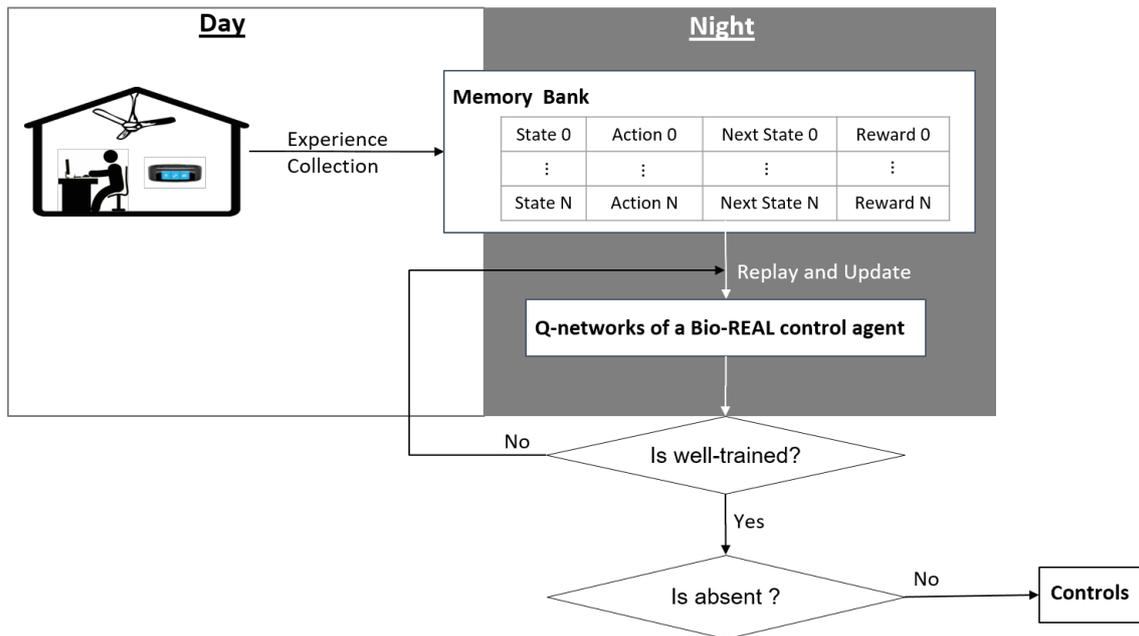


FIGURE 7.12: The training process in practice

7.4 Training Experiment

The objective of the training experiment was to collect data samples to train the 14 Bio-REAL gents.

7.4.1 Experiment Process

The experiment had 13 sessions. One session lasted around 2 hours to avoid the effects caused by long-time exposure. There were 4-6 subjects participated in each session based on their own time schedules. Four different action sequences were designed to investigate the possible differences in subjects' responses, physiological or psychologically, caused by the sequence differences. Table 7.5 shows the action sequences and the duration of each action. Each sequence was repeated several times so that each of the 14 subjects can expose to all four sequences. Subject8 only participated in three sequences due to his/her schedule. Before each session, subjects had 5 minutes to wear wristbands and adapt to the environment in the experiment room. Each session had a 10-minute half-time break. All sessions started with Fan speed 0. Each above-zero fan speed lasted 20-30 minutes to investigate the possible effects caused by longer-time exposure to the air-movement of a specific fan speed. During the sessions, the data mentioned in Table 7.3 were collected, processed, and saved in the memory bank of each agent.

7.4.2 Experimental Conditions

During the 13 sessions, the air temperature was maintained around 27 °C, varied slightly between 27.2 °C and 26.8 °C. The radiant temperature was the same as or 0.1 °C higher than the air temperature. The relative humidity fluctuated between 63.5 % and 74.8%, the higher values of which were due

TABLE 7.5: Action sequences of the experience collection experiment at NUS

Sequence	Fan Speed (Duration)				
1	0 (10 min)	2 (30 min)	0 (20 min)	4 (30 min)	0 (10 min)
2	0 (10 min)	6 (30 min)	4 (30 min)	2 (30 min)	0 (10 min)
3	0 (10 min)	2 (20 min)	4 (20 min)	6 (20 min)	0 (20 min)
4	0 (20 min)	6 (20 min)	2 (20 min)	6 (10 min/20 min)	4 (10 min/20 min)

to the raining days. The statistics of the experimental conditions are summarized in Table 7.6. Thus, the major contributor to the change in biological and subjective responses was the variation in air velocity caused by fan speed adjustment.

7.4.3 Experiment Data Analysis

There was 576 subjective feedback collected from the 14 experiment subjects, each of whom had around 40 feedback, as shown in Table 7.7. The number of wrist temperature samples collected from each subject was about 1,000. With the fan speed varied among speed0, speed2, speed4, and speed6 at temperature 27 °C, the 14 experiment subjects voted mostly "acceptable" (0) and "slightly unacceptable"(-1) to the air movement. The "acceptable" voting was more than the "slightly unacceptable" one for all subjects excepting for subject9 and subject10. Subject5 always voted "acceptable" regardless of the fan

TABLE 7.6: Experimental condition during the experience collection experiment

	Air Temperature	Radiant Temperature	Relative Humidity
Max	27.2 °C	27.3	74.8 %
Min	26.8 °C	26.8	63.6 %
Mean	26.9 °C	27.0	70.1 %

speed.

The wrist temperature of the experiment subjects was very responsive to the fan speeds. As shown in Figure 7.13, 7.14, 7.15, and 7.16, it immediately decreased when the fan was on and increased when the fan was off. Long-time exposure to a fan speed resulted in continuous drops in wrist temperature and sometime increased unacceptability in air movement. Moreover, the variation of wrist temperature depends not only on the fan speed but also on previous wrist temperature. For example, in Sequence1 of subject3 in Figure 7.13b, the wrist temperature dropped at fan speed2 because the initial temperature was high, while, in Sequence2, the wrist temperature raised at speed2 because the previous wrist temperature was low. Besides, if the initial wrist temperature is the same, the higher the fan speed, the larger the decrease rate of the wrist temperature.

However, the causes of acceptability to the air-movement were distinct. Some accepted high air-movement and disliked low air-movement, while others favored high air-movement. The wrist temperature was correlated

TABLE 7.7: air movement acceptability and wrist temperature collected from the experiment

Subject	air movement acceptability					Wrist Temperature
	Total	0	-1	-2	-3	
1	41	31	10	0	0	987
2	40	28	12	0	0	909
3	39	33	5	1	0	1069
4	42	28	14	0	0	981
5	44	44	0	0	0	1055
6	41	26	14	1	0	966
7	43	36	7	0	0	1050
8	32	22	10	0	0	723
9	40	19	21	0	0	896
10	44	21	22	1	0	1049
11	41	23	8	4	6	927
12	44	27	15	2	0	982
13	44	28	16	0	0	1003
14	41	24	14	3	0	996

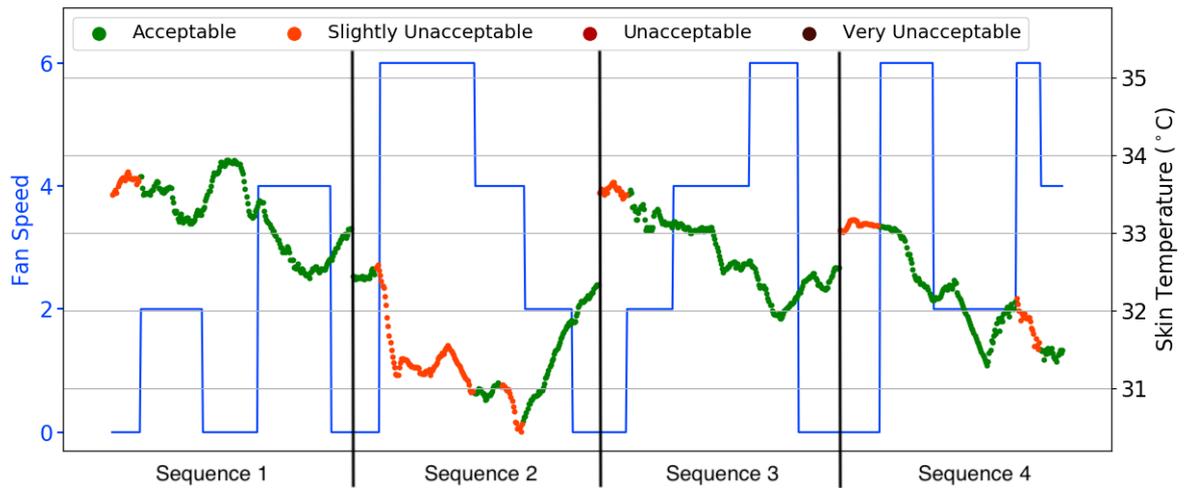
with air movement acceptability, but the correlations were different for different subjects. Based on the correlation differences, the 14 subjects can be grouped into five types.

The first type is neutral-preferred subjects, such as subject2, subject3, and subject4. They preferred air movement neither too high nor too low. As shown in Figure 7.13, most of the time, they felt air movement "slight unacceptable" or "unacceptable" when their wrist temperature was high or remarkably low. There were occasional situations that the neutral-preferred subjects were acceptable even when their wrist temperature was low, such as sequence2 of subject3. These could be caused by the reasons that can not be captured by the sensor data, like unknown psychological factors.

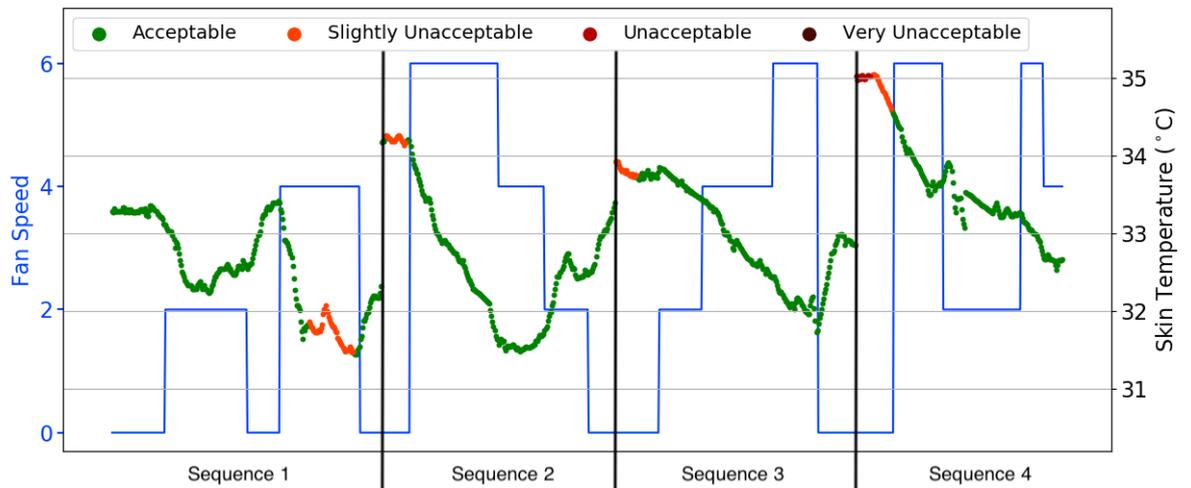
The second type is warm-preferred subjects, who preferred low fan speeds or staying with a higher wrist temperature, such as subject9 and subject10. As shown in Figure 7.14a, subject10 was always not acceptable with fan speed greater than 2. Although subject9 disliked fan speed0 or high wrist temperature in Sequence1, he/she disliked staying with a low skin temperature for the rest three sequences, as shown in Figure 7.14b. Therefore, there was a higher probability that subject 9 was a warm-preferred subject.

The third type is cool-preferred subjects, such as subject6 and subject11, as shown in Figure 7.15. They preferred high fan speeds or staying with a lower wrist temperature. For example, subject6 and subject11 were not pleasant as long as the fan speed was 0. Moreover, not every cool-preferred subject dislikes the low air movement the same. Subject11 especially disliked low fan speed since he/she reported "very unacceptable" for fan speed0.

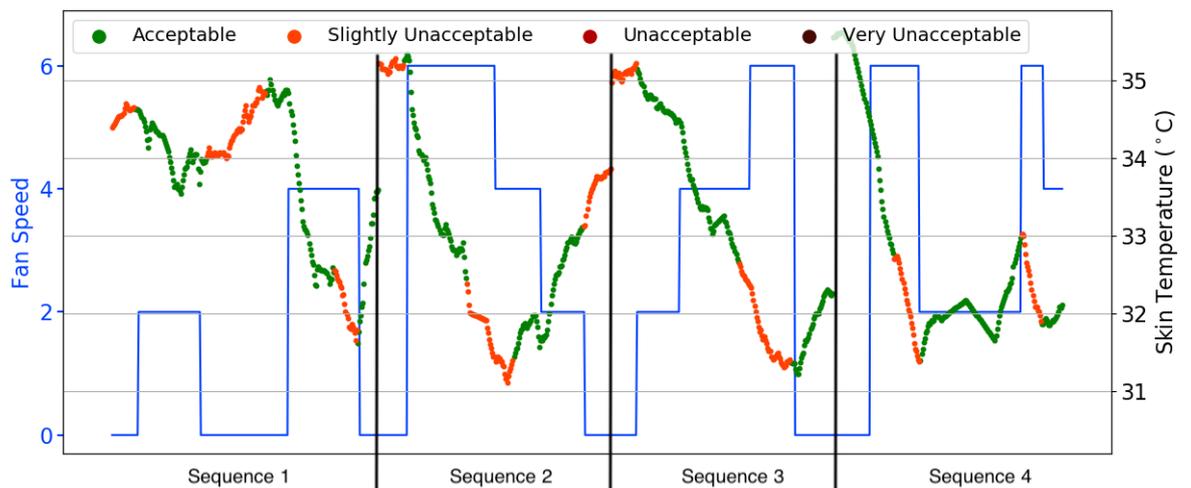
The fourth type is no preference subjects, who always feel acceptable regardless of the fan speeds and wrist temperature. As shown in Figure 7.16, although the wrist temperature of subject5 varied between 33 °C and 35 °C with the change of fan speeds, the variation degree was not as high as other subjects. Subject5 voted acceptable all the time, which could be explained by his/her excellent thermoregulation mechanism.



(A) Subject 2

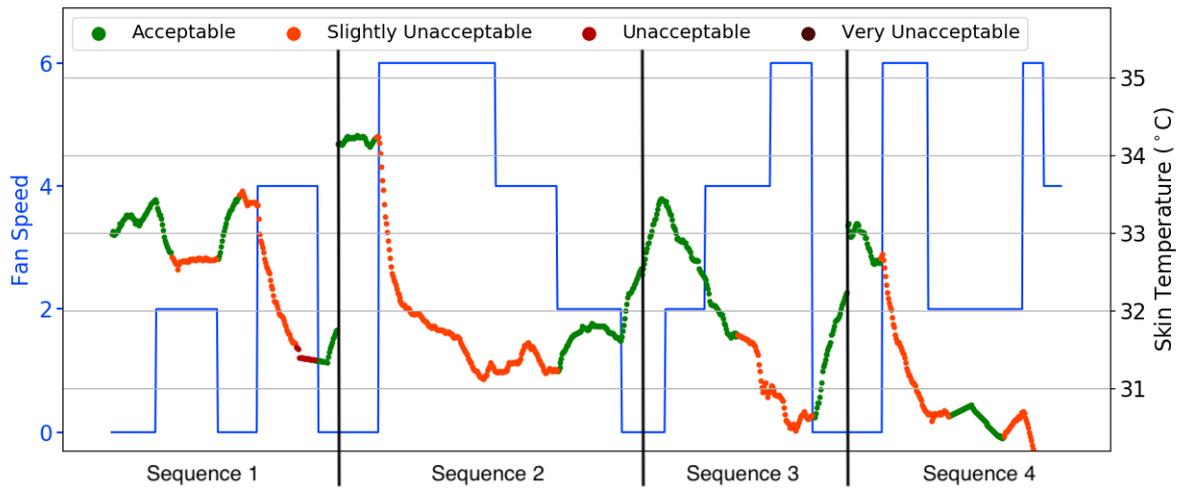


(B) Subject 3

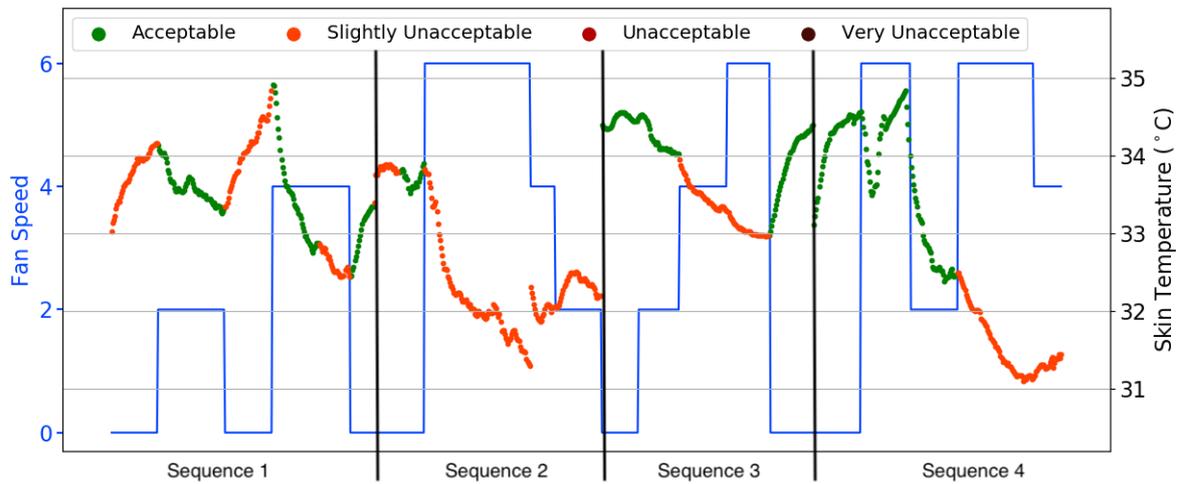


(C) Subject 4

FIGURE 7.13: air movement acceptability and wrist temperature variation of neutral-preferred subjects in the four action sequences. Dotted line = wrist temperature at 27 °C and variations in fan speeds. Colors = acceptance to the air movement.

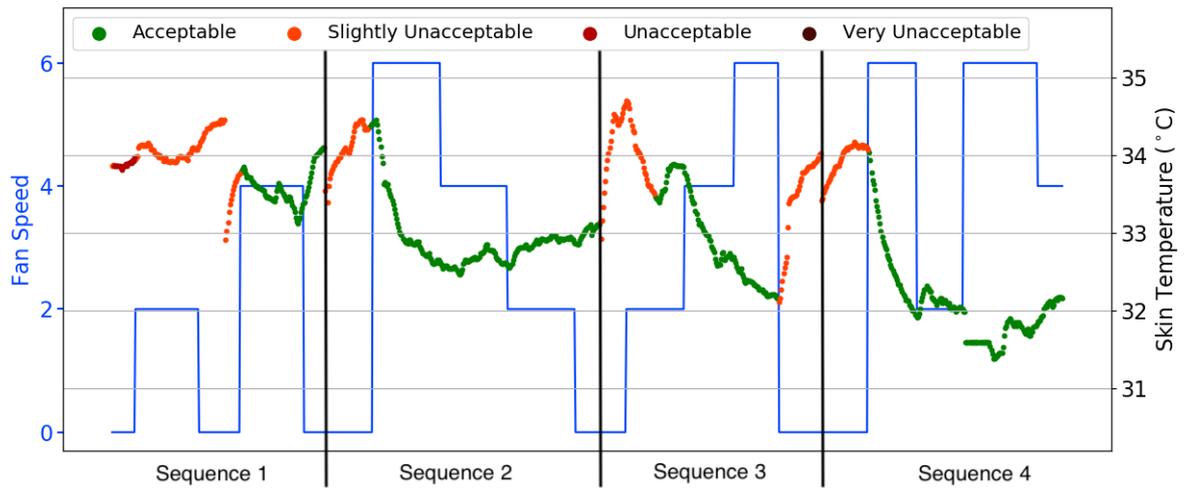


(A) Subject 10

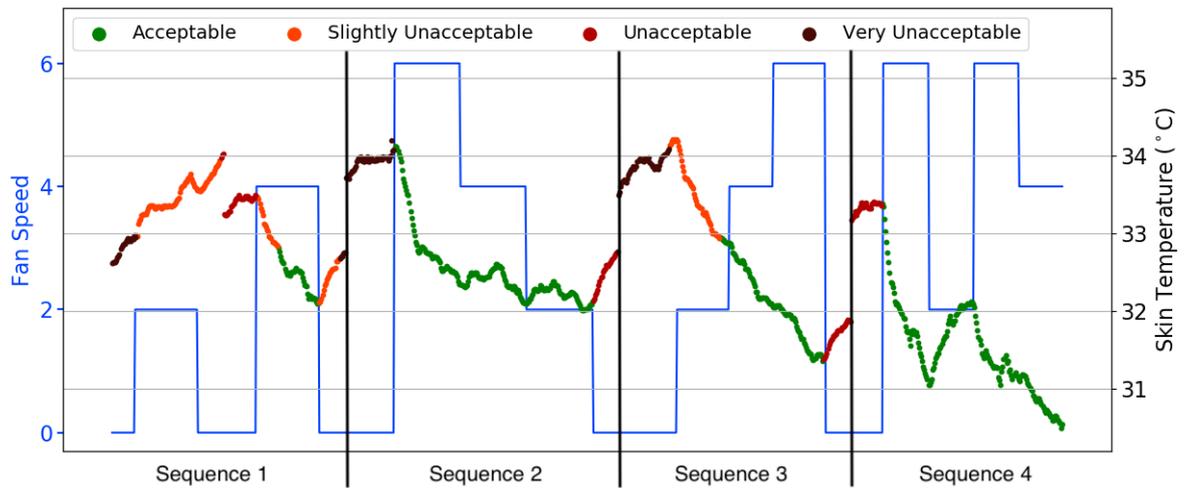


(B) Subject 9

FIGURE 7.14: air movement acceptability and wrist temperature variation of warm-preferred subjects in the four action sequences.



(A) Subject 6



(B) Subject 11

FIGURE 7.15: air movement acceptability and wrist temperature variation of cool-preferred subjects in the four action sequences.

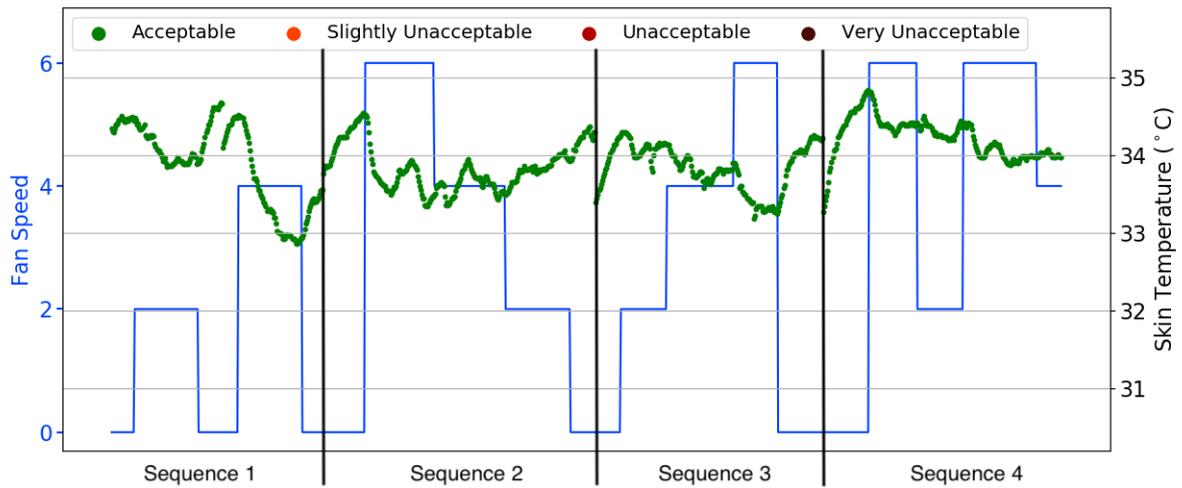
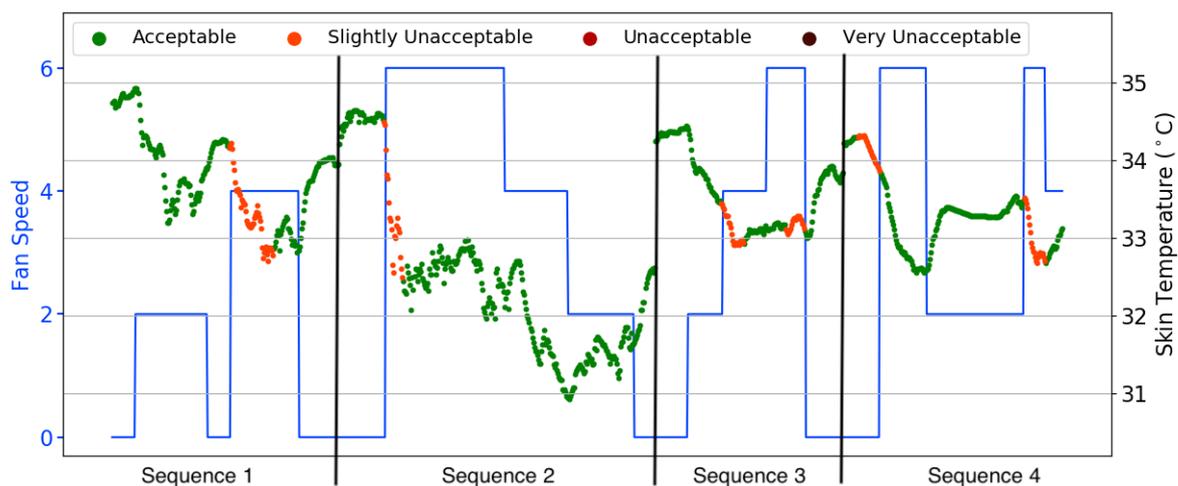


FIGURE 7.16: air movement acceptability and wrist temperature variation of no preference subjects in the four action sequences.

The last type is irregular-preference subjects, whose preferences can not be explained by their wrist temperature or the fan speed. As shown in Figure 7.17a, subject7 were not acceptable when his/her wrist temperature neither too high nor too low, which seems not reasonable. Moreover, the subjects whose wrist temperature varied irregularly was also grouped in the last type. As shown in Figure 7.17b, it is odd that wrist temperature of subject8 raised with the increased fan speed, although the relationship of wrist temperature and air movement acceptability was sound.



(A) Subject 7



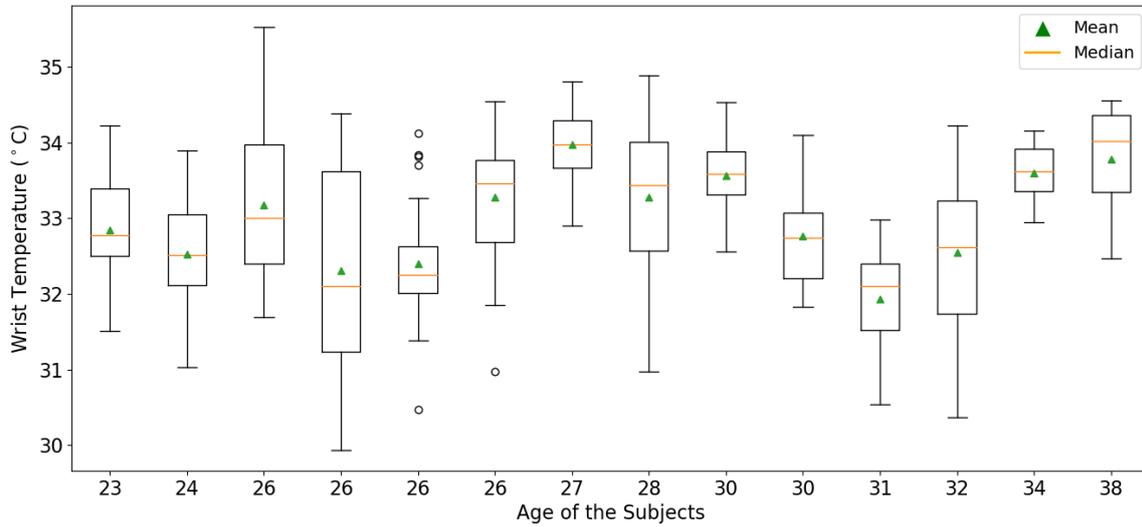
(B) Subject 8

FIGURE 7.17: air movement acceptability and wrist temperature variation of irregular-preference subjects (e.g., subject 5) in the four action sequences.

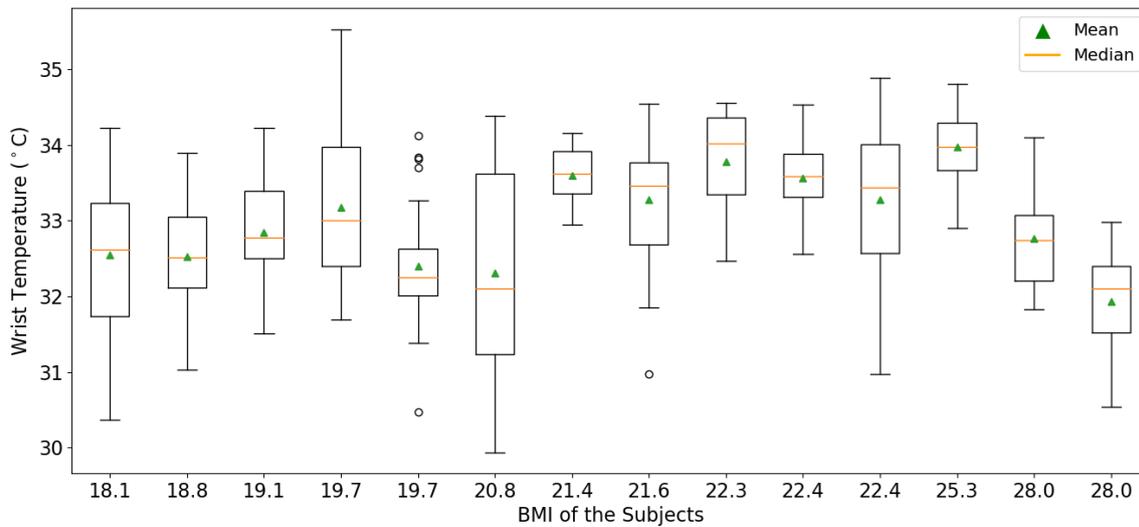
In addition to the relationships among fan speed, wrist temperature, and air movement acceptability, the thesis also investigated the correlations between wrist temperature and age/BMI. Figure 7.18 shows that wrist temperature distribution of the 14 subjects voted acceptable to air movement. As shown in the box-plots, the interquartile range was quite different for different subjects, indicating different wrist temperature variation when voting

acceptable. The mean, median, max, min, and quarterlies of the wrist temperature didn't have any observed correlation with both the age and BMI.

The subjects at age 27 and BMI 25.3 had the highest mean wrist temperature.



(A)



(B)

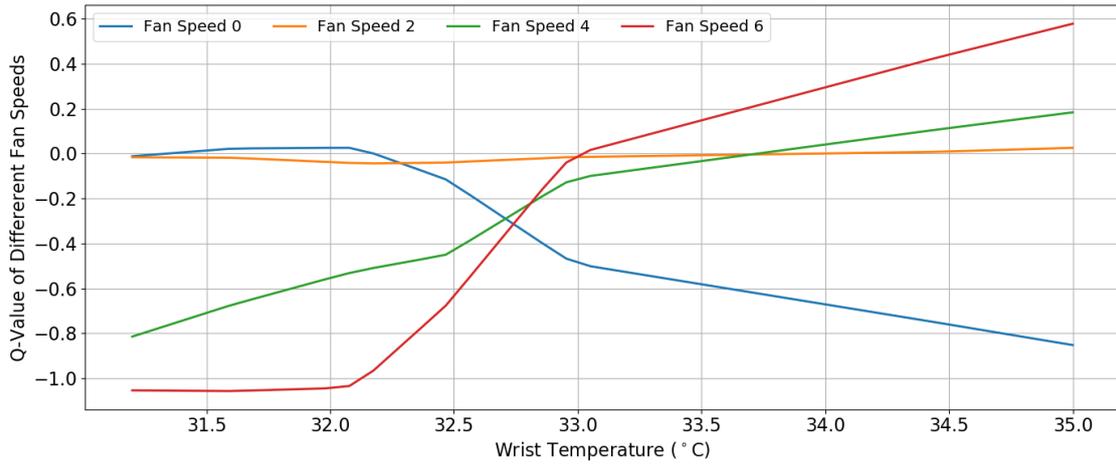
FIGURE 7.18: Wrist temperature distribution when voted acceptable to the air movement among the 14 subjects across different (A) age and (B) BMI

7.4.4 The Results for Agent Training

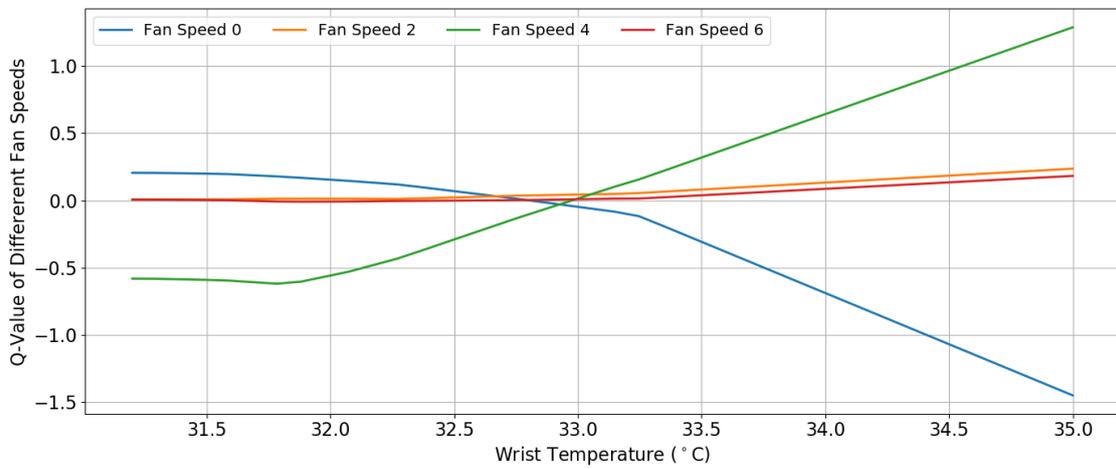
14 Bio-REAL agents were trained by the collected data described in 7.4.3 according to the process described in 7.3.2. The Q-networks of the 14 Bio-REAL agents were updated individually. Whether the agents were well-trained or not was visually examined based on the Q-value or the relationship between predicted best fan speed and skin temperature.

As mentioned in 7.4.3, there were five types of subjects. The predicted Q-value and best fan speed confirmed that, among the 14 experiment subjects, there were five neutral-preferred subjects (subject1, subject2, subject3, subject4, and subject12), four warm-preferred subjects (subject9, subject10, subject13, and subject14), two cool-preferred subjects (subject6 and subject11), one no preference subject (subject5), and two irregular-preference subjects (subject7 and subject8).

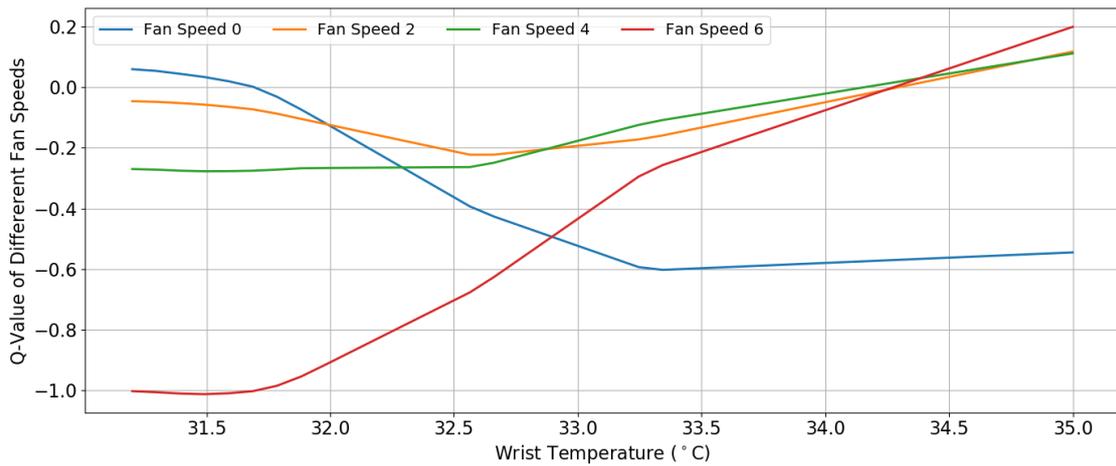
For the neutral-preferred subjects, the lower fan speed at lower wrist temperature had higher Q-value. With the increases in the wrist temperature, the Q-value of lower fan speeds was decreased and that of higher fan speeds was increased, as shown in Figure 7.19. Figure 7.20 shows the predicted preferred fan speeds in response to the wrist temperature for subject2, subject3, and subject4. When the wrist temperature was greater than 33 °C, the three subjects all preferred the higher fan speeds, either speed4 or speed6.



(A) Subject 2

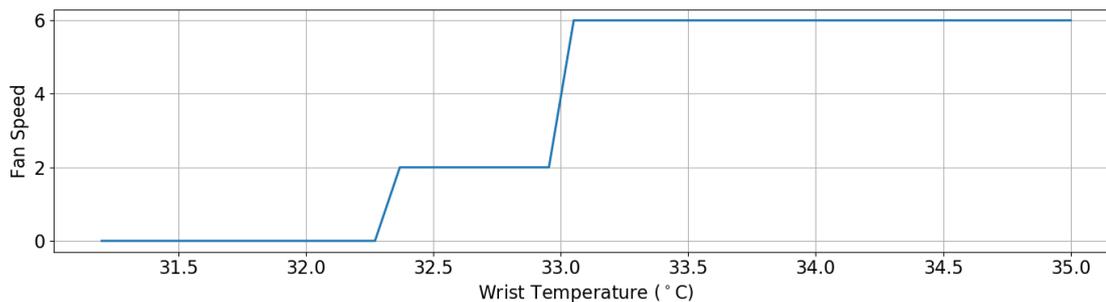


(B) Subject 3

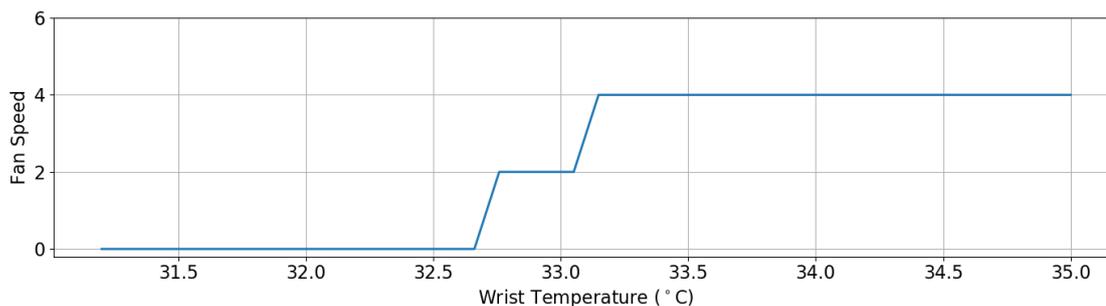


(C) Subject 4

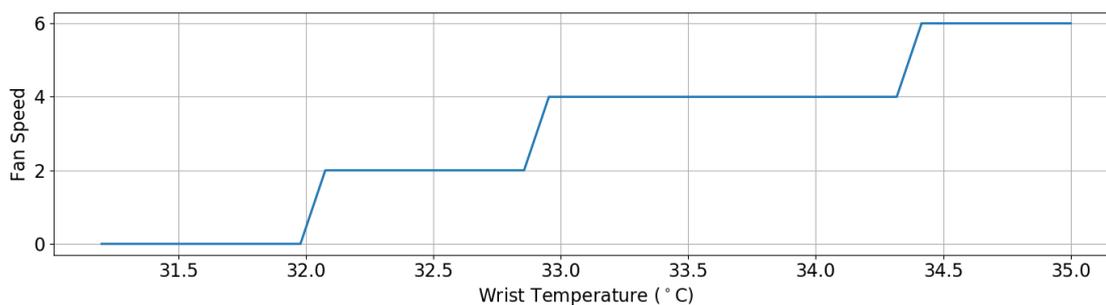
FIGURE 7.19: Q-value (expected air movement acceptability) of fan speeds at different wrist temperature for neutral-preferred subjects



(A) Subject 2



(B) Subject 3

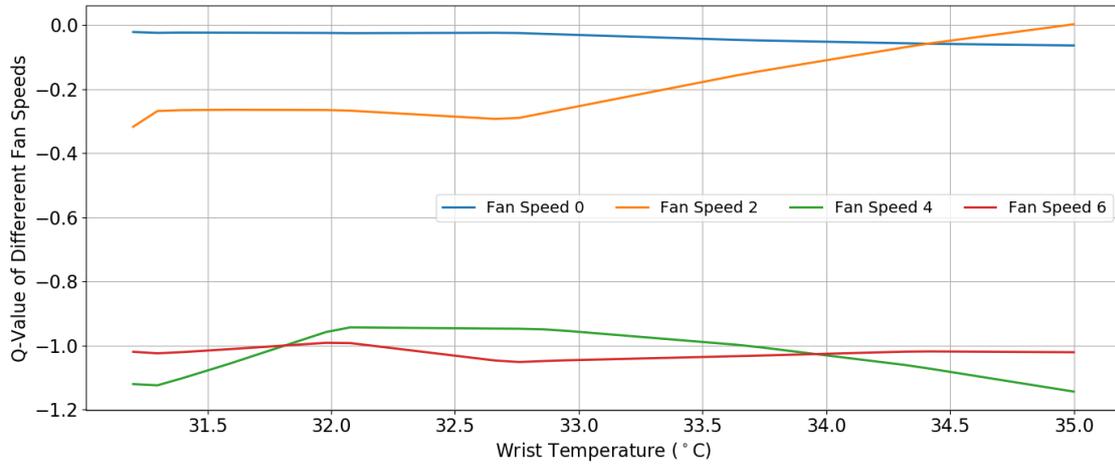


(C) Subject 4

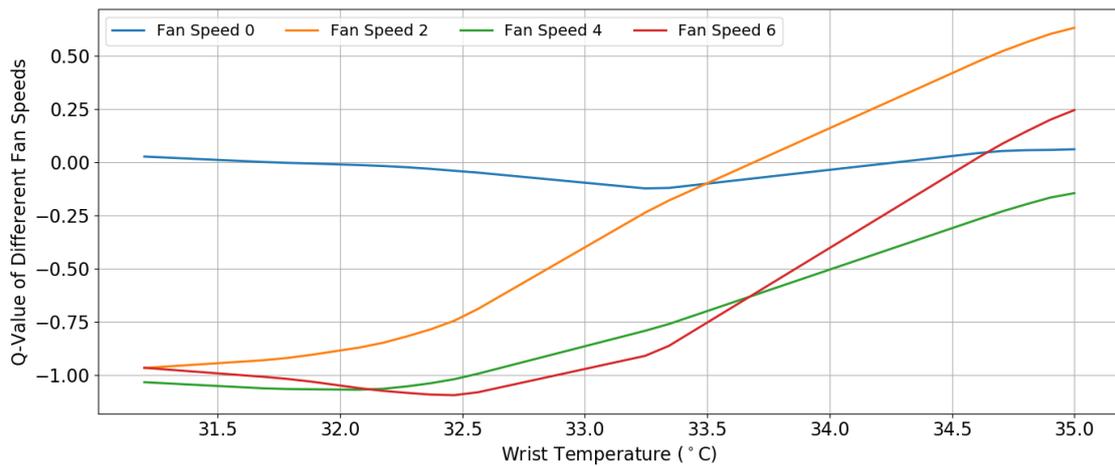
FIGURE 7.20: Predicted best fan speed (action) at different wrist temperature for neutral-preferred subjects

The predicted preferred fan speeds of the warm-preferred subjects were mostly fan speed0 and speed2, as shown in Figure 7.21 and 7.22. For Subject10, the speed4 and speed6 had low Q-value regardless of the wrist temperature. The speed0 had higher Q-value most of the time. Only when the wrist temperature was higher than 34.2°C, the Q-value of speed2 was higher than speed0. Although the Q-value of speed4 and speed6 raised with the increases in wrist temperature, it was always lower than that of speed2 for

Subject9. Fan speed2 was preferable than speed0 for subject9 when the wrist temperature was higher than 33.5 °C.

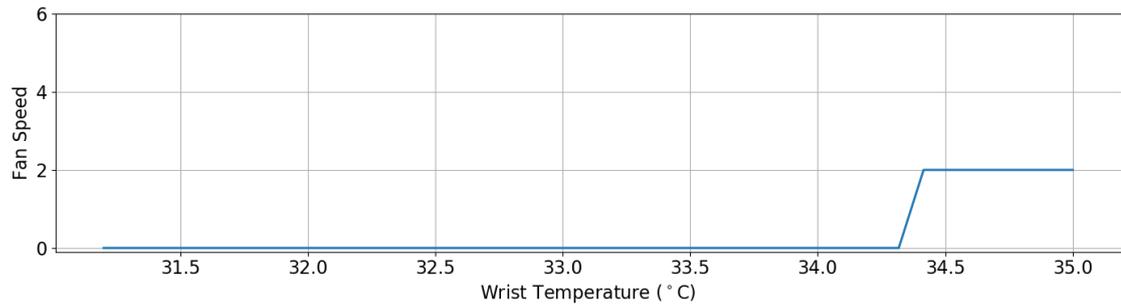


(A) Subject 10

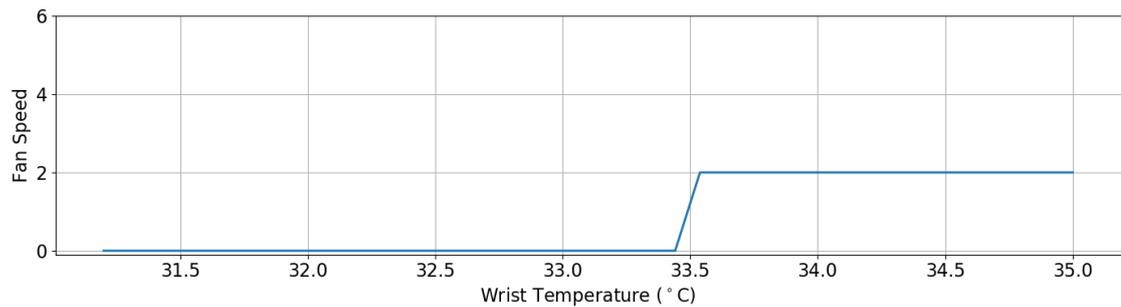


(B) Subject 9

FIGURE 7.21: Q-value (expected air movement acceptability) of fan speeds at different wrist temperature for the warm-preferred subjects



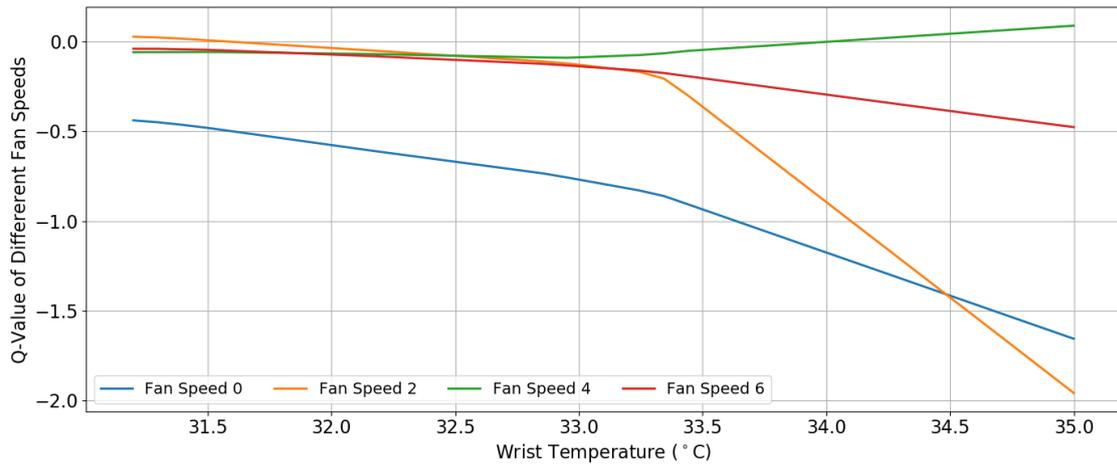
(A) Subject 10



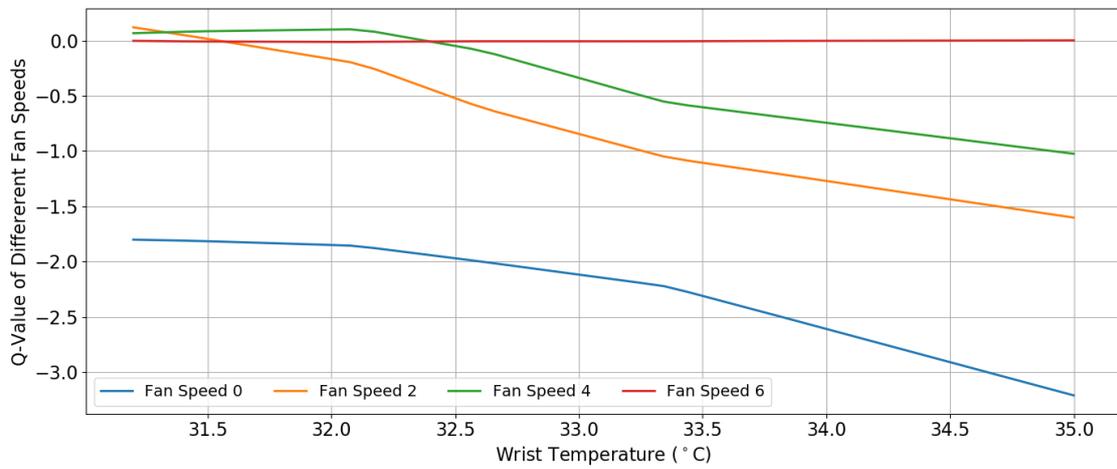
(B) Subject 9

FIGURE 7.22: Predicted best fan speed at different wrist temperature for the warm-preferred subjects

The predicted preferred fan speeds of the cool-preferred subjects were always higher than fan speed0, as shown in Figure 7.23 and 7.24. The Q-value of speed0 was lower than other fan speeds across the wrist temperature range. Subject11 preferred a colder condition more than Subject6. He/she preferred speed6 as soon as the wrist temperature was higher than 32.5 °C.

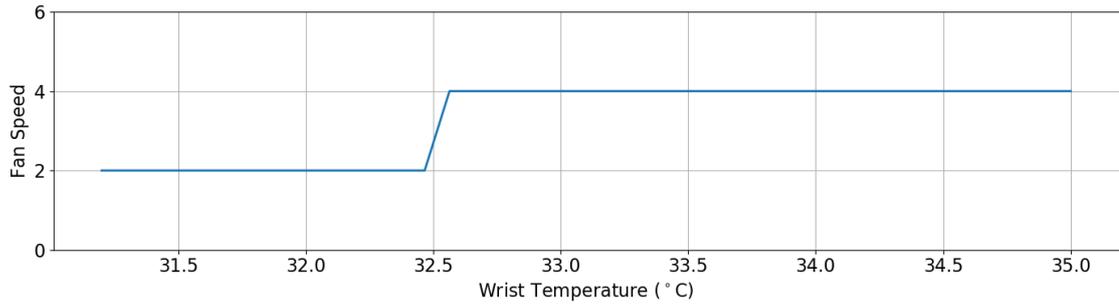


(A) Subject 6

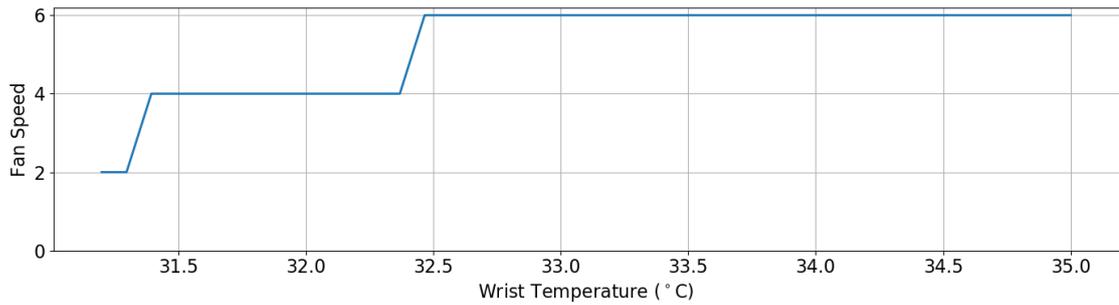


(B) Subject 11

FIGURE 7.23: Q-value (expected air movement acceptability) of fan speeds at different wrist temperature for the cool-preferred subjects



(A) Subject 6



(B) Subject 11

FIGURE 7.24: Predicted best fan speed at different wrist temperature for the cool-preferred subjects

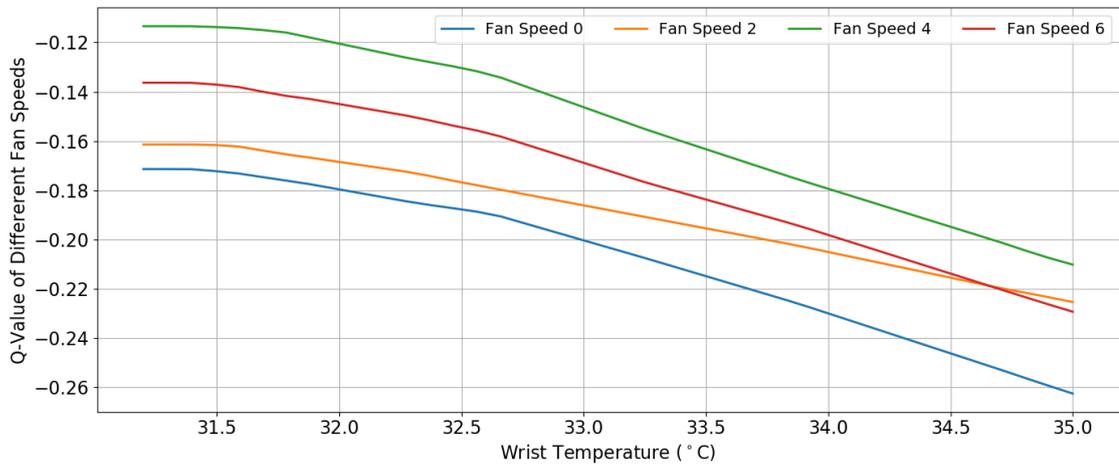
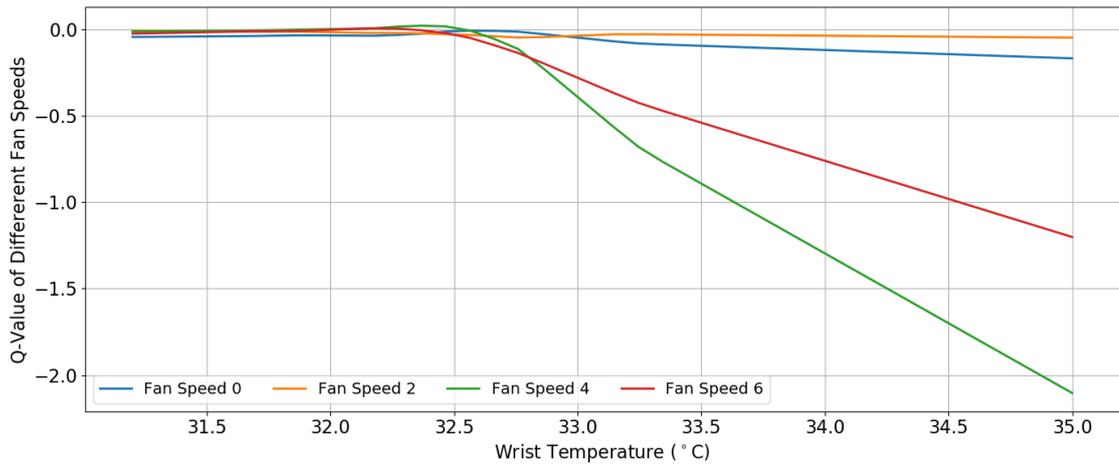


FIGURE 7.25: Q-value (expected air movement acceptability) of fan speeds at different wrist temperature for the no-preferences subjects

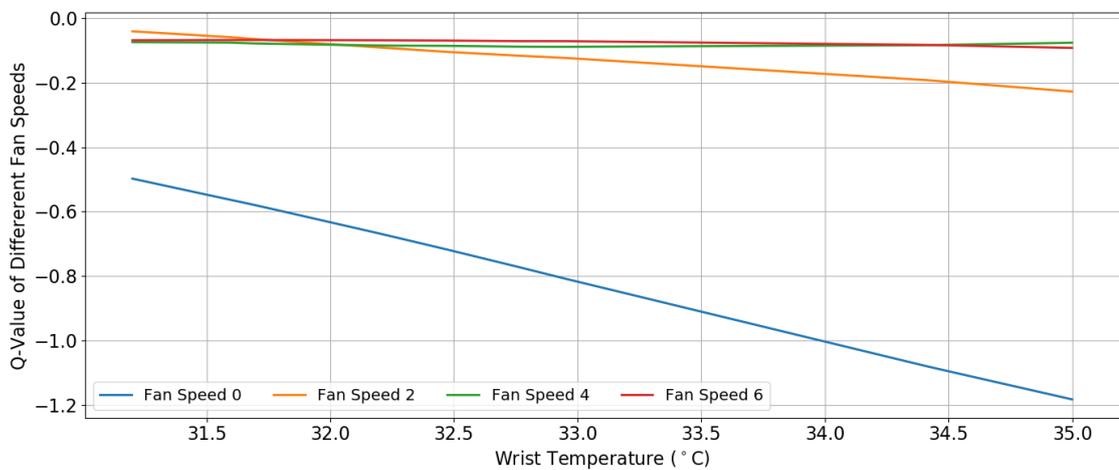
The no-preference subject (Subject5) voted acceptable to all fan speeds, which means his/her data collected from the experiment provided no useful information. Therefore, the training result or the Q-value was meaningless,

as shown in 7.25. During the control process, the predicted preferences of Subject5 were not considered since he/she would be happy no matter which fan speed is selected.

Both the Q-value and the predicted preferred fan speed of the irregular-preference couldn't present any patterns or characteristics, as shown in Figure 7.26 and 7.27. The preferred fan speeds and wrist temperature was not positively correlated. The irregular patterns were possible due to lacking data. If there could be more data collected from the subjects, the prediction could show regular preferences.

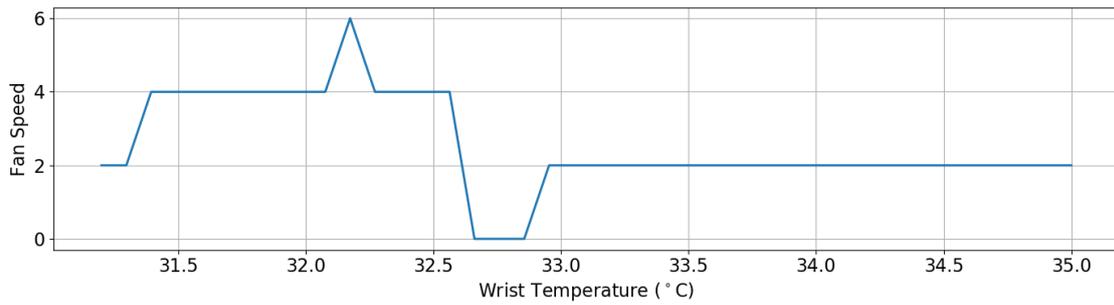


(A) Subject 7

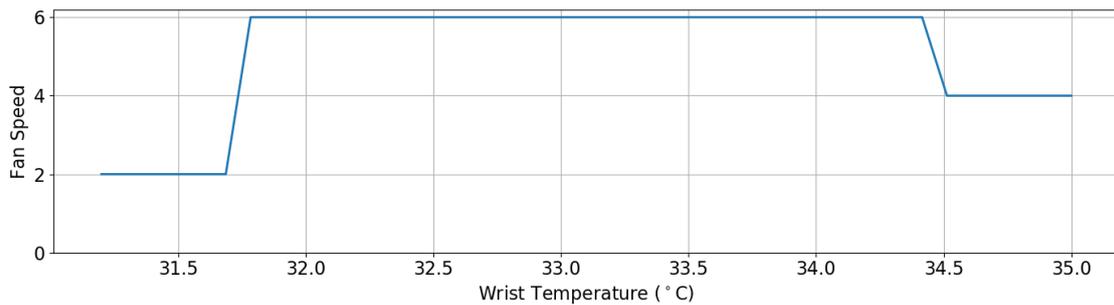


(B) Subject 8

FIGURE 7.26: Q-value (expected air movement acceptability) of different fan speeds at different wrist temperature for the irregular-preference subjects



(A) Subject 7



(B) Subject 8

FIGURE 7.27: Predicted best fan speed at different wrist temperature for the irregular-preference subjects

7.4.5 Multi-agent Negotiation

As shown in the training results, the 14 subjects had 14 different predicted preferences in fan speeds with regards to wrist temperature, so the negotiation was inevitable in the office where occupants shared the ceiling fans. As mention in Section 3.2, the Bio-REAL control system executes negotiation according to the Q-value of all the agents. This experiment promoted the negotiation method into two steps:

- If there are more than half occupants prefer one action, it will be the negotiated action.
- Otherwise, the negotiated action (fan speed) is be the one that can maximum weighted sum Q-value. w^i is the importance or weight assigned

to each subject.

$$a_{negotiated} = \operatorname{argmax}_{a^i \in A} \sum_i w^i Q^i(S_t^i, A) \quad (7.2)$$

Section 7.5.2 demonstrates an example of the multi-agent negotiation.

7.5 Evaluation Experiment

The evaluation experiment was to test the control performance of the Bio-REAL control system with well-trained agents and compare it to the baseline.

7.5.1 Experiment Design

The 14 subjects were separated into three groups for the evaluation experiment. There were five subjects in each group, as shown in Table 7.8. Since the subject serial number was assigned at random, the subjects were randomly assigned to the three groups. Subject10 was in both Group2 and Group3 to ensure the number of subjects was equal for each group. During the experiment, the five subjects sited on the right side of the experiment room. Three sited underneath one ceiling fan and two underneath another ceiling fan, as shown in Figure 7.9. The two ceiling fans behaved as a single ceiling fans as they actuated the same fan speed every 10-minutes.

Six experiment sessions were undertaken to evaluate both baseline control and Bio-REAL control system. The baseline was the static fan speed2 (air velocity ranged from 0.028 m/s to 0.693 m/s at different vertical locations)

TABLE 7.8: The evaluation experiment at NUS

Session	Subjects Group	Control Strategy
1	Group1: Subject 1, 2, 3, 4, 5	
2	Group2: Subject 6, 7, 8, 9, 10	Baseline control : static fan speed2
3	Group3: Subject 10, 11, 12, 13, 14	
4	Group1: Subject 1, 2, 3, 4, 5	
5	Group2: Subject 6, 7, 8, 9, 10	Bio-REAL control: dynamic fan speed
6	Group3: Subject 10, 11, 12, 13, 14	

at temperature 27 °C. It was determined based on the ASHRAE standard 55 ASHRAE 55, 2013. The standard declared that the airspeed should be in the range of 0.1m/s to 0.7 m/s at operative temperature 27 °C to achieve thermal comfort, as shown in Figure 7.28.

During the Session4-6, the Bio-REAL control system with trained agents was deployed in the experiment room to automatically control the ceiling fan based on subjects' wrist temperature. Each session lasted around 3 hours. There were two less than 10-minute breaks every 50 minutes.

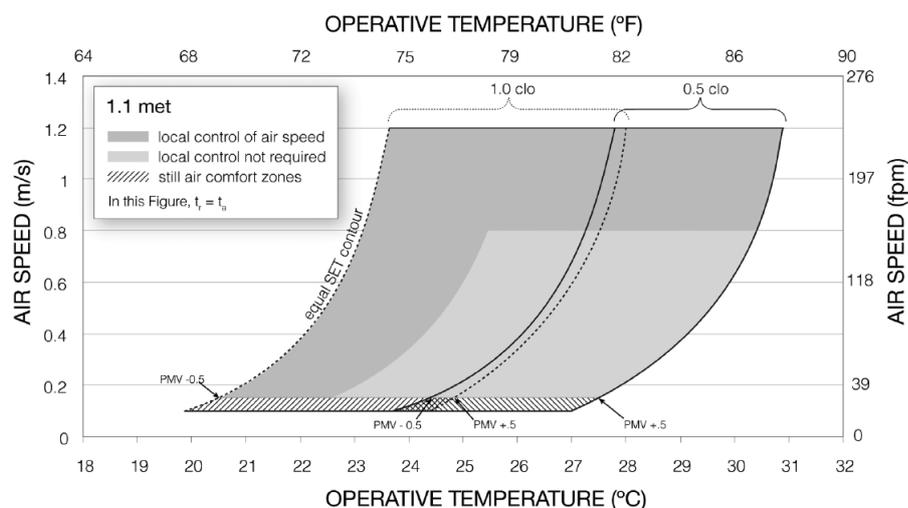


FIGURE 7.28: Acceptable range of operative temperature and air speeds for the comfort zone at humidity ratio 0.010 in ASHRAE standard 55 .

7.5.2 Negotiation Example

A snapshot of Session5 was presented to explain the process of negotiation and control. At one point of Session5, the wrist temperature of the five subjects was collected and read by the Bio-REAL agents. As shown in Table 7.9, subject6, subject7, and subject8 had a higher wrist temperature, while subject9 and subject10 had a lower one. The Q-network of the Bio-REAL agents for the five subjects predicted the Q-value of each of the four fan speeds given the wrist temperature. The Q-value explained not only the expected acceptability but also the acceptability levels of each fan speed. For subject6 and subject8, fan speed4 was the best since it had the highest Q-value. Subject7 preferred fan speed0 the most. The highest Q-value for subject9 and subject10 lied on fan speed 0.

Since there was no fan speed preferred by more than half occupants, the weighted sum Q-value of the five subjects was computed and shown in Table

TABLE 7.9: Wrist temperature and Q-value snapshot for the five subjects in Session5

Subject	Wrist temperature	Q-value			
		Fan Speed 0	Fan Speed 2	Fan Speed 4	Fan Speed 6
6	34.16	-1.306	-1.186	-0.230	-0.352
7	34.26	-0.093	0.135	-0.008	-0.787
8	34.46	-0.624	-0.005	0.055	-0.430
9	32.36	-0.043	-0.752	-1.015	-1.050
10	32.76	-0.280	-0.470	-0.787	-0.993

TABLE 7.10: Weighted sum Q-Value

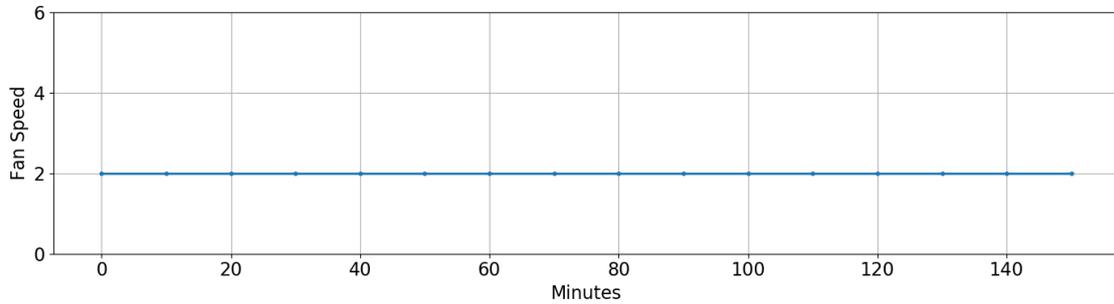
Fan Speed 0	Fan Speed 2	Fan Speed 4	Fan Speed 6
-0.469	-0.456	-0.305	-0.722

7.10. In this experiment, the weights of all occupants were the same. Since fan speed4 had the highest weighted sum Q-value, it was the negotiated fan speed that can maximize the air movement acceptability of Group5 give the wrist temperature. Therefore, the two ceiling fans run on fan speed4.

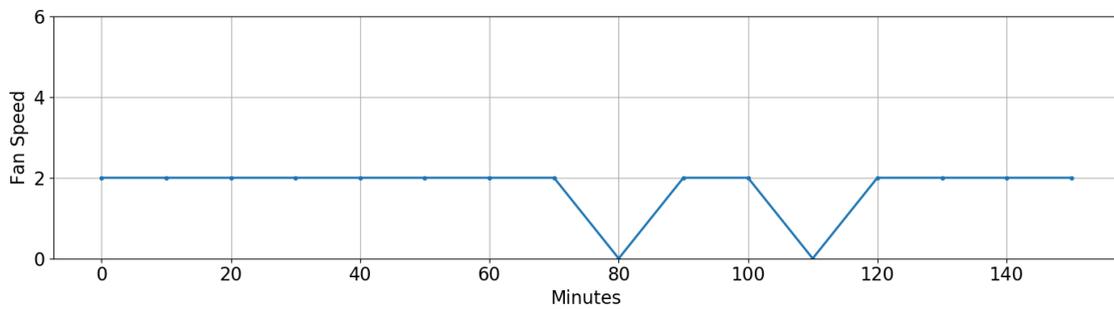
7.5.3 Evaluation Result

The baseline sessions were operated by the static fan speed schedule (Figure 7.29a). During the Bio-REAL control sessions, the fan speed schedules were dynamic and generated in response to occupants' wrist temperature. Each Bio-REAL control session had different fan speed schedules due to the differences in their participants, as shown in Figure 7.29b, 7.29c, and 7.29d. Since Group1 had four neutral-preferred subjects and one no-preference subject, the fan speed generated for them were mostly speed2. The two-times speed0 in the latter part of the Session4 is because the wrist temperature was slightly low for some subjects after long-time exposure to speed2. Group2 had two warm-preferred subjects, one cool-preferred subject, and two irregular-preference subjects. The higher fan speed in the latter part of Session5 was due to the irregular preferences of Subject8. There were three warm-preferred, one neutral-preferred, and one cool-preferred subjects in Group3. Since this group had more warm-preferred subjects, there were more times of speed0 in Session6. Moreover, the high fan speed at the beginning of the Session5 and Session6 was because the occupants just came from outside with a higher wrist temperature. Besides, there was one subject left the room in the middle

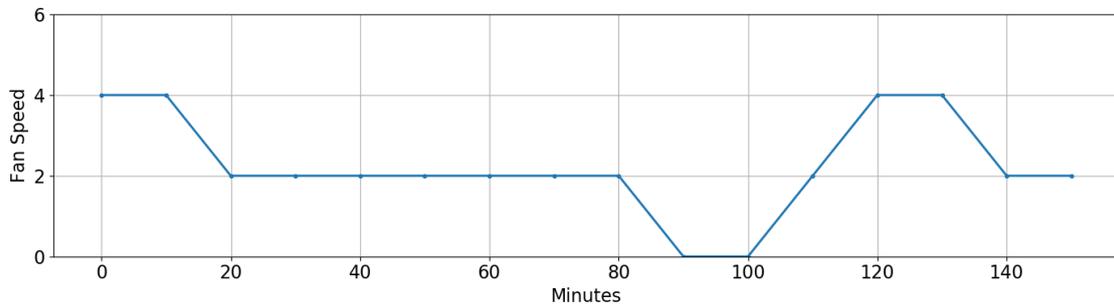
in Session5. The Bio-REAL control system can leave out the agent for the absent subject and work well with the occupancy disturbance.



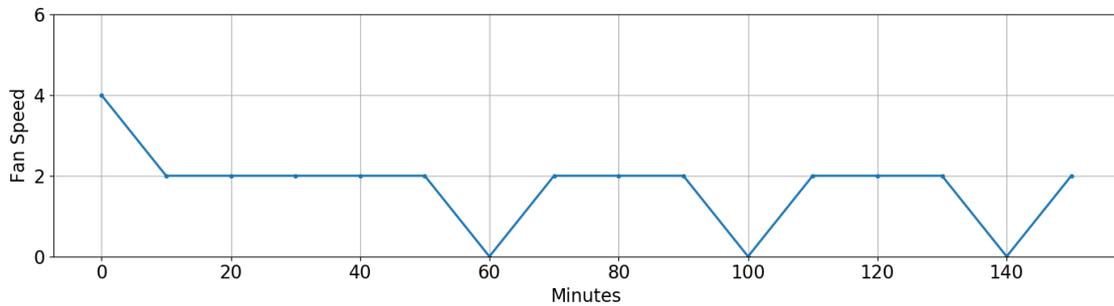
(A) Fan speed schedule of baseline sessions (session 1,2, and 3)



(B) Bio-REAL control session for Group1 (Session4)



(C) Bio-REAL control session for Group2 (Session5)



(D) Bio-REAL control session for Group3 (Session6)

FIGURE 7.29: Fan speed schedules for the evaluation experiment

The performance of the control strategies was quantified by the rate of voting comfortable and acceptable. There were 219 subjective votes collected from the baseline sessions and 223 from the Bio-REAL control sessions. The rate was calculated using Equation 7.3. N is the number of votes of baseline sessions or Bio-REAL control sessions. Figure 7.30 and 7.31 compared the performance of the baseline and the Bio-REAL controls. The Baseline met the ASHRAE's goal of 80% comfort rate (ASHRAE 55, 2010) just right if speed2 setting is ensured with 27 °C. The Bio-REAL control system achieved 94% of comfort rate, which was 14% higher than the Baseline. The air movement acceptability of Bio-REAL control was 11 % higher than the Baseline with fixed fan speed 2.

$$rate = \frac{\text{number of votes from each comfort (acceptability) level}}{N} \times 100\% \quad (7.3)$$

Besides, the Bio-REAL control also successfully saved the fan energy due to the dynamic fan speeds. Figure 7.29b, 7.29c, and 7.29d show that, except for speed2, there were speed0 seven times and speed4 five times in the Bio-REAL control sessions. The power of different fan speed was different, ranging from 1.9W to 26.8W. If Assuming the experiment room running the four Haiku ceiling fan 7.5 hours per day, the estimated yearly fan energy consumption of baseline control will be 66.5 kWh and that of Bio-REAL control will be 63.3 kWh. The estimated yearly fan energy saving of the Bio-REAL control will be 4.5 % as compared to the baseline control.

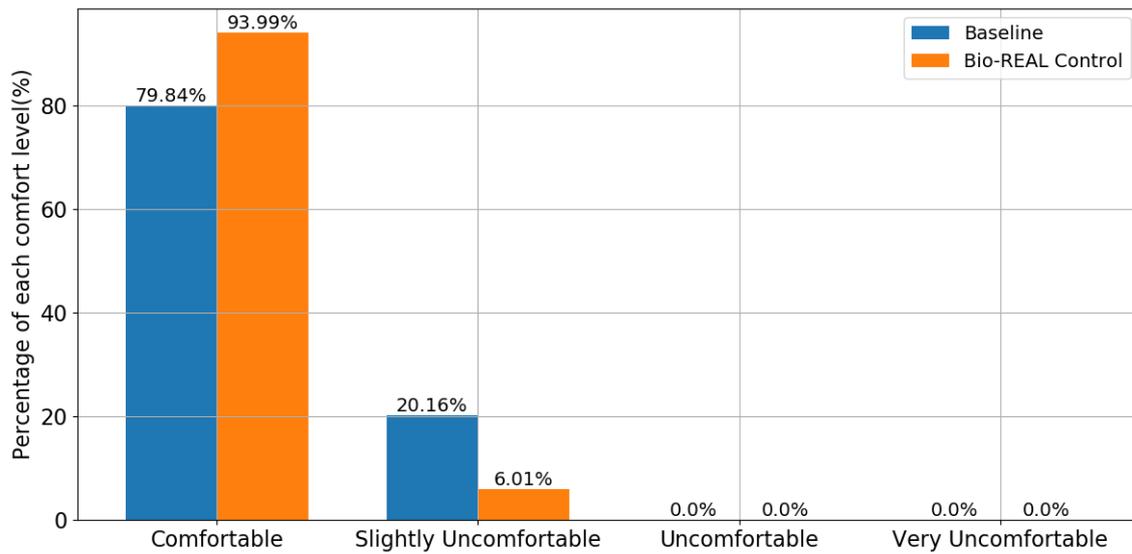


FIGURE 7.30: Comparison in thermal comfort rate between baseline and Bio-REAL control

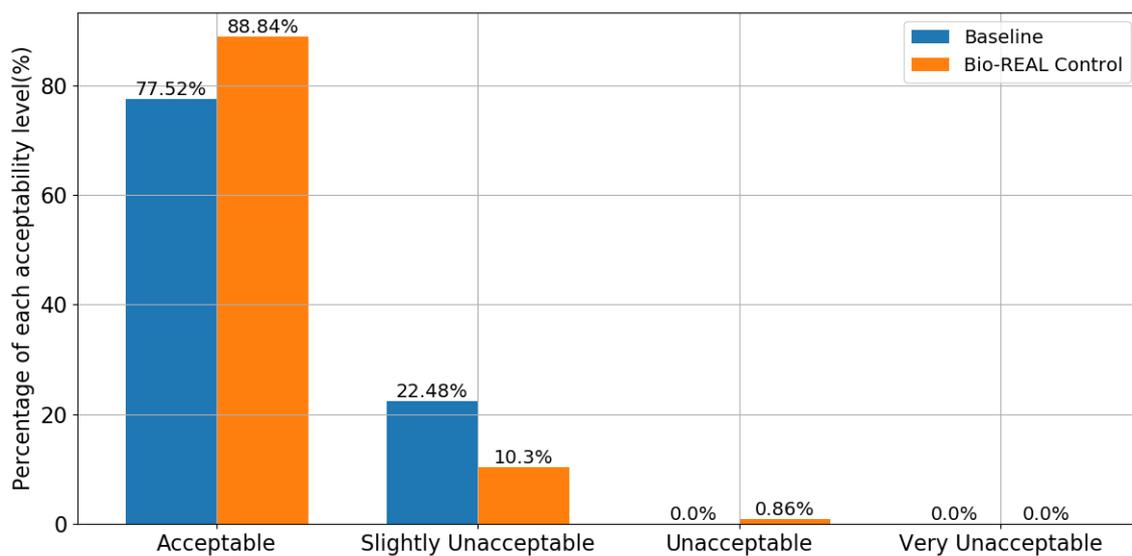


FIGURE 7.31: Comparison in air movement acceptability rate between baseline and Bio-REAL control

The result also demonstrated that, if convective cooling is provided with ceiling fans, raising the indoor temperature in a tropical climate to 27 °C can achieve comfort and save a huge amount of energy. As indicated by the energy temperature correlation model shown in Equation ?? (Yuan et al., 2013),

the HVAC energy consumption is proportional to the difference between indoor and outdoor temperature. In equation ??, λ is the conductivity of a room, M is energy transformation ratio of an HVAC system, T_i is the indoor temperature, and T_o is the average outdoor temperature. The Singapore average outdoor temperature is 28 °C. Based on the Equation refenergy-nus, the HVAC energy at indoor temperature 27 °C is 5-times lower than that at 22 °C.

$$HVAC\ Energy = \left| \frac{\lambda}{M} (T_i - T_o) \right| \quad (7.4)$$

7.6 Conclusion and Discussion

The field experiments demonstrated the real-world performance of the Bio-REAL control system with 14 subjects in the net-zero energy building at NUS. There were 14 Bio-REAL agents trained by the data collected from the training experiment, including wrist temperature, fan speed, and air movement acceptability. The training data showed that, although there were individual differences, the wrist temperature had a close correlation with air movement acceptability. Using the training samples, the agents updated their Q-networks daily rather than real-time to ensure learning stability.

The testing results showed that the ASHRAE standard for fan speed2 (air velocity ranged from 0.028 m/s to 0.693 m/s at different vertical locations) at 27 °C provided 80% comfortable rate across the diverse occupants, indicating that raising the temperature to 27 °C could achieve comfort if convective

cooling is presented by ceiling fans. The Bio-REAL ceiling fan controls in a 10-minute interval can achieve 14 % higher comfort rates than the baseline with a static fan speed² at 27 °C, with 4.5 % yearly fan energy saving as compared to baseline with fan speed² and 5 times less HVAC energy consumption as compared to VAV with setpoint 22 °C.

One limitation of the Bio-REAL control system in this experiment is that the Bio-REAL agents are examined manually to determine whether they are well-trained or not. An automatic examination mechanism should be designed to achieve a completely automated control. Moreover, the experiments tested the Bio-REAL control system with ceiling fans, a promptly responsive cooling system. The real-world performance of the Bio-REAL control system working with the typical HVAC system was not investigated. More training data and time will be needed for HVAC controls.

Nevertheless, to the author's knowledge, the experiments were the first example of implementing reinforcement learning in building thermal controls without the pre-training by simulation. The Bio-REAL control system is capable of being introduced in any building with shared controls of thermal conditioning, requiring 5-10 days of training based on the frequency of subjective responses and the type of controls (more responsive system take fewer days).

Chapter 8

Conclusion

This thesis developed a bio-sensing and reinforcement learning control (Bio-REAL) system to personalized improve thermal comfort and energy efficiency. The Bio-REAL system comprises a bio-sensing network, multiple personalized Bio-REAL agents, and a negotiator. The bio-sensing network uses Microsoft smart band to measure occupants wrist temperature in real-time. The Bio-REAL agents initiate control decisions in response to wrist temperature, subjective feedback, and environmental conditions. The negotiator resolves conflicts in the decisions initiated by different Bio-REAL agents. The state-of-art RL algorithm, double Q learning with experiment replay and neural network approximation, was applied to train the Bio-REAL agents. The Bio-REAL system were trained and evaluated using three experiment techniques: simulation experiments, preliminary field and simulation experiments, and field experiments.

8.1 Summary of Experimental Findings

8.1.1 Simulation Experiment: VAV with Electric Reheat for Heating

The simulation experiment tested the feasibility of the Bio-REAL system with a simple simulated learning environment and determined the structure and the hyper-parameters of the Bio-REAL systems. The simulated learning environment was comprised of classic occupant models, including the PMV model and Pierce two-node model, and the EnergyPlus model of a room heated by a VAV system with electric reheat, which was the Northern zone of a one-story DOE reference building. Three personalized Bio-REAL agents and a negotiator was created for the Bio-REAL system.

The results of the simulation experiment showed that the Bio-REAL system converged to an optimum after learning with 112 episodes. The optimal dynamic control policy created by the Bio-REAL system can achieve 0.49% thermal comfort improvement as compared to classic thermal comfort model based static control schedule (25 °C). The Bio-REAL system had 52% better performance than the standard static control schedule (22 °C) in winter.

8.1.2 Preliminary Field and Simulation Experiment at CMU: Water Sourced Radiators for Heating

The field and simulation experiment evaluated the performance of the Bio-REAL control system with a learning environments simulated reality. The

learning environment was comprised of occupant models developed using data collected from a human subject experiment and an EnergyPlus model of the experimental room with water sourced radiators. The occupant models contained personalized thermal comfort models and skin response models. The personalized thermal comfort models created six individual thermal comfort zones, which were different from the classic PMV comfort zone. Six personalized Bio-REAL agents and a negotiator was created for the Bio-REAL system.

The learning environment for training and testing was different in terms of their outdoor weather conditions. The experiment included four cases with two different scenarios. Case1 was baseline with standard static schedule. Case2 was personalized thermal comfort mode based controls. In Case3 and Case4, the Bio-REAL system was deployed to control the temperature setpoint. The objective function of the Bio-REAL control system in Case3 weighted more on thermal satisfaction. Case4 balanced the weighting on thermal satisfaction and energy efficiency. There were six subjects in Scenario 1 and five subjects in Scenario 2. The Bio-REAL system converged less than 600 episodes (simulation run) for all cases. The time taken for training on Windows 10 system with random access memory 16GB and processor speed 3.4GHZ was about 10 hours.

The testing results of the experiment showed that the dynamic control policy derived by the Bio-REAL system had more advantage in improving thermal satisfaction and energy efficiency than static schedules. In the case of balancing comfort and energy for six subjects, the Bio-REAL controls can

reduce dissatisfaction rate 0.37% and save energy consumption 2.1% as compared to the personalized thermal comfort based controls (fixed 23 °C).

8.1.3 Field Experiment at NUS : Smart Ceiling Fans for Cooling

The field experiments demonstrated the real-world performance of the Bio-REAL control system with 14 subjects in a room of the Net-zero energy building at NUS. The room has four Haiku smart ceiling fans. During the experiment, the indoor temperature was maintained at 27 °C, but the ceiling fan speeds were varied. There were 14 Bio-REAL agents trained by the data collected from the training experiment, including wrist temperature, fan speed, and air movement acceptability. The objective of the Bio-REAL system is to find an optimized ceiling fan schedule at room temperature 27 °C.

The testing results showed that the ASHRAE standard for fan speed2 (air velocity ranged from 0.028 m/s to 0.693 m/s at different vertical locations) at 27 °C provided 80% comfortable rate across the diverse occupants, indicating that raising the temperature to 27 °C could achieve comfort if convective cooling is presented by ceiling fans. The Bio-REAL ceiling fan controls in a 10-minute interval can achieve 14 % higher comfort rates than the baseline with a static fan speed2 at 27 °C, with 4.5 % yearly fan energy saving as compared to baseline with fan speed2 and 5 times less HVAC energy consumption as compared to VAV with setpoint 22 °C.

8.2 Contribution

The Bio-REAL control system and the experimental findings contribute to the domain of thermal comfort, building controls, the application of IoT and artificial intelligence in buildings, as described below.

8.2.1 Comfort and Energy Benefit

The findings and results of the experiments demonstrated that the Bio-REAL system can provide comfort for each individual with low energy consumption, addressing individual differences in thermal comfort for multi-occupant spaces with no individual controls. The Bio-REAL system also addresses a range of heating and cooling choices from ambient to task systems. With the achievement in comfort improvement and energy saving, the Bio-REAL system contributes to occupant health and productivity, as well as sustainability.

8.2.2 The Application Of Internet of Things (IoT) in Buildings

The Bio-REAL system interrelated occupant, environment, digital devices, and mechanical systems in buildings and contributes to the applications of IoT in buildings, as shown in Figure . The Bio-REAL system uses Bio-sensing, participatory sensing, and environmental sensing to interrelate occupant biological response, occupant subjective responses, environmental conditions. The integration of Bio-sensing and RL control system connected occupant and environment to mechanical systems in buildings.

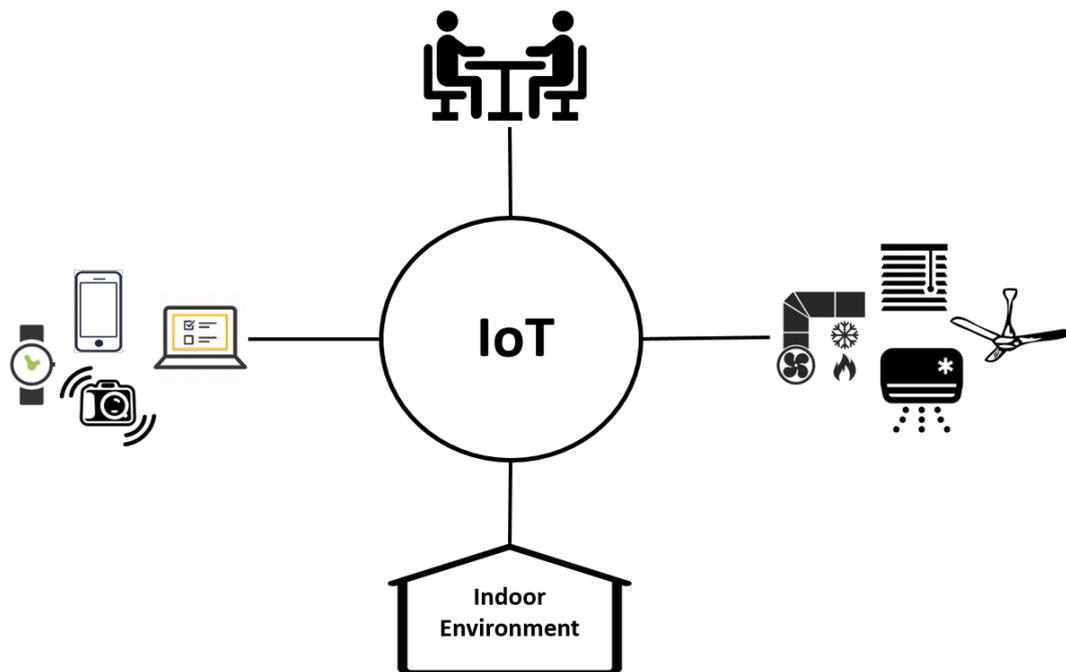


FIGURE 8.1: The Bio-REAL system contributes to the application of IoT in buildings

Besides, The smart-bands adopted by this thesis advanced the bio-sensing technology for its acceptability (none-intrusiveness) and affordability. The smart-bands collect wrist temperature, which has been proved as the bio-signal closely correlated with personalized thermal comfort. The experiments data in this thesis verified the correlation by showing the relationship between wrist temperature and thermal sensation and the relationship between wrist temperature and air movement acceptability.

Moreover, wrist temperature can be a substitution of the environmental variables that are not easy to measure. The experiment at CMU presented a strong correlation between wrist temperature and operative temperature. The experiment at NUS showed that wrist temperature is very responsive to the air moment. Since radiant temperature and air movement are not easy to

measure in practice, wrist temperature can be their replacement. Replacing the environmental variables that highly correlated with the wrist temperature can simplify the sensing network in buildings and reduce the implementation cost of the personalized thermal comfort controls.

8.2.3 The Application of Artificial Intelligence in Buildings

This thesis is a showcase of implementing reinforcement learning (RL) in occupant centric thermal control. There were innovations in the design and the learning process of the Bio-REAL system.

The recent approach of implementing RL in building controls is deploying the RL control systems in real buildings after sufficiently training them with simulation. This approach was hard to be applied to personalized thermal control because most existing physics-based simulation tools are not supportive of personalized thermal comfort modeling. This thesis overcame the issue by integrating personalized occupant thermal models with the building models in the co-simulation. The personalized thermal comfort models were the synergy of the data simulated from the PMV model and the field data. The model established an individual thermal comfort zone for each occupant. The approaches of combining simulated data and field data for thermal comfort model development and integrating physics-based simulation tool and data-driven models contributes to the application of artificial intelligence in buildings.

The experiment at NUS further advanced the reinforcement learning implementation in occupant-centric thermal controls by introducing the approaches of learning without simulation. The learning process was split to experience collection, agent updating, and agent checking to guarantee learning stability. Only the well-trained Bio-REAL agent plays a part in controlling the building systems so that the control performance won't be impaired by the RL exploration. This experiment is a brand-new example of real-world RL application in building controls. The approaches of training and testing RL agents with the real-world environment also contributes to the application of artificial intelligence in buildings.

8.2.4 The application Of Digital Twin in Buildings

The idea of creating one agent for one occupant contributes the application of Digital Twin in buildings. An Bio-REAL agent can be considered as a digital replica of an occupant to control building systems. Besides, by introducing the negotiator that resolves the conflicts in personal differences, the application of the Bio-REAL system can be extended to the shared environment with multiple occupants easily. The structure of the Bio-REAL system with multiple agents and negotiation decomposes the tasks of thermal comfort optimization, energy efficiency optimization, and conflict negotiation. The decomposition saves computational resources and time.

8.3 Limitation and Future Work

With the contributions made, the thesis still have several limitations. Future work can focus on the following improvement to make the Bio-REAL control system more intelligent and versatile.

8.3.1 Large-scale Application in Buildings

The thesis evaluated the performance of the Bio-REAL control system at a zone level, e.g., adjusting zone temperature setpoint or ceiling fans in a zone, but no at a building level. For a large-scale application in buildings with multiple zones, each zone may have one Bio-REAL control system to optimize the thermal comfort of the zone occupants and save energy at a zone level. For a building with a centralized HVAC system, different zones usually share other components of the HVAC system, although the zone terminal units can be controlled independently. Therefore, a building negotiator will be needed for the large-scale application to coordinate the operation of HVAC components. The building negotiator will communicate with the zone negotiators back and forth to optimize whole-building energy efficiency. An example of the architecture of the Bio-REAL control system for the large-scale application is shown in Figure 8.2

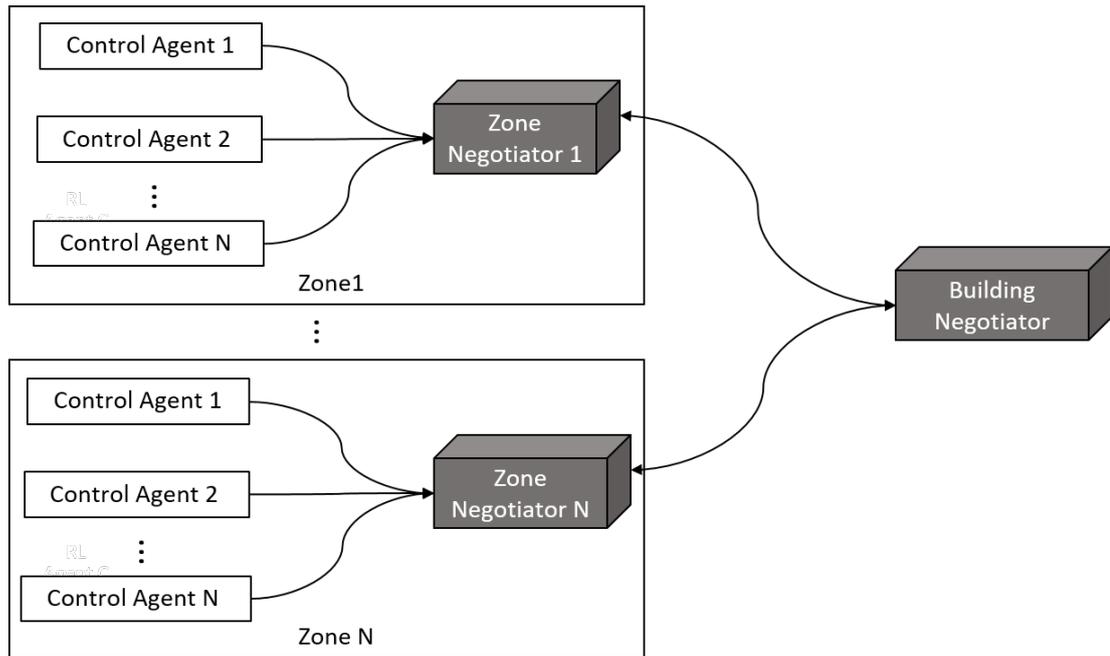


FIGURE 8.2: The Bio-REAL control system for large-scale application

8.3.2 Multi-agent Communication Network

The thesis designed the multi-agent structure for the Bio-REAL control system, but the agents learn independently without communicating with each other. In the future, researchers could build a connected network so that the agents can communicate, share information, and learning collaboratively, as shown in Figure 8.3. With the connected network, a negotiator handling the conflicts will not be needed in the control system. Via interaction, the agents can asymptotically reach a consensus and establish a shared control policy (Tan, 1993).

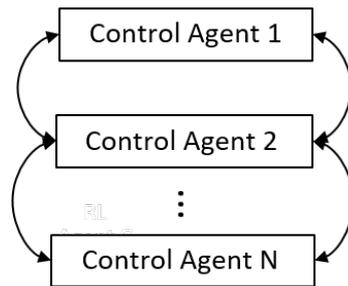


FIGURE 8.3: Multi-agent communication network

8.3.3 The Ensemble of Multiple Learning Algorithms

The thesis selected the "model-free" reinforcement learning (RL) algorithms due to its idea of learning through interaction. The results of this kind of learning is not impaired by the bias of non-representative models or data samples. In contrast, the performance of the "model-based" learning was determined by the models. The results of supervise learning was depended on the collected data samples. However, "model-free" RL is a trial and error learning algorithm, indicating that a learning agent will make lots of errors before being a mature agent. Further work can integrate different learning algorithms to circumvent the weakness of different kinds of learning and achieve better performance (sutton1991dyna), as shown in Figure 8.4. By ensembling different learning methods, RL agents can still learn by interacting with the environment. The interaction experiences can be used to update the MDP (Markov Decision Process) models. The RL agents can solve the MDP models for planning. The collected historical samples can be processed to train neural network via supervised learning. The Q-networks of RL agents and the neural networks of the supervised learning can work together to generate the policy (Silver and Hassabis, 2016).

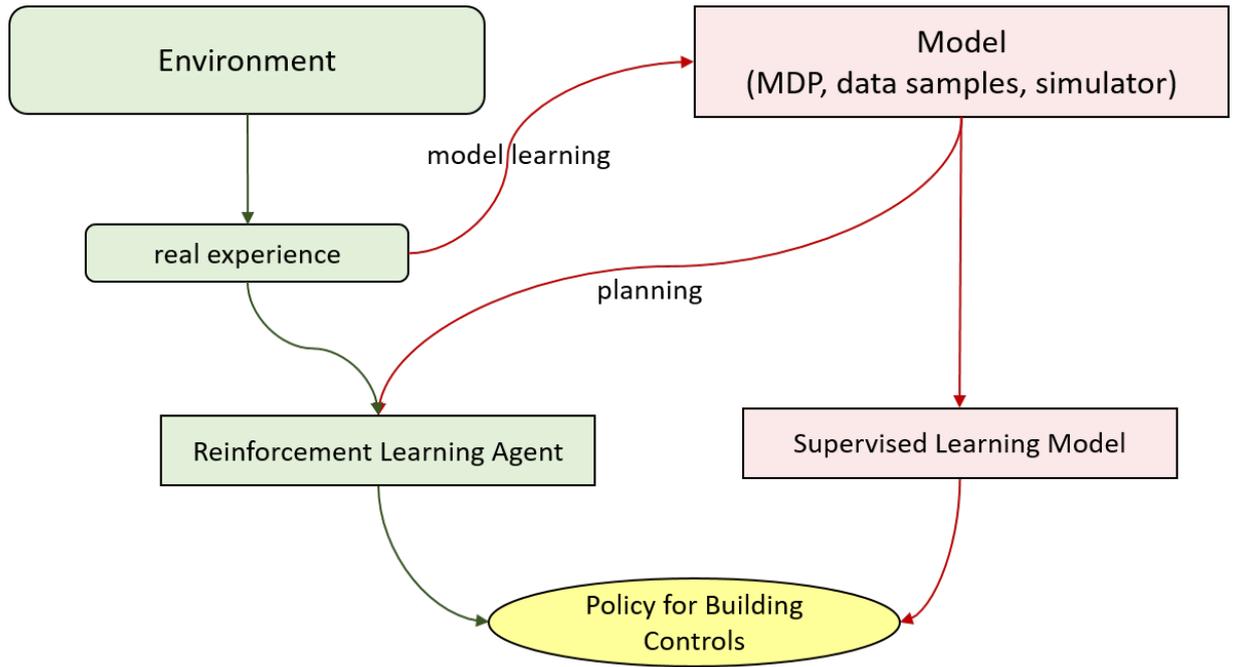


FIGURE 8.4: The ensemble of multiple learning algorithms

Bibliography

- 10551, ISO (2019). “Ergonomics of the physical environment – Subjective judgement scales for assessing physical environments”. In:
- Ahmad, Muhammad Waseem et al. (2016). “Building energy metering and environmental monitoring—A state-of-the-art review and directions for future research”. In: *Energy and Buildings* 120, pp. 85–102.
- Alcalá, Rafael et al. (2003). “Fuzzy control of HVAC systems optimized by genetic algorithms”. In: *Applied Intelligence* 18.2, pp. 155–177.
- ASHRAE. (2017). *2017 ASHRAE Handbook: Fundamentals*. ASHRAE.
- ASHRAE 55, ASHRAE Standard (2010). “Standard 55-2010”. In: *Thermal environmental conditions for human occupancy*, p. 12.
- (2013). “Standard 55-2013”. In: *Thermal environmental conditions for human occupancy*, p. 12.
- Auffenberg, Frederik, Sebastian Stein, and Alex Rogers (2015). “A personalised thermal comfort model using a Bayesian network”. In: *Twenty-Fourth International Joint Conference on Artificial Intelligence*.
- Bauer, Eric, Daphne Koller, and Yoram Singer (1997). “Update rules for parameter estimation in Bayesian networks”. In: *Proceedings of the Thirteenth conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc., pp. 3–13.

- Bauman, Fred S, Edward A Arens, et al. (1996). "Task/ambient conditioning systems: engineering and application guidelines". In:
- Bhatia, A (2011). "Centralized Vs Decentralized Air Conditioning Systems". In: *Course No: M05-012, Continuing Education and Development, Inc., NY, info@cedengineering.com.*
- Boman, Magnus et al. (1998). "Energy saving and added customer value in intelligent buildings". In: *Third International Conference on the Practical Application of Intelligent Agents and Multi-Agent Technology*, pp. 505–517.
- Chaudhuri, Tanaya et al. (2017). "Machine learning based prediction of thermal comfort in buildings of equatorial Singapore". In: *2017 IEEE International Conference on Smart Grid and Smart Cities (ICSGSC)*. IEEE, pp. 72–77.
- Chaudhuri, Tanaya et al. (2018). "Random forest based thermal comfort prediction from gender-specific physiological parameters using wearable sensing technology". In: *Energy and Buildings* 166, pp. 391–406.
- Choi, Joon Ho (2010). "CoBi: bio-sensing building mechanical system controls for sustainably enhancing individual thermal comfort". PhD thesis. figshare.
- Choi, Joon-Ho and Vivian Loftness (2012). "Investigation of human body skin temperatures as a bio-signal to indicate overall thermal sensations". In: *Building and Environment* 58, pp. 258–269.
- Choi, Joon-Ho, Vivian Loftness, and Azizan Aziz (2012). "Post-occupancy evaluation of 20 office buildings as basis for future IEQ standards and guidelines". In: *Energy and buildings* 46, pp. 167–175.

- Cook, Diane J, Juan C Augusto, and Vikramaditya R Jakkula (2009). "Ambient intelligence: Technologies, applications, and opportunities". In: *Pervasive and Mobile Computing* 5.4, pp. 277–298.
- Crawley, Drury B et al. (2001). "EnergyPlus: creating a new-generation building energy simulation program". In: *Energy and buildings* 33.4, pp. 319–331.
- Dai, Changzhi et al. (2017). "Machine learning approaches to predict thermal demands using skin temperatures: Steady-state conditions". In: *Building and Environment* 114, pp. 1–10.
- Dalamagkidis, Konstantinos et al. (2007). "Reinforcement learning for energy conservation and comfort in buildings". In: *Building and environment* 42.7, pp. 2686–2698.
- Daum, David, Frédéric Haldi, and Nicolas Morel (2011). "A personalized measure of thermal comfort for building controls". In: *Building and Environment* 46.1, pp. 3–11.
- De Dear, Richard and Gail Schiller Brager (1998). "Developing an adaptive model of thermal comfort and preference". In:
- Doherty, T and Edward A Arens (1988). "Evaluation of the physiological bases of thermal comfort models". In:
- Dounis, AI and DE Manolakis (2001). "Design of a fuzzy system for living space thermal-comfort regulation". In: *Applied Energy* 69.2, pp. 119–144.

- Dounis, Anastasios I and Christos Caraiscos (2009). "Advanced control systems engineering for energy and comfort management in a building environment—A review". In: *Renewable and Sustainable Energy Reviews* 13.6-7, pp. 1246–1261.
- EIA (2019). *How much energy is consumed in U.S. residential and commercial buildings?* URL: <https://www.eia.gov/tools/faqs/faq.php?id=86&t=1> (visited on 05/14/2019).
- Fanger, P Ole and Jørn Toftum (2002). "Extension of the PMV model to non-air-conditioned buildings in warm climates". In: *Energy and buildings* 34.6, pp. 533–536.
- Fanger, Poul O et al. (1970). "Thermal comfort. Analysis and applications in environmental engineering." In: *Thermal comfort. Analysis and applications in environmental engineering*.
- François-Lavet, Vincent et al. (2018). "An introduction to deep reinforcement learning". In: *Foundations and Trends® in Machine Learning* 11.3-4, pp. 219–354.
- Frontczak, Monika and Pawel Wargocki (2011). "Literature survey on how different factors influence human comfort in indoor environments". In: *Building and environment* 46.4, pp. 922–937.
- Gagge Fobelets, Berglund (1986). "A standard predictive index of human response to the thermal environment". In: *ASHRAE Trans.; (United States)* 92:2B.

- Gandhi, Dhiraj, Lerrel Pinto, and Abhinav Gupta (2017). "Learning to fly by crashing". In: *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, pp. 3948–3955.
- Ghahramani, Ali, Farrokh Jazizadeh, and Burcin Becerik-Gerber (2014). "A knowledge based approach for selecting energy-aware and comfort-driven HVAC temperature set points". In: *Energy and Buildings* 85, pp. 536–548.
- Ghahramani, Ali, Chao Tang, and Burcin Becerik-Gerber (2015). "An online learning approach for quantifying personalized thermal comfort via adaptive stochastic modeling". In: *Building and Environment* 92, pp. 86–96.
- Ghahramani, Ali et al. (2016). "Infrared thermography of human face for monitoring thermoregulation performance and estimating personal thermal comfort". In: *Building and Environment* 109, pp. 1–11.
- Hagras, Hani et al. (2003). "A hierarchical fuzzy-genetic multi-agent architecture for intelligent buildings online learning, adaptation and control". In: *Information Sciences* 150.1-2, pp. 33–57.
- Hasan, Mohammad H, Fadi Alsaleem, and Mostafa Rafaie (2016). "Sensitivity study for the PMV thermal comfort model and the use of wearable devices biometric data for metabolic rate estimation". In: *Building and Environment* 110, pp. 173–183.
- Höppe, Peter (1999). "The physiological equivalent temperature—a universal index for the biometeorological assessment of the thermal environment". In: *International journal of Biometeorology* 43.2, pp. 71–75.
- Hoyt, Tyler et al. (2005). "Energy savings from extended air temperature set-points and reductions in room air mixing". In:

- Huizenga, Charlie et al. (2006). "Air quality and thermal comfort in office buildings: results of a large indoor environmental quality survey". In:
- Indraganti, Madhavi, Ryoza Ooka, and Hom B Rijal (2015). "Thermal comfort in offices in India: Behavioral adaptation and the effect of age and gender". In: *Energy and Buildings* 103, pp. 284–295.
- Indraganti, Madhavi and Kavita Daryani Rao (2010). "Effect of age, gender, economic group and tenure on thermal comfort: a field study in residential buildings in hot and dry climate with seasonal variations". In: *Energy and buildings* 42.3, pp. 273–281.
- Jazizadeh, Farrokh and Wooyoung Jung (2018). "Personalized thermal comfort inference using RGB video images for distributed HVAC control". In: *Applied Energy* 220, pp. 829–841.
- Jazizadeh, Farrokh, Franco Moiso Marin, and Burcin Becerik-Gerber (2013). "A thermal preference scale for personalized comfort profile identification via participatory sensing". In: *Building and Environment* 68, pp. 140–149.
- Jazizadeh, Farrokh et al. (2011). "Continuous sensing of occupant perception of indoor ambient factors". In: *Computing in Civil Engineering (2011)*, pp. 161–168.
- Jazizadeh, Farrokh et al. (2012). "Human-building interaction for energy conservation in office buildings". In: *Construction Research Congress 2012: Construction Challenges in a Flat World*, pp. 1830–1839.
- Johnson, Curtis D (1999). *Process control instrumentation technology*. Prentice Hall PTR.

- Kanarachos, A and K Geramanis (1998). "Multivariable control of single zone hydronic heating systems with neural networks". In: *Energy Conversion and Management* 39.13, pp. 1317–1336.
- Karjalainen, Sami (2012). "Thermal comfort and gender: a literature review". In: *Indoor air* 22.2, pp. 96–109.
- Keras (2019). *Keras: The Python Deep Learning library*. URL: <https://keras.io/>.
- Kim, Joyce, Stefano Schiavon, and Gail Brager (2018). "Personal comfort models—A new paradigm in thermal comfort for occupant-centric environmental control". In: *Building and Environment* 132, pp. 114–124.
- Kojima, Kazuyuki (2011). "Sensor network for detecting human's thermal comfort considering of individuals". In: *2011 International Conference on Electronic Devices, Systems and Applications (ICEDSA)*. IEEE, pp. 176–181.
- Kolokotsa, D et al. (2002). "Genetic algorithms optimized fuzzy controller for the indoor environmental management in buildings implemented using PLC and local operating networks". In: *Engineering Applications of Artificial Intelligence* 15.5, pp. 417–428.
- Kruse, Rudolf (2008). *Fuzzy neural network*. URL: http://www.scholarpedia.org/article/Fuzzy_neural_network (visited on 10/24/2008).
- Kurz, Andrea (2008). "Physiology of thermoregulation". In: *Best Practice & Research Clinical Anaesthesiology* 22.4, pp. 627–644.

- Lam, Abraham Hang-yat, Yi Yuan, and Dan Wang (2014). “An occupant-participatory approach for thermal comfort enhancement and energy conservation in buildings”. In: *Proceedings of the 5th international conference on Future energy systems*. ACM, pp. 133–143.
- Lamoudi, Mohamed Yacine, Mazen Alamir, and Patrick Béguey (2011). “Distributed constrained model predictive control based on bundle method for building energy management”. In: *2011 50th IEEE Conference on Decision and Control and European Control Conference*. IEEE, pp. 8118–8124.
- learn, scikit (2019). *scikit-learn*. URL: <https://scikit-learn.org/stable/>.
- Li, Da, Carol C Menassa, and Vineet R Kamat (2018). “Non-intrusive interpretation of human thermal comfort through analysis of facial infrared thermography”. In: *Energy and Buildings* 176, pp. 246–261.
- Lin, Long-Ji (1993). *Reinforcement learning for robots using neural networks*. Tech. rep. Carnegie-Mellon Univ Pittsburgh PA School of Computer Science.
- Liu, Hong et al. (2014). “The response of human thermal perception and skin temperature to step-change transient thermal environments”. In: *Building and environment* 73, pp. 232–238.
- Liu, Shichao (2018). “Personal thermal comfort models based on physiological parameters measured by wearable sensors”. In:
- Liu, Simeng and Gregor P Henze (2006). “Experimental analysis of simulated reinforcement learning control for active and passive building thermal storage inventory: Part 2: Results and analysis”. In: *Energy and buildings* 38.2, pp. 148–161.

- Liu, Weiwei, Zhiwei Lian, and Bo Zhao (2007). "A neural network evaluation model for individual thermal comfort". In: *Energy and Buildings* 39.10, pp. 1115–1122.
- Loftness, Vivian et al. (2009). "The value of post-occupancy evaluation for building occupants and facility managers". In: *Intelligent Buildings International* 1.4, pp. 249–268.
- Lu, Lu et al. (2005). "HVAC system optimization—in-building section". In: *Energy and Buildings* 37.1, pp. 11–22.
- Ma, Yudong, Garrett Anderson, and Francesco Borrelli (2011). "A distributed predictive control approach to building temperature regulation". In: *Proceedings of the 2011 American Control Conference*. IEEE, pp. 2089–2094.
- Mnih, Volodymyr et al. (2013). "Playing atari with deep reinforcement learning". In: *arXiv preprint arXiv:1312.5602*.
- Mnih, Volodymyr et al. (2015). "Human-level control through deep reinforcement learning". In: *Nature* 518.7540, p. 529.
- Mnih, Volodymyr et al. (2016). "Asynchronous methods for deep reinforcement learning". In: *International conference on machine learning*, pp. 1928–1937.
- Moroşan, Petru-Daniel et al. (2010). "Building temperature regulation using a distributed model predictive control". In: *Energy and Buildings* 42.9, pp. 1445–1452.
- Nouvel, Romain and Franck Alessi (2012). "A novel personalized thermal comfort control, responding to user sensation feedbacks". In: *Building Simulation*. Vol. 5. 3. Springer, pp. 191–202.

- OpenAI (2019). *A Toolkit for Developing and Comparing Reinforcement Learning Algorithms*. URL: <https://gym.openai.com/>.
- Pan, Xinlei et al. (2017). "Virtual to real reinforcement learning for autonomous driving". In: *arXiv preprint arXiv:1704.03952*.
- Parsons, Ken (2014). *Human thermal environments: the effects of hot, moderate, and cold environments on human health, comfort, and performance*. CRC press.
- Pérez-Lombard, Luis, José Ortiz, and Christine Pout (2008). "A review on buildings energy consumption information". In: *Energy and buildings* 40.3, pp. 394–398.
- Samet, Jonathan M and John D Spengler (2003). "Indoor environments and health: moving into the 21st century". In: *American Journal of Public Health* 93.9, pp. 1489–1493.
- SDE (2018). *Net-Zero Energy Building - SDE4*. URL: <https://www.arch.nus.edu.sg/about/facilities/net-zero-energy-building-sde-4/> (visited on 05/14/2019).
- Shaikh, Pervez Hameed et al. (2014). "A review on optimized control systems for building energy and comfort management of smart sustainable buildings". In: *Renewable and Sustainable Energy Reviews* 34, pp. 409–429.
- Silver, David and Demis Hassabis (2016). "Alphago: Mastering the ancient game of go with machine learning". In: *Research Blog* 9.
- Silver, David et al. (2016). "Mastering the game of Go with deep neural networks and tree search". In: *nature* 529.7587, p. 484.

- Sim, Jai Kyoung, Sunghyun Yoon, and Young-Ho Cho (2018). “Wearable sweat rate sensors for human thermal comfort monitoring”. In: *Scientific reports* 8.1, p. 1181.
- Sim, Soo et al. (2016). “Estimation of thermal sensation based on wrist skin temperatures”. In: *Sensors* 16.4, p. 420.
- Stavropoulos, Thanos G et al. (2012). “BOnSAI: a smart building ontology for ambient intelligence”. In: *Proceedings of the 2nd international conference on web intelligence, mining and semantics*. ACM, p. 30.
- Sutton, Richard S, Andrew G Barto, Francis Bach, et al. (2018). *Reinforcement learning: An introduction, Second edition, in progress*. MIT press.
- Szeliski, Richard (2010). *Computer vision: algorithms and applications*. Springer Science & Business Media.
- Tan, Ming (1993). “Multi-agent reinforcement learning: Independent vs. cooperative agents”. In: *Proceedings of the tenth international conference on machine learning*, pp. 330–337.
- Van Hasselt, Hado, Arthur Guez, and David Silver (2016). “Deep reinforcement learning with double q-learning”. In: *Thirtieth AAAI conference on artificial intelligence*.
- Vázquez-Canteli, José, Jérôme Kämpf, and Zoltán Nagy (2017). “Balancing comfort and energy consumption of a heat pump using batch reinforcement learning with fitted Q-iteration”. In: *Energy Procedia* 122, pp. 415–420.

- Wang, Shengwei and Zhenjun Ma (2008). "Supervisory and optimal control of building HVAC systems: A review". In: *HVAC&R Research* 14.1, pp. 3–32.
- Wang, Ziyu et al. (2015). "Dueling network architectures for deep reinforcement learning". In: *arXiv preprint arXiv:1511.06581*.
- Watkins, Christopher JCH and Peter Dayan (1992). "Q-learning". In: *Machine learning* 8.3-4, pp. 279–292.
- Yang, Rayoung and Mark W Newman (2012). "Living with an intelligent thermostat: advanced control for heating and cooling systems". In: *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*. ACM, pp. 1102–1107.
- Yuan, Yi et al. (2013). "A study towards applying thermal inertia for energy conservation in rooms". In: *ACM Transactions on Sensor Networks (TOSN)* 10.1, p. 7.
- Zhang, Chenlu, Zhiang Zhang, and Vivian Loftness (2019). "Bio-sensing and Reinforcement Learning Approaches for Occupant-centric Control". In: *2019 ASHRAE Annual Conference*.
- Zhang, Hui et al. (2010a). "Comfort, perceived air quality, and work performance in a low-power task–ambient conditioning system". In: *Building and Environment* 45.1, pp. 29–39.
- Zhang, Hui et al. (2010b). "Thermal sensation and comfort models for non-uniform and transient environments: Part I: Local sensation of individual body parts". In: *Building and Environment* 45.2, pp. 380–388.

-
- Zhang, Zhiang et al. (2019). "Whole Building Energy Model for HVAC Optimal Control: A Practical Framework based on Deep Reinforcement Learning". In: *Energy and Buildings*.
- Zhao, Hai-xiang and Frédéric Magoulès (2012). "A review on the prediction of building energy consumption". In: *Renewable and Sustainable Energy Reviews* 16.6, pp. 3586–3592.
- Zhao, Jie et al. (2014). "Occupant behavior and schedule modeling for building energy simulation through office appliance power consumption data mining". In: *Energy and Buildings* 82, pp. 341–355.
- Zhong, Jinghua (2006). "PID controller tuning: A short tutorial". In: *Mechanical Engineering, Purdue University*, pp. 1–10.