

IWIN2024



International Workshop on Informatics

Proceedings of
International Workshop on Informatics

September 1-4, 2024
Yanagawa, Fukuoka



Sponsored by Informatics Society

IWIN2024



International Workshop on Informatics

Proceedings of
International Workshop on Informatics

September 1-4, 2024
Yanagawa, Fukuoka



Sponsored by Informatics Society

Publication office:

Informatics Laboratory

3-41, Tsujimachi, Kitaku, Nagoya 462-0032, Japan

Publisher:

Tadanori Mizuno, President of Informatics

ISBN:

978-4-902523-51-5

Printed in Japan

Table of Contents

Session 1: Systems and Applications 1 1

(Chair: Kei Hiroi) (9:10 - 11:10, Sep. 2)

- (1) Design and Implementation of Power Distribution Simulator with Electric Vehicles 3
Masashi Saito, Wataru Kunimatsu, Hikaru Akutsu, Chiaki Kojima, Kazutoshi Sakakibara, and Hiranao Kawamura
- (2) Function Calling in the Latest Large Language Models for Automated Classification of Non-Functional Requirements in Information Systems ... 9
Kazuhiro Mukaida, Seiji Fukui, Takeshi Nagaoka, Takayuki Kitagawa, Shinpei Ogata, and Kozo Okano
- (3) A Method for Automatic Creation of State Transition Diagrams from Software Requirement Specifications 17
Takeki Ninomiya, Maiko Onishi, Shinpei Ogata, and Kozo Okano
- (4) A Study of Web System for Comparison and Analysis of Cooking Actions Based on Activity Sensing 23
Fuma Kato, Ayato Kumazawa, Katsuhiko Kaji, Kimihiro Mizuno, Katsuhiko Naito, Naoya Chujo, and Tadanori Mizuno

Keynote Speech 1 35

(Chair: Yuichi Tokunaga) (11:20 - 12:20, Sep. 2)

- (I) New World that the Digital Connects and Opens 37
Dr. Seishi Hanaoka

Session 2: Sensing and Analysis 67

(Chair: Tomoya Kitani) (13:20 - 15:50, Sep. 2)

- (5) Design and Implementation of Smartphone Beacons in Stay Estimation System using BLE69
Kota Togawa and Katsuhiko Kaji
- (6) Anomaly Detection for Weather Sensors Based on Relationship Among Multiple Sensors 75
Kiyoto Kimura and Takuya Yoshihiro
- (7) A Study on Accuracy Management Methods for IoT Sensors Considering Spatio-temporal Correlations 83
Taiki Yamada and Takuya Yoshihiro
- (8) A Bicycle Caution Spot Detection Architecture using Brainwave and Driving Data 89
Hayato Tomisu, Naoto Kai, and Tomoki Yoshihisa
- (9) A Social Sensor Simulation Platform for Sensor Evaluation and Tuning95
Kei Hiroi, Akihito Kohiga, Sho Fukaya, and Yoichi Shinoda

Session 3: Systems and Applications 2..... 105

(Chair: Katsuhiko Kaji) (16:10 - 18:10, Sep. 2)

- (10) Individual Information Presentation for Autonomous Mobility107
Yuichi Yaguchi, Takumi Yotsuji, Kouta Imoto, Kazuki Tamegai, Seiji Komiya, and Toshihiro Wakita

- (11) A Method for Automated Viewpoint Adjustment to Eliminate Blind Spots in Drone Remote Operation Using Virtual TPV113
Reo Akamine, Noriaki Takenoue, and Yuichi Tokunaga

- (12) An Efficient Route Planning Method for 3D Modeling of Utility Poles.....119
Koji Yamagishi and Yuichi Tokunaga

- (13) Investigation on Access Performance of Table Partitioning in PostgreSQL125
Yoshiharu Kikuchi, Sora Terao, Takuro Sakiyama, and Mikiko Sode

Session 4: Multimedia and Communication 131

(Chair: Kozo Okano) (8:40 - 11:10, Sep. 3)

- (14) Audio Signal Compression in Surround Environments Applying Wavelet Transform 133
Miyuki Shirai, Yuhi Shuno, Hiroki Yamamoto, Sho Ishikawa, and Mikiko Sode
- (15) A Study on Effectiveness of Emphasis Methods for MR Stamps in 360-degree Internet Live Broadcasting 139
Yoshia Saito
- (16) "It's Still My Turn!": Visualizing Speech Intervals for Enhanced Conversational Satisfaction. 143
Atsuto Kurokochi, Yo Kuwamiya, Anri Nakano, and Minoru Kobayashi
- (17) Empirical Experiment of Sales Promotion Using Parametric Array Loudspeaker and Sales Pitch Audio as Auditory Stimulation 151
Kohdai Nohara, Shigemi Ishida, Hiroshi Inamura, and Junko Nohara
- (18) Controlling Difficulty Levels in Automated Question Generation Using Large Language Models: An Initial Study for Educational Video with ChatGPT 157
Xingyu Cai and Tomoo Inoue

Keynote Speech 2 163

(Chair: Koichi Ishibashi) (11:20 - 12:20, Sep. 3)

- (II) Privacy Issues on Smartglass AI 165
Prof. Masahiko Tsukamoto

Session 5: Network and Security 183

(Chair: Takuya Yoshihiro) (13:20 - 15:20, Sep. 3)

- (19) Machine Learning for Anomaly Detection Based on Communication Traffic Analysis 185

Neo Miyama and Koichi Ishibashi

- (20) Study of application to IoT DEP based on quantitative indicators using AHP 189

Masaki Mitsuuchi, Tetsuya Yokotani, Koichi Ishibashi, Midori Sugaya, and Li Yanzhi

- (21) Estimating Communication Quality in Wireless Sensor Network Using Time-Series Traffic Trend Similarity 195

Yui Kobayashi, Ryoza Kiyohara, and Yoshiaki Terashima

- (22) Analysis of How Documentation Contributes to Coherence-Based Malicious Code Detection 201

Tomoya Sudo and Yoshiaki Terashima

Session 6: Systems and Applications 3 209

(Chair: Yoshia Saito) (15:30 - 17:30, Sep. 3)

- (23) Automated Generation of Extraction Rules for Japanese Functional Requirements 211
Maiko Onishi and Kozo Okano
- (24) An Evaluation of the Effectiveness of Olfactory Stimulation Notifications for Motorcycle Riders 217
Le Thi Thao Van, Li Ao, and Tomoya Kitani
- (25) Competitive Programming Game to Promote Programming Motivation for Non-Programmers by Highlighting Programmer Coolness 225
Hiroki Oka, Ayumi Ohnishi, Takeshi Nishida, Tsutomu Terada, and Masahiko Tsukamoto
- (26) Proposal for a Work Classification System in Cell Production Manufacturing Sites Utilizing Surveillance Cameras 233
Takao Inoue, Kazuki Fukae, and Toru Kobayashi

Message from the General Chairs



It is our great pleasure to welcome all of you to Yanagawa, Japan, for the 18th International Workshop on Informatics (IWIN 2024). This workshop has been held annually by the Informatics Society. Since 2007, the workshops were held in Naples in Italy, Wien in Austria, Hawaii in the USA, Edinburgh in Scotland, Venice in Italy, Chamonix in France, Stockholm in Sweden, Prague in Czech Republic, Amsterdam in Netherlands, Riga in Latvia, Zagreb in Croatia, Salzburg in Austria, and Hamburg in Germany, Wakayama in Japan (virtually), Fukui in Japan (virtually), Wakayama in Japan, and Hokkaido in Japan respectively.

In IWIN 2024, 26 papers were accepted after peer reviewing by the program committee. Based on the papers, six technical sessions were organized in a single-track format, which highlighted the latest research results in the areas such as Systems and Applications, Sensing and Analysis, Multimedia and Communication, and Network and Security. IWIN2024 will also welcome two keynote speakers: Dr. Seishi Hanaoka, General Manager of Center for Digital Services, Hitachi Ltd., Research & Development Group. Prof. Dr. Masahiko Tsukamoto, Department of Electrical and Electronics Engineering, Graduate School of Engineering, Kobe University. We greatly appreciate their participation in our workshop.

We would like to thank all the participants and contributors who made the workshop possible. It is indeed an honor to work with a large group of professionals around the world for making the workshop a great success. We are looking forward to seeing you all in the workshop. We hope you enjoy IWIN 2024.

September 2024

Koichi Ishibashi
Yuichi Tokunaga

Organizing Committee

General Co-Chairs

Yuichi Tokunaga (Kanazawa Institute of Technology, Japan)
Koichi Ishibashi (Kanazawa Institute of Technology, Japan)

Steering Committee

Hitoshi Aida (The University of Tokyo, Japan)
Toru Hasegawa (Osaka University, Japan)
Teruo Higashino (Kyoto Tachibana University, Japan)
Tadanori Mizuno (Aichi Institute of Technology, Japan)
Jun Munemori (The Open University of Japan, Japan)
Yuko Murayama (Tsuda University, Japan)
Ken-ichi Okada (Keio University, Japan)
Norio Shiratori (Chuo University / Tohoku University, Japan)
Osamu Takahashi (Future University Hakodate, Japan)

Program Co-Chair

Yoshitaka Nakamura (Kyoto Tachibana University, Japan)
Yu Enokibori (Nagoya University, Japan)

Financial Chair

Tomoya Kitani (Shizuoka University, Japan)

Publicity Chair

Yoshitaka Nakamura (Kyoto Tachibana University, Japan)

Program Committee

Akihito Hiromori (Osaka University, Japan)	Hironobu Abe (Tokyo Denki University, Japan)
Akira Uchiyama (Osaka University, Japan)	Hiroshi Horikawa (Shizuoka University, Japan)
Fumiaki Sato (Toho University, Japan)	Hiroshi Mineno (Shizuoka University, Japan)
Hideyuki Takahashi (Tohoku Gakuin University, Japan)	Hiroshi Sugimura (Kanagawa Institute of Technology, Japan)

Kanae Matsui (Tokyo Denki University, Japan)	Tetsuya Shigeyasu (Prefectural University of Hiroshima, Japan)
Katsuhiro Naito (Aichi Institute of Technology, Japan)	Tetsuya Yokotani (Kanazawa Institute of Technology, Japan)
Kei Utsugi (Hitachi, Ltd, Japan)	Tomoo Inoue (University of Tsukuba, Japan)
Keiichi Abe (Kanagawa Institute of Technology, Japan)	Tomoya Kitani (Shizuoka University, Japan)
Ken Ohta (NTT DOCOMO, Japan)	Tomoyuki Yashiro (Chiba Institute of Technology, Japan)
Koji Akiyama (Tokyo Denki University, Japan)	Toshihiro Wakita (Kanagawa Institute of Technology, Japan)
Makoto Imamura (Tokai University, Japan)	Yoh Shiraishi (Future University Hakodate, Japan)
Manato Fujimoto (Osaka Metropolitan University, Japan)	Yoshiaki Terashima (Soka University, Japan)
Masaaki Shirase (Future University Hakodate, Japan)	Yoshinobu Kawabe (Aichi Institute of Technology, Japan)
Masaji Katagiri (Professional University of Information and Management for Innovation, Japan)	Yuhichiro Nomura (Shizuoka University, Japan)
Masakatsu Nishigaki (Shizuoka University, Japan)	Yuichi Bannai (Kanagawa Institute of Technology, Japan)
Masaki Endo (Polytechnic University, Japan)	Yuki Koizumi (Osaka University, Japan)
Masaki Nagata (Shizuoka University, Japan)	Yukimasa Nagai (Mitsubishi Electric Corporation, Japan)
Masashi Saito (Kanazawa Institute of Technology, Japan)	Yusuke Gotoh (Okayama University, Japan)
Mikiko Sode (Kanazawa Institute of Technology, Japan)	
Minoru Kobayashi (Meiji University, Japan)	
Nobuhiro Kobayashi (University of Nagasaki, Japan)	
Orie Abe (Information Communication Media Laboratory, Japan)	
Satoru Matsumoto (Osaka University, Japan)	
Shigemi Ishida (Future University Hakodate, Japan)	
Shigeyoshi Ohno (Polytechnic University, Japan)	
Shinichiro Mori (Chiba Institute of Technology, Japan)	
Shinji Kitagami (Fukui University of Technology, Japan)	
Shoji Sano (Kanazawa Institute of Technology, Japan)	
Takaaki Umedu (Shiga University, Japan)	
Takaya Yuizono (Japan Advanced Institute of Science and Technology)	
Takayasu Yamaguchi (Akita Prefectural University, Japan)	
Tetsushi Ohki (Shizuoka University, Japan)	

Session 1:
Systems and Applications 1
(Chair: Kei Hiroi)

Design and Implementation of Power Distribution Simulator with Electric Vehicles

Masashi Saito*, Wataru Kunimatsu*, Hikaru Akutsu**, Chiaki Kojima**,
Kazutoshi Sakakibara†, Hironao Kawamura‡

* Graduate School of Engineering, Kanazawa Institute of Technology, Japan

** Department of Electrical and Electronic Engineering, Toyama Prefectural University, Japan

† Department of Data Science, Toyama Prefectural University, Japan

‡ Innovation Laboratory, Hokuriku Electric Power Company

msaito@neptune.kanazawa-it.ac.jp, b1846432@planet.kanazawa-it.ac.jp

hikaru_akutsu@pu-toyama.ac.jp, chiaki@pu-toyama.ac.jp, sakakibara@pu-toyama.ac.jp
h.kawamura@rikuden.co.jp

Abstract - We have developed a simulator to establish energy management algorithms for realizing power distribution over wide areas using electric vehicles. This simulator is for a service called a virtual power distribution network, which not only eliminates waste power from solar power generation, but also cuts peak power demand.

This simulator supports the area being 20km square, with roads running north-south and east-west every 1km. Using this simulator, we report the first simulation results, which is to apply the designing of EVs' behavior algorithm.

The simulation results show that we should prepare enough but not excessive power charging/discharging points to support power distribution using EVs.

Keywords: Power Distribution, EV, Wide Area Simulation

1 INTRODUCTION

The introduction of renewable energy is accelerating toward a decarbonized society, and photovoltaic power generation (PV) is attracting attention. However, due to the shutdown of some power plants due to natural disasters such as earthquakes, and the increased use of air conditioning equipment due to falling temperatures, electricity demand has increased, and it is becoming an issue that daily peak demand for electricity cannot be met [1] [2].

On the other hand, the amount of electricity generated by renewable energy in Japan increased about four times in fiscal 2020 compared to fiscal 2012, resulting in a surplus of electricity supply, and output control from April to September 2023 reached 194 times, about three times the previous year. In areas with high electricity demand, such as the Tokyo metropolitan area, requests to save electricity were made, and there were concerns about a shortage of power supply. Such unstable supply and demand of electricity could lead to power outages.

In addition, there is a situation where new construction of power transmission network facilities is not progressing due to opposition from surrounding areas, and measures to deal with the aging of existing power transmission networks are also a heavy burden.

The Japanese government plans to increase the proportion of renewable energy to 36-38% in FY2030, which is about double the proportion in FY2019, and has estimated that 6

trillion to 7 trillion yen will be needed to develop the power grid in 2022 [3]. Therefore, it can be said that conventional power generation and transmission capacity is not enough to meet the demand.

Sustainable and stable supply of energy, including electricity, as well as decarbonization and low carbonization are important issues, and from this perspective, the introduction of renewable energy in the power sector and the introduction of EVs (electric vehicles) in the transportation sector are expanding. EVs are equipped with large-capacity storage batteries as a power source, so they are expected to be used not only as mobile vehicles but also in the power sector. Assuming that the penetration rate of EVs increases to 50% of new car sales by 2050, it is estimated that the power consumption used by EVs at that time will increase to 46.3 billion kWh, up 164.3% from 2016 [4].

EVs are equipped with large-capacity storage batteries as a power source, so they are expected to be used not only as mobile vehicles but also in the power sector. Representative examples include (i) providing energy services to the power grid and consumers by connecting EVs to the grid and charging/discharging their storage batteries, and (ii) serving as an emergency power source during disasters or grid blackouts.

In this paper, to realize the role of (i) above, we have developed a simulator that uses EVs to transport electricity to places where there is a power shortage. Using the developed simulator, we report the first evaluation results to see whether a power transmission system can be established without changing the current grid system.

2 RELATED WORKS

Ichii et al. [5] examined the possibility of EVs replacing privately operated power lines as a means of making up for power shortages in smart towns, which is an important issue in the use of renewable energy. As a result, the possibility of EVs replacing privately operated power lines was low in satellite smart towns (SSTs) close to distributed power sources, but in small, remote SSTs, the transportation costs of EVs were lower than those of privately operated power lines, indicating the possibility of EVs replacing privately operated power lines. In addition, with improvements in electricity self-sufficiency and advances in DC power supply

technology, it is expected that costs will be reduced by using EVs. It was also suggested that there is a high possibility of low-cost use of EVs for electricity transportation. This suggests that costs can be reduced for SSTs by having EVs take on the role of electricity transportation, and that transmission separation could be an opportunity to promote the use of renewable energy.

Tanoue et al. [6] aimed to build a support system for power transportation by EVs. Although the power consumption characteristics of EVs vary depending on the driving state, distance, and altitude; to facilitate the use of geographic information systems, they proposed a method to estimate the power consumption of EVs while driving by separately expressing the power consumption of distance and altitude difference. As a result, they derived an equation for estimating the power consumption of EVs while driving using geographic information. They also showed that it is possible to visually grasp the EV power by visualizing it using the derived estimation equation.

Urabe et al. [7] designed a new EV infrastructure system that realizes peak shift of power by providing power (batteries) to users within the power supply area using EVs and considered the leveling of power demand. As a result, they showed the effectiveness of efficiently using the batteries provided to users from three perspectives: the optimal number of EV infrastructure systems to be placed within the area, the optimal facility placement of EV infrastructure systems within the area, and a plan for providing power to users within the area using the EV infrastructure system.

Murakami [8] used agent-based simulation to examine the problem of power load concentration due to the spread of electric vehicles (EVs). He showed that random charging start times can sometimes achieve high power load leveling effects. He showed that autonomous information exchange and interactions among EV agents can achieve higher power load leveling. He focused on information exchange within small groups and suggested that power load leveling can be achieved even in large-scale systems without information exchange between independent groups.

Takagi et al. [9] focused on autonomous decentralized control that did not require additional investment in communication infrastructure, and proposed two measures to mitigate the impact on the grid that reflected the distribution characteristics of daily driving distances and ensured convenience for EV users. As a result of the evaluation, it was found that the magnitude of the steep peak was proportional to the number of EVs charging at the same time, so along with the EV penetration rate, how frequently EV users charge was also an important factor. However, since the frequency of charging depended on various uncertain factors such as the battery capacity of EVs that would become more widespread in the future and the specifications of chargers, it was shown that even under the conditions assumed in this paper where charging impacts were most likely to occur, it was possible to mitigate the peak due to EV charging load by adopting a quadratic function case, etc.

Moriya et al. [10] proposed a method of power trading using virtual power plants in smart grids and the use of PHEVs and EVs as storage batteries to reduce peak power demand across

the entire power grid, but they did not consider the driving of PHEVs and EVs. They also proposed a method of power leveling that considered the driving of PHEVs and EVs. As a result, when comparing timer charging and bottom charging, it was found that timer charging had a greater leveling effect when there were many EVs, and they showed that the power load could be reduced by using EVs.

Mustapha Aachiq et al. [11] considered electric vehicles (EVs) as a means of adjusting demand in homes and created a model to simulate what kind of EV battery operation would be optimal for consumers when linking EV batteries with PV systems under the electricity pricing system that will become more diverse in the future. The results showed that the impact of FIT prices on EV battery operation and the economics of each household is that when the FIT PV purchase price falls, the amount of reverse power flow is suppressed and the amount of charging from PV to EV increases. In addition, it is suggested that V2G can contribute to load leveling in the grid if a large number of PHEVs are introduced and electricity prices in homes are determined in conjunction with the marginal fuel cost of the grid.

These studies showed the possibility of transmitting power over a small area and meeting electricity demand.

In small-scale areas, it is easy to link EVs and power storage systems, and it is possible to respond to peak power demand even after power demand has increased. However, in medium-scale or larger areas, it is difficult to link EVs and energy storage systems, so it is necessary to evaluate whether it is possible to meet the power demand in areas with a power shortage through evaluation. Therefore, evaluation is carried out in a medium-scale area assuming an increased number of vehicles and a map of 20 km x 20 km.

3 OVERVIEW OF ELECTRIC POWER TRANSMISSION USING EVs

3.1 Electric power transmission by EV

We have set up a use case to realize a charging/discharging service using EVs in medium-sized area energy management and evaluate it using a simulator to investigate the impact on the supply and demand balance. Specifically, we have built a system in which surplus electricity generated by PV is charged into EVs that stop at charging/discharging points, and then transmitted to other charging/discharging points for discharging. An overview of power transmission is shown in Figure 1.

3.2 Overview of our EV energy management system simulator

The EV energy management system is positioned between existing energy management (power control) systems and MaaS (Mobility as a Service) services, which will continue to develop in the future, and proposes charging and discharging services to moving EVs through road network. Figure 2 shows the overview of the EV energy management system.

Unfortunately, there is no MaaS system that can be freely used in an EV energy management system, and the energy management system cannot be used. Therefore, we have developed the "EV Energy Management System Simulator (hereinafter EVEMSS)" to simulate electricity transportation. EVEMSS is composed of a vehicle information simulator instead of a MaaS system, an EV energy management system, and an existing energy management system that provides data on the grid Database that manages the power demand and PV power generation managed by the energy management system (Figure 3).

By using EVEMSS to investigate and analyze the impact in various use cases, it is possible to use it as a basic tool for verifying and proposing new power supply systems and their services.

The EV energy management system is a subsystem that realizes customer value by customers who drive EVs stopping at places called charging/discharging points. This subsystem provides a buying and selling intermediation service by effectively utilizing surplus electricity generated by solar power generation, realizing a win-win relationship in which EV owners can purchase cheap electricity and PV owners can sell waste electricity that they are forced to discard. Through such intermediation, we hope to determine the possibility of creating new service businesses.

3.3 EVEMSS system configuration

The system configuration of the EVEMSS is shown in Figure 4. The vehicle information simulator drives the vehicle using data for vehicle movement, road networks, and charging/discharging point data. When charging/discharging becomes necessary, it queries the EV energy management system to determine the route and destination. The EV energy management system manages the amount of electricity and buying/selling prices at charging/discharging points, as well as managing the power logs of the charging/discharging points. The existing energy management system stores data

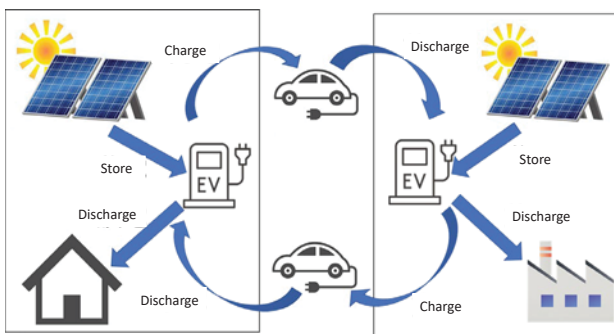


Figure 1: An overview of electricity transportation

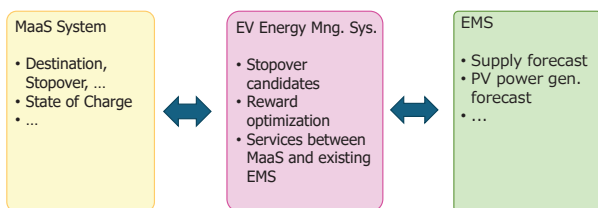


Figure 2: EV Energy Management System Overview

on the amount of electricity generated and demand at each location.

3.4 Road Network in EVEMSS

The system has been currently built assuming a Manhattan model road network. A tertiary mesh code is used for this purpose. The distance of a tertiary mesh code is not constant because the Earth is a sphere, but in the Hokuriku region it is 926 m north to south and 1,118 m east to west. EV operation is simulated on a 20 x 20 grid road network divided into tertiary meshes [13].

The roads defined above have a graph (network) structure. In order to make it possible to represent any road in any location, we are using Neo4j, a NoSQL database that makes it easy to model graph structures.

We assume that the data on charging and discharging points is held by the MaaS service. Each node stores the amount of battery power held in the battery in that area, the price at the time of buying and selling and so on.

4 SIMULATION SETTINGS AND RESULTS

4.1 Use cases used in the simulation

To design EVs' behavior algorithm, we start from simple use cases with respect to charging/discharging points, the number of EVs, and power demand/generation volume.

The number of charging/discharging points was set to {5, 10, 30, 50, 100}. Figure 5 shows the layout of the number of charging/discharging points being 100, and the parts colored marked represent their positions.

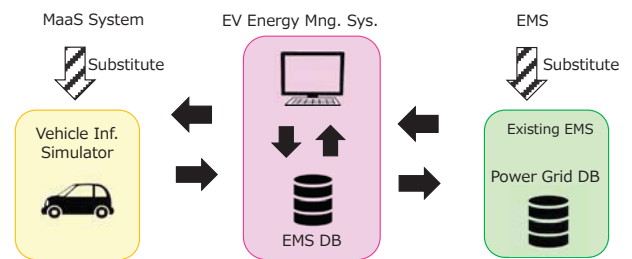


Figure 3: EV Energy Management System Simulator Overview

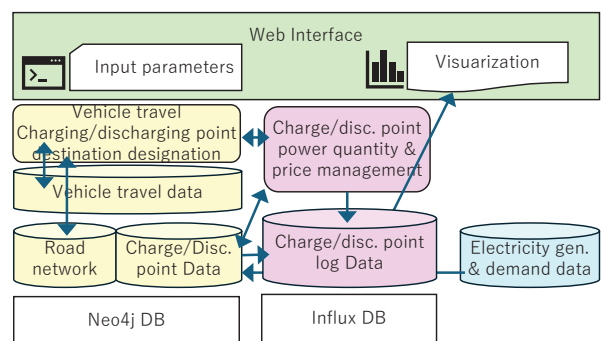


Figure 4: EVEMSS System Configuration

The layout settings of 5, 10, 30, and 50 are set in order of proximity to the central charging/discharging point.

Regarding the use cases for power demand and power generation, power demand was set to 0 and power generation to infinity in order to grasp the PV infrastructure of the minimum number of charging/discharging points required for the number of EVs.

Each vehicle is prepared to make two laps around the perimeter within the set range, and it drives along the red line and in the direction of the arrow. The initial battery value of the vehicle is set to a random value between 16 and 30 kWh. The first vehicle departs at 6:00 a.m., and after that, each vehicle departs every one minute, completing two laps around the perimeter (Figure 6). Simulations are performed by changing the number of vehicles to {10, 30, 50, 100, 300, 500, 1000}. The departure time of the 1,000th vehicle is 5:40pm.

One lap around the perimeter is about 80 km, and the EV runs at 40 km/h, so that it takes about 4 hours for two laps.

The timing for the vehicle to charge is when the remaining battery level of the vehicle falls below 20%, and the charging/discharging point with the cheapest electricity price is selected, and the vehicle's battery is charged until it becomes full. The price of each charging/discharging points are predefined, center areas are cheap, and perimeter areas are a little bit expensive.

The timing for the vehicle to discharge is when the remaining battery level of the vehicle falls between 80% and 90%. The charging/discharging point with the highest electricity price is selected, and the vehicle's battery is discharged until it is less than 50%.

The vehicle charges and discharges at a rate of 1kWh per minute.

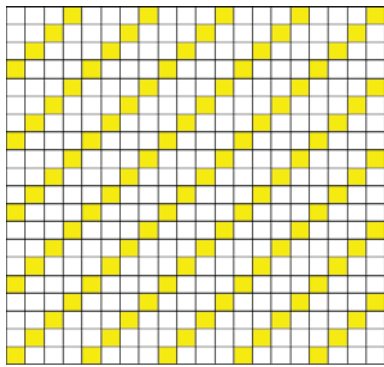


Figure 5: Location of 100 charging/discharging points

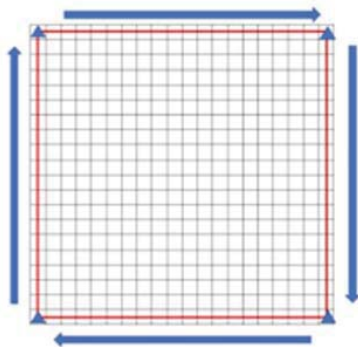


Figure 6: EV use cases

4.2 Vehicle behavior settings

The following conditions are set as vehicle behavior settings:

- When a vehicle charges, it heads to a point with a low price for buying electricity, which is set in advance,
- When discharging, it heads to a point with a high price,
- A charging/discharging point is selected from among locations where charging/discharging points are not in use,
- If the waiting time for charging/discharging is more than 1 hour and 20 minutes (the number of vehicles waiting at the charging/discharging point is more than 2), the vehicle does not head to that location,
- If none of these conditions apply, the vehicle continues to drive toward its destination.

As described above, the EV automatically selects the a kind of optimal charging/discharging point according to the price of electricity and will not head to a point with a long waiting time.

4.3 Simulation results

Table 1 shows the simulation results for the number of vehicles that complete the trip according to the number of charging/discharging points. It can be seen that 300 vehicles completely run out of power when there are 5 charging/discharging points, and 500 vehicles completely run out of power when there were 10 charging/discharging points.

The reason for this might be that if the waiting time for charging/discharging is more than 1 hour and 20 minutes, that is the number of vehicles waiting at the charging/discharging point is more than 2, it continues to drive to the destination. Since the time when charging is required from the start of the first vehicle is within 2 to 4 hours, if the number of vehicles exceeds a certain number, the charging/discharging point run out of power and no longer be in operation. This phenomenon occurs when the number of EVs is extremely large compared to the number of charging/discharging points.

When the number of charging/discharging points is 30, 50, or 100, and the number of vehicles running is 300, there is almost no difference in the number of vehicles that have completed the trip. Also, when the number of charging/discharging points is 50, 100, and the number of vehicles running is 500 or 1,000, there is almost no difference in the number of vehicles that have completed the trip, regardless of the number of charging/discharging points. From this, there is no change in the number of vehicles that

Table 1: The number of completed runs relative to charging/discharging points

# of spots / # of veh.	5	10	30	50	100
10	10	10	10	10	10
30	29	30	30	30	30
50	36	43	50	50	50
100	48	51	65	95	100
300	0	9	272	293	292
500	0	0	395	460	483
1,000	0	0	558	720	682

have completed the trip when a certain number of charging/discharging points are installed. We believe that the reason for this is that the number of times that a charging/discharging point can provide service in one day is 48, and the charging/discharging time stipulates the upper limit.

Figure 7 is a graph of the number of completed runs relative to the number of charging/discharging points. When the number of charging/discharging points is 5 and 10, and the number of running vehicles is 100, only about 50% of the vehicles complete the run at both points. This is due to the EV's behavior settings causing it to run out of power and become unable to run. When the number of charging/discharging points is 50 and 100, and the number of running vehicles is 300, there is almost no difference in the number of completed runs. The number of times that a charge/discharge point can be serviced in a day is 48, and we believe that the charge/discharge time stipulates the upper limit.

Figure 8 is a graph showing the maximum number of completed runs at each charging/discharging point. When expressed as a logarithmic approximation, it is as follows:

$$y = 274.7\log(x) - 447.2$$

In this setting, the number of vehicles that have completed their trips converge once the number of charging/discharging points reaches a certain number.

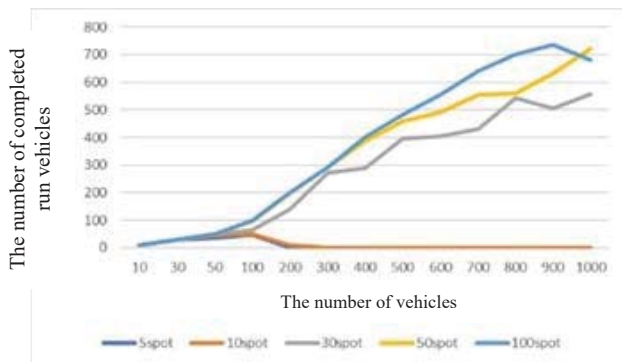


Figure 7: The number of completed run vehicles when the number of charge/discharge points being 5, 10, 30, 50, 100

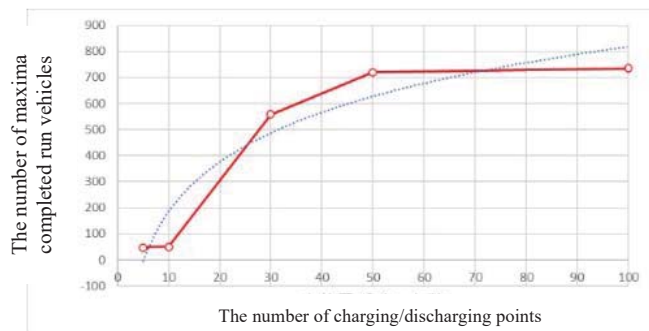


Figure 8: Maximum number of vehicles completed at each charge/discharge point

5 CONCLUDING REMARKS AND FUTURE WORKS

In this paper, we set up specific use cases required for regional energy management that realizes charging and discharging services using EVs and evaluate them using a simulator to investigate the impact on supply and demand balance. The simulator assumes commonly used vehicles, allows free movement, and has specifications that allow the number of vehicles and charging/discharging points to be variable and given as parameters, allowing for various use cases to be set.

In this study, we assume an area of 20×20 squares and examine the feasibility of a service that transports surplus PV power using EVs. As a precursor to this, we reported on the number of charging and discharging facilities required as the number of EVs increases.

It is found that the system run out of power completely at 300 EVs when there are 5 charging /discharging points, and at 500 EVs when there are 10 charging/discharging points. It can be said that the above phenomenon occurs when the number of EVs is extremely high compared to the number of charging/discharging points.

When the number of charging/discharging points is 30, 50, and 100, and the number of EVs traveling is 300, there is almost no difference in the number of vehicles that complete the trip. In addition, there is almost no difference in the number of vehicles that completed the trip, regardless of the number of charging/discharging points, even when the number of charging/discharging points is 50 or 100, and the number of vehicles traveling is 500 or 1000. The number of times that a charging/discharging point can service 48 times in a day, and the charging/discharging time stipulates the upper limit.

The optimal number of charging/discharging points in the settings of this report was shown using the logarithmic approximation of the approximate graph.

In the current simulator, charging/discharging points are set so that only one vehicle can be charged/discharged at each point, so by allowing multiple EVs to charge/dischARGE, it is possible to eliminate the need for charging/discharging points and EV infrastructure. To do this, it is necessary to improve the current simulator and expand the system that manages the number of vehicles that can be charged/discharged.

Using the demand for residential, commercial, and industrial areas, we will use a simulator to verify whether it is possible to transport electricity. Power transportation part-time job mode might help a lot to for peak demand and surplus power disposal.

REFERENCES

- [1] Ministry of Economy, Trade and Industry Japan, "Draft report on the tight electricity supply and demand situation in eastern Japan in March 2022", pp3-7 (2022).
- [2] A. Nakajima et al., "Impact of the shutdown of large-scale generators due to the Great East Japan Earthquake on JEPX electricity spot prices", Proceedings of the Tohoku Branch Joint Conference of the Institute of Electrical Engineers of Japan (2012).
- [3] Yomiuri Shimbun, Electricity Crisis <2> Poor power transmission, renewable energy waste, 2023/11/21.
- [4] M. Fujiyama, "Electricity consumption in 2050 will decrease by 20% compared to 2016 - Population decline and progress in energy conservation will significantly push down electricity consumption", Japan Research Institute Research Focus, No.2018-003 (2018).
- [5] K. Ichii et al., "How to transport electricity to satellite-type smart towns", Urban Planning Papers, No.3, Vol.55, pp569-576 (2020).
- [6] K. Tanoue et al., "Estimation of Electric Vehicle Power Transport Capacity Using Geographic Information Based on Vertical Power Cost Calculation", Proceedings of Kyushu Joint Conference of Institutes of Electrical Engineers of Japan (2015).
- [7] S. Urabe et al., "Optimal Design Method of EV Infrastructure System Considering Power Demand Leveling - By Providing Batteries to Users in Power Supply Areas -", Information Processing Society of Japan, No. 3, 2012-MPS-91, pp1-7 (2012).
- [8] T. Murakami et al., "Agent Simulation of Power Load Leveling When Electric Vehicles are Connected to the Grid", IEEEJ Transactions on Electrical Engineering, Vol. C (Electronics, Information and Systems Division), No. 9, Vol. 133, pp1658-1662 (2013).
- [9] M. Takagi et al., "Nighttime Charging Load Leveling Strategy Considering User Convenience of Electric Vehicles," IEEEJ Transactions on Energy and Power Systems, Vol. 11, Vol. 134, pp. 908-916 (2014).
- [10] T. Moriya et al., "Research and Use of Electricity Trading Using Virtual Power Plants in Smart Grids," Proceedings of the 75th National Conference, Vol. 1, Vol. 201, pp. 783-784 (2013).
- [11] M. Aachiq et al., "Evaluation of Residential PV-EV Systems for Supply-Demand Balance of Power Systems," IEEEJ Transactions on Energy and Power Systems, Vol. 1, Vol. 135, pp. 27-34 (2015).
- [12] The Institute of Electrical Engineers of Japan, "Regional Supply System Model," https://www.iee.jp/pes/ele_systems/base_model/overview/, (accessed Sep. 1, 2023).
- [13] Ministry of Land, Infrastructure, Transport and Tourism, "Third Mesh Land Use Data - "National Land Numerical Information", <https://nlftp.mlit.go.jp/ksj/gml/datalist/KsjTmplt-L03-a.html>, (accessed Jan. 25, 2024)..

Function Calling in the Latest Large Language Models for Automated Classification of Non-Functional Requirements in Information Systems

Kazuhiro Mukaida^{*}, Seiji Fukui^{**}, Takeshi Nagaoka^{**}, Takayuki Kitagawa^{**}, Shinpei Ogata^{*}
and Kozo Okano^{*}

^{*}Graduate School of Science and Technology, Shinshu University, Japan

^{**}Toshiba Corporation, Japan

22w2805g@shinshu-u.ac.jp, {Fukui.Seiji, Nagaoka.Takeshi, Kitagawa.Takayuki}@toshiba-sol.co.jp,
{ogata, okano}@cs.shinshu-u.ac.jp

Abstract - We focus on non-functional requirements (NFRs), which are often overlooked in requirement definitions of system developments and propose a method that allows individuals without extensive expertise to efficiently extract and classify NFRs from requirement specifications. Previously, the authors experimented with classification using pre-trained Transformer models like BERT and GPT-2, trained on human-annotated datasets. However, recent advancements, such as tools like ChatGPT, enable classification via prompt interactions alone. In this paper, we explore the capabilities of ChatGPT's latest Function Calling feature, aiming to demonstrate its superior classification performance compared to prompt-only responses and conventional classification methods which need large training data. Using Function Calling, we can extract structured data directly from conversational prompts, utilizing this feature for precise classification. We also assess the impact of fine-tuning ChatGPT on our specific dataset to determine if customized adjustments further enhance classification accuracy. Our results show that this approach significantly shortens the process of model creation and training while achieving accuracy equal to or surpassing that of traditional methods. The findings highlight the potential of advanced natural language models like ChatGPT in making NFRs classification more efficient and precise.

Keywords: Function Calling, ChatGPT, GPT-4-Toutbo, GPT-4o, Non-Functional Requirements, Documents Classification

1 INTRODUCTION

In the development of information systems, requirements are broadly categorized into functional and non-functional requirements [1]. Functional requirements define the specific functions that the system must perform, while NFRs describe the overall characteristics of the system, such as quality, performance, reliability, and efficiency. NFRs set the expectations for the system's operational environment and are crucial in ensuring the overall quality of the software [2]. The "Non-Functional Requirements Grade" provided by the Information-Technology Promotion Agency (IPA) offers a comprehensive guide line for classifying NFRs, along with graded levels for each requirement, allowing developers and users to confirm and set NFRs based on a common standard.

However, due to the abstract nature of NFRs, their clear identification and classification pose challenges. Consistent

and efficient classification has traditionally required the involvement of experienced experts. To address this issue, automated classification techniques have gained attention, particularly those leveraging deep learning and large language models (LLMs) for document classification. Since the introduction of Transformer-based models such as BERT (Bidirectional Encoder Representations Transformers) and GPT (Generative Pre-trained Transformer) in 2018, there have been significant improvements in classification accuracy. These improvements are achieved by fine-tuning the pre-trained models on large text corpora for specific tasks. Nevertheless, the process of fine-tuning for specific tasks has been burdensome due to the need for task-specific annotated data.

The authors have previously experimented with automated classification developing models incorporating BERT and GPT-2, as documented in [3]. These models were able to achieve a certain level of accuracy of approximately 0.8. Recently, with the proliferation of interactive models like ChatGPT, it has become possible to perform inference through simple prompt interactions. However, there are still limitations to the performance of those models, specifically mistakes arising from insufficient understanding of the context in classification tasks. This paper explores the use of the Function Calling feature associated with GPT models to classify NFRs. Function Calling is primarily intended for invoking other APIs or functions, but in the process, the structured data is generated for suggesting necessary arguments for these calls, utilizing this feature for classification. Additionally, since LLMs can still be fine-tuned by users with small amounts of data, we explored this effect. By labeling the training data with "the Non-Functional Requirements Grade", which serves as the classification criteria, we also facilitated the annotation process.

2 RELATED WORK

2.1 Classification of Non-Functional Requirements

Kinoshita et al. [4] discuss a method for extracting non-functional requirements from requirement documents. Specifically, they set keywords and apply them to the requirement documents to identify non-functional requirements. Furthermore, they propose an approach to

identify errors in non-functional requirements. This method involves defining case frames for the actions described in the requirements and verifying the requirements using criteria for identifying omissions, ambiguities, redundancies, and inconsistencies.

2.2 Classification Using Deep Learning

Gnanasekaran et al. [5] studied and developed a recurrent neural network model, recognized for its effectiveness in processing continuous natural language text. Using this model, they classified non-functional requirements described in natural language into five categories: maintainability, operability, performance, security, and usability. Experimental validation was conducted based on two datasets encompassing approximately 1,000 non-functional requirements.

Kitagawa et al. [6] proposed an automatic classification method using Word2Vec and CNN. This method extends beyond non-functional requirements and project management sentences to include functional requirements and other sentences. It targets all sentences contained in Japanese Requests for Proposal documents.

2.3 Classification Using Large Language Models

Since the introduction of the Transformer by Vaswani et al. [7], research in natural language processing tasks has increasingly leveraged large language models based on the Transformer, replacing recursive neural networks such as RNNs, LSTMs, and GRUs. Fine-tuning large pre-trained language models like BERT [8] for specific tasks has often resulted in optimized models achieving higher accuracy.

Following the emergence of interactive models like ChatGPT in 2022, which allow users to easily issue questions and commands in natural language, these models have demonstrated flexibility across various tasks and become widely used in numerous natural language processing tasks. Interactive models provide a user-friendly interface, enabling users to interact with the models using free-form text without being constrained to specific formats. This allows individuals without programming skills or specialized knowledge in machine learning to utilize these models.

However, models specifically built and trained for particular tasks can understand the unique characteristics of those tasks and make more precise predictions. This is due to the model's deeper understanding of domain-specific language nuances and terminologies. Gutierrez et al. [9] reported that fine-tuned BERT models achieved higher accuracy than interactive GPT-3 models for named entity recognition and relation extraction tasks in the medical field. Sun et al. [10] encouraged interactive models to find superficial cues and then derive final judgments through a diagnostic inference process but found that fine-tuned models for text classification still outperformed interactive models in accuracy. Ibe et al. [11] conducted experiments using large generative language models to classify requirement specifications with minimal examples by adjusting prompts,

and similarly found that BERT-based classification outperformed interactive models.

On the other hand, the performance of ChatGPT as a question answering system (QAS) has been reported to match or even exceed that of traditional task specific QAS. Tan et al. [12] demonstrated that ChatGPT offers superior conversational robustness and explainability compared to traditional QAS.

3 PROPOSED METHOD

This paper proposes an automatic classification method for NFRs in system development specification documents. This method leverages the Function Calling feature of GPT models, a type of large language model, and is based on IPA standards.

3.1 Function Calling

Function Calling is a feature provided by the API of large language models, enabling the model to indirectly interact with external APIs or systems and generate specific actions. This technology involves parsing natural language queries, selecting appropriate functions, and generating JSON responses. For example, in response to the query "Tell me about the weather in Tokyo," the model suggests calling a function that uses a weather API with the region name as an argument. The argument "Tokyo" is prepared for the weather API call. The program then retrieves Tokyo's weather through the weather API and passes this information back to the model as an additional prompt. The model then provides the final answer, significantly expanding the potential for interactive applications and services using the model. This feature bridges the gap between natural language understanding and the execution of specific actions, enhancing the model's practicality and allowing it to flexibly respond to user requests. Function Calling not only streamlines the interaction between the model and external systems but also improves the overall efficiency of handling user queries by automating the process of converting natural language into actionable commands.

3.2 Classification using Function Call

A derivative use of Function Calling involves extracting structured data from text. For example, a function like `extract_data` (name: string, birthday: string) can be defined and invoked as needed, extracting a person's name and birthday from the text as arguments. This functionality allows for the rapid and accurate extraction of necessary information from large volumes of text data, facilitating efficient data analysis and information management. This method allows for a more structured and organized approach to handling text data, ensuring that the extracted information is consistent and accurate.

Function Calling can indeed help in selecting the most appropriate argument based on the overall context of a query, even if the specific term isn't explicitly mentioned. The technology leverages the model's natural language understanding to infer the intended meaning and relevant information from the user's input.

In this paper, instead of extracting information from documents, we performed document classification using this feature. We forced the invocation of a dummy function from prompts providing evaluation sentences, suggesting classification categories as arguments. By specifying the argument type as an enum, we enumerated the classification categories, ensuring that suggestions were made only from specific choices. By leveraging Function Calling for classification, we can automate and streamline the process, reducing the potential for human error and increasing overall efficiency.

3.3 Fine Tuning for GPT-3.5

Furthermore, this paper attempts to improve models with classification accuracy by additional training of GPT-3.5 Turbo with a small amount of training data. Fine-tuning enables the model to learn language expressions specific to NFRs tasks or domains. This helps the model understand specialized terminology and contexts that are challenging for general language models, achieving higher classification accuracy. Moreover, fine-tuning with a small amount of training data significantly reduces the time and cost associated with preparing annotated large datasets, compared to traditional methods. For complex tasks like NFRs classification, it is crucial to learn efficiently from limited examples.

Fine-tuning not only tailors the model to specific tasks but also enhances its ability to handle nuanced and domain-specific language, resulting in more precise and reliable classifications. This approach is feasible even in scenarios where annotated data is scarce, enabling effective learning and adaptation with minimal resources.

4 EVALUATION EXPERIMENT

4.1 Models and Test Data Set

In this paper, we utilized the ChatGPT API with the following models:

GPT-4 Variants

- gpt-4o-2024-05-13
- gpt-4-turbo-2024-04-09

GPT-3.5 Turbo

- gpt-3.5-turbo-0125

The gpt-4o-2024-05-13 model represents the latest iteration of the GPT-4 series, incorporating the most recent advancements in language understanding and generation capabilities.

The gpt-4-turbo-2024-04-09 model is offering optimized performance and a popular choice for a wide range of practical applications.

The gpt-3.5-turbo-0125 model, although from a previous generation, provides a valuable comparison point. It offers cost-effective performance and supports fine-tuning, allowing for customization to specific tasks or domains. This makes the gpt-3.5-turbo-0125 model advantageous for scenarios where budget constraints and the need for tailored solutions are paramount.

Table 1 Number of Test Data Instances for Each Classification

Availability	81
Performance/Scalability	70
Operability/Maintainability	90
Migratability	60
Security	90
System Environment/Ecology	23
Total	414

We specified that the argument type for function calling properties is enum. For enum types, we defined "availability, performance/scalability, operability/maintainability, migratability, security, and system environment" based on the classification of "non-functional requirements grades." Each model provided responses for each evaluation sentence. The temperature was set to 0 to ensure deterministic outputs.

The primary focus of the experiment is to verify the superiority of Function Calling over the baseline case, which relies solely on prompt interactions for classification. At the same time, we are able to observe the differences between models. Using these models enables us to compare the advantages of Function Calling across different model generations. By assessing the performance of Function Calling among GPT-3.5 Turbo and GPT-4 variants, we can determine how advancements in the models contribute to enhanced Function Calling capabilities. This comparison provides valuable insights into their practical applications and efficiency, highlighting the strengths and improvements of the newer models over their predecessors.

The test dataset used in this paper was constructed based on the existing requirement specifications of public tenders announced by government agencies. The requirement specifications were randomly selected from documents created by the Ministry of Economy, Trade and Industry, the Digital Agency, and the Ministry of Health, Labor and Welfare. The data was randomly extracted from these documents, and labeling was performed based on the IPA's "non-functional requirements grades." The detailed distribution of sentences for each label is shown in Table 1.

4.2 Baseline Classification Method Without Function Calling

As a baseline, we evaluated a classification method using prompts only, without Function Calling. This method involved using a pair of prompts consisting of the definition of the classification task and the sentence to be classified and assessing the model's response accuracy. This was conducted in a zero-shot state without any in-context learning, providing only the classification description and no examples. The task definition used was: "In the IPA's 'non-functional requirement grades', NFRs are classified into six categories: availability, performance/scalability, operability/maintainability, migratability, security, and system environment. Answer with the name of only one of these six categories." This approach aimed to observe how effectively the model could classify sentences without Function Calling, establishing a baseline for accuracy.

Table 2 Classification Performance Metrics for GPT-3.5 Turbo, GPT-4 Turbo, and GPT-4o using Function Calling and Prompt-Only Methods

	GPT-3.5 Turbo		GPT-4 Turbo		GPT-4o	
	Function Calling	Prompt only	Function Calling	Prompt only	Function Calling	Prompt only
Accuracy	0.705	0.696	0.775	0.740	0.792	0.783
Precision	0.719	0.608	0.811	0.766	0.785	0.787
Recall	0.648	0.618	0.734	0.674	0.779	0.741
F1-score	0.651	0.587	0.753	0.681	0.780	0.754

4.3 Effect of Function Calling

Table 2 compares the classification results using Function Calling and prompt-only methods for each model. The evaluation metrics include accuracy, precision, recall, and F1-score, with the averages calculated using macro-averaging. In most models, the method using Function Calling achieved higher accuracy compared to the prompt-only method. For instance, GPT-3.5 Turbo shows an accuracy of 0.705 with Function Calling versus 0.696 with prompt only. GPT-4 Turbo demonstrates even more significant gains with Function Calling, achieving 0.775 compared to 0.740 with prompt-only. The GPT-4o model also exhibits superior performance with Function Calling, reaching an accuracy of 0.792 compared to 0.783 with prompt only. For GPT-4o, the performance improvement with prompt-only methods is remarkable, reducing the advantage of Function Calling

Overall, these results suggest that the use of Function Calling significantly enhances the performance of language models across various evaluation metrics. While there are some nuances, particularly with GPT-4o, the general trend highlights the practical advantages of this approach in classification tasks.

4.4 Misclassification Analysis

In this section, we analyze the misclassification issues encountered across various categories, with a particular focus on the "System Environment/Ecology" category. As typical examples, Figures 1 and 2 present the recall values of different categories using Function Calling and prompting methods, respectively, highlighting the areas where misclassifications are most prevalent.

As shown in Figures 1 and 2, the "System Environment/Ecology" category exhibits notably low recall values in both Function Calling and prompting methods, with the prompting method performing particularly poorly. This significant challenge in accurately classifying instances related to this category underscores the complexity and ambiguity inherent in the "System Environment/Ecology" category, making it a key area for focused analysis.

The "System Environment/Ecology" category is defined by several key elements, including "system constraints/premises," which are the agreed-upon conditions

under which the system is installed, "system characteristics," which involve the users and the regional scope surrounding the system, "compliance standards," and "equipment installation environmental conditions." These elements collectively describe the broader environmental and contextual factors impacting the system's operation and integration.

Possible reasons for low recall include the broad and complex nature of the "System Environment/Ecology" category, which encompasses a wide range of factors from technical constraints to regional and user characteristics. This introduces ambiguity, making it difficult for models to accurately distinguish relevant terms from those of other categories. Additionally, elements of system environment and ecology often overlap with other categories. For example, Environmental conditions might relate to security concerns. This overlap can lead to misclassification as models struggle to differentiate between closely related concepts. In addition, the interpretation of terms related to system environment and ecology is highly context dependent. Terms that are relevant in one context may have different implications in another, making it difficult for models to generalize and accurately classify.

Overall, misclassification is a significant challenge across several categories, but it is particularly pronounced in the "System Environment/Ecology" category due to its complexity and overlap with other factors.

4.5 Fine Tuning Models

4.5.1. Training Data and Embedding for Fine Tuning

GPT-3.5 Turbo is capable of fine-tuning with small amounts of data. Fine-tuning is greatly influenced by the training dataset used, making its selection extremely important. In this paper, we utilized descriptions of "non-functional requirement grades" as the training dataset instead of using sentences from actual requirement specifications. The primary reason for this choice is that it allows for more efficient annotation, as the classification of sentences is more straightforward. Additionally, one reason is the hypothesis that it may be difficult to learn the diversity of NFRs with a small amount of data with small numbers of training data. The

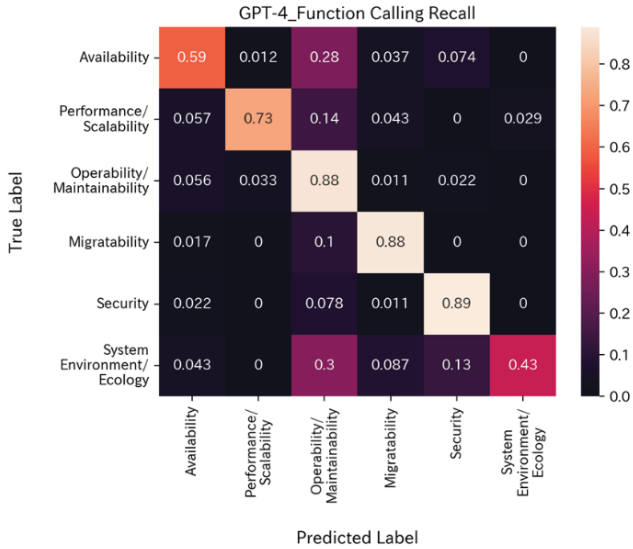


Figure 1 Recall Matrix for GPT-4 Turbo Function Calling and Prompting Performance Across 6 Categories

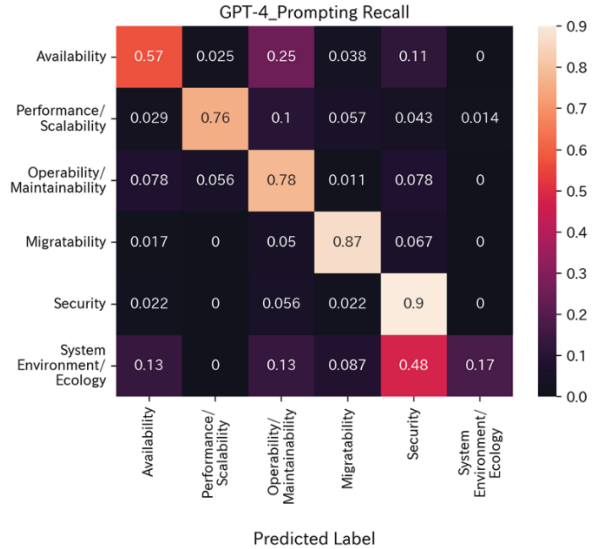


Figure 2 Recall Matrix for GPT-4 Turbo Prompting only Performance Across 6 Categories

"non-functional requirement grades" classify NFRs into six categories: availability, performance/scalability, operability/maintainability, migratability, security, and system environment/ecology. Descriptions related to each category were labeled with the respective classification and used as the training data. As shown in Table 3, the training dataset comprises a total of 933 instances, distributed across the six NFR categories.

For each experiment, a necessary number of instances was randomly extracted from this dataset. This approach ensured that the model could be fine-tuned effectively while utilizing a representative sample of the data for each specific experiment. Figure 3 shows the t-SNE plot of descriptions for each "non-functional requirement grades" after embedding them using the text-embedding-3-large model provided by OpenAI, which converts the text into 3072-dimensional vectors. Each color represents one of the six NFR categories: availability, performance/scalability, operability/maintainability, migratability, security, and system environment/ecology. The clear clustering of some categories, like security, suggests that the classification task for these categories might be relatively straightforward for the model, likely resulting in higher accuracy, precision, and recall. Conversely, the dispersion observed in some categories, like the system environment/ecology categories, indicates potential challenges by enhancing context understanding, in achieving high classification performance. These categories might require more sophisticated models or

additional context to improve classification accuracy. The t-SNE visualization serves as a validation tool, demonstrating that the embedding model can capture and represent the similarities and differences among the NFR descriptions to a significant extent. This visual validation supports the choice of using for fine-tuning and classification tasks.

In conclusion, the t-SNE plot effectively illustrates the clustering behavior of the NFR descriptions based on their embeddings. The distinct clusters for some categories and the overlapping regions for others provide valuable insights into the complexity of the classification task. These insights highlight the strengths and potential context understanding in using the current model for classifying non-functional requirements, guiding future improvements, and fine-tuning efforts.

4.5.2. Fine Tuning Effect

In order to comprehensively evaluate the progression of loss and its impact on accuracy, three different fine-tuning models

Table3 Number of Training Data Instances for NFRs Grades

Availability	168
Performance/Scalability	175
Operability/Maintainability	158
Migratability	101
Security	163
System Environment/Ecology	168
Total	933

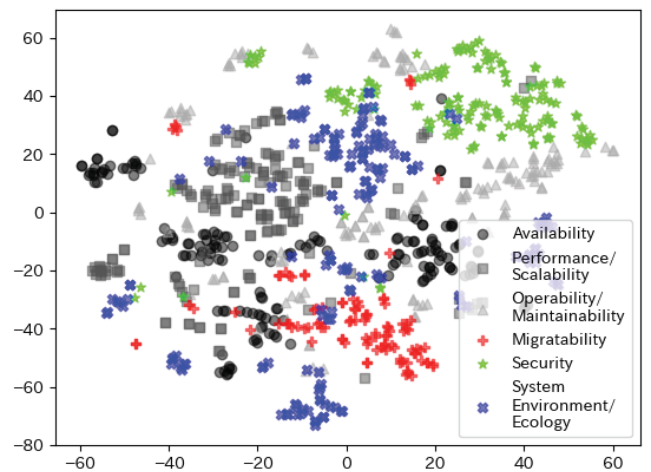


Figure 3 t-SNE Visualization of Embedded Training Data

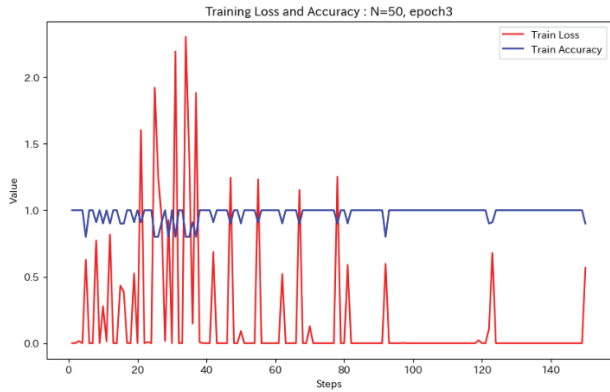


Figure 4-1 Training Loss and Accuracy over Epochs: N=50, Epoch=3 (red: train loss, blue: train Accuracy)

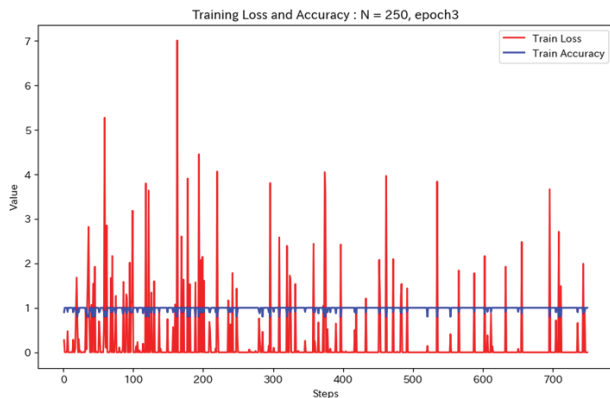


Figure 4-2 Training Loss and Accuracy over Epochs: N=250, Epoch=3 (red: train loss, blue: train Accuracy)

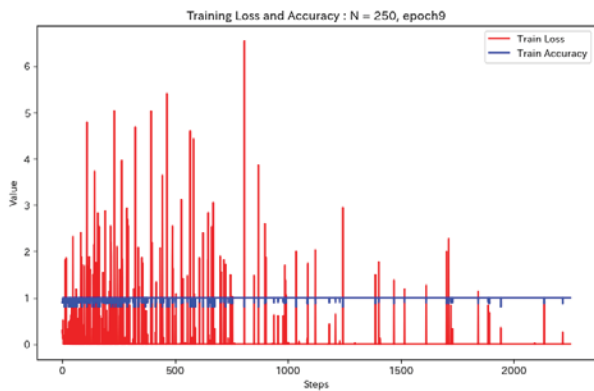


Figure 4-3 Training Loss and Accuracy over Epochs: N=250, Epoch=9 (red: train loss, blue: train Accuracy)

were tested, each designed to address varying conditions in terms of dataset size and number of epochs, thus providing a robust analysis.

In the first scenario, illustrated in Figure 4-1, training was conducted over three epochs using 50 randomly selected training data samples. The loss progression in this scenario exhibits significant fluctuations in the initial stages but stabilizes in the later stages. Considering the small size of the training dataset, this early stabilization is presumed to lead to insufficient accuracy.

In the second scenario, depicted in Figure 4-2, the training dataset was increased to 250 randomly selected samples, and training was conducted over three epochs. Compared to the

first scenario, the adaptation to a more diverse training dataset results in greater and more prolonged fluctuations. However, it is inferred that this leads to more advanced learning.

In the third scenario, shown in Figure 4-3, 250 training data samples were used, but the model was trained over nine epochs. In this case, the loss stabilizes within one epoch, suggesting that increasing the number of epochs does not necessarily aid in learning more from the data.

From these observations, it is evident that the size of the training data and the number of epochs have a significant impact on the learning process.

Figure 5 compares the classification accuracy of these fine-tuned GPT-3 Turbo models. Among the metrics of Accuracy, Precision, Recall, and F1-score, we specifically chose the F1-score due to its balanced representation of both Precision and Recall. In both scenarios, where the training data consists of either 50 or 250 randomly selected samples, fine-tuning significantly improves the F1 score,

reaching levels comparable to the accuracy achieved using the two types of Function Calling in GPT-4. Increasing the training data from 50 to 250 samples results in a slight improvement in the F1 score; however, no additional improvements were observed with an increase in the number of epochs.

While using Function Calling without fine-tuning is effective, combining it with fine-tuning achieves higher classification accuracy. Furthermore, fine-tuning reduces the performance gap between using Function Calling and using only prompts. These findings provide valuable insights for selecting the optimal model design and training strategy for complex tasks such as non-functional requirements classification.

5 DISCUSSION

The proposed method for automatic classification of NFRs using Function Calling with large language models has demonstrated significantly improved efficiency while maintaining F-1 score comparable to traditional methods. Notably, the results utilizing both GPT-4 Turbo and GPT-4o models were remarkable. Even in a complete zero-shot learning state, without any fine-tuning or in-context learning, these models achieved accuracy levels comparable to specialized training models from previous studies.

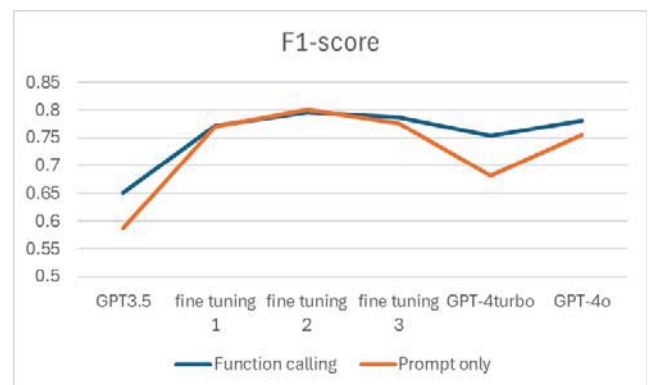


Figure 5 Comparison of F1-Scores for Function Calling and Prompt-Only Approaches Across Different Models and Fine-Tuning Stages

This underscores the advanced language understanding and reasoning capabilities of GPT-4 Turbo and GPT-4o, which can potentially match the outcomes of domain-specific models. This is particularly useful in scenarios where preparing extensive training data is challenging or impractical.

Fine-tuning GPT-3.5 Turbo results in a significant increase in F1-score for both Function Calling and prompt-only approaches, effectively eliminating the difference between the two. Depending on the number of data points and the number of epochs, fine-tuned GPT-3.5 Turbo can achieve results comparable to or even exceeding those of the Function Calling of GPT-4 Turbo and GPT-4o. This demonstrates that, with adequate fine-tuning, GPT-3.5 Turbo remains a viable and highly effective option for achieving top-tier classification performance. While the Function Calling of GPT-4 Turbo and GPT-4o achieves high performance efficiently, fine-tuning GPT-3.5, although requiring the preparation of training data, can yield equal or superior performance.

The improvement in F1-score with fine-tuning highlights the significance of the training dataset's quantity and quality. Fine-tuning enhances the model's contextual understanding and classification accuracy, even with a small amount of data, making it a cost-effective method.

6 SUMMARY OF CONTRIBUTION AND CONCLUSIONS

This paper proposed and demonstrated the effectiveness of an automatic classification method for NFRs using Function Calling with large language models. The Function Calling capabilities of GPT-4 Turbo and GPT-4o achieved excellent F-1 score comparable to traditional specialized models even in a zero-shot learning state, showcasing their high capability. Additionally, GPT-3.5 Turbo, with minimal fine-tuning, achieved results comparable to GPT-4 variants, emphasizing the importance and efficiency of fine-tuning. The significance of fine-tuning is evident, as it enhances the model's contextual understanding and classification accuracy, making it relatively cost-effective method. Notably, the performance of the prompt-only approach also improves significantly with fine-tuning, ultimately eliminating the difference in performance between the two methods.

By leveraging Function Calling and fine-tuning based on the available training data, we efficiently achieved high F1-score classification based on IPA's "non-functional requirements grades", potentially streamlining the analysis of requirement specifications in system development. However, instances of misclassification were observed, indicating the need for improvements in the model's contextual understanding abilities and the selection of appropriate training data.

Future research should address these challenges by focusing on enhancing the model's contextual understanding and developing more sophisticated training datasets. Such efforts are expected to further improve the accuracy of automatic classification, contributing to more efficient and precise software quality management.

ACKNOWLEDGEMENT

This research is being partially conducted as Grant-in-Aid for Scientific Research C (21K11826).

REFERENCES

- [1] Japan Information Service Industry Association REBOK Planning WG, Knowledge System of Requirements Engineering, Kindaikagaku-sha (2011).
- [2] Trade and Industry, Guidelines for Improving the Reliability of Information Systems, 2nd Edition (2009).
- [3] K. Mukaida, S. Fukui, T. Nagaoka, T. Kitagawa, S. Ogata, and K. Okano, Efficient Automatic Classification of Non-Functional Requirements in Information Systems Using Deep Learning, IEICE Technical Report, vol. 123, pp. 13-18 (2023).
- [4] T. Kinoshita, T. Omori, and A. Onishi, Extraction and Validation of Non-Functional Requirements from Japanese Requirement Documents, IPSJ Research Report (SE), vol. 2021.15, pp. 1-6 (2021).
- [5] R. K. Gnanasekaran, S. Chakraborty, J. Dehlinger, and L. Deng, Using Recurrent Neural Networks for Classification of Natural Language-based Non-functional Requirements, in Proceedings of the 4th Workshop on Natural Language Processing for Requirements Engineering, CEUR Workshop, vol. 2857 (2021).
- [6] T. Kitagawa, and T. Nagaoka, Proposal and Evaluation of Automatic Classification Method for Requirements Specifications Using Deep Learning, IPSJ Transactions, vol. 61, no. 4, pp. 842-852 (2020).
- [7] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, and I. Polosukhin, Attention is All You Need, Advances in Neural Information Processing Systems, pp. 6000-6010 (2017).
- [8] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, in Proceedings of the Conf. of the North American Chapter of the Association for Computational Linguistics, vol. 1, pp. 4171-4186 (2019).
- [9] B. J. Gutierrez, N. McNeal, C. Washington, Y. Chen, L. Li, H. Sun, and Y. Su, Thinking about GPT-3 In-Context Learning for Biomedical IE? Think Again, Findings of the Association for Computational Linguistics: EMNLP, pp. 4497-4512 (2022).
- [10] X. Sun, X. Li, J. Li, F. Wu, S. Guo, T. Zhang, and G. Wang, Text Classification via Large Language Models, Findings of the Association for Computational Linguistics: EMNLP, pp. 8990-9005 (2023).
- [11] S. Ibe, S. Kurata, T. Nagaoka, A. Furuhashi, K. Goto, S. Fukui, and T. Kitagawa, Automatic Classification of Requirements Specifications Using Large Language Models, Proceedings of the Software Engineering Symposium 2023, pp. 86-92 (2023).
- [12] Y. Tan, D. Min, Y. Li, W. Li, N. Hu, Y. Chen, and G. Qi, Can ChatGPT Replace Traditional KBQA Models? An In-Depth Analysis of the Question Answering Performance of the GPT LLM Family, Proceedings of

the International Semantic Web Conference, pp. 348-367 (2023).

A Method for Automatic Creation of State Transition Diagrams from Software Requirement Specifications

Takeki Ninomiya[†], Maiko Onishi[‡], Shinpei Ogata^{*}, and Kozo Okano^{*}

[†]Graduate School of Science and Engineering, Shinshu University, Japan

[‡]Graduate School of Humanities and Sciences, Ochanomizu University, Japan

^{*} Faculty of Engineering, Shinshu University, Japan

{23w2074b}@shinshu-u.ac.jp

Abstract - In software development, development proceeds using requirement specifications that describe software requirements in natural language. However, ambiguities in the description of natural words may cause unintended behavior in the system. To solve such problems, software developers create state transition diagrams from requirement specifications and perform model checking. The purpose of this paper is to automatically convert requirement specifications written in natural language to state transition diagrams. Using a large-scale language model as the conversion method, state transition diagrams are created by extracting state transitions from requirement statements and converting them to PlantUML description format. Experiments showed that adding the process of creating state transition diagrams to the prompts made the state transition diagrams produced by ChatGPT more accurate. As a result, the generated state transition diagrams correctly reflected the information about states, transitions, and events described in the requirement statements. On the other hand, there were some omissions and misrepresentations of information on semantics, such as detailed condition content and actions.

Keywords: Requirement Specification, State Transition Diagrams, LLM, PlantUML, Chain of Thought Prompting

1 INTRODUCTION

In software development, the process often proceeds by using requirement specifications written in natural language [1]. Those requirements are written on the assumption that the product will behave as the developer expects it to behave. However, when reflecting requirements, ambiguities in natural language and inconsistencies in requirements can cause the system to behave differently from the specification and in unintended ways[2]. Design errors due to ambiguous or inconsistent wording are often discovered in the testing process later in the development process, and these ambiguous statements force the developer to go back to the design process again[3][4]. Rework caused by the testing process errors increases a great deal of extra costs. One of the methods to prevent such rework is to create state transition diagrams from requirement specifications and perform model checking. Model checking based on state transition diagrams allows developers to check for unrecoverable and undesirable system states caused by unintended behavior

during the design phase. However, creating state transition diagrams and inspecting models requires specialized knowledge. Therefore, it is difficult for beginners to handle. In addition, if the system is complex and consists of many components, manually extracting all the states of the components can be a labor-intensive task. Therefore, the goal of this research is the automatic conversion of state transition diagrams from requirement specifications containing state transition descriptions written in Japanese. This research is expected to help designers and developers share system specifications without conflicts at low cost. Previous research has proposed a method for extracting state transition descriptions from requirement statements using syntactic analysis based on the rules of natural language notation, and creating state transition diagrams based on the extracted elements[5][6][7][8]. A rule-based method using dependency analysis can extract the name of a state variable and its state from requirement statements[9]. However, simply applying the extraction results to a state transition diagram template was not enough to create a correct diagram. This is because different names may be assigned to the same state. In other words, while the extraction of elements for state transition descriptions was automated, the complete automation of diagram creation was not achieved. On the other hand, natural language processing technology using Large Language Models (LLMs) is rapidly developing and its usefulness is being confirmed. Many studies in the field of software modeling have also explored the use of LLMs such as ChatGPT. In some studies, LLM was used to convert requirement statements into UML descriptions such as plantUM and to create UML diagrams[10]. Prior research confirmed that ChatGPT understands most UML diagrams, including class diagrams, use cases, state transition diagrams, sequence diagrams, and activity diagrams[10]. This research uses ChatGPT to create state transition diagrams by extracting state transitions from requirement statements described in natural language and converting them to PlantUML description format. This enables automation up to the creation of state transition diagrams, which was previously impossible.

This paper is organized as follows. Section 2 presents the technology used in this study and related research. Section 3 presents the proposed methodology. Sections 4 and 5 present the experiments and results, respectively. Section 6 discusses the method based on the results. Finally, Section 7 concludes

the paper.

2 PREPARE

2.1 State Transition Diagrams

A state transition diagram represents the behavior of a model consisting of a combination of states, transitions, and events. In model-driven development, state transition diagrams are widely used for purposes such as checking implementation specifications and analyzing scenarios[11]. In UML 2.0, state transition diagrams are called state machine diagrams. State transition diagrams have a notation that specifies the method of operation, called semantics. The semantics include conditions and transition actions related to transitions, actions and activities inside states, and structures such as composite and parallel states[11].

2.2 Modeling with ChatGPT

Large Language Models (LLMs) are models of natural language processing that have been trained on large amounts of text data. Typical examples of large-scale language models include BERT, announced by Google in 2018, and GPT-3, announced by OpenAI in 2020. In May 2024, OpenAI introduced the new GPT-4o, a model with better performance in languages other than English. This study used the GPT3.5 and GPT-4o models for its experiments. Since ChatGPT is a language model, it cannot generate graphical models, but it can generate models using text-based UML notation. ChatGPT understands several UML notations such as PlantUML, Mermaid, Markdown UML. Prior studies have shown several characteristics of software modeling using ChatGPT[10].

- ChatGPT can represent models in multiple UML notations. In general, PlantUML tends to have fewer syntax errors.
- Previous conversation history will cause fluctuations in the generated results.
- Variation in ChatGPT responses to the same prompt.
- The problem domain affects the structure, content, and level of abstraction of the generated model.

Considering the above, this method selected PLantUML as the UML notation. PlantUML is a tool that allows code-based description of UML and has been used in previous studies for automated creation of UML diagrams[12]. Also, prior research using ChatGPT recommends starting a new chat each time a new model is generated. This is because previous conversation history influences the results generated. Therefore, this experiment also switched to a new chat in each experiment.

3 PROPOSED METHOD

This section provides an overview of the methodology. An overview of this method is shown in Figure 1. Figure 1 represents giving ChatGPT requirement statements and prompts explaining how to create a state transition diagram, and receiving a state transition diagram described in PlantUML format as a response from ChatGPT. This is executed in PlantUML and converted into a state transition diagram image. Requirement statements are generally classified into functional and non-functional requirements. Functional requirements include elements such as UI, database, processes, and context. In this method, the primary input statements are those that describe the processes in the functional requirements. Statements describing processes consist of elements such as activities, parameters, and rules, and they represent the processing of transitions and actions. A transition statement is a statement that describe a change of state by a pair of states and events[13]. An active statement is a statement that, based on a combination of state and event, instructs the execution of operations[13]. Or it instructs the interactions with the external environment through actuators or other means.

To enable ChatGPT to appropriately convert these requirements into state transition diagrams, careful crafting of prompts is essential. This method used Chain of Thought Prompting as the prompt description method. It has been found that ChatGPT responses are more concrete when specific examples of inputs and outputs are given. Chain of Thought Prompting is a method that improves LLM capabilities by including intermediate reasoning steps before solving the problem in the prompt[14]. As an intermediate inference step, this method describes the process of extracting the state transition description from the requirement statement and converting it to PlantUML format. As an example, ChatGPT receives requirement statements, a PlantUML description created from the requirement statements, and the procedure for converting the requirement statements to PlantUML. Figure 2 shows part of a prompt that illustrates the process of converting requirement statements to PlantUML. Usually, a requirement statement describing the behavior of the system indicates that when an event occurs, the system will go from state A to state B and take action. Therefore, this study defines the elements that constitute state transitions as the pre-state, post-state, event, guard condition, and action. The prompt in Figure 2 explains how to convert each of those elements to PlantUML format when they are included in requirement statements. This allows ChatGPT to understand conversion patterns according to the rules.

There are two ways to describe actions in a state transition diagram:

1. Actions executed upon entering or exiting a state, or while staying within that state.
2. Conditional actions executed when conditions are satisfied, and the transition is successfully completed.

The former type of action is documented within the state using

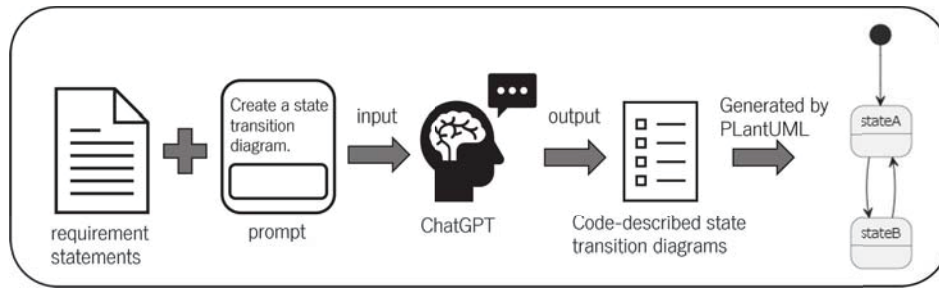


Figure 1: Overview of the proposed method

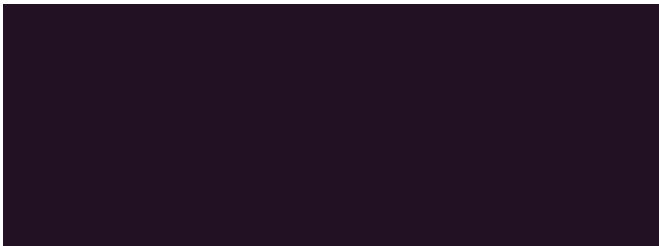


Figure 2: Part of the process of converting requirement statements into PlantUML

entry, exit, or do, depending on the execution timing. Entry is an action executed once upon transitioning into a state, exit is an action executed once upon leaving a state, and do is an activity continuously executed while remaining in the state. An action refers to a process that is executed without interruption or suspension, whereas an activity refers to a process that is expected to allow interruptions or suspensions during execution. The prompt explains that, depending on the execution timing, one of entry, exit, or do should be specified for the action. On the other hand, the latter type of conditional action is documented within the transition. In a state transition diagram, conditional actions are represented as “event [guard condition]/conditional action”. When the requirement statement contains a conditional action, the prompt instructs ChatGPT to convert it as follows:

- pre-state → post-state : event [guard condition]/conditional action

For example, a timer requirement statement is converted by ChatGPT into a PlantUML description as follows.

- 「タイマ起動中に、タイマボタンを 3sec 以上続けて長押ししたら、ブザーを 100msec 鳴らした後、0min0sec にリセットされ、タイマが停止する。」
- 起動中 → 停止中 : タイマボタンを 3sec 以上長押し /100msec ブザー鳴動後、0min0sec にリセット

Since conditional actions and entry actions within a state are executed within the same cycle, it is recommended to consolidate operations involving the same target or variable into either one[11]. It is also recommended that conditional actions

include processes that cannot be interrupted and can be completed in a short amount of time. Whether an action is documented as a state action or a conditional action is determined by the factors mentioned above, as well as the common design rules established by the development team.

In this method, the prompt for creating a state transition diagram includes, in addition to the input statements, the instruction to “indicate from which statements you have created which transition, together with the rationale for its derivation. This instruction causes the ChatGPT response to output the process of creating a transition from requirement statements with the result. If the generated state transition diagram differs from the desired result, the developer can review the derivation process to identify whether the issue lies in the inadequacy of the input requirement statement or an error in ChatGPT’s conversion.

4 EXPERIMENTS

The evaluation experiment investigated the quality of the state transition diagrams produced by the proposed method. The evaluation experiment also investigated whether the model and prompts used affected the results generated. This experiment uses GPT3.5 and GPT4o, and compares the results using the chain of thought prompting method and the method in which only the requirement statements are given (no prompting device). Two types of requirement statements were prepared for the experiment. The first is a requirement statement for a CD player control panel in Astah’s “UML State Machine Diagram and State Transition Table Tutorial[15]. The requirements for the CD player operation panel are described as transition statements in a basic syntactic format, such as “When in the previous state, if an event occurs, it transitions to the subsequent state and performs an action.” There are eight requirement sentences in total. The Astah tutorial includes a state transition diagram created from the request statement, and this experiment treats it as the correct diagram. The second is the requirement statement for the timer button in Chapter 2, Operation Requests, of the “Hotpot, Version 7” published by SESSAME[16]. Compared to the requirements for the CD player, the requirement statements for the Hotpot are more complex, with more detailed descriptions of transition conditions and actions. There are a to-

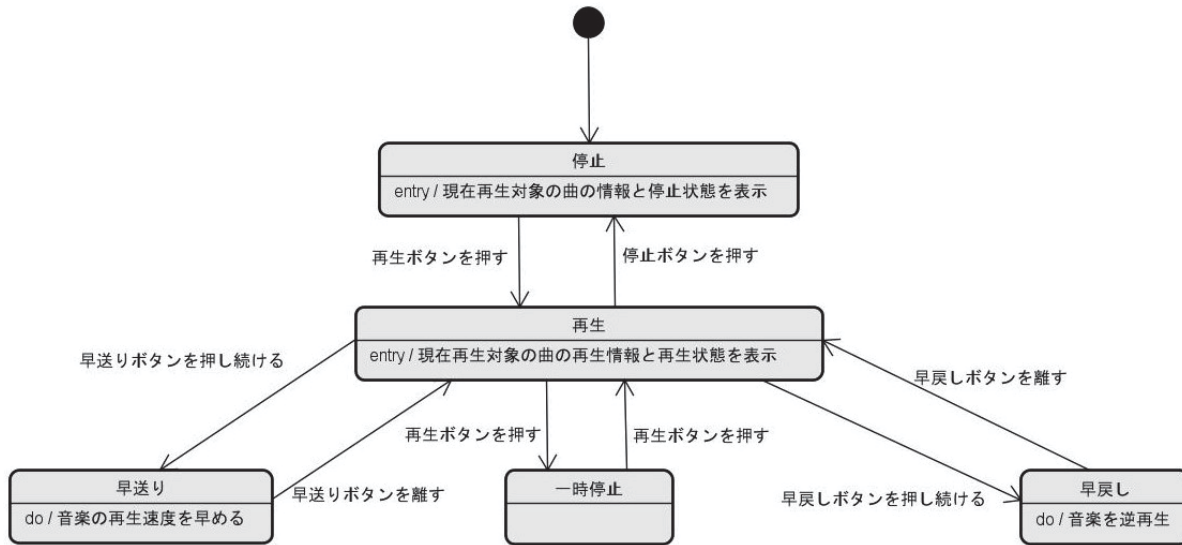


Figure 3: The state transition diagram of the CD player control panel created using chain of thought prompting in GPT4o.

tal of seven requirement sentences. The state transition diagram for the correct answer was created by ourself in advance.

The evaluation experiment compares the state transition diagram created by the proposed method with the correct diagram to determine whether it is correct or incorrect. The experiment defined states, transitions, events, and semantics as the evaluation criteria for the state transition diagrams created. Semantics include guard conditions and condition actions, state actions, and activities. The experiment checks that the state, transitions, and events accurately reflect what should be described in the state transition diagram without deficiencies. The experiment checks whether the created state transition diagrams correctly represent the semantics.

5 RESULT

Figure 3 shows the state transition diagram of the CD player control panel created using chain of thought prompting in GPT4o.

Figure 4 shows the state transition diagram of the timer created using chain of thought prompting in GPT4o.

5.1 CD Player Control Panel

The state transition diagram of a CD player control panel has five states, nine transitions including the transition from the start state to the initial state, and eight events for the transitions. In the experiments, when GPT-4 or chain-of-thought prompting was used, these states, transitions, and events were accurately reflected in the state transition diagram without any omissions or excesses. On the other hand, when GPT3.5 was used without the prompt device, the defined state names were correctly reflected, but the state structure had unnecessary composite states and transitions could not be connected correctly. In other words,

a crappy diagram was output that differed from the correct state transition diagram.

5.2 Timer

The timer's state transition diagram has three states, eight transitions including the transition from the starting state to the initial state, and seven events for the transitions. When using chain of thought prompting in GPT4o, the state, transitions, and events were accurately reflected in the state transition diagram, except for one transition and its events. The transition that could not be reflected was the self-transition caused by the countdown when the timer is activated. However, since this self-transition was not explicitly described in the requirement statement, this omission is understandable. On the other hand, when verifying the semantic behavior represented by the state transition diagram in Figure 4, considering the actions and guard conditions, some errors were identified. The details are discussed in the considerations presented later. When using chain of thought prompting in GPT 3.5, one extra transition was output in addition to the missing countdown transition. However, the other states, transitions, and events were accurately reflected in the state transition diagram. When GPT3.5 was used without prompt devising, only the state names were correctly reflected, but the transitions were completely random and the events could not be properly reflected.

6 CONSIDERATION

A common feature of GPT3 and GPT4o is that PlantUML syntax rules are rarely wrong. The experiment also confirmed that GPT properly understands and processes Japanese words

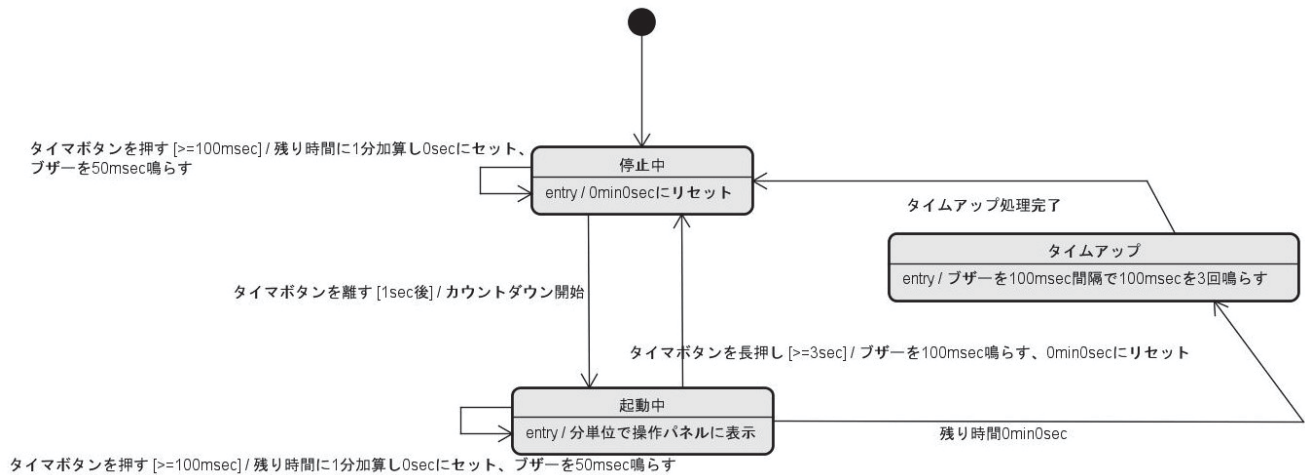


Figure 4: The state transition diagram of the timer created using chain of thought prompting in GPT4o

that represent logical relationships. For example, when a requirement sentence contains the words “または” to indicate a disjunction relationship, GPT understands the relationship and reflects it appropriately in the state transition diagram. On the other hand, when GPT3 and GPT4o were compared, the quality of the generated state transition diagrams differed significantly. The difference in results was particularly apparent when no prompt was devised. This result suggests that GPT3.5 understands the PlantUML syntax rules for describing state transition diagrams, but does not understand how to create state transition diagrams from requirement statements. However, by using chain of thought prompting and providing a process of transformation patterns from requirement statements to PlantUML descriptions, GPT3.5 can determine the elements of states, transitions, and events from requirement statements and reflect them in state transition diagrams. Furthermore, GPT3.5 does not fully understand how to describe state actions, conditional actions, or guard conditions in PlantUML. Therefore, it was necessary to use chain of thought prompting to explain these aspects as well. On the other hand, GPT4o can create a state transition diagram that satisfies the description of the requirement statement better than the diagram generated by GPT3.5, even without any prompting devices. In fact, GPT4o was able to create a state transition diagram that accurately satisfied all of the states, transitions, and events in the CD player’s control panel requirements. However, even with GPT-4, it was unable to semantically accurately describe some actions and certain guard conditions. In particular, determining whether a behavior should be categorized as an action or an activity and appropriately labeling it as entry or do seems to be a difficult task. In both models, the use of chain of thought prompting to create state transition diagrams was effective. Even GPT4o, which shows excellent performance, can describe guard conditions and state actions more semantically accurate by using

chain of thought prompting.

The practicality of this method is then discussed. State transition diagrams are used in software development for purposes such as scenario analysis and implementation specifications[11]. This method is considered to be effectively applicable in scenario analysis during requirements definition and requirement analysis. This is because experimental results have shown that it can accurately extract states, transitions, and events from requirement statements and convert them into state transition diagrams. During the requirements definition phase, this method can be used to automatically create state transition diagrams from use cases and requirements. This helps in verifying the system’s behavior and ensures that no states or events are overlooked. On the other hand, when using this method for implementation or analysis of specifications, additional manual adjustments to the state transitions created by ChatGPT would be necessary. This is because there may be errors or omissions related to action symbols of states and guard conditions of transitions that involve execution timing. For example, the timer’s experimental results depicted in figure 4 might appear correct at first glance, but the model does not exhibit the correct system behavior for a timer. In Figure 4, pressing the timer button in the stop state increments the time and then re-enters the stop state, resetting the time to 0 with the entry action. This model does not allow for setting the time from the stop state. It is likely challenging at present to have ChatGPT create system behavior that is completely and systematically correct.

In this method, ChatGPT essentially follows the transformation rules described in the prompt to convert requirement statements into the PlantUML format. In the process, there were few self-indulgent additions by ChatGPT to states, transitions, or events. Most of the time when a change or additional element is included by ChatGPT, it is considered to be a case where the status is omitted or not stated in the requirement statement. Some

related papers also recommend explicitly stating pre-state, post-state, and event in the requirement statements[17]. In fact, the requirement statements for the CD player operation panel, which were created with high quality, clearly included all these elements. Conversely, when the requirement statements omit significant details about states, it becomes challenging to create high-quality state transition diagrams. Therefore, moving forward, it is considered necessary to devise input templates to create more accurate state transition diagrams.

This experiment focused on models with a small number of states. In the future, experiments will be conducted on systems with a larger number of states and more complex structures.

7 CONCLUSION

This study used a large-scale language model as a transformation method to create state transition diagrams by extracting state transitions from requirement statements and transforming them into PlantUML description format. Experiments showed that the state transition diagrams produced by ChatGPT were more accurate when the process of creating the state transition diagrams was added to the prompts. When GPT3 and GPT4o were compared, it was confirmed that GPT4o was superior in creating state transition diagrams. As a result, the state transition diagrams created using chain-of-thought prompting correctly reflected the information about states, transitions, and events described in the requirement statements. On the other hand, some of the semantic information, such as detailed condition contents and actions, were missing or misrepresented.

Acknowledgement

This research is being partially conducted as Grant-in-Aid for Scientific Research C (21K11826).

REFERENCES

- [1] Tetsuo Tamai: "Fundamentals of Software Engineering," Iwanami bookstore (2004)
- [2] Daniel Aceituna, Hyunsook Do and Seok-Won Lee: "Interactive Requirements Validation for Reactive Systems through Virtual Requirements Prototype," Model-Driven Requirements Engineering Workshop, Trento, Italy, pp. 1-10 (2011)
- [3] John Fitzgerald and Peater Gorm Larsen: "Modelling Systems: Practical Tools and Techniques in Software Development," Cambridge University Press (2009)
- [4] Yoichi Omori and Keiojiro Araki: "Toward a Quality Improvement of Specifications in Natural Language Based on the Semi-equivalent Formal Models," Journal of Information Processing Society of Japan, Vol.3, No.5, pp. 18-28 (2010)
- [5] Maiko Onishi, Shinpei Ogata, Kozo Okano and Daisuke Bekki: "A Method for Matching Patterns Based on Event Semantics with Requirements," Proceedings of 14th International Joint Conference on Knowledge-Based Software Engineering (JCKBSE), pp. 181-192 (2022)
- [6] Maiko Onishi, Shinpei Ogata, Kozo Okano and Daisuke Bekki: "Reducing Syntactic Complexity for Information Extraction from Japanese Requirement Specifications," Proceedings of 29th Asia-Pacific Software Engineering Conference (APSEC), pp. 387-396 (2022)
- [7] Naru Nakamura, Ryota Yamamoto, Norihiro Yoshida and Hiroaki Takada: "An Extraction of State Transition Descriptions in a Requirements Specification Document for an Embedded System," IPSJ SIG Technical Report, vol.118, no.230, pp. 25-30(2018)
- [8] Muhammad Waseem Anwa, Imran Ahsan, Farooque Azam and Wasi Haider: "A Natural Language Processing (NLP) Framework for Embedded Systems to Automatically Extract Verification Aspects from Textual Design Requirements," 12th International Conference on Computer and Automation Engineering, pp. 7-12(2020)
- [9] Masanosuke Ohto, Hiroya Ii, Kozo Okano and Shinpei Ogata: "Proposal of Extracting State Variables and Values from Requirement Specifications in Japanese by using Dependency Analysis," Proceedings of the 25th International Conference on Knowledge-Based and Intelligent Information Engineering Systems, pp. 1649-1657 (2021)
- [10] Javier Cámara, Javier Troya, Lola Burgueño and Antonio Vallecillo: "On the assessment of generative AI in modeling tasks: an experience report with ChatGPT and UML," Software and Systems Modeling (SoSyM), Vol.22, No.3, pp. 781-793(2023)
- [11] Takayuki Kubo: "State Transition Design Techniques for Embedded Engineers - State Transition Diagram and State Transition Table description techniques for use in the field -," TechShare Corporation(2012)
- [12] Abdelkareem M. Alashqar: "Automatic Generation of UML Diagrams from Scenario-Based User Requirements," Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 07, No.02 (2021)
- [13] Mukuta Yamamoto, Norihiro Yoshida and Hiroaki Takada: "Survey of model extraction techniques based on embedded systems development artifacts," Computer Software Vol.39, No.3, pp.4-16(2022)
- [14] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed H. Chi, Quoc V. Le and Denny Zhou: "Chain-of-thought prompting elicits reasoning in large language models," Advances in Neural Information Processing Systems(2022)
- [15] Change Vision, Tutorial UML State Machine Diagrams - State Transition Tables, <https://astah.change-vision.com/ja/tutorial/tut-state-machine-diagram-state-transition-table-cd.html> (2024.06.11access)
- [16] SESSAMI, Hot Pot, <https://www.sesame.jp>
- [17] Daniel Aceituna and Gursimran Walia: "Model-based requirements verification method: Conclusions from two controlled experiments," Information and Software Technology Volume 56, Issue 3, pp. 321-334(2014)

A Study of Web System for Comparison and Analysis of Cooking Actions Based on Activity Sensing

Fuma Kato[†], Ayato Kumazawa[†], Katsuhiko Kaji[‡],
Kimihiko Mizuno^{*}, Katsuhiko Naito[‡], Naoya Chujo[‡], and
Tadanori Mizuno[‡]

[†]Graduate School of Business Administration and Computer Science, Aichi Institute of Technology

[‡]Department of Information Science, Aichi Institute of Technology

^{*} Mitsubishi Electric Engineering Co., Ltd.

{b24707bb, kaji}@aitech.ac.jp

Abstract – There are various methods for sharing cooking recipes. Consequently, systems that allow users to share their impressions of the outcome and taste through photos and text have become widespread. However, there is no mechanism for comparing cooking actions with others or with one’s past self. This study aims to realize a system that allows users to share and compare cooking actions, enabling self-analysis and reflection through the visualization of actions during cooking. We extracted features related to stationary actions and movements between fixed points. For stationary actions, specifically cucumber slicing, we obtained the average pace, average acceleration, and the standard deviation of acceleration. For movements between fixed points, focusing on hamburger steak preparation, we extracted features such as dwell time and location transitions. We investigated whether self-analysis could be conducted and whether there were any changes in cooking skills through the comparison of cooking actions using graph displays. As a result, while the graphical display prompted self-analysis and led to a change in awareness, no growth trend was observed with a displaying only once.

Keywords: Analysis Cooking Actions, Behavioral Visualization, Action Recognition, Operational Analysis

1 INTRODUCTION

There are various methods for sharing cooking recipes. With the widespread use of the internet, sharing and searching for recipes has become easier. Services such as Cookpad and Kurashiru are examples of platforms that provide recipes via photos and text. These services also feature functionalities that allow users to report the results of the recipes they have used. In Cookpad, this function is known as “Tsukurepo,” and in Kurashiru, it is called “Taberepo.” Users can share their feedback on the results and taste of the recipes through photos and brief text.

However, these platforms lack a system that allows users to compare their cooking skills and efficiency with those

of others or their past selves. Such a system could help users realize their improvement and analyze their cooking abilities objectively through comparisons with others. By collecting multiple cooking data points, users can understand how their cooking practices differ from others. Over a long period, users can track their cooking progress compared to their past performance.

Currently, wearable sensors such as smartphones and smartwatches are widely used. These wearable sensors come equipped with various sensors, including accelerometers, gyroscopes, magnetometers, heart rate monitors, barometers, and GPS. Sensing can be performed using dedicated devices or general-purpose devices equipped with sensors like smartphones. Dedicated sensing devices tend to be more accurate than general-purpose ones. However, when requesting sensing services, dedicated devices involve preparation time and monetary costs. General-purpose devices might not always meet the required accuracy for estimation but are more cost-effective. In this study, we judged that the accuracy of general-purpose sensors is sufficient for estimating cooking actions. Moreover, collecting a large amount of feature data makes comparison easier, so it is ideal to widely gather data. Therefore, we use general-purpose wearable sensors in this study.

Work analysis, which involves checking and analyzing work tasks, is widely practiced. Generally, the time taken for each process step is measured. For example, when examining the efficiency of line work, the time taken for each task such as assembly and quality inspection is measured with a stopwatch and used as an evaluation criterion. Wearable sensors can quantify actions and use these quantified movements as evaluation criteria in work analysis. In assembly work, for instance, the hand movements involved in assembly can be quantified and used as evaluation criteria. Quantifying movements also allows the efficient imitation of actions to be conveyed in a non-intuitive manner. By numerically reducing the difference between the desired movement and one’s own movement, similar actions can be achieved.

This study aims to assist in self-analysis of cooking actions using behavioral sensing data obtained from wearable sensors. The approach involves sensing cooking actions using wearable sensors and extracting features. We collect and visualize feature data from multiple people cooking the same recipe and share these visualizations. Users can recognize their improvement points and growth areas through graph comparisons. The evaluation targets are the cooking actions during the process, not the results such as appearance or taste. While appearance and taste play important roles in evaluating cooking, they are outcomes of the cooking process and do not capture the detailed actions during cooking. By comparing the cooking actions, users can identify improvement and growth points in their movements, promoting self-analysis. Therefore, the focus is on evaluating cooking actions. In this study, we extracted feature quantities from slicing actions, estimated actions from point-to-point movements, and investigated the impact of graph visualization on users.

This paper is structured as follows. Chapter 2 describes related research. Chapter 3 outlines the overall configuration of the system designed to promote self-analysis. Chapter 4 details the feature extraction of fixed-point actions using acceleration. Chapter 5 describes the feature extraction of actions between fixed points using location information. Chapter 6 discusses the visualization of actions that encourage self-analysis, based on the feature extraction results obtained in Chapter 4. Chapter 7 summarizes the research and discusses future challenges.

2 RELATED STUDY

There are studies that estimate actions based on sensing results. Sensing methods can be broadly classified into two types: sensors fixed in place and sensors attached to people or objects. Sensors fixed in place include Wi-Fi, BLE, camera images, and 3D point clouds. For example, BLE sensors can be deployed in an environment to recognize activities [1]. Sensors attached to people or objects include accelerometers, gyroscopes, magnetometers, temperature and humidity sensors, and light sensors. Accelerometers can recognize simple tasks [2] and estimate human actions [3]. Gyroscopes can be used to estimate foot twisting actions [4]. There are studies that estimate actions based on sensing results. The methods of sensing can be broadly divided into two types: those that fix sensors to a location and those that attach sensors to people or objects. The type that fixes sensors to a location includes Wi-Fi, BLE, camera images, and 3D point clouds. In this study, we use only commonly available wearable sensors and a sensing method that does not depend on the cooking environment to estimate actions. Therefore, the type that requires installation in the environment is not suitable for this study. Since wearable sensors are sensors that people carry or wear, this study uses wearable sensors. To recognize body movements, we use accelerometers and gyroscopes.

There is existing study on position estimation. For example, GPS is commonly used for position estimation [5].

While GPS is a standard method for location tracking, it is not suitable for indoor position estimation. Given that our study focuses on estimating movements during cooking, GPS is not appropriate. Indoor positioning typically relies on data collected from sensors. Sensors used for indoor positioning include IMU, magnetic sensors, IMES, Wi-Fi [6], RFID [8], infrared sensor [9], and image sensors [10]. One method that uses IMU is PDR (Pedestrian Dead Reckoning), which detects human walking using IMU data to estimate position [11]. Instead of relying on instantaneous values, PDR uses accumulated historical data for position estimation. Since our study uses wearable sensors, methods requiring multiple sensor installations, such as IMES or Wi-Fi, are not suitable. PDR is appropriate for this study because it can utilize the accelerometers and gyroscopes commonly found in wearable sensors. Therefore, this study employs PDR for position estimation.

There is study on the visualization of action analysis results. For instance, some studies focus on displaying the movement trajectories of individuals [12][13]. However, in this study, we aim to extract features from the fine movements of the arms, making such display methods unsuitable. There is also study on visualizing data using dedicated devices [14], which allows for more flexible displays. Nevertheless, dedicated devices involve significant costs and are not practical for general use. Therefore, we developed a web system to present feature data to users and encourage comparisons. To ensure that the visualization can be accessed using the devices that users already own, we opted for a web-based system. This approach facilitates identifying areas for improvement and recognizing growth by comparing overall feature data. While some studies explore AR displays or visualizations superimposed on images [15], our study does not employ such methods. This is because environments like kitchens vary for each user, and adapting to this variability would require users to perform 3D mapping, which is cumbersome. Instead, we visualize feature data extracted from sensor data through graphs.

A study has been conducted on various aspects of cooking, and research on cooking is being conducted. One area of study involves proposing recipes based on available ingredients, where appropriate recipes are suggested based on the input ingredients [16]. Another area focuses on analyzing recipes and videos, such as examining the relationship between different recipes [17]. These studies aim to support the cooking process itself. In contrast, our study aims to promote self-analysis after cooking. Additionally, there are studies that analyze and estimate cooking processes. Some of these studies include recognizing cooking activities using methods like OpenPose [18], sensing cutting actions by attaching sensors to chopping boards and knives [19], detecting fried foods by sensing the air [20], and analyzing cooking processes using accelerometers [21]. These studies use sensors to make estimations. In our study, we use accelerometers and gyroscopes. For displaying the results, we employ graphs.

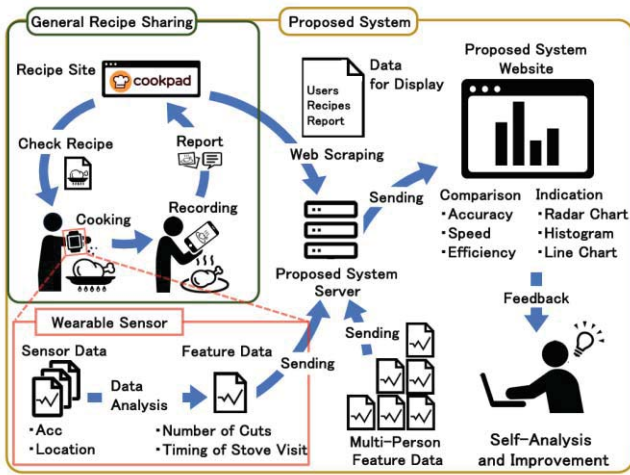


Figure 1: System configuration

3 COOKING ACTION COMPARISON SYSTEM USING WEARABLE SENSORS

In this chapter, we discuss the requirements for achieving the goals of this study and the system configuration based on these requirements. The system configuration diagram is shown in Figure 1.

3.1 Requirements Definition

We define the requirements to fulfill the objective of this study, which is to assist in the self-analysis of cooking action using behavioral sensing data obtained from wearable sensors. To meet these requirements, the following are necessary: Extraction of features that positively influence cooking skills through visualization, Effective display of feature data, A system that facilitates easy sharing and comparison of data.

Cooking involves various processes, such as: Washing dishes and ingredients, Measuring with spoons and scales, Mixing with chopsticks or whisks, Cutting and prepping ingredients (e.g., slicing, filleting), Cooking (e.g., frying, boiling), Moving between different locations (e.g., stove, sink). Each of these processes has distinct features. For instance, in cutting actions, features could include speed, force, knife angle, the neatness of the cut surface, and the shape and size of the ingredients.

However, it is crucial to determine if displaying these features has a positive impact on cooking outcomes. For example, comparing the speed of cutting actions between different people might not reveal differences in taste or appearance. Directly related factors to taste and appearance include the shape and size of ingredients and the degree of cooking, which are difficult to capture with wearable sensors. However, features indirectly related to the quality of cooking, such as how carefully ingredients are handled, can be captured. It can be inferred that careful handling results in uniformly sized pieces. Therefore, it is essential to identify features that positively influence cooking skills through visualization.

Furthermore, there are two types of evaluation criteria for what is considered "skilled": "Better without limit" – for example, faster cutting with a knife is considered better. "Closer to specific data is better" – for example, movements closer to those of a professional are considered better.

The "better without limit" approach is straightforward and easy to understand but might lead to undesirable consequences, such as injury from focusing too much on speed. The "closer to specific data" approach provides a realistic goal, as it is based on actual data, but it requires preparing target data and defining what constitutes a skilled individual. It is necessary to consider whether it is more appropriate to define skilled individuals based on efficient movements or the ability to produce delicious food.

Effective display of feature data aims to promote self-analysis by comparing cooking actions with others. Therefore, the display method must facilitate comparison. This study adopts graphs to display feature data, as multiple graphs corresponding to different features can facilitate easier comparisons. Given that the kitchen environment varies, only the feature data is displayed in the graphs. The content users derive from comparisons differs between comparing with others and with their past selves. Comparing with others helps users understand how proficient their actions are and how close they are to skilled cooks. Comparing with past performance helps users recognize their growth. Hence, the graph display should reflect these aspects.

3.2 System Configuration

The proposed system shown in Figure 1 extends common recipe-sharing systems to display the analysis and comparison of cooking actions.

Common recipe-sharing systems follow the flow of "viewing the recipe," "cooking," "recording," and "posting." Users view shared recipes on a recipe site, cook referring to the recipe, take pictures of the finished dish for posting, and summarize their thoughts. Finally, users share the images and comments on the recipe site.

The proposed system consists of four components: Sensing Cooking Actions with Wearable Sensors, Extracting Feature Data on Cooking Skills Using Sensor Data, Collecting Feature Data on Cooking Skills, Visualizing and Sharing the Comparison of Collected Feature Data.

In the sensing phase, wearable sensors capture cooking actions, generating sensor data. Actions such as cutting and washing, as well as movement trajectories during cooking, are recorded using smartphones and smartwatches. Sensor data might include acceleration, angular velocity, and positional information.

The captured sensor data is analyzed to extract feature data, identifying characteristics of cooking actions. Examples of features obtained from sensor data include the number of cuts estimated from peak detection in acceleration data and the force applied during cutting. Smartphones use acceleration and angular velocity data for position estimation, which helps in estimating cooking processes. For instance, actions near the stove involve

heating, while actions near the refrigerator involve opening and retrieving ingredients, allowing the estimation of refrigerator visits and door openings. Combining cooking actions with position estimation helps in estimating cooking processes.

Extracted feature data is shared on a server, linked to the recipe information. Multiple individuals cooking the same recipe upload their feature data to the server, enabling comparisons with others. A larger dataset increases the reliability of the average values.

Collected feature data is visualized in graphs via a web system for comparison. Graphical displays facilitate easier comparisons with others and with one's past performance. Comparison metrics include speed, accuracy, and periodicity of actions. Comparison targets might include professional cooks, individuals at the same skill level, friends, and past selves. Comparing with professionals highlights the gap to the ultimate goal, while comparisons with peers reveal areas of strength and weakness. Comparing with friends can increase competitiveness and provide conversational topics during social interactions. Comparing with past performance reveals personal growth and areas needing improvement.

4 ANALYSIS AND COMPARISON OF FIXED POINT ACTIONS USING ACCELERATION

In this chapter, we discuss the analysis and comparison of cucumber slicing as an example of fixed point actions. Cooking actions can be divided into fixed point actions, such as cutting, grilling, and washing, and movements between points, such as between the sink and the refrigerator. For fixed point actions, we extract features of the slicing action using accelerometers.

4.1 Investigation of Acceleration During Cucumber Slicing

To analyze and compare fixed point actions, we focused on the slicing process, specifically the slicing of cucumbers. The study involved sensing the slicing process to collect data for analysis. Participants included one housewife (Subject A) and 14 university students (Subjects B to O). The length of the cucumbers and the thickness of the slices were not standardized, considering that it would be challenging to standardize these in general usage of the system.

Subjects wore smartwatches (TicWatch E3 and Google Pixel Watch) on the arm holding the knife, and accelerometer data, including gravity, was collected. The experimental setup is shown in Figure 2. As ground truth data, the slicing process was also recorded on video. Sensor data was normalized, and sensor noise was removed using a moving average filter. From this data, feature data for comparison was extracted. An example of the normalized acceleration data is shown in Figure 3.

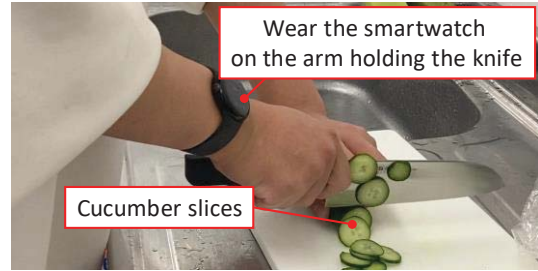


Figure 2: Sensing cucumber slicing

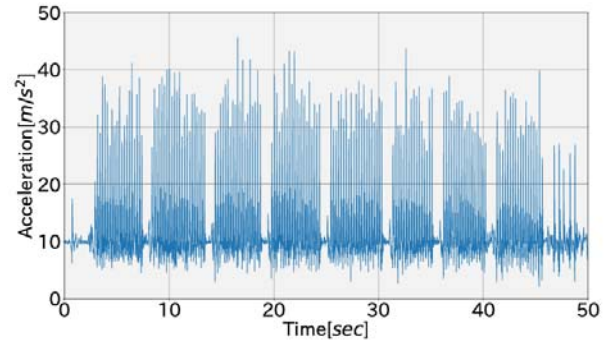


Figure 3: Example of cut acceleration

4.2 Feature Extraction from Cucumber Slicing

From the collected sensor data, the average pace, average acceleration, and standard deviation of acceleration were extracted as features. The results for each feature are shown in Table 1.

Table 1: Features by subject

Subject	Ave pace [sec]	Ave acc [m/s^2]	Standard deviation of acc [m/s^2]
A	0.52	32.6	5.59
B1	0.96	18.1	2.87
B2	0.93	18.0	3.49
B3	0.42	16.6	2.95
C1	0.90	26.8	9.45
C2	0.72	15.1	2.18
D1	0.76	13.1	0.86
D2	0.88	13.4	1.01
E1	0.43	17.7	4.92
E2	1.00	17.0	3.83
F1	0.75	12.7	0.70
F2	1.05	12.3	0.24
G1	0.25	13.2	1.25
G2	0.58	13.2	1.56
H1	1.21	13.1	1.07
H2	0.61	13.4	1.04
I	1.01	17.3	2.14
J	0.94	17.2	3.02
K	0.92	13.0	0.71
L	1.13	14.0	1.62
M	0.86	12.6	0.71
N	0.52	14.0	1.45
O	0.68	15.7	2.66
Ave	0.79	16.1	2.40

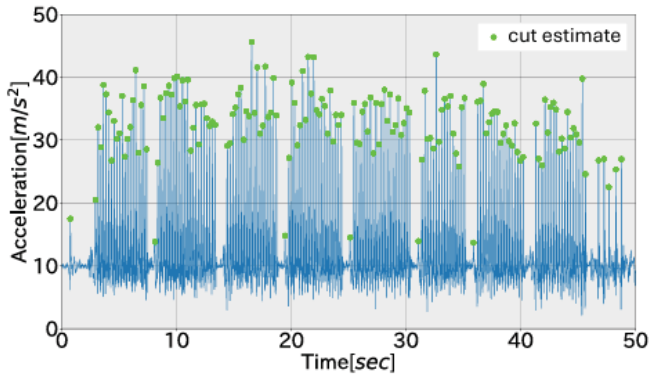


Figure 4: Example of cut acceleration

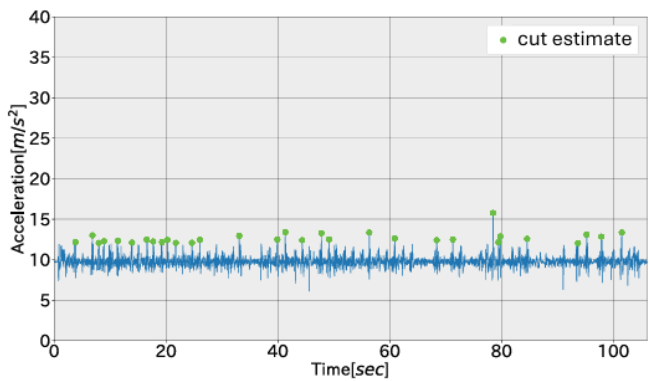


Figure 5: Example of acceleration not exceeding threshold

Average pace represents the time taken to make one cut. This can be used for speed comparison. In this survey, the length of the cucumber and the thickness of the slices were not specified. Therefore, it is considered difficult to compare the number of cuts and the time taken individually, as they depend on these factors. Given this, average pace is used as the indicator of cutting speed.

To determine the average pace, we estimate the number of cuts and the time taken. From the recorded video and sensor data, we confirmed that the acceleration changes significantly when the knife strikes the cutting board. Therefore, we use peak detection to estimate the number of cuts. A cut is estimated at the maximum value between when the acceleration exceeds the threshold and when it falls below it. Additionally, no new estimates are made for 0.2 seconds after the threshold is exceeded to prevent false detections. The estimated results of the cuts are shown in Figure 4.

The estimation of cut counts was also performed with other participants, and the estimation accuracy was examined. The estimation results showed errors between the actual and estimated counts, with an average error of 15.4 cuts. The errors varied among participants. The reasons for this include instances where the acceleration did not exceed the threshold and instances where the acceleration exceeded the threshold due to movements other than cutting. An example where the acceleration does not exceed the threshold is shown in Figure 5. Figure 5 shows an example where the estimated count is 31, but the actual count is 57,

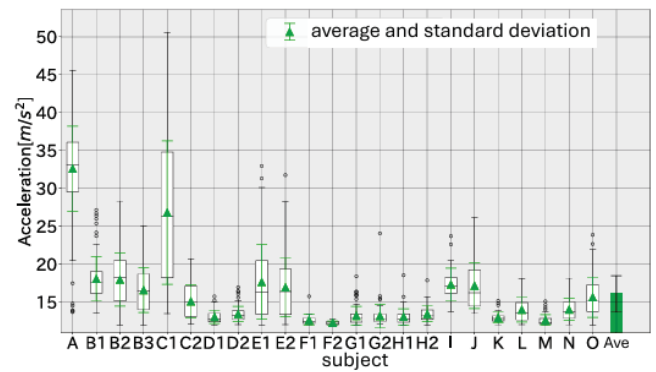


Figure 6: Acceleration of cutting by subject

resulting in an error of 26 cuts. The issue of acceleration not exceeding the threshold can be resolved by appropriately setting thresholds for each participant's acceleration. Regarding the issue of acceleration exceeding the threshold due to non-cutting movements, such as stabbing and discarding the cucumber ends or moving the cut cucumbers, a possible solution is to use only the gravitational component from the three-axis acceleration.

The time spent cutting is determined by the difference between the first and last timestamps of the estimated cuts. Any interval where the cutting is paused for more than 1.5 seconds is excluded from the calculation of the average pace. This exclusion accounts for temporary interruptions in the cutting process, such as removing cucumber pieces stuck to the knife or adjusting the position of the holding hand. The average pace is calculated using these estimated counts and times.

The average and standard deviation of the acceleration at the moment of cutting can reveal information about the applied force and its consistency. The graph of the acceleration at the moment of cutting is shown in Figure 6. The average acceleration is calculated as the mean of the acceleration values at the moment of cutting. Since stronger force results in higher acceleration, this measurement indicates the applied force. The standard deviation of the acceleration is calculated as the standard deviation of the acceleration values at the moment of cutting. When the applied force is consistent, the standard deviation is small, and when there is variation in the applied force, the standard deviation is large. This measurement indicates the consistency of the applied force.

5 ANALYSIS AND COMPARISON OF MOVEMENT BETWEEN FIXED POINTS USING LOCATION INFORMATION

This chapter describes the analysis and comparison of movement between fixed points. In the investigation of movement between fixed points, the steps involved in cooking a hamburger steak are analyzed to extract relevant features. To examine movement-related features, it is necessary to analyze a cooking process that involves movement. The hamburger steak cooking process includes movement between locations such as the cutting board, stove, and sink, making it suitable for this investigation.

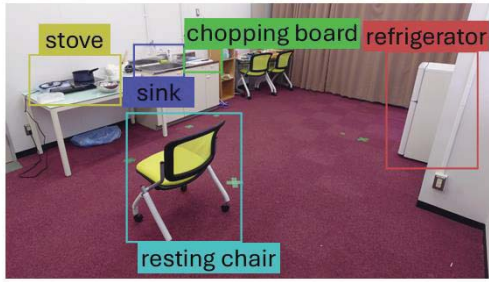


Figure 7: Cooking Environment

5.1 Investigation of Movement During Hamburger Steak Cooking Process

To analyze and compare movement between fixed points, we examined the cooking process of making hamburger steaks to identify relevant features.

For data collection, we performed sensing during the hamburger steak cooking process. The subjects were four students. The layout of the cooking environment is shown in Figure 7. The hamburger steak cooking process consists of the following steps. First, ingredients are taken out from the refrigerator. Next, the onions are finely chopped. The chopped onions, minced meat, beaten egg, milk, breadcrumbs, and seasonings are mixed in a bowl. After kneading the mixture, it is shaped, and the air is removed. The formed patties are then cooked over medium heat for 10 minutes. During the cooking process, there is a waiting period while the hamburger steaks are being cooked. During this waiting time, the participants often engage in other tasks such as washing the used utensils. The activities during this waiting time can vary between skilled and inexperienced individuals, making it a useful point of comparison. For this data collection, the following locations were used: refrigerator, cutting board, sink, stove, and a resting chair. The cooking process involves steps such as chopping, mixing, and cooking, which require at least the cutting board and the stove. Additionally, a refrigerator for storing ingredients and a sink for washing utensils are needed. A resting chair was also provided for waiting periods, such as during the cooking time. Specific work positions were designated for each location.

For sensing, we used mocopi¹ to collect movement history. Mocopi is a device that can perform motion capture by attaching inertial sensors to six locations: the head, waist, both arms, and legs. To achieve the objectives of this study, it would be preferable to use a smartphone or smartwatch for position estimation using PDR. However, position estimation using these devices is difficult and does not yield satisfactory results. This study focuses more on extracting features from movement rather than position estimation. Therefore, mocopi, which easily acquires positional information, was used. Although mocopi is not as widely used as smartphones or smartwatches, it is commercially available to the general public and is easier to obtain compared to other dedicated sensing devices.

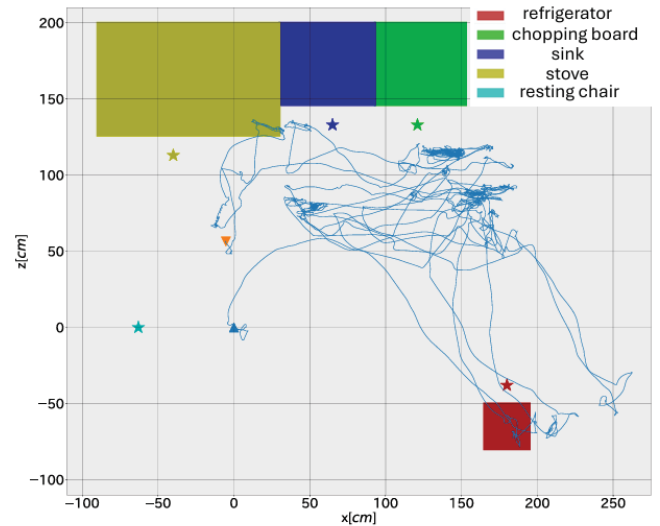


Figure 8: Recorded movement trajectory

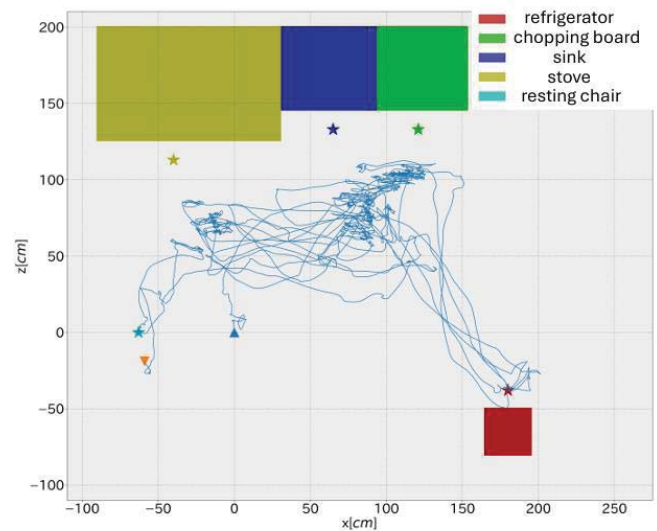


Figure 9: Corrected movement trajectory (Subject A)

5.2 Movement Estimation Using Wearable Sensors

We performed position estimation using wearable sensors to investigate the overall accuracy. Temporal changes in position estimation were treated as movement trajectories. The obtained movement trajectories are shown in Figure 8. The "▲" marks the starting point, the "▼" marks the endpoint, and the "★" indicates the working positions. The colored areas represent non-entry zones.

To correct the cumulative errors, we utilized specific stopping points and waist height data. Predefined stopping locations were established during data collection. Among these, the refrigerator and the rest chair are notable because the waist height becomes lower than usual when stopping at these points. Therefore, when the waist height is detected to be lower, it is presumed that the subject is at either the refrigerator or the rest chair. This information is used to

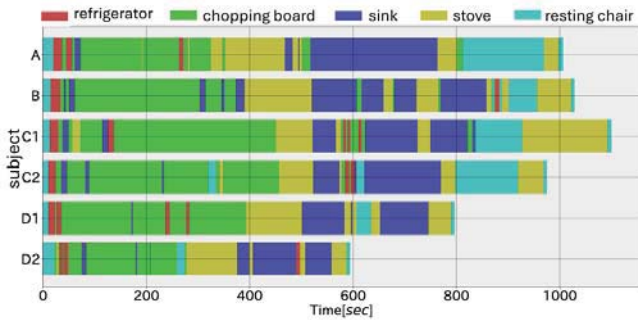


Figure 10: Correct answer data by subject

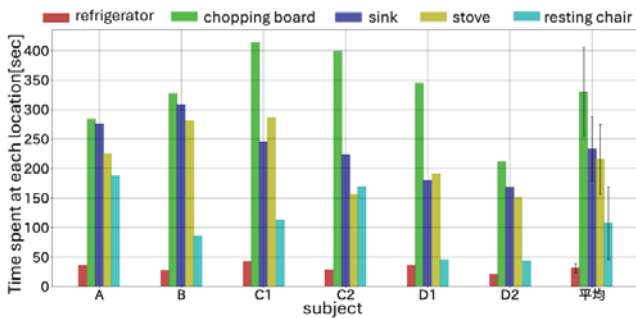


Figure 11: Cumulative stay time per location

correct the position by assuming the subject is at one of these two locations when the waist height drops. The correction applied at this point is then used to adjust the accumulated trajectory errors from the previous correction point or the starting point.

The actual results of the corrected movement trajectory are shown in Figure 9. The location estimation was performed by assuming that the nearest stopping point from the current location was the actual current location. In the case of Subject A, the corrected trajectory provided a reasonably accurate estimation. The same correction method was applied to other subjects as well. However, the accuracy obtained for other subjects was not as high as that for Subject A. For example, in the correct data, there were instances where the subject was sitting on a chair, but the trajectory did not reflect a correction towards the chair. Upon examining the actual data, it was found that the waist height did not fall below the threshold. The overall skeletal data suggested that the subject was leaning against a wall rather than sitting. This indicates that the system failed to recognize the sitting posture.

Additionally, the estimation accuracy was found to be weak for locations that are close to each other, such as the sink and the cutting board. To improve estimation accuracy, strategies such as making the subject sit periodically to increase correction frequency, and increasing the number of correction points, can be considered. Furthermore, while this correction only addressed coordinates, it is also necessary to correct for rotation. The movement trajectories obtained in this study showed significant variability in accuracy among subjects. There is still room for improvement in accuracy. For the comparisons that follow, the correct data will be treated as the movement trajectory.

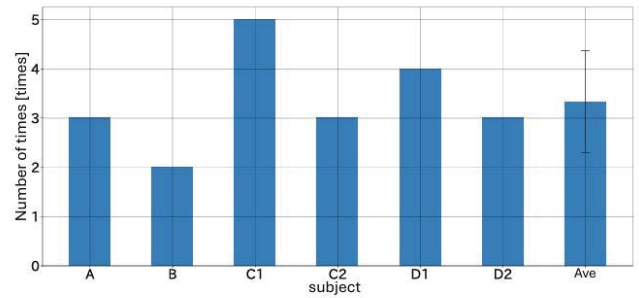


Figure 12: Number of trips to the refrigerator

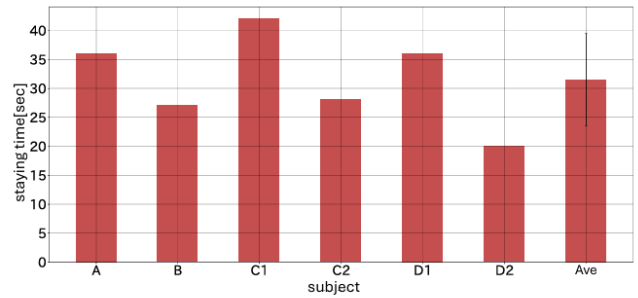


Figure 13: Estimated time spent in refrigerator

5.3 Analysis and Comparison of Movement Trajectories

Using the ground truth data, we analyzed and compared the data to examine the differences among participants. Figure 10 shows the ground truth data for each participant.

First, we analyzed and compared the residence time at each location. Figure 11 shows the cumulative residence time at each location. As an example, we compared the time spent at the sink. Since most of the time spent at the sink involves washing dishes, we compared how quickly each participant completed this task. From Figure 11, we can see that Participant 4's second attempt was the fastest, while Participant 2 took the longest. This allows us to analyze whether a participant is faster or slower than the overall average when comparing the time spent on dishwashing to the overall average.

Furthermore, we analyze and compare the transitions between locations. The transitions between locations are calculated by counting the number of times a subject moves from the current location to the next location. The transitions from A to B and from B to A are counted separately. The method of counting considers a transition when the subject stays at the current location for more than one second before moving to the next location. An example of the obtained data is shown in Figure 12. In this instance, we analyze and compare the number of times the subject goes to the refrigerator. Figure 13 shows the number of times each subject goes to the refrigerator and the overall average. At least, the subject should go to the refrigerator twice: once to take out the ingredients and once to put them back. However, a high number of visits may indicate that the subject is taking out and putting back ingredients as needed, or they

may be making multiple trips because they cannot carry everything at once. Therefore, we do not analyze skillfulness or unfamiliarity as these judgments vary among users. From this data, users can see how frequently or infrequently they open the refrigerator compared to the overall average. If the number is low, it indicates efficient movement; if the number is high, the user might try to reduce it.

In this way, comparisons of stay durations and transitions between locations revealed the skillfulness or unfamiliarity of users when compared with others, and showed growth when compared with their past performances. While we obtained these features, the current system is not yet viable as it relies on ground truth data. If we can improve the accuracy of location estimation for all subjects, it will be possible to estimate the processes users are performing.

6 A WEB SYSTEM FOR PROMOTING SELF-ANALYSIS THROUGH THE COMPARISON AND SHARING OF COOKING ACTIONS

In this chapter, we discuss methods for promoting self-analysis. We propose a web system for displaying and sharing feature data through graphs, and we investigate the impact of graph displays on users.

6.1 Methods for Collecting, Displaying, and Sharing Feature Data

For comparison, it is necessary to have feature data from multiple individuals who have cooked the same recipe. Therefore, the proposed system wraps around Cookpad. Cookpad has a function called "tsukurepo" which allows users to share the results and impressions of a recipe they tried. This tsukurepo data is linked and stored with feature data. On Cookpad, recipes, tsukurepo, and users are assigned variable-length integer IDs. These IDs are obtained via web scraping and stored in a database. When displaying recipes and tsukurepo in the proposed system, data is retrieved using these IDs. The sensor data registration screen is shown in Figure 14. A list of posted Tsukurepo and data uploaded from the wearable sensor are displayed. By selecting each and pressing the register button, the data is registered.

In displaying feature data, methods that allow comparisons with past performances and other users' data need to be considered. As mentioned in Chapter 3, comparison with others should reveal how proficient one's actions are and whether they resemble those of skilled cooks. Comparisons with one's past should show the degree of personal growth. To meet these needs, bar graphs and histograms are effective for comparisons with others. Sorting bar graphs allows users to see their relative position within the group, while histograms show how much their abilities differ from others. Line graphs are effective for comparisons with one's past, as plotting over time reveals the degree of improvement in cooking skills. Histograms can also accommodate both types of comparisons, allowing for overlapping displays of data to show differences

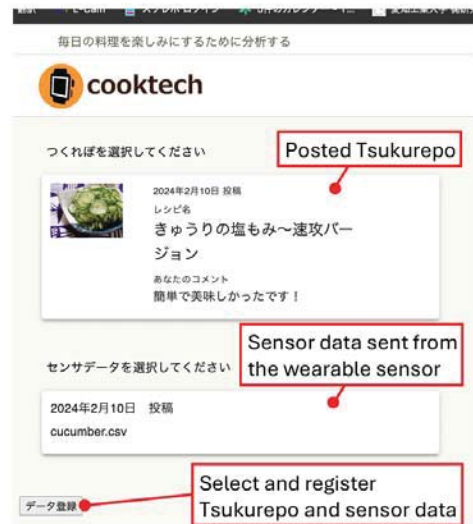


Figure 14: Sensor data registration

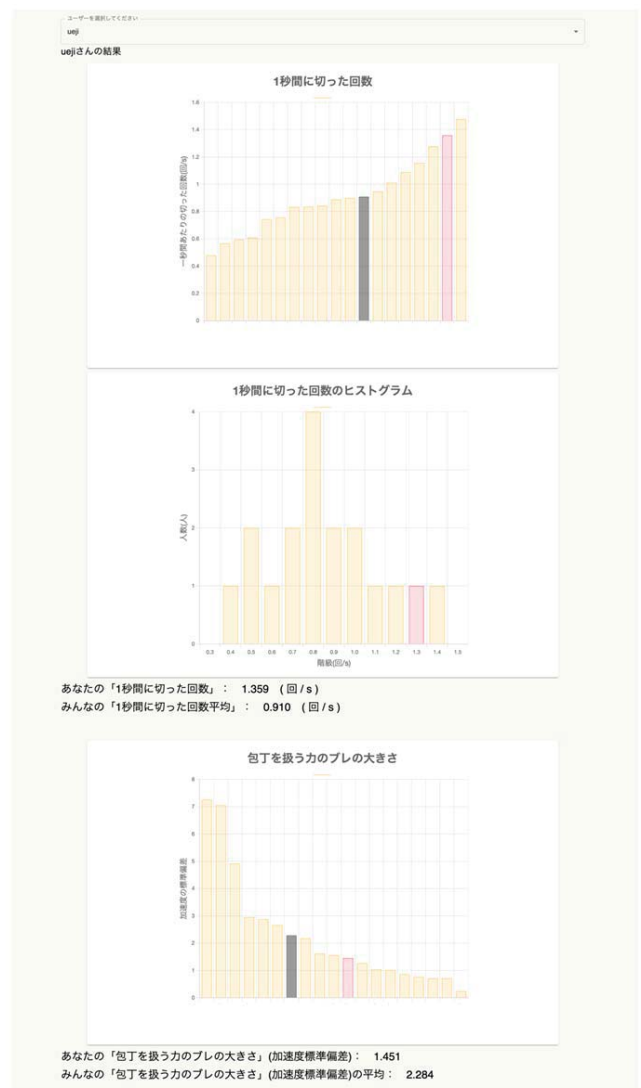


Figure 15: Example of feedback to the user from a web system

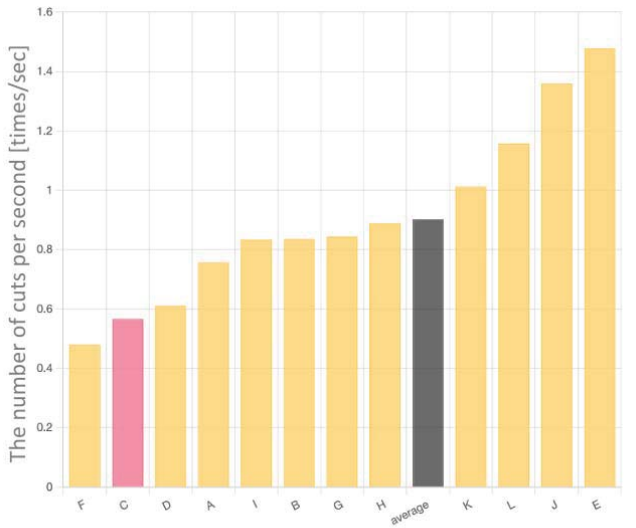


Figure 16: Graph presented to the participants

in specific metrics. The graph display screen is shown in Figure 15. This screen shows the speed and strength of the cuts.

Self-analysis is primarily conducted to identify areas for improvement and growth. Therefore, if actual improvements lead to enhanced abilities, methods to visualize this progress must also be considered.

6.2 Investigating Whether Displaying Graphs of Feature Data During Cooking Can Promote Self-Analysis

We conducted an experiment to investigate the impact of visualizing feature data through graphs. The first experiment aimed to determine whether graph displays of feature data during cooking could assist in self-analysis. We hypothesized that comparing and visualizing feature data for the same task through graphs would help identify differences in cooking actions between individuals. The subjects were 10 men and women in their twenties. Each participant wore a smartwatch on the arm holding the knife while slicing cucumbers. The cutting actions were sensed using the smartwatch. Feature data was extracted from the accelerometer values. In this experiment, the features were the number of slices made per second and the variance in force when the knife cut through the cucumber. The criteria during this experiment were that a higher number of slices per second and a smaller variance in force were considered better.

After sensing, a graph comparing the feature data of the participant with that of others was displayed. Figure 16 shows the graph presented to the participants. The graphs display the data of individual participants and the average of all participants using bar charts and histograms. The bar charts are sorted in ascending or descending order for each item, while the histograms show the number of people in each class and the class to which the individual belongs. The graph displays the average pace of slicing cucumbers and the variance in the strength of the cuts.

Table 2: Survey results on the impact of feature data graph display on users(n=10)

Was there any change in your evaluation of your own cooking before and after comparing the graphs?	
Yes	30 %
A few	40 %
Neutral	20 %
Not many	10 %
None	0 %
Did you find any points for improvement in your cooking?	
Yes	50 %
A few	40 %
Neutral	0 %
Not many	10 %
None	0 %
Did you discover anything new about your cooking through the graph comparison?	
Yes	70 %
A few	20 %
Neutral	0 %
Not many	0 %
None	20 %

Before and after the graph display, participants perform self-analysis and answer a questionnaire. The questionnaire is rated on a 5-point scale for each question. Participants also provide reasons for each of their answers. However, the purpose of this experiment is to investigate the impact of graph displays on users, and data accuracy is not considered.

From the experimental results, it was found that displaying data in a graph encouraged self-analysis. Table 2 shows the results of the questionnaire. In response to the question, "Did your evaluation of your own cooking change before and after comparing with the graph?" 70% of the respondents answered that it changed or slightly changed. The opinions of these subjects included statements such as "I realized I wasn't putting in enough effort compared to others," "My average pace was the highest," and "I was surprised to be below average." These opinions indicate that the graphical display led to comparisons between oneself and others or the average. Additionally, one subject noted, "I didn't know there was such a large variation in the force used with the knife until now," indicating that they learned something new. Other responses included, "I didn't understand what 'degree of variation' and 'mean and standard deviation' represented and what needed to be improved. I only understood that my cutting pace was slow." These responses suggest that if the display method or the content of the graph labels is unclear, it does not result in effective visualization.

In response to the question, "Were there any points you could improve in your cooking?" 90% of the respondents answered that there were or there were some. The opinions of these subjects included, "The speed was good, but there was some inconsistency," "I realized my knife handling was not great and decided to improve it," and "After looking at the graph, I thought it might be better if I could cut the ingredients a bit faster." From these responses, it is clear that the graph display helps in identifying areas for improvement. Other opinions included, "There were some micuts and thicker cuts," which are not related to the graph

display, suggesting that the graph may have provided little insight for these participants. Additionally, one opinion stated, "According to the graph, I'm the best, so I don't know what to improve." This indicates that in small groups, when someone is rated the best, they may not be able to identify areas for improvement.

In response to the question, "Were there any points you could improve in your cooking?" 90% of the respondents answered that there were or there were some. The opinions of these subjects included, "The speed was good, but there was some inconsistency," "I realized my knife handling was not great and decided to improve it," and "After looking at the graph, I thought it might be better if I could cut the ingredients a bit faster." From these responses, it is clear that the graph display helps in identifying areas for improvement. Other opinions included, "There were some miscuts and thicker cuts," which are not related to the graph display, suggesting that the graph may have provided little insight for these participants. Additionally, one opinion stated, "According to the graph, I'm the best, so I don't know what to improve." This indicates that in small groups, when someone is rated the best, they may not be able to identify areas for improvement.

In response to the question, "Did you make any new discoveries about your cooking through the comparison with the graph?" 90% of the respondents answered that they did or somewhat did. The opinions of these subjects included, "My cutting rhythm was surprisingly consistent," "My cutting speed was faster compared to others," and "I had never compared my cooking with someone else's, so it was very refreshing to know where I stood." These responses indicate that many people were able to make new discoveries through the graph display. Additionally, one participant mentioned, "I realized that I am not good at cooking," indicating that merely displaying the data can lead to negative perceptions about one's cooking. Among those who answered that they did not make any new discoveries, one opinion was, "I realized from the graph that my cutting speed was slow, but I was already aware of that. I didn't understand what 'degree of variation' and 'mean and standard deviation' represented." This suggests that those who are already aware of their cooking skills may not find new insights, and that unclear display methods or graph labels do not result in effective visualization. From these survey results, it is evident that displaying data in graphs encourages self-analysis.

6.3 Investigating the Impact of Graph Display on Ability Change and the Effect of Multiple Graph Displays on Awareness

As a second evaluation experiment, the impact of multiple graph displays on skill improvement was investigated. The experiment aimed to determine whether data visualization through graphs can enhance cooking skills. The subjects included six individuals who participated in the first experiment. These six subjects underwent the same sensing process in the same conditions as the first experiment. The second experiment was conducted approximately one week after the first experiment to ensure that the practice of

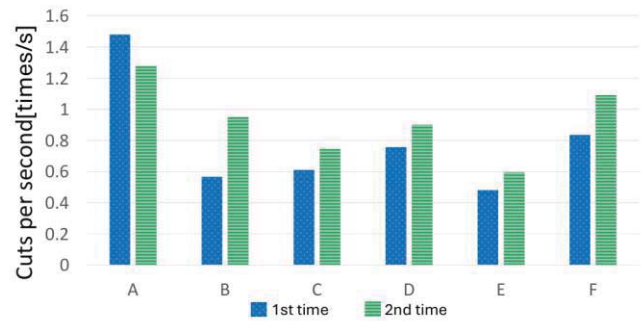


Figure 17: Change in cuts per second between first and second attempt

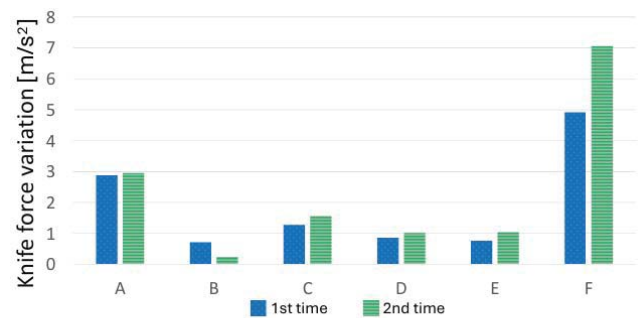


Figure 18: Change in knife force variation between first and second attempt

slicing cucumbers from the previous session did not influence the results. This setup was designed to examine whether the subjects' awareness changed due to the graph display and how the results differed due to changes in awareness. The results of the subjects' cooking skill progression are shown in Figures 17,18.

While changes in cooking ability were observed, there was no evident growth trend. Five out of six participants increased the number of cuts per second. Additionally, the variation in the force applied while using the knife decreased for one out of six participants. It is speculated that the subjects focused on speed when slicing cucumbers. This focus on speed likely led to an increase in the number of cuts per second for all participants, but resulted in uneven application of force. When a particular metric appears easier to improve, participants might prioritize changing that specific aspect. Therefore, the displayed content should be designed to avoid bias towards any single metric and instead encourage an awareness of overall skill development across all steps of the process.

The results of the questionnaire are shown in Table 3. In response to the question, "Did you pay attention to anything in particular compared to the first time?" 83.3% of the respondents answered that they did or somewhat did. This indicates that the first graph display helped participants identify areas for improvement. The comments included, "I wanted to increase my pace," "I focused on adding more force since I lacked it last time," and "I tried to cut at a steady tempo based on the graph results." These comments suggest that participants focused on speed after the first display. Additionally, comments such as "I tried not to miss my cuts" and "I thought my cutting pace and cucumber

Table 3: Survey Results on the Impact of Two Graph Displays on Users(n=6)

Did you notice anything specific compared to the first time?	
Yes	50.0 %
A few	33.3 %
Neutral	0.0 %
Not many	0.0 %
None	16.7 %
Were there any changes when comparing the results of the first and second times?	
Yes	33.3 %
A few	16.7 %
Neutral	0.0 %
Not many	16.7 %
None	33.3 %

thickness were not very good” indicate that participants were mindful of aspects not explicitly shown on the graph.

In response to the question, “Did you notice any changes when comparing the results of the first and second experiments?” 50% of the respondents answered that they did or somewhat did. This suggests that a displaying only once did not result in significant changes. Comments from those who noticed changes included, “My speed increased, and the variation in force decreased compared to the first time,” and “The pace became somewhat faster.” These responses show that the graph display highlighted differences between the first and second attempts. On the other hand, those who answered that they did not notice changes commented, “It was almost the same” and “No.” This indicates that a displaying only once did not make them aware of changes.

These results demonstrate that while graph displays can influence awareness even after one viewing, a displaying only once is insufficient to show a growth trend in cooking ability.

7 CONCLUSION

This study aimed to assist in self-analysis through the visualization and sharing of features extracted from sensor data of recipes and cooking actions. The approach involved using wearable sensors to sense actions during cooking, extracting features, collecting feature data from multiple individuals cooking the same recipe, visualizing these features in graphs, and sharing them. Users could then identify their areas of improvement and strengths by comparing the graphs.

For fixed-point actions, the analysis was conducted using the example of slicing cucumbers. A smartwatch was worn on the arm holding the knife to capture acceleration data. Features such as average pace, average acceleration, and the standard deviation of acceleration were obtained. For actions involving movement between fixed points, the analysis was conducted using the example of cooking hamburger steaks. Movement trajectories were determined using mocopi, but due to insufficient accuracy, the analysis was based on correct labels. Features such as duration of stay, cooking time, and location transitions were obtained. Various methods for visualizing and sharing the feature data comparisons were explored.

Experiments were conducted to investigate whether graph visualization facilitated comparisons. The results showed that the graph displays enabled participants to compare themselves with others and identify their strengths and areas for improvement. An additional experiment investigated the impact of multiple graph displays. The results indicated that while the graph displays altered participants’ awareness and resulted in changes in their cooking abilities, a displaying only once did not show a growth trend. From the above, it was found that self-analysis can be assisted through the visualization of graph displays of the results of sensing cooking activities using wearable sensors.

Future challenges include investigating cooking actions other than slicing, such as frying or mixing. This can provide clues for improving multiple actions. Currently, only acceleration is used for estimation. By using other sensors such as angular velocity, new features can be discovered.

It is necessary to classify each process by process estimation. Currently, only single processes such as cutting are targeted. In reality, multiple processes are included in a single cooking session. If processes can be estimated from a series of cooking actions, it can accommodate various recipes.

The accuracy of action estimation needs to be improved. Currently, the accuracy of action estimation is poor, and correct data is obtained from videos taken with a camera. This process is manual and time-consuming. Moreover, the results of action estimation could potentially be useful for process estimation.

The display method needs improvement. One participant noted in the free comment section of the evaluation experiment, “Even if I know what’s wrong, I don’t know how to fix it. Specific advice would be helpful.” The content displayed should also avoid emphasizing any one aspect disproportionately. As indicated by the second evaluation experiment, when an easily improvable aspect like “speed” is apparent, more difficult aspects like “reducing variation” might

be neglected. Providing specific methods for improving “reducing variation” could encourage improvements in multiple areas. This study promotes self-analysis but does not support the actual improvement process. Since the purpose of self-analysis is to facilitate improvement from the current state, providing specific improvement methods could better assist in self-analysis and improvement.

To investigate the impact of multiple, continuous graph displays over a long period, it is necessary to increase the number of trials in the experiments. In the second evaluation experiment, the graph was displayed after the first experiment, and then the experiment was conducted again. Although there was a difference in participants’ awareness, no significant difference in skills was found. Therefore, a long-term experiment is needed to examine the impact of graph displays. Additionally, to determine whether the impact is due to the graphs, a control group that does not receive graph displays should be included for comparison.

REFERENCES

- [1] De, Debraj and Bharti, Pratoool and Das, Sajal K. and Chellappan, Sriram, Multimodal Wearable Sensing for Fine-Grained Activity Recognition in Healthcare, *IEEE Internet Computing* 19(5), pp.26-35, 2015.
- [2] Murakami Tomoko, Tori Kentaro, Tyo Kenta, Uchihira Naoshi, Estimation of nursing work behavior using topic models from sensor data and work knowledge, *Transactions of the Japanese Society for Artificial Intelligence : AI*, Vol.29, No.5, pp.427-435, 2014.
- [3] Uchida Yasuhiro, Sawamoto Jun, Sugino Eiji, A system to recognize human activity using acceleration senso, *IEICE Technical Report*, Vol.115, No.232, pp1-6, 2015- 9.
- [4] Shimauchi Takeaki, Katsuki Takashi, Toyoda Osamu, Behavior Sensing using Shoes Equipped with Small Sensor Module, *MES Papers* 28, pp.307-310, 2018.
- [5] Ogawara Ryo, Hatano Hiroyuki, Fuji Masahiro, Itoh At-sushi, Watanabe Yu, Location Estimation System Based on GPS Positioning and Sensor Information, *IPSJ Journal* 56(1), pp.2-12, 2015-01-15.
- [6] Fujita Susumu, Kaji Katsuhiko, Kawaguchi Nobuo, A Wireless LAN Location Estimation Method Based on Gaussian Mixture Model, *Journal of Information Processing*, Vol.52, No.3, pp.1069-1081, 2011.
- [7] Bahl, Paramvir and Padmanabhan, Venkata N, RADAR: An in-building RF-based user location and tracking system, *Proceedings IEEE INFOCOM 2000. Conference on computer communications. Nineteenth annual joint conference of the IEEE computer and communications societies (Cat. No. 00CH37064)*, Vol.2, pp.775-784, 2000.
- [8] Subedi, Santosh and Pyun, Jae-Young, A Survey of Smartphone-Based Indoor Positioning System Using RF-Based Wireless Technologies, *Sensors* 2020, 20 (24), 2020.
- [9] Omae Manabu, Use of LiDAR in autonomous driving, *Journal of the Society of Instrument and Control Engineers*, Vol.59, No.5, pp.311-315, 2020.
- [10] Iwasa Hidehiko, Aihara Nobuhiro, Yokoya Naokazu, Takemura Haruo, Memory-Based Self-Localization Using Omnidirectional Images, *The transactions of the Institute of Electronics, Information and Communication Engineers*, Vol.J84-D2, No.2, pp.310-320, 2001.
- [11] Wang, Lei and Deng, Jianzhi and Zhang, Fengming, Trajectory Estimation Method of People in Forest Based on Inertial Sensor, *EBIMCS '20: Proceedings of the 2020 3rd International Conference on E-Business, Information Management and Computer Science*, pp.582-586, 2021-12.
- [12] Ueda Kenki, Ohki Hiromu, Mizumoto Teruhiro, Tamai Morihiko, yasumoto keiichi, *DPSWS papers* 2013, No.6, pp.116-118, 2013-11-27.
- [13] Miyata Masahiro, Tamaru Eriko, Futai Hiroyuki, Hayashi Koichi, An Attempt to Visualize Student Behavior during Class Activity using Wide-Area Sensing Methods, *Research Report Computers and Education (CE)*, Vol.2023, No.7, pp.1-6, 2023.
- [14] Urano Kenta, Hiroi kei, Yonezawa Takuro, Kawaguchi Nobuo, A LED Device for Visualizing Heart Beat, *Proceedings of the 2248 Symposium on Multimedia, Distributed Coordination and Mobile*, Vol.2020, pp.1616-1622, 2020.
- [15] Inaba Hiroshi, Taki Tsuyoshi, Miyazaki Shin-ya, Hasegawa Jun-ichi, Koeda Mitsuhiro, Yamamoto Hide- hiro, Kitagawa Kaoru, Visualization of Human Body Sensing for Supporting Sports Motion Analysis, *Papers for Special Issues "2003 NICOGRAPH Spring"*, Vol.2, No.3, pp.94-100, 2023.
- [16] Lim, Hyoyoung and Huang, Xiaolei and Miller, Samuel and Edelmann, Joshua and Euken, Timothy and Voids, Stephen, Smart Cook: Making Cooking Easier with Multimodal Learning, *djunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers*, pp.129-132, 2019-9.
- [17] Ninomiya Akari, Ozaki Tomonobu, Cooking Recipe Analysis Based on Sequences of Distributed Representation on Procedure Texts and Associated Images, *Proceedings of the 12th Workshop on Multimedia for Cooking and Eating Activities*, No.6, pp.13-18, 2020.
- [18] Okumura, Tsukasa and Urabe, Shuichi and Inoue, Katsufumi and Yoshioka, Michifumi, Cooking Activities Recognition in Egocentric Videos Using Hand Shape Feature with Openpose, *CEA/MADiMa '18: Proceedings of the Joint Workshop on Multimedia for Cooking and Eating Activities and Multimedia Assisted Dietary Management*, pp.42-45, 2018-7.
- [19] Kobayasi Hanano, Kato Takehiro, Yokokubo Anna, Guillaume Lopez, Proposing a System to Supporting for Kitchen Knife Skill Improvement using Acceleration Sensor, *Proceedings of the Symposium on Multimedia, Distributed, Cooperative and Mobile 2022*, pp.1000- 1003, 2020.
- [20] Khaloo, Pooya and Oubre, Brandon and Yang, Jeremy and Rahman, Tauhidur and Lee, Sunghoon Ivan, NOSE: A Novel Odor Sensing Engine for Ambient Monitoring of the Frying Cooking Method in Kitchen Environments, *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, Vol.3, No.2, pp.1-25, 2019.
- [21] Lago, Paula and Takeda, Shingo and Alia, Sayeda Shamma and Adachi, Kohei and Benaissa, Brahim and Charpillet, François and Inoue, Sozo, A dataset for com- plex activity recognition with micro and macro activities in a cooking scenario, preprint, 2020.

Keynote Speech 1:

Dr. Seishi Hanaoka

(General Manager, Hitachi, Ltd.,
R&D Group, Center for Digital
Services)

(Chair: Yuichi Tokunaga)

New World that the Digital Connects and Opens

Seishi Hanaoka, Ph.D.
General Manager of Center for Digital Services
Hitachi, Ltd., Research and Development Group

September 2nd, 2024

© Hitachi, Ltd. 2024. All rights reserved.

Contents

1 | Background and introduction of Society 5.0 and CPS

2 | 6 Key factors for Smart City

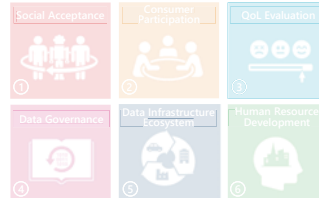


3 | Next step and for future – new world

Contents

1 Background and introduction of Society 5.0 and CPS

2 6 Key factors for Smart City



3 Next step and for future – new world

1-1. Social Issues Facing Humanity

The collage features three main sections with text overlays:

- Urbanization:** 72% increase in urban populations worldwide (2050)
- Changing demography:** Changes in workforce; The ratio of population aged 60 and over to double in 2050 from 2000
- Climate change and resource shortages:** Global demand for energy to be increased by 50% and water usage as well by 40% in 2030

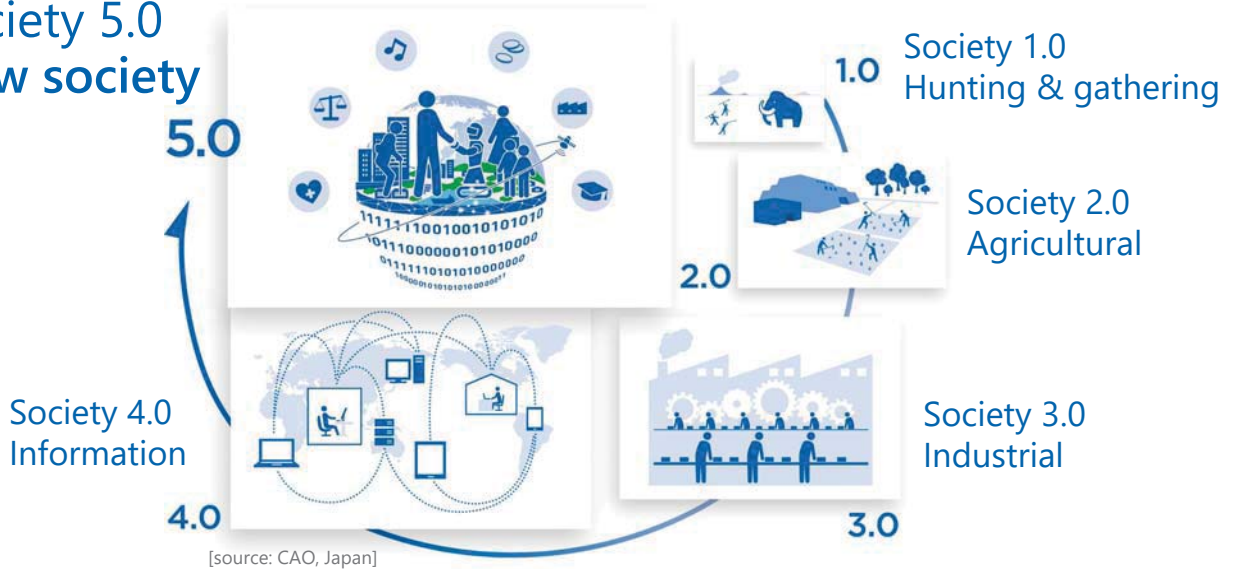
At the bottom, there is a grid of 17 Sustainable Development Goals (SDGs) icons. The text "Society 5.0" is on the left and "SDGs" is on the right.

*Source: PwC Japan homepage; "Megatrends: Five megatrends and potential effects"

1-2. Society 5.0

A human-centered society in which economic development and the resolution of social issues are compatible with each other through a highly integrated system of cyberspace and physical space.

Society 5.0 New society



© Hitachi, Ltd. 2024. All rights reserved.

5

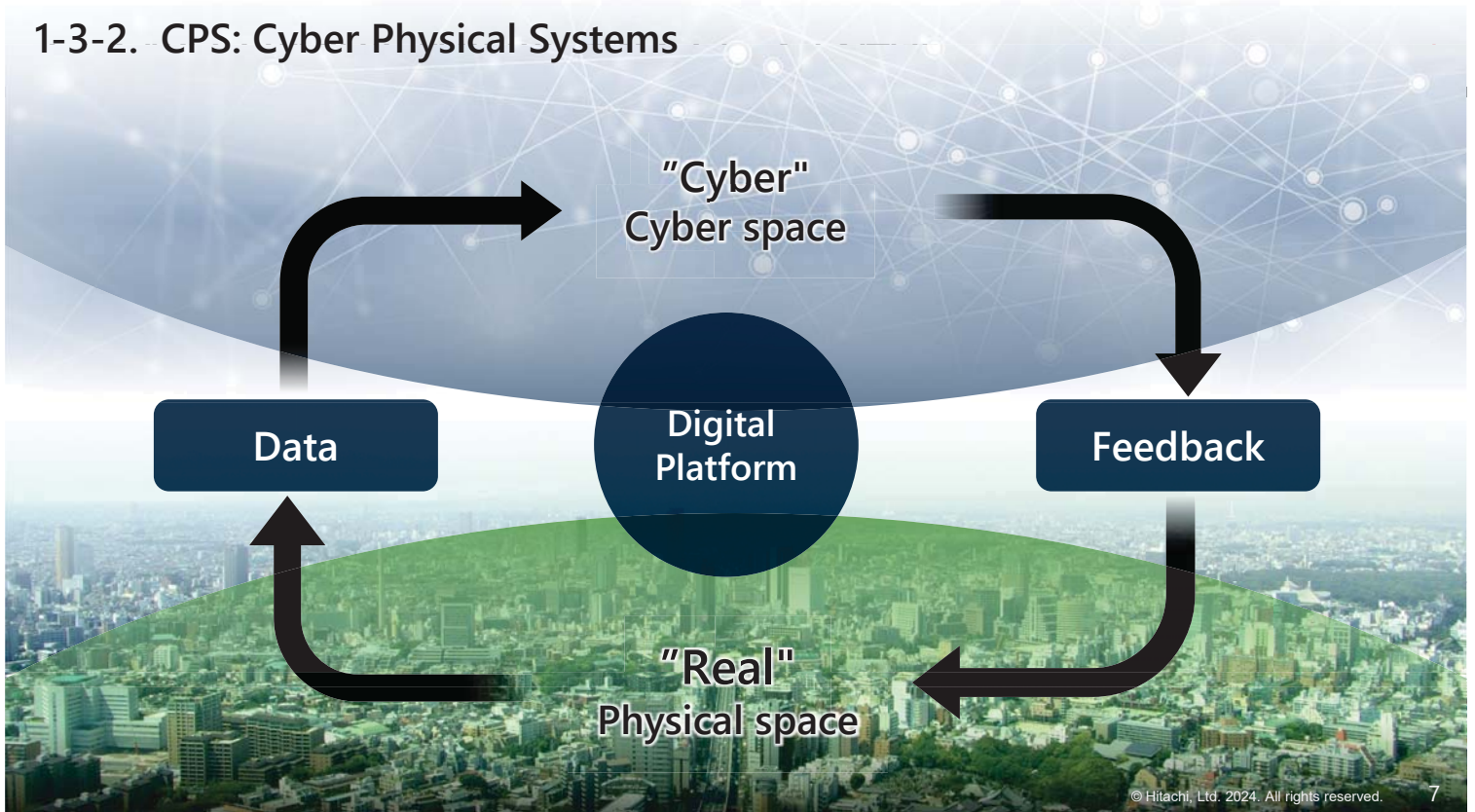
1-3-1. 6 Key factors for Smart City (Society 5.0)



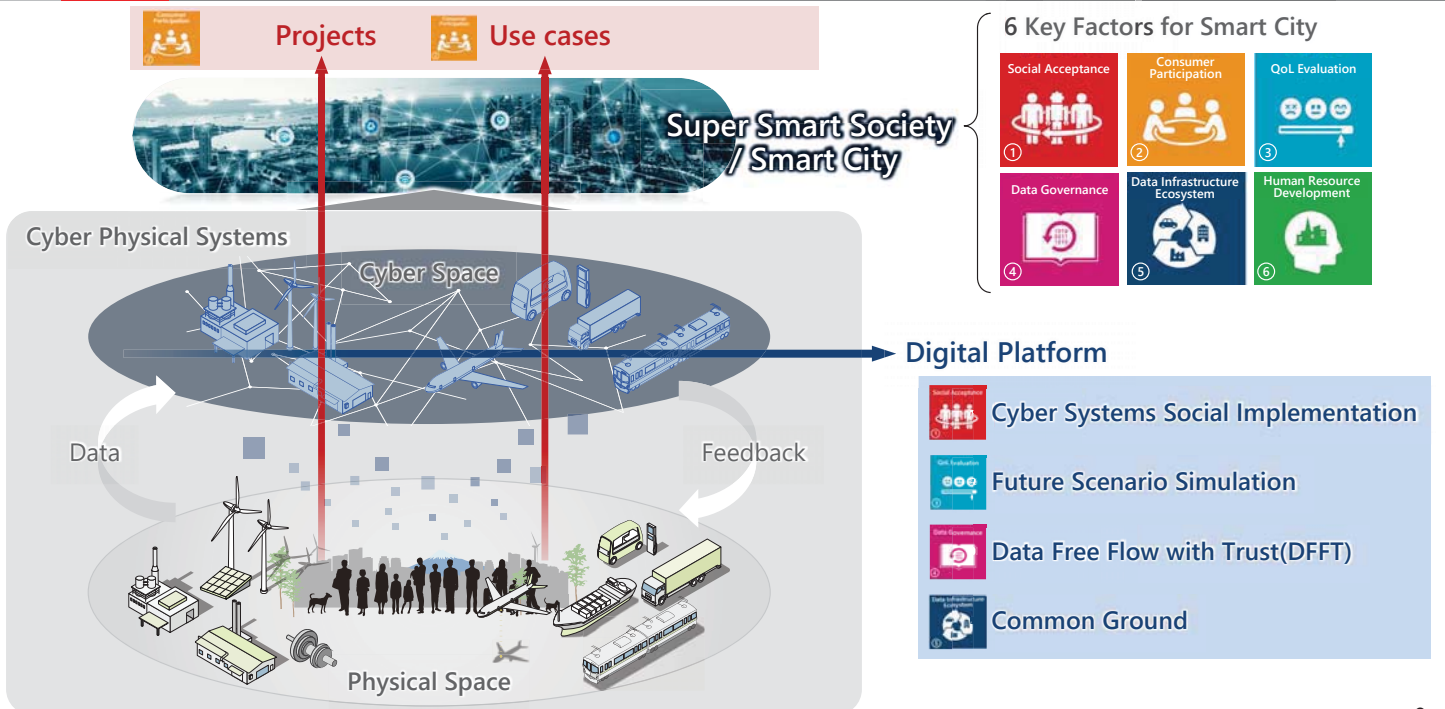
OT: Operational Technology

© Hitachi, Ltd. 2024. All rights reserved.

1-3-2. CPS: Cyber Physical Systems



1-3-3. CPS: New world that the digital connects and opens



Contents

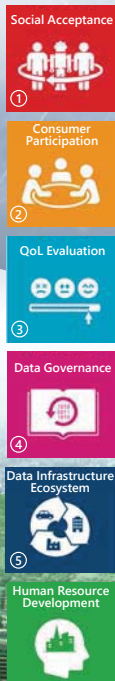
1 Background and introduction of Society 5.0 and CPS

2 6 Key factors for Smart City



3 Next step and for future – new world

6 Key factors for Smart City (Society 5.0)



Social Acceptance

Customer Participation

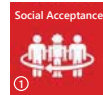
QoL Evaluation

Data Governance

Data Infrastructure Ecosystem

Human Resource Development

2-1-1. Key Factor 1 : Social Acceptance

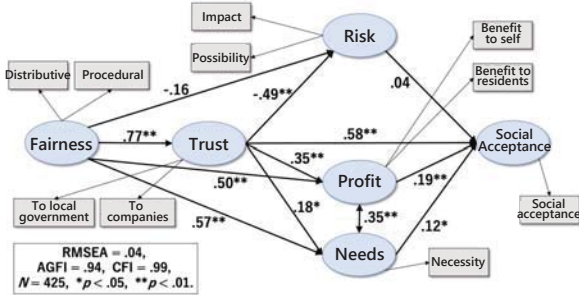


Although detailed services are expected using personal information, there are concerns that the collection and use of personal information will become a problem.

Importance of social acceptance

= local communities deepen their understanding, make autonomous decisions and accept measures.

Understanding the general nature of social acceptance

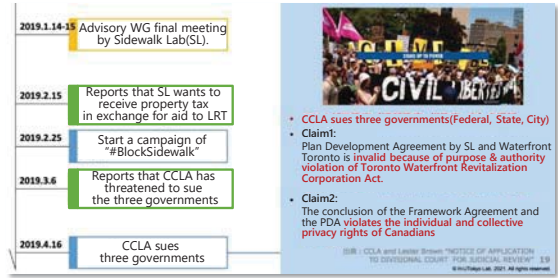


Attitudinal model of social acceptance

Importance of psychological factors related to technology implementation and operation, such as "trust" and "fairness"

LRT: Light Rail Transit, CCLA: The Canadian Civil Liberties Association

Understanding the local history of social acceptance

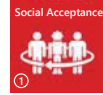


History of citizen protests at Sidewalk Toronto

Organizing the local implementation process for domestic cases and Toronto cases

© Hitachi, Ltd. 2024. All rights reserved. 11

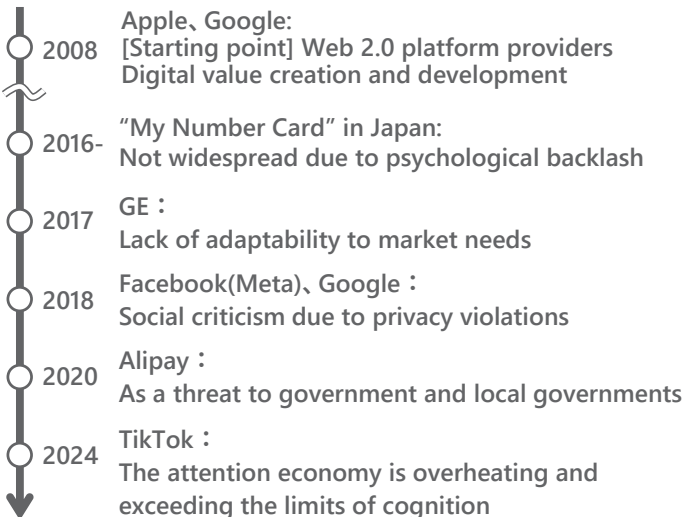
2-1-2. Past developments in cyber systems and challenges



The penetration of cyber systems into society requires consideration of culture, laws, etc.

Examples of issues related to social implementation

One of the issues at stake is the shift to stakeholder capitalism



© Hitachi, Ltd. 2024. All rights reserved. 12

2-1-3. Changing tide in cyber social implementation



Shift from "Creating new services" to "Establishing them in society"

in the previous page

- As GAFA has faced, the introduction of digital technology requires it to be in harmony with laws and social norms.
- There is a need to change behavior based on user needs and internal factors

"Mechanism" for social acceptance and expansion is required.

Precedent cases



Making Sense (Barcelona, 2015-2017)
Formation of a community of "smart citizens" who create new ways of creating and using data. This started with the introduction of an IoT platform called "Sentilo," which installs sensors all over the city, but has expanded to include social considerations.
<http://making-sense.eu/>



SIP (Japan, 2022)
To translate research results into social implementation, we will develop strategies and utilize indicators for social implementation from five perspectives: not only technology, but also business, systems, social acceptability, and human resources.
https://www.keidanren.or.jp/journal/times/2022/1201_09.html

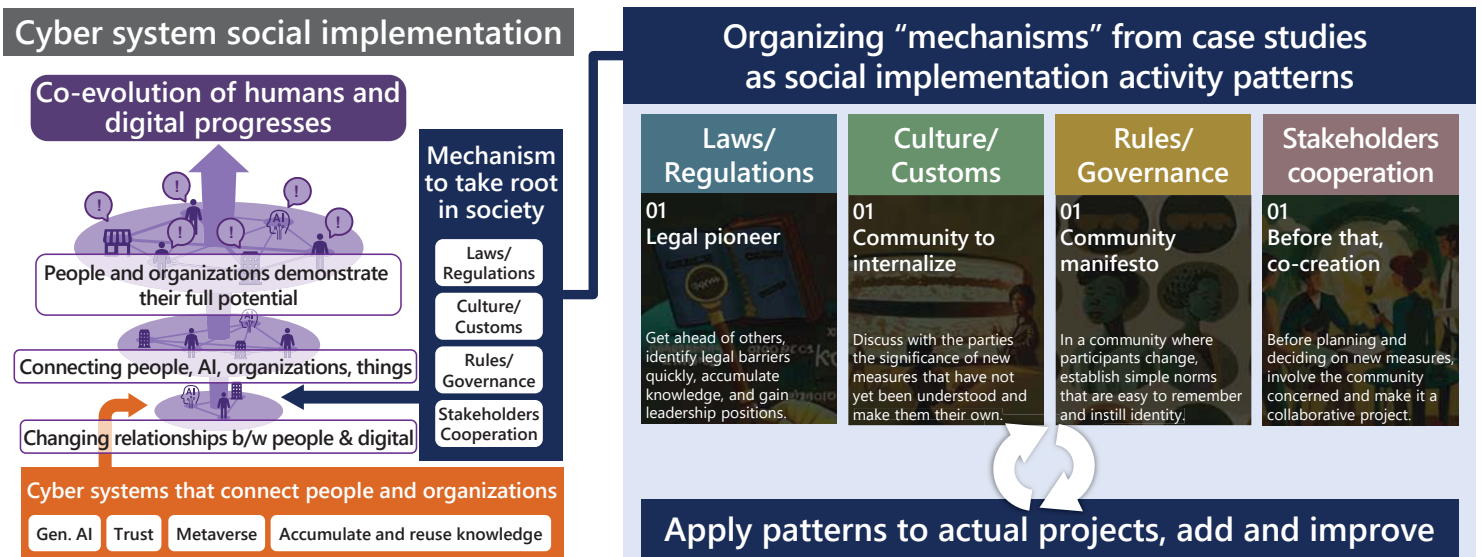


VINNOVA (Sweden, 2022年)
"Designing Missions" outlines a design process for realizing a sustainable society, rather than developing technology as a means to an end. It's a kind of playbook.
<https://www.vinnova.se/>

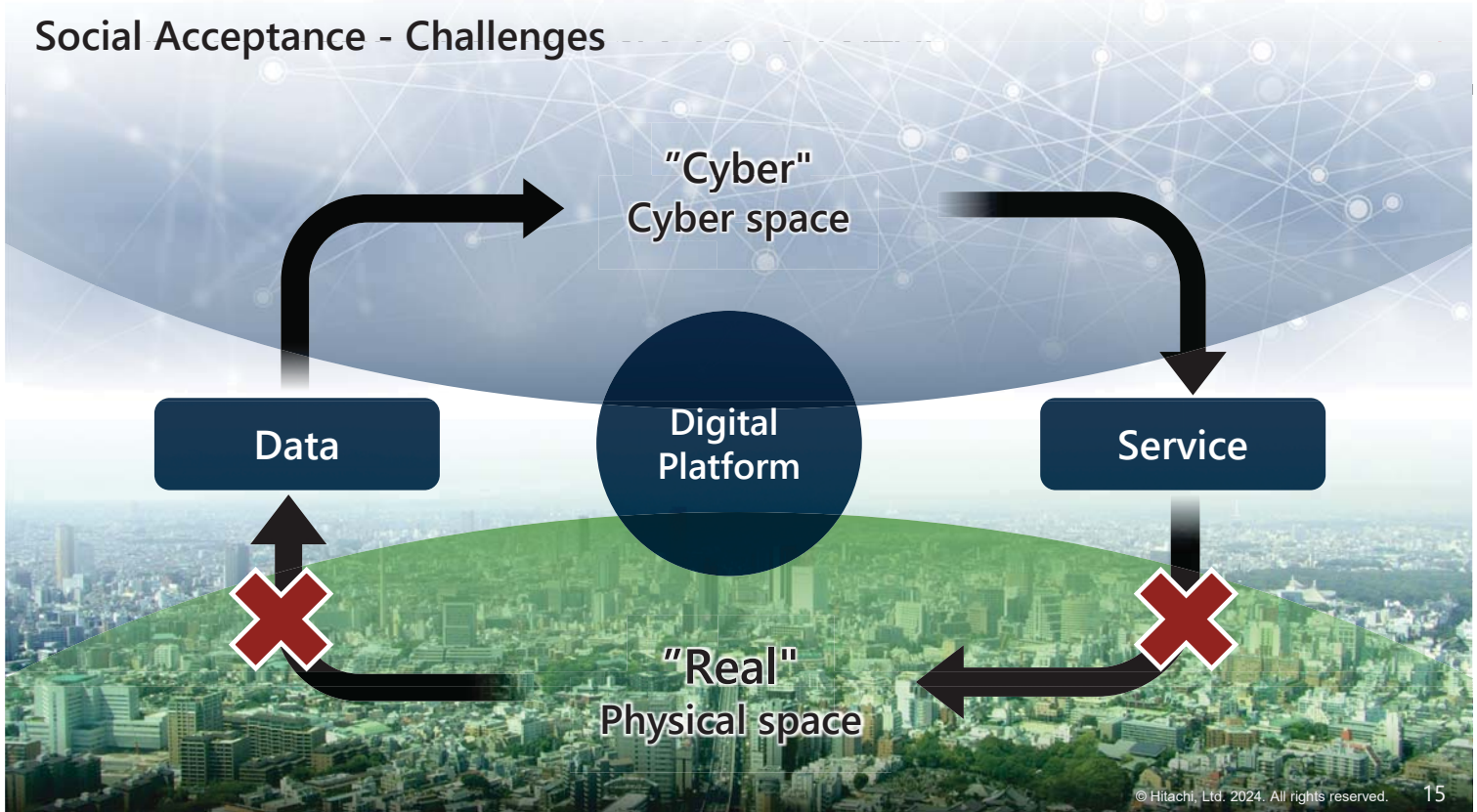
2-1-4. Mechanism: Social implementation activity pattern



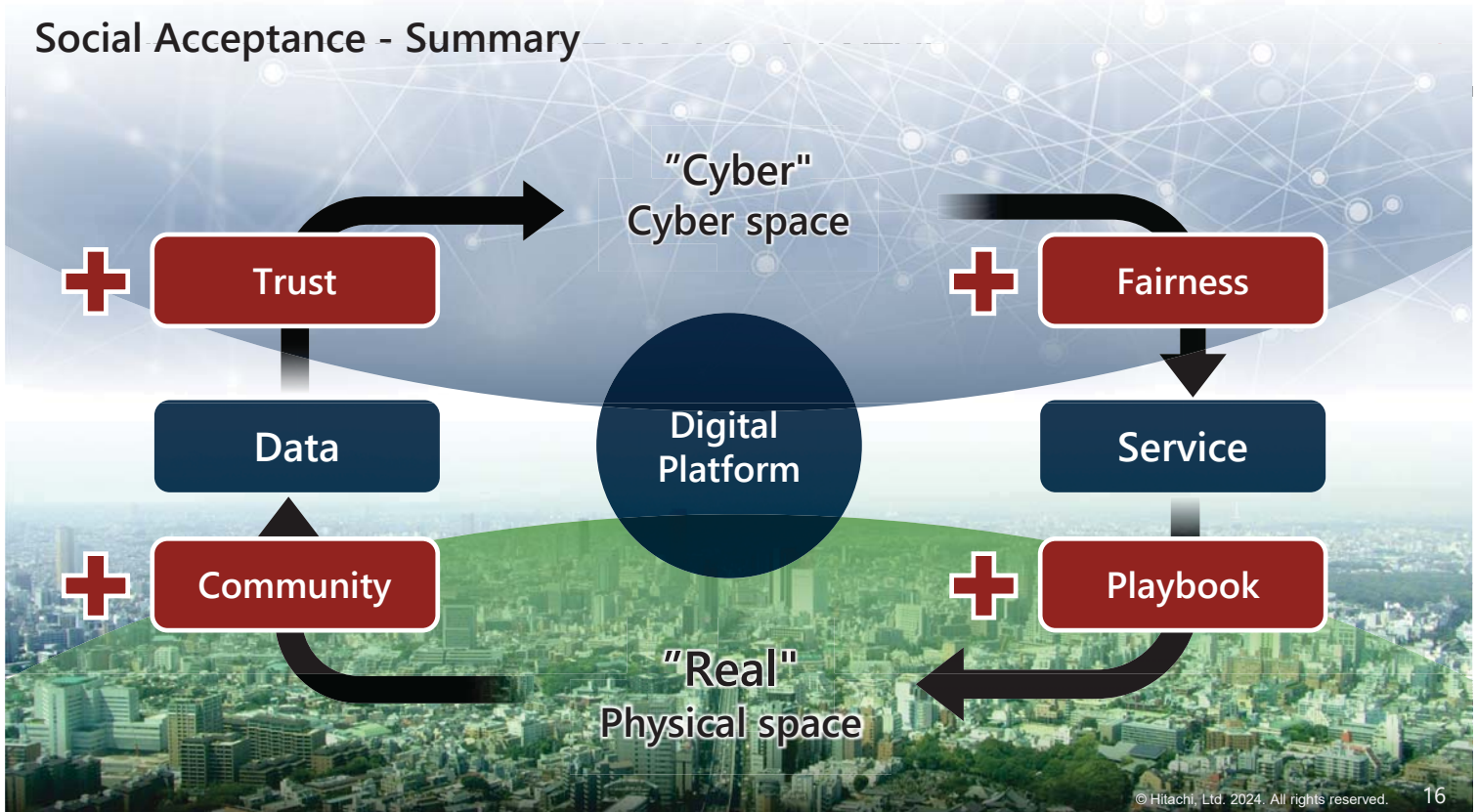
"Mechanisms" that take root in society are necessary to realize the co-evolution of people and the digital. Patterning knowledge for social implementation, such as cooperation among stakeholders



Social Acceptance - Challenges



Social Acceptance - Summary



6 Key factors for Smart City (Society 5.0)



OT: Operational Technology

© Hitachi, Ltd. 2024. All rights reserved.

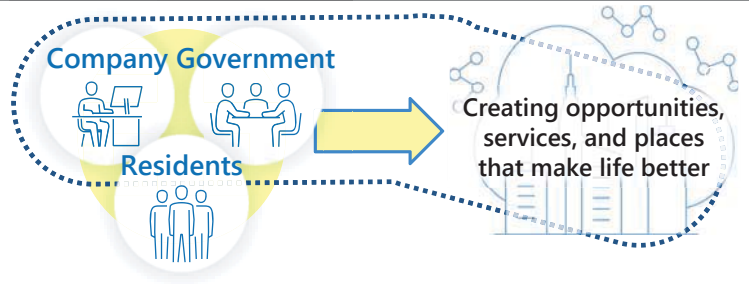
2-2-1. Key Factor 2 : Consumer Participation



HITACHI
Inspire the Next

What is Living Lab...

A system to promote participation of residents as part of smart city development and to conduct research, experiments, and demonstrations together in their daily lives.



◆Participate and apply to the case of Kashiwanoha Smart City



Town Planning for Everyone Studio Kashiwanoha
(Ph.1: Mechanism for gathering voices)

Collect, organize, and visualize residents' opinions on issues related to their daily lives. Consideration of project themes to be created from the Living Lab based on feedback from residents and city users



Under the theme idea generation

Managed the process of generating specific issues and ideas from residents and users of the city. Conduct residents' workshops, etc.

Frail Prevention AI living laboratory

Implement and introduce the developed AI technology as a service together with users

© Hitachi, Ltd. 2024. All rights reserved. 18

2-2-2. Future Living Lab. – Activity with Kokubunji in Tokyo



HITACHI
Inspire the Next

Through local produce delivery by residents, digital payments that create connections, and evacuation support under the corona that connects peacetime and contingency,

Demonstrating how to foster local communities for the future at local events



What we have confirmed:

- ✓ **Changes in local residents' awareness of local assets such as urban agriculture and local stores**
- ✓ **The connections between people within the community should lead to the community's ability to be active.**



Activity Results

1. Take Me, Eat Me, My Vegetables (2018)
2. Bunji Bar (2018, 2019 fiscal year)
3. Home evacuation support application (FY2020, 2021)



© Hitachi, Ltd. 2024. All rights reserved. 19

2-2-3. Future Living Lab. – Activity with Miura in Kanagawa



HITACHI
Inspire the Next

While implementing conceptual activities in the community in collaboration with farmers, local businesses, universities, and other organizations that share our vision for future value change,

Explore infrastructure mechanisms to bring "involvement" to the community



What we have confirmed:

- ✓ **Derive perspectives necessary for infrastructure to create "involvement," such as "Create mechanisms for citizens to feel each other and the people of the city."**
- ✓ **The value that exists in a community cannot be assumed in advance, but can only be spun out through practice.**



Activity: Hi Miura Project

1. Slow Payments (FY 2019)
2. Future Vegetable Seeds (FY2020)
3. Map of Miura Coast to be created by everyone (FY2021)

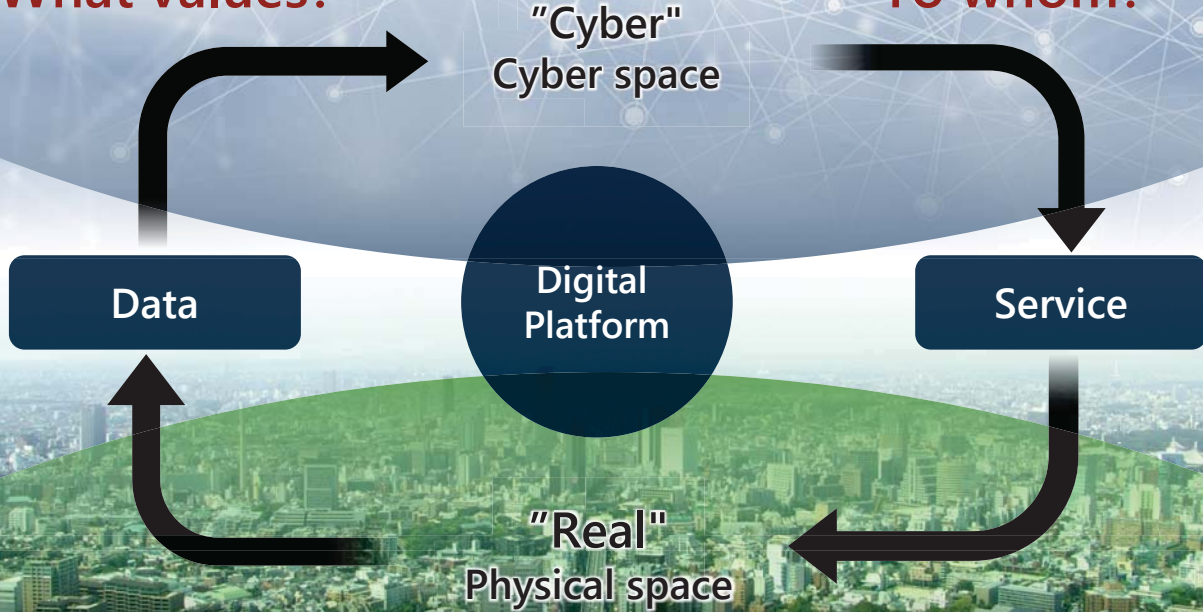


© Hitachi, Ltd. 2024. All rights reserved. 20

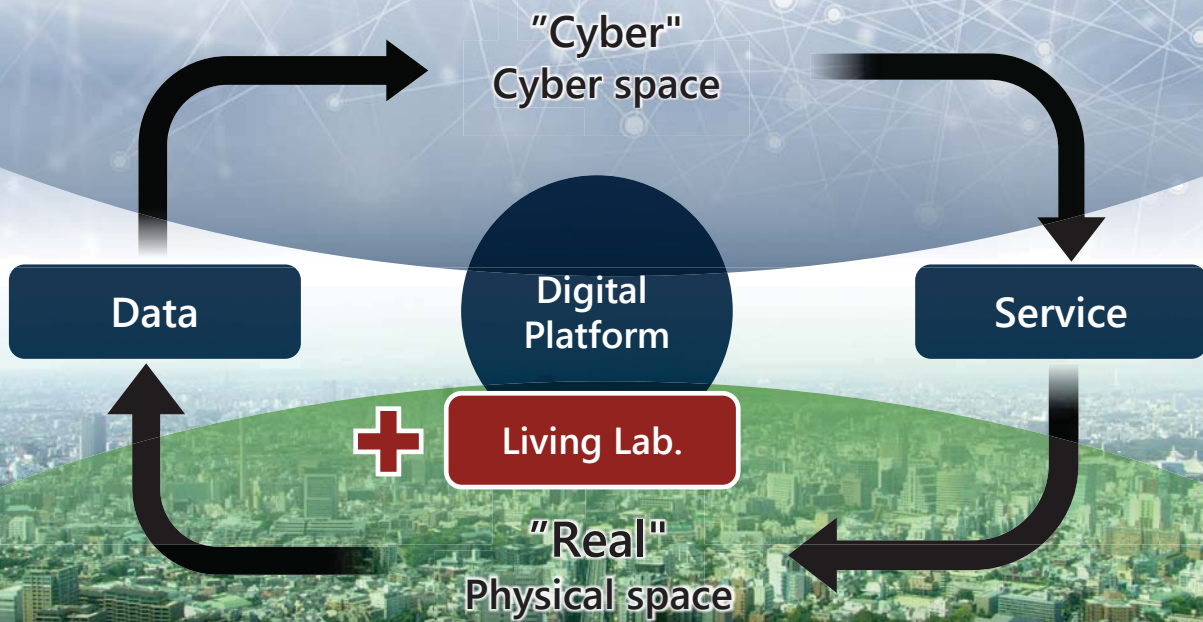
Customer Participation - Challenges

What values?

To whom?



Customer Participation - Summary



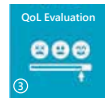
6 Key factors for Smart City (Society 5.0)



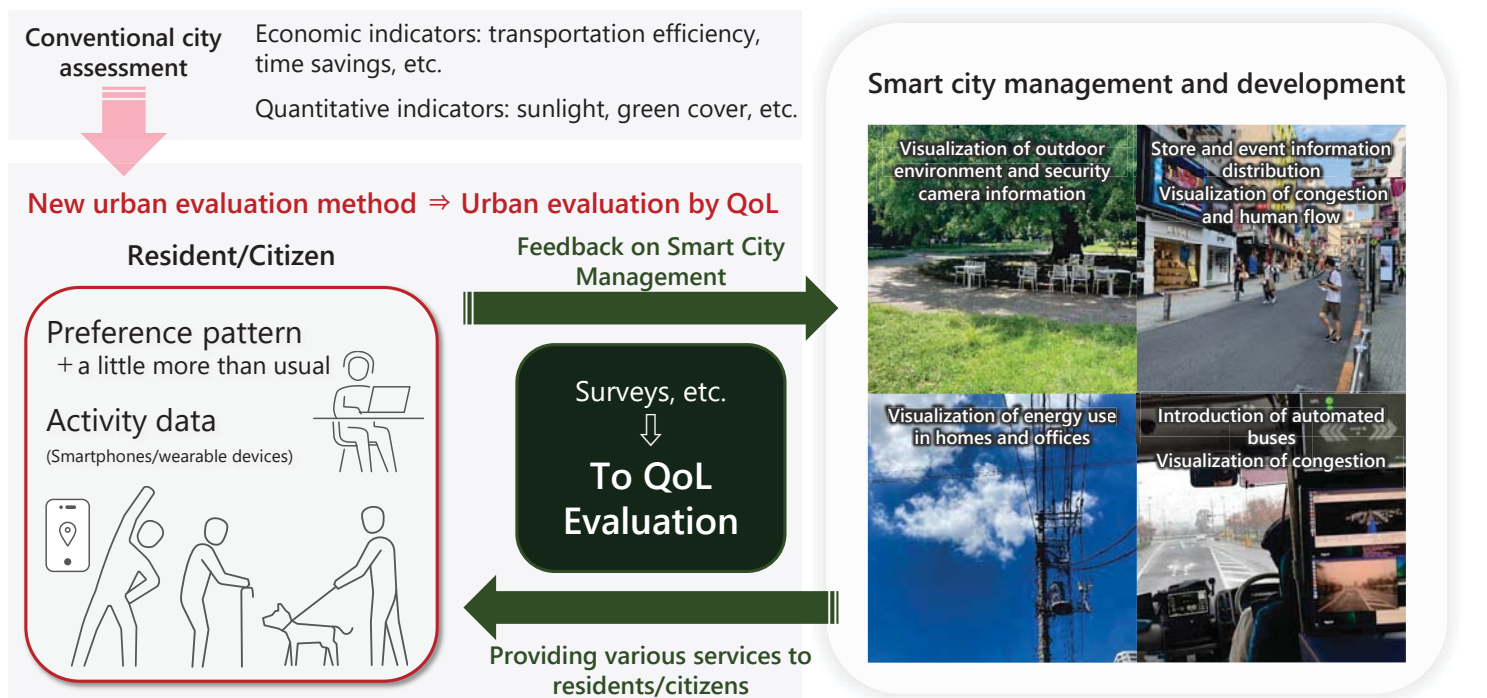
OT: Operational Technology

© Hitachi, Ltd. 2024. All rights reserved.

2-3-1. Key Factor 3 : QoL Evaluation

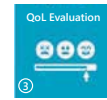


HITACHI
Inspire the Next

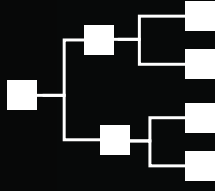


© Hitachi, Ltd. 2024. All rights reserved.

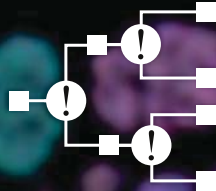
2-3-2. Future Scenario Simulation



Utilized to solve complex issues where multiple trade-offs exist.
Patterns of possible future scenarios and decision critical future junctures.



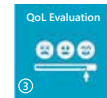
1. Generation of a large number of future scenarios



2. Identification of future junctures

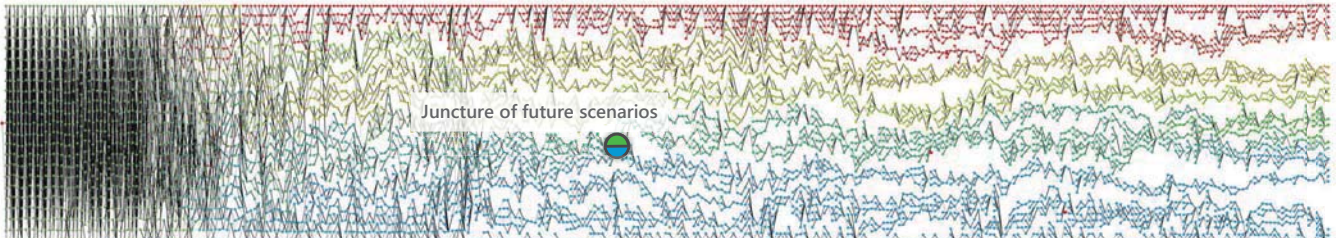
It is impossible to predict exactly one future.
However, it is possible to estimate multiple possible scenarios.

2-3-3. Case1 - Local government decarbonization roadmap



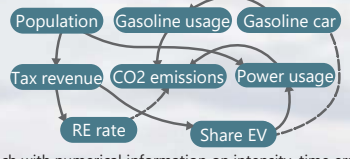
How can local governments facing diverse issues achieve decarbonization in 2050?
Identify future scenario patterns and important future turning points, and formulate a long-term roadmap.

2022
2050



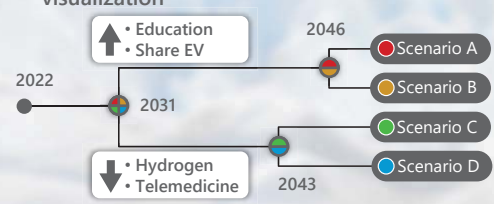
1. Information gathering and modeling Human Operations

Solid line: Positive Dashed line: Negative




Each with numerical information on intensity, time error, and their variation

2. Scenario and junction point visualization AI Work

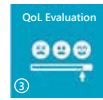


3. Scenario Evaluation and Policy Consideration Human Operations



Let's educate your family about the environment in the shared EV parking lot!

2-3-4. Case1 - Decarbonization roadmap in Hokkaido



A roadmap proposal for achieving decarbonization in Ishikari City and Obihiro City, Hokkaido will be materialized through a collaborative creation workshop using this technology

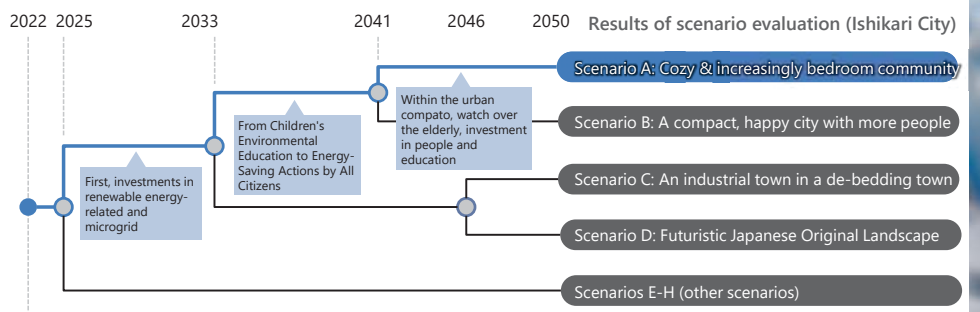
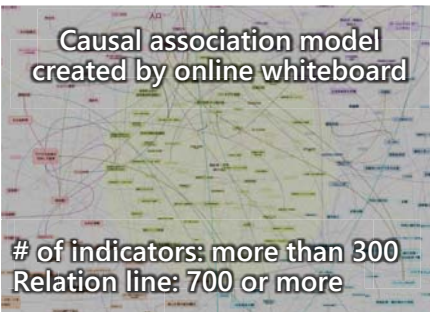
1. Information gathering and modeling



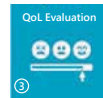
2. Scenario and junction point visualization



3. Scenario Evaluation and Policy Creation



2-3-5. Case 2 - Sustainability of Japanese society



The NHK Special "2024 Our Choices: Six Futures by AI and Experts" broadcast on January 4, 2024, introduced some of the results of the latest version of the simulation.

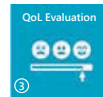


Number of indicators	Relationship line	Input indicator data
378	> 1000	20 years

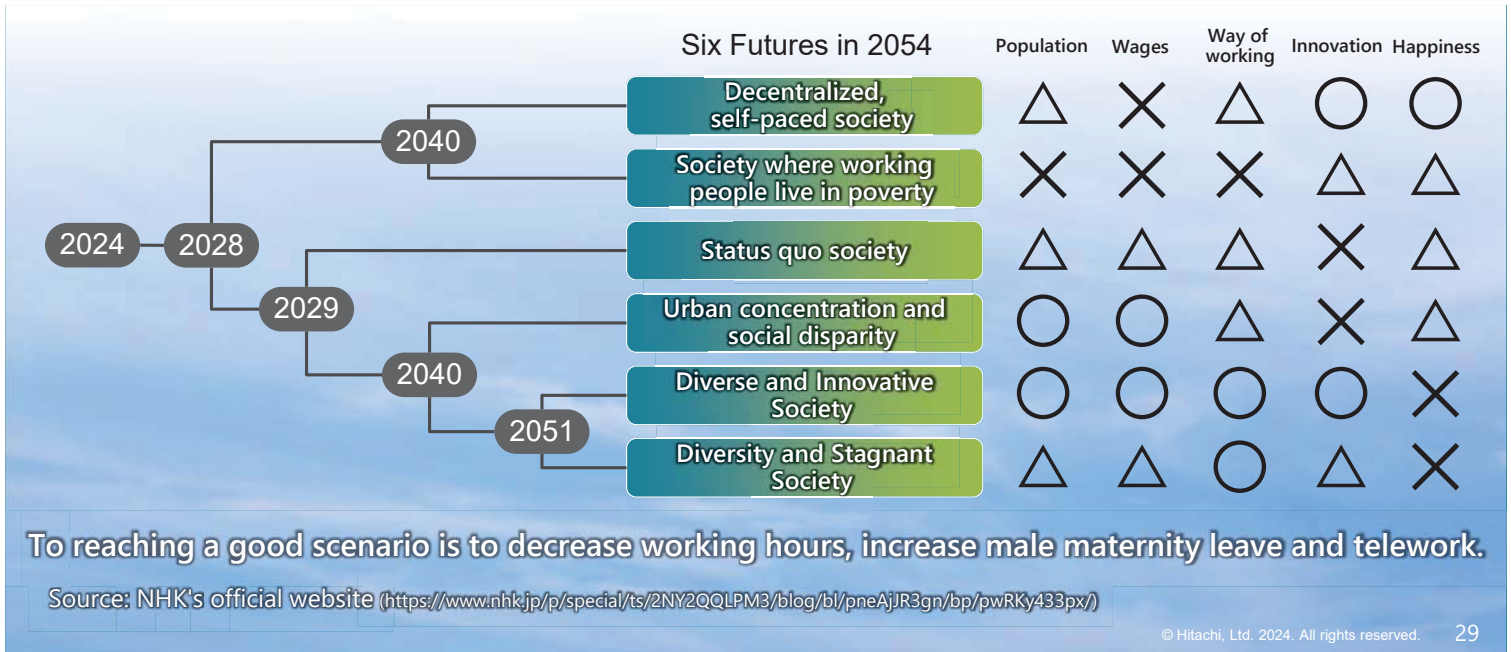
With the advice of professors from Kyoto University and Tokyo Metropolitan University, the simulation model was supplemented with simulation models related to the declining birthrate problem, low wages, and innovation.

Source: NHK's official website (<https://www.nhk.jp/p/special/ts/2NY2QQLPM3/blog/bl/pneAjR3gn/bp/pwRky433px/>)

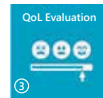
2-3-6. Case 2 - Sustainability of Japanese society



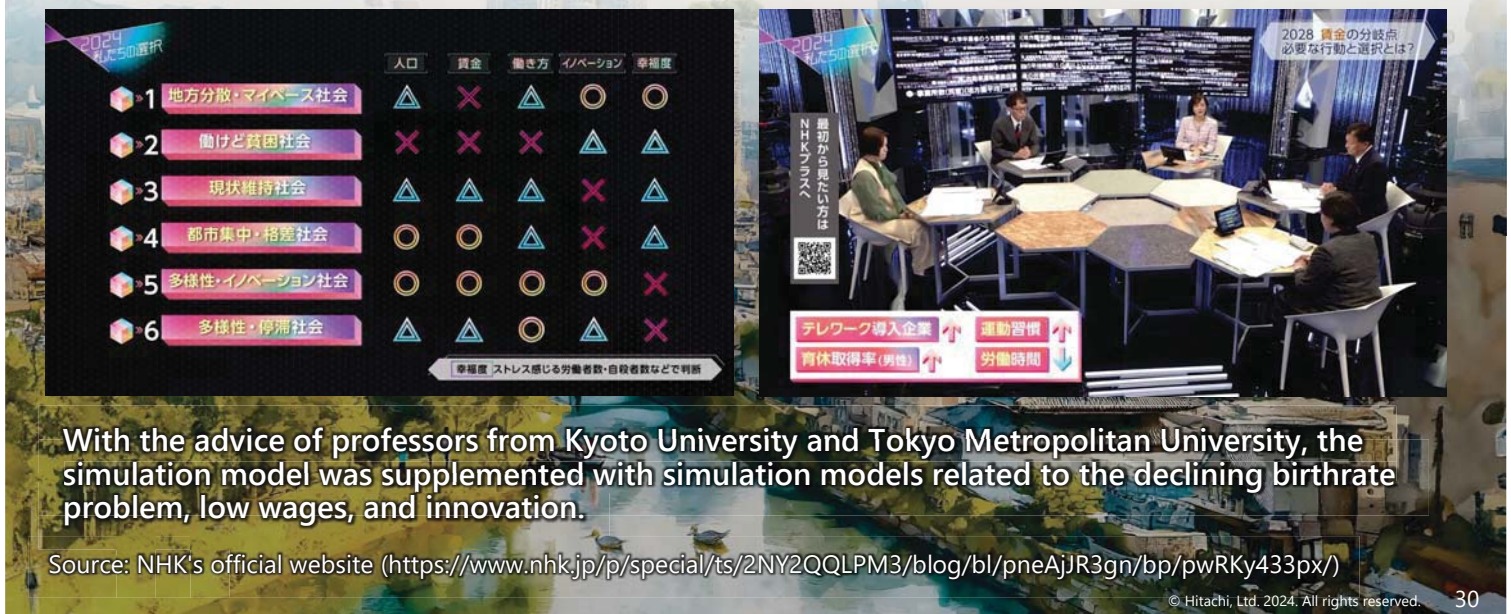
A "diverse and innovative society" appears to be the best scenario, but a "decentralized and fast-paced society" does not seem to be a bad idea.



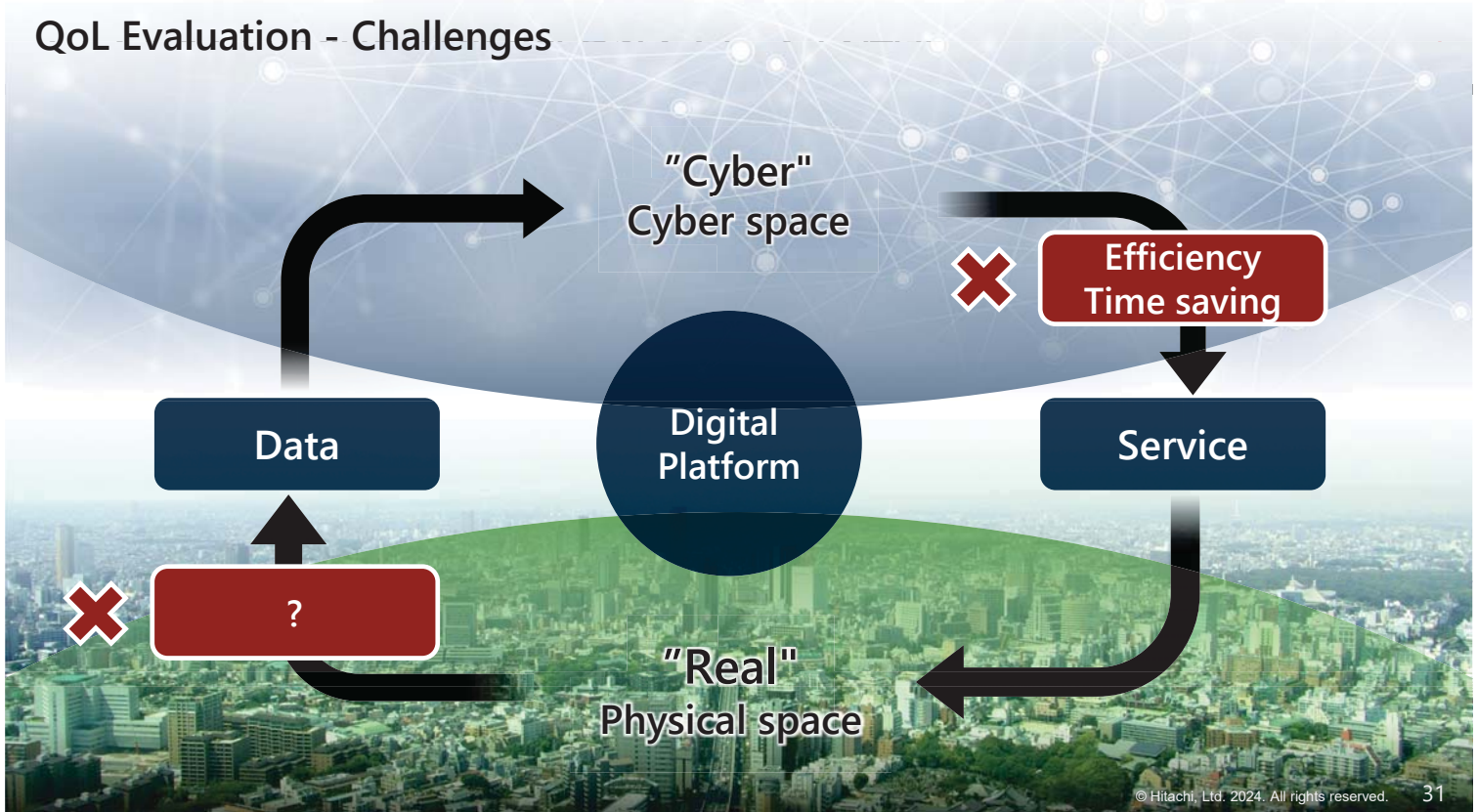
2-3-7. Case 2 - Sustainability of Japanese society



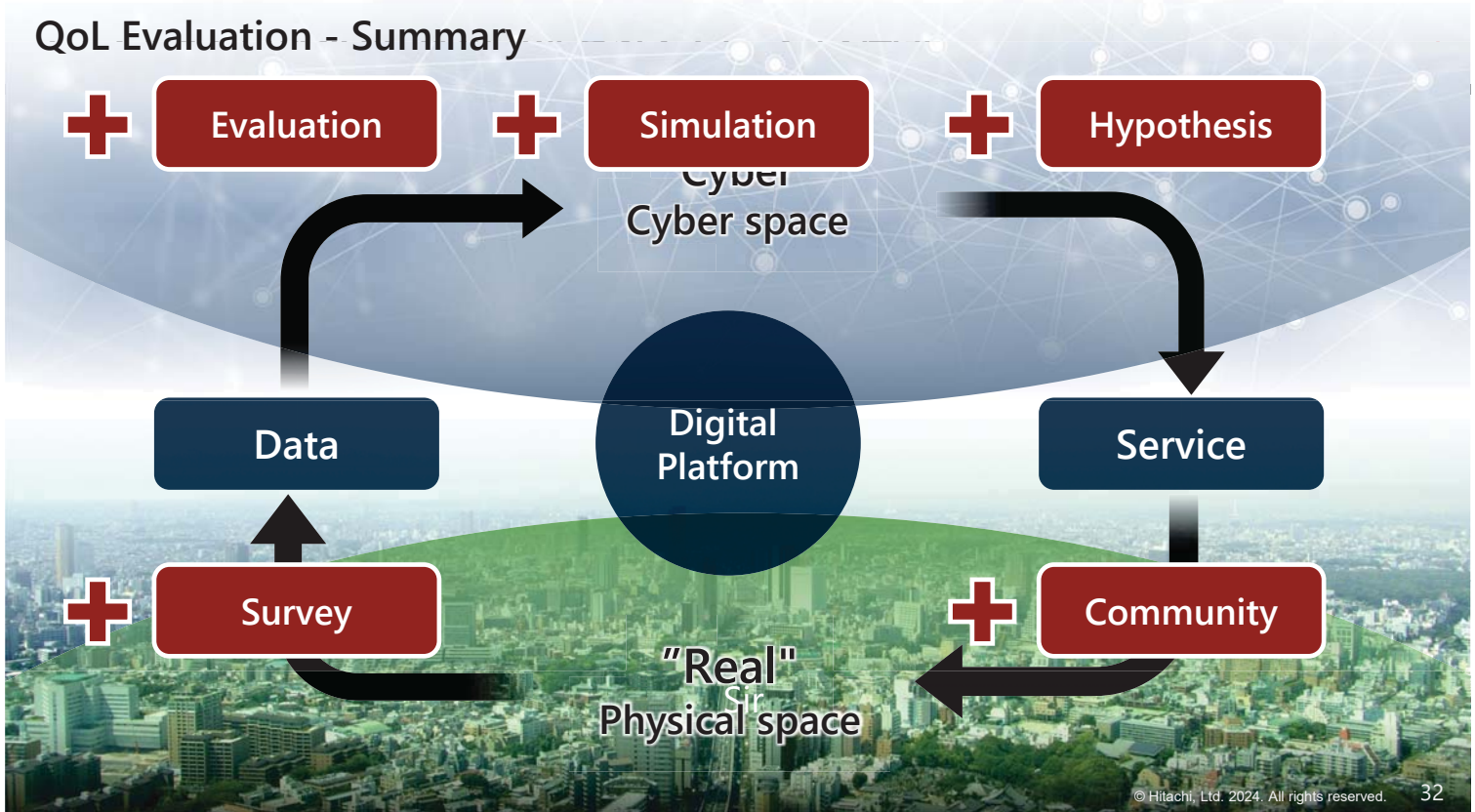
The NHK Special "2024 Our Choices: Six Futures by AI and Experts" broadcast on January 4, 2024, introduced some of the results of the latest version of the simulation.



QoL Evaluation - Challenges



QoL Evaluation - Summary



6 Key factors for Smart City (Society 5.0)



OT: Operational Technology

© Hitachi, Ltd. 2024. All rights reserved.

2-4-1. Key Factor 4 : Data Governance



HITACHI
Inspire the Next

What is "Data Governance"? – Risk of Data Utilization

Viewpoints	Overview
Compliance with related laws and regulations	Are relevant laws and regulations, soft laws, etc. appropriately understood and complied with?
Privacy protection	When handling personal information, etc., is it handled in a way that does not cause anxiety or discomfort?
Respect for intellectual property and confidential information	Are intellectual property and confidential information held by organizations participating in smart cities properly protected?
Ensuring security	Is security management in place to prevent problems such as information leaks and service tampering?
Proper operation	Is there a risk that smart city services will impair fairness or cause damage to citizens?
Maintaining overall governance	Is there a system in place to reduce the risks mentioned above among all the stakeholders that make up smart city services?



2-4-2. Data Free Flow with Trust (DFFT)

Concept

Proposed by Japanese government in 2019

International free flow of data to generate higher productivity, greater innovation, and improved sustainable development, while addressing concerns over privacy, data protection, intellectual property rights, and security

Challenges

Cross Border Data Flows

Open and Free Data Flows

Data Governance

2-4-3. Trends and challenges in DFFT



Trends

- In Europe, GAIA-X was launched to interconnect data spaces ('20)
- Automotive industry supply chain field data space Catena-X also begins full operation ('23)
- IAP was established under the OECD as an international framework to materialize DFFT ('24)
- In Japan, an interdisciplinary data space (DATA-EX) based on CADDE that transcends industry, academia, and government is started by the Data Society Promotion Council (DSA) ('24)

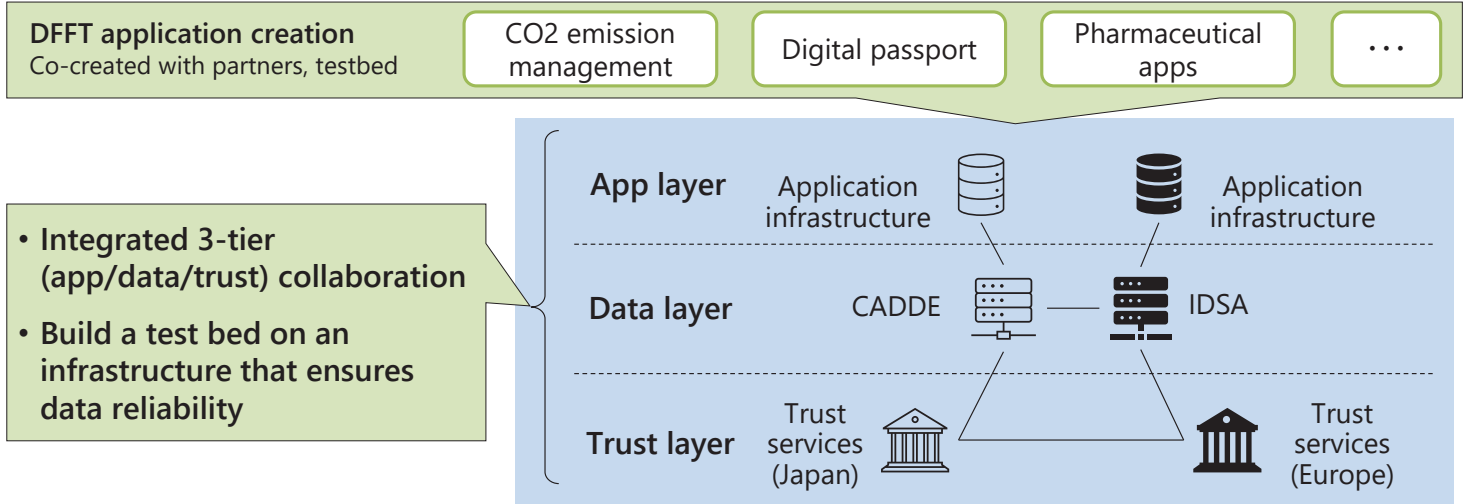
Challenge

- Ecosystem that distributes highly reliable data on a large scale has not yet been realized.
- It is necessary to identify and launch a group of applications that will support the creation/development of the ecosystem.

2-4-4. DFFT research overview



DFFT ecosystem through industry-government-academia collaboration Provide reliable data and analyze global data to solve social issues



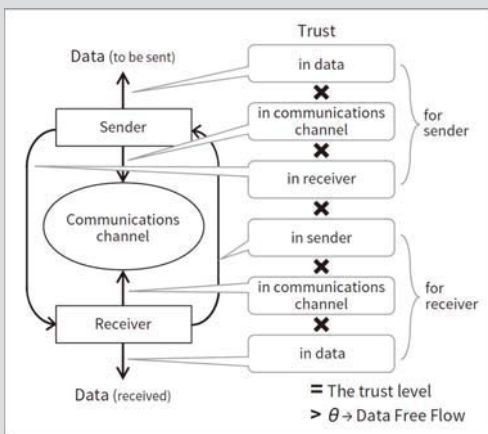
DFFT infrastructure testbed

2-4-5. DFFT research theme



DFFT Model

A free flow of data occurs if trust level of sender, receiver and communication channel exceeds a certain threshold

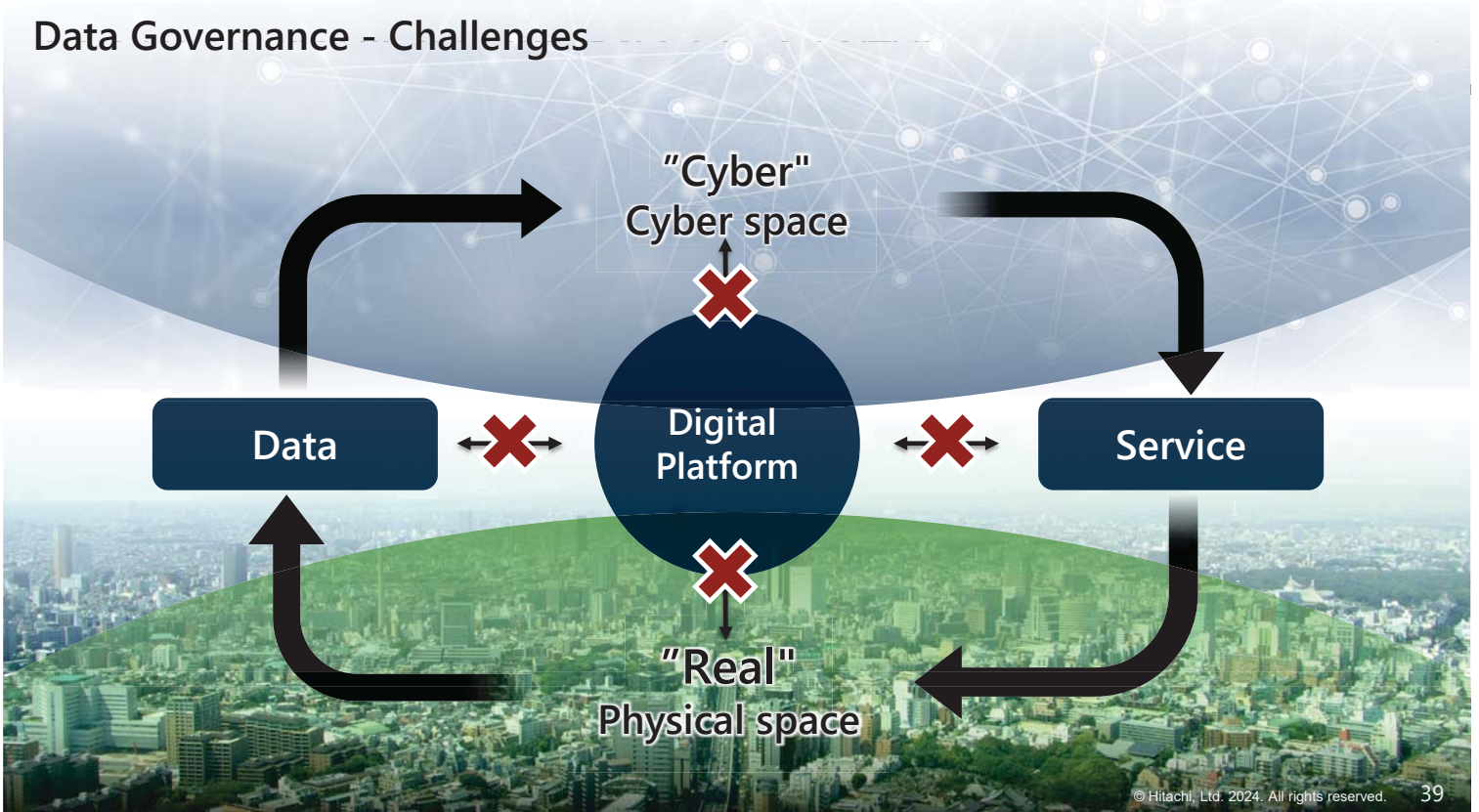


Self-Sovereign Identity (SSI) by PBI

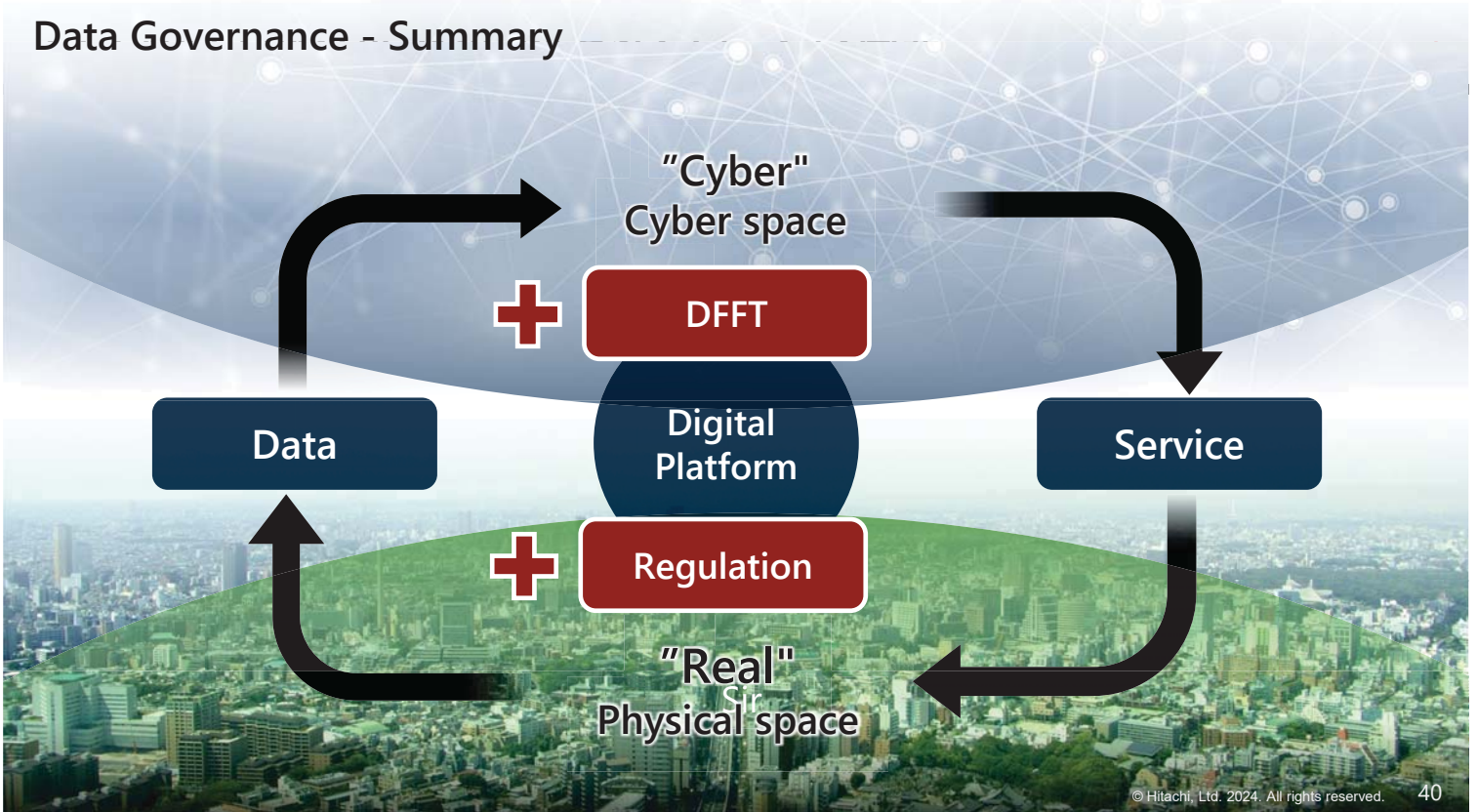
SSI by PBI allows individuals own and control their identity under strong protection of security and privacy



Data Governance - Challenges



Data Governance - Summary



6 Key factors for Smart City (Society 5.0)



OT: Operational Technology

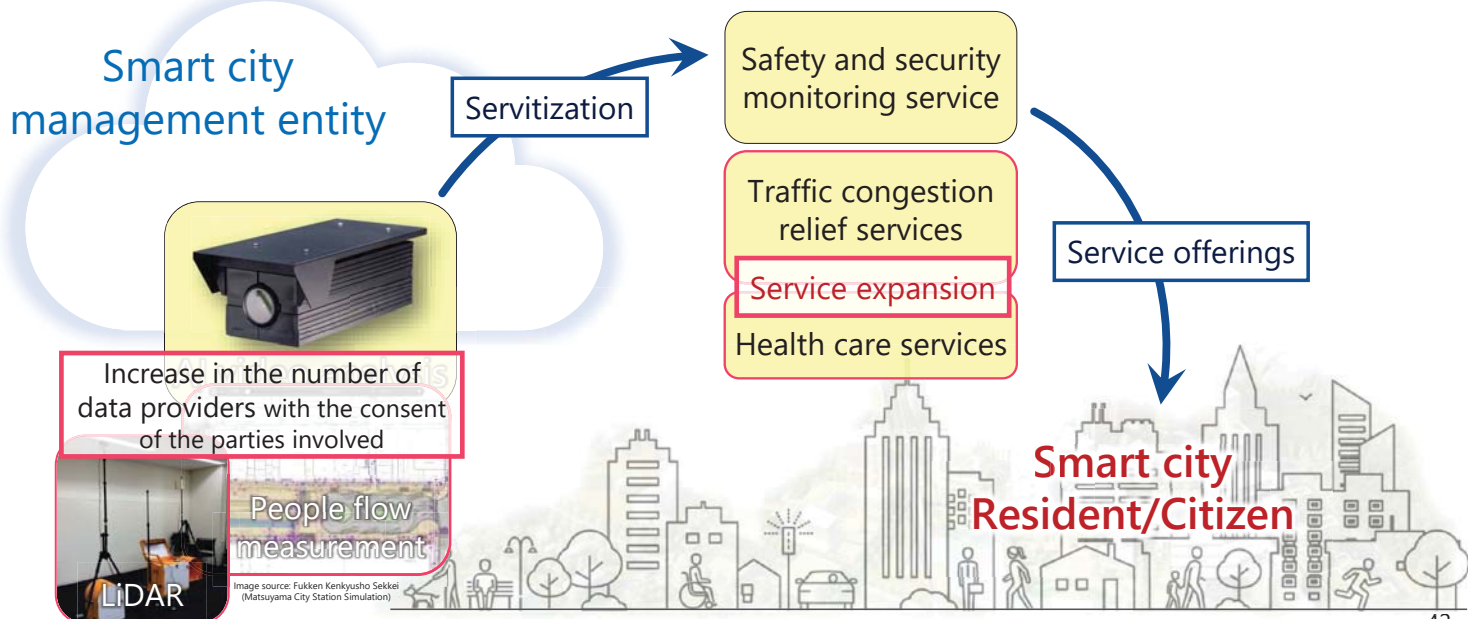
© Hitachi, Ltd. 2024. All rights reserved.

2-5-1. Key Factor 5 : Data Infrastructure Ecosystem



HITACHI
Inspire the Next

Ecosystem that utilizes increasing data to expand the services provided



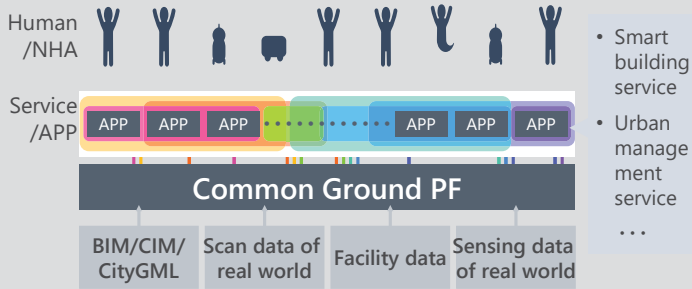
© Hitachi, Ltd. 2024. All rights reserved.

2-5-2. Common Ground



Common Ground Platform (CGPF)

Data is stored centrally in CGPF and processed appropriately for each user, facilitating real-time data linkage between companies in different industries and between different types of devices.



NHA : Non-Human Agent. Robots, drones, or AR and VR avatars that have no physical reality

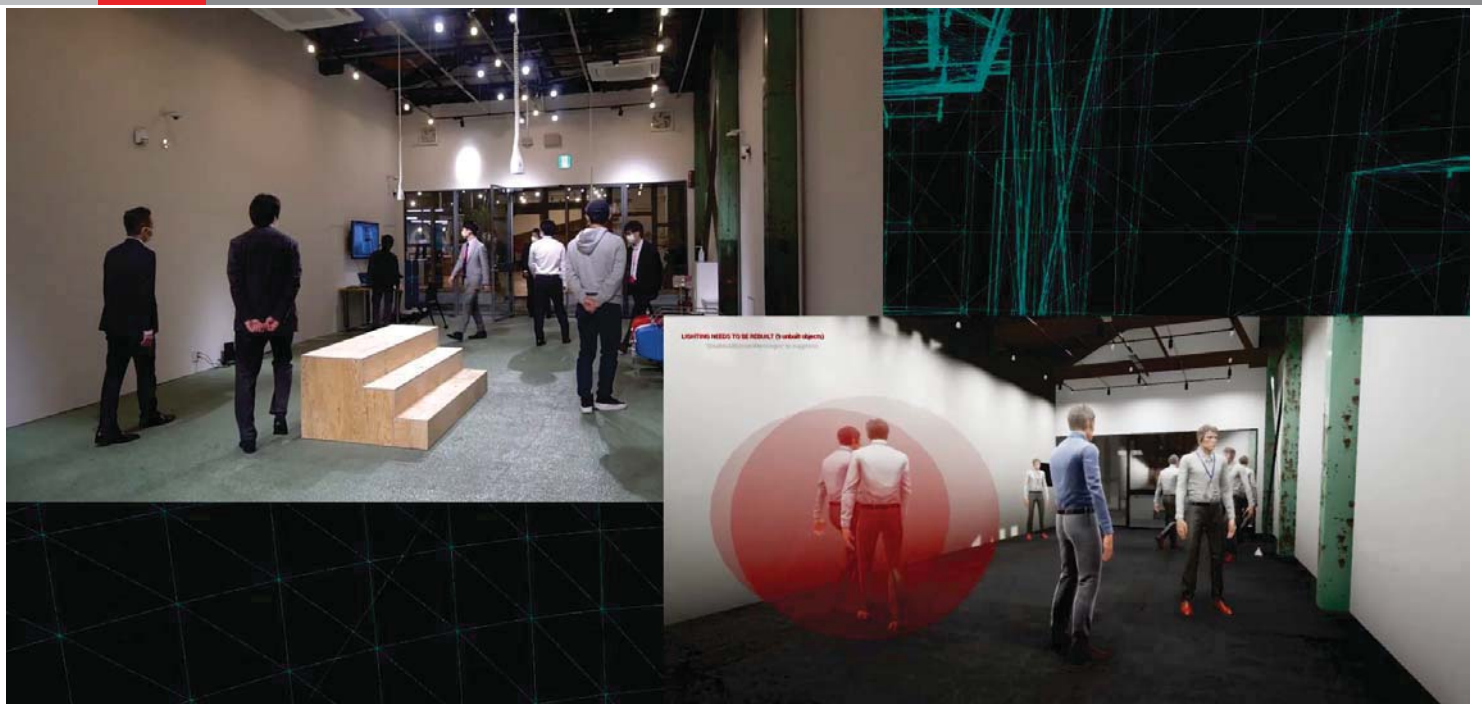
Common Ground Living Lab (CGLL)

Opened Living Lab @Osaka, with the Osaka Chamber of Commerce and Industry (July 1st, 2021)



© Hitachi, Ltd. 2024. All rights reserved. 43

2-5-3. CGPF example: Real-time people position tracking



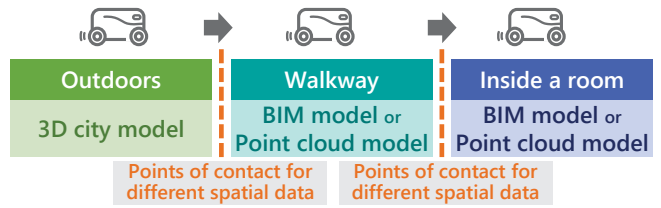
© Hitachi, Ltd. 2024. All rights reserved. 44



2-5-4. Experiment example #1 at CGLL

Participated in the 3D city model development and utilization project "PLATEAU" led by the Ministry of Land, Infrastructure, Transport and Tourism

Smooth indoor and outdoor mobility movement even with 3D models of different spatial data formats and accuracies

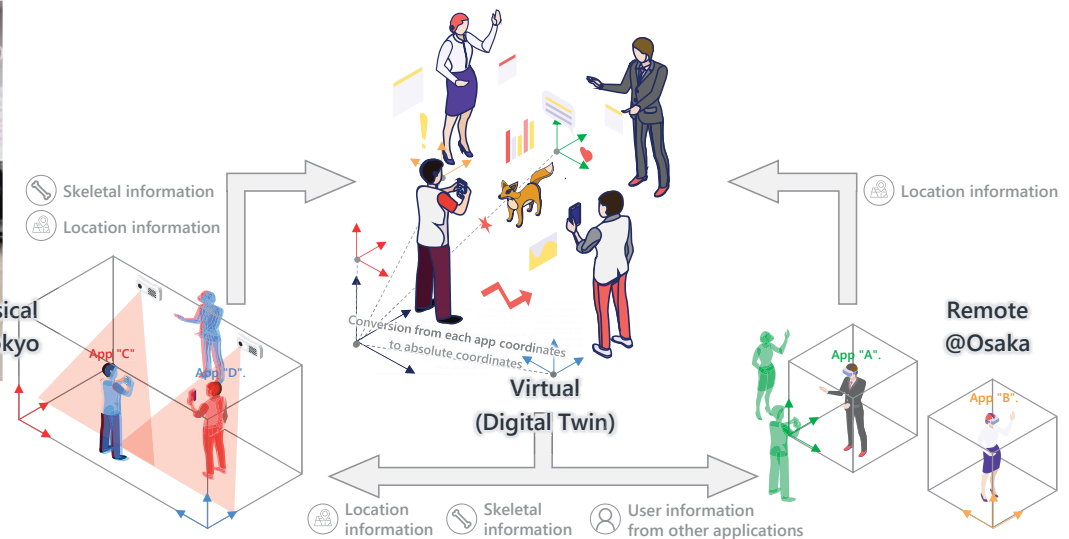
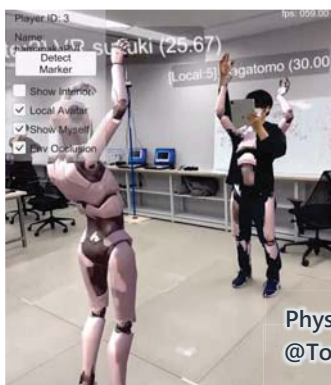


Reference: Hitachi Press Release <https://www.hitachi.co.jp/New/cnews/month/2022/08/0805.html>

2-5-5. Experiment example #2 at CGLL

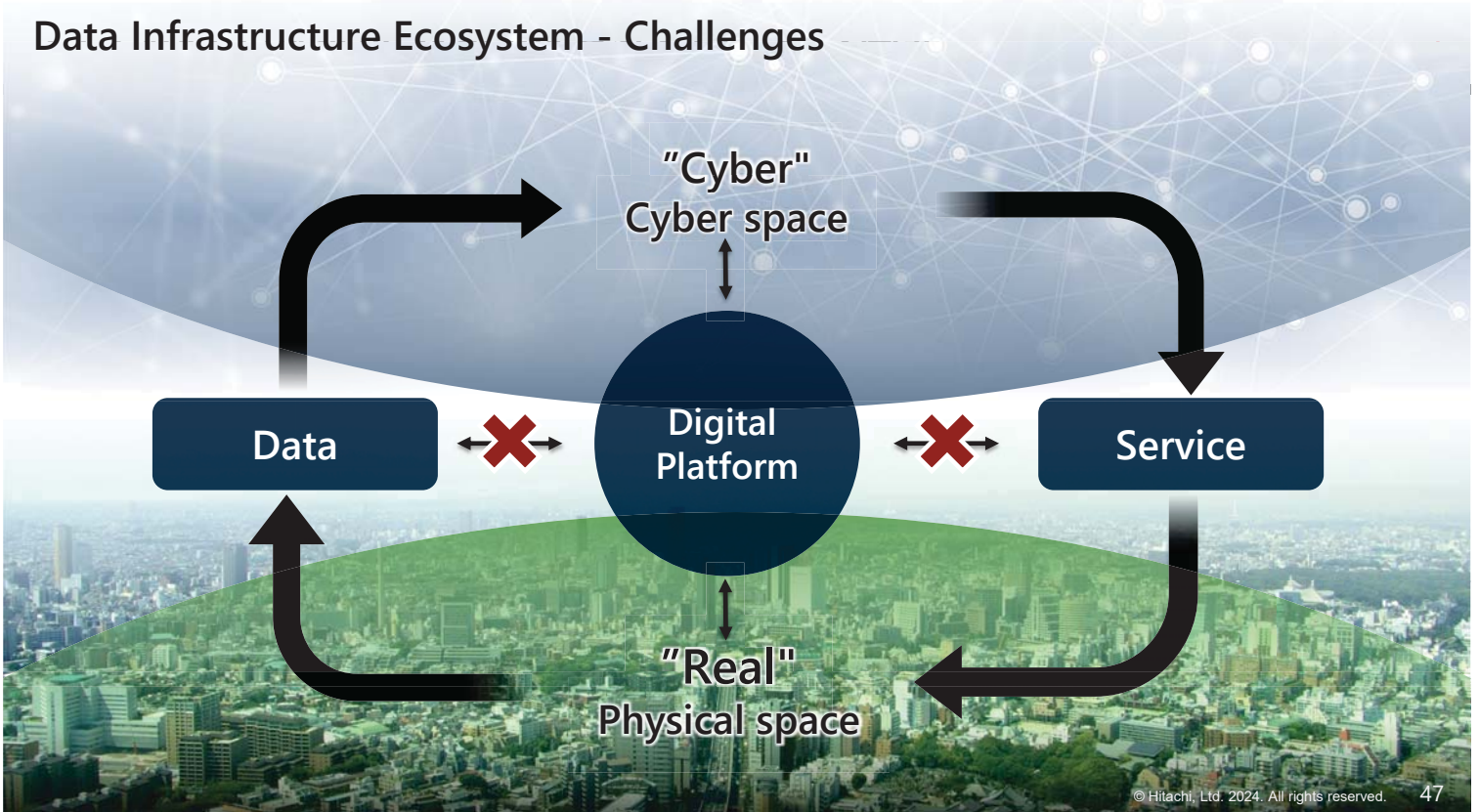


Access CGPF from different applications and share gestures, postures, etc. between all users in real time. Contracted Cabinet Office research project with CGLL participating companies.

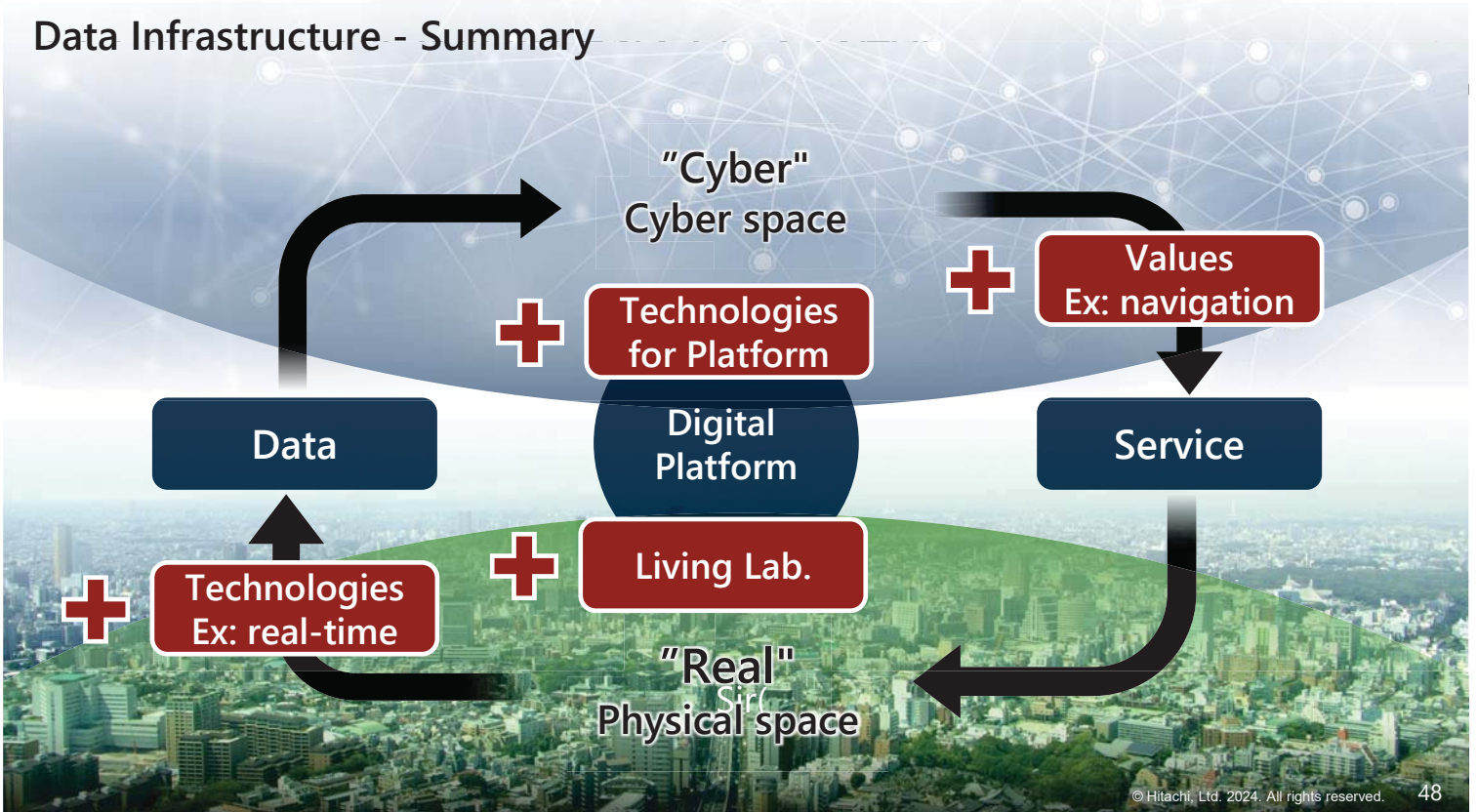


Reference: Cabinet Office, Government of Japan, Research Project on the Development and Establishment of Advanced Services in Super Cities and Digital Rural Health Special Zones

Data Infrastructure Ecosystem - Challenges



Data Infrastructure - Summary



6 Key factors for Smart City (Society 5.0)



OT: Operational Technology

© Hitachi, Ltd. 2024. All rights reserved.

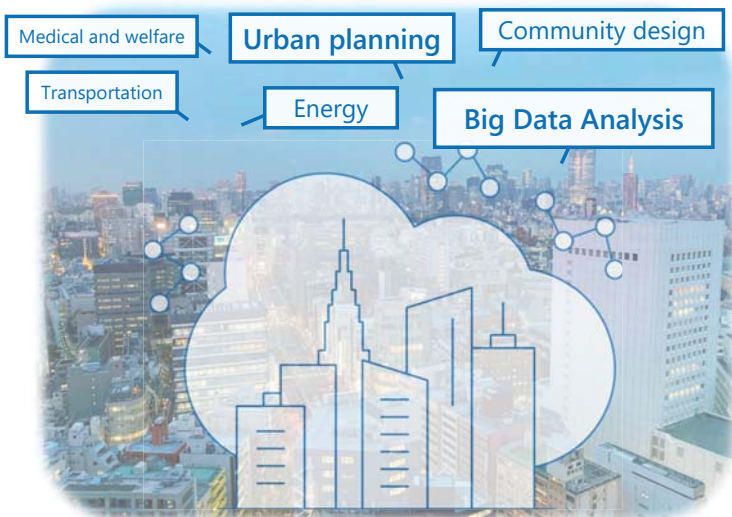
2-6. Key Factor 6 : Human Resource Development



HITACHI
Inspire the Next

Urgent need to develop human resources for smart city development

Diverse areas of expertise in smart city implementation and development

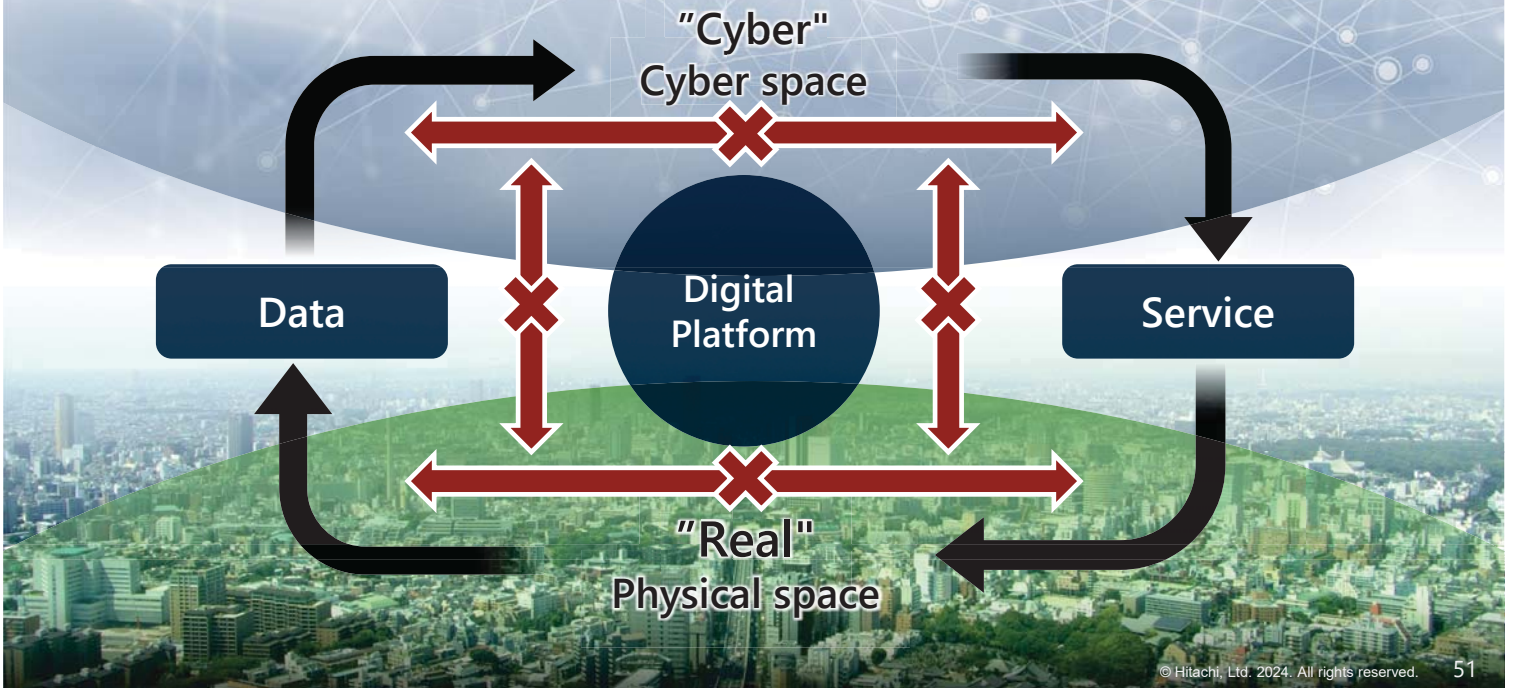


Developing human resources with cross-disciplinary and cross-sector **organizational coordination skills** and **knowledge of elemental technology**

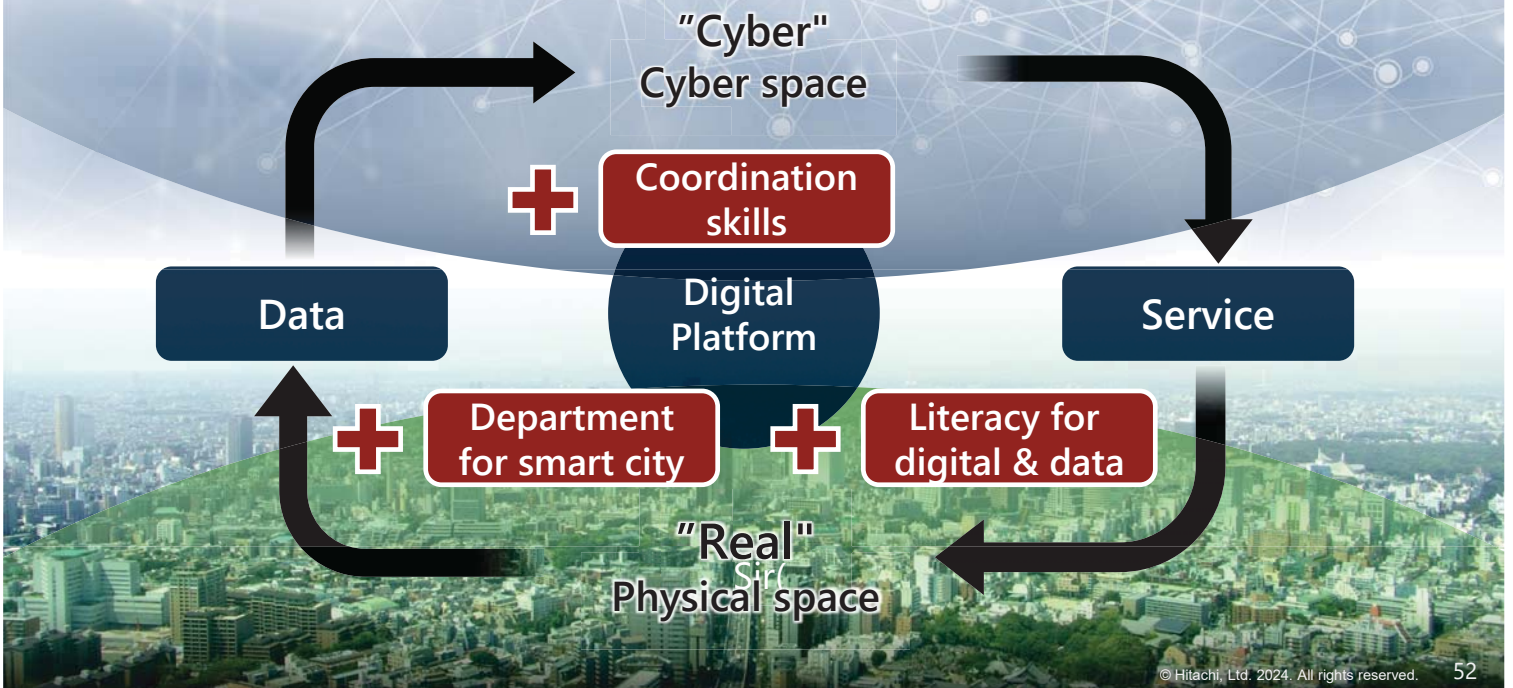
Specifically ...

Establishment of an internal administrative structure	<ul style="list-style-type: none"> • Creation of positions that can be directed across disciplines • Establishment of a department in charge of smart city promotion with combined fields
Improve literacy of residents and citizens	<ul style="list-style-type: none"> • Promote correct understanding of information technology, data, etc. used in smart cities

Human Resource Development - Challenges



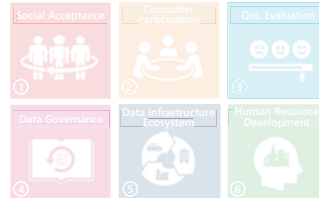
Human Resource Development - Summary



Contents

1 Background and introduction of Society 5.0 and CPS

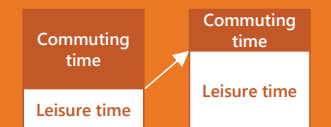
2 6 Key factors for Smart City



3 Next step and for future – new world

3-1. Next step: Toward Well-being City

Well-being of Individuals

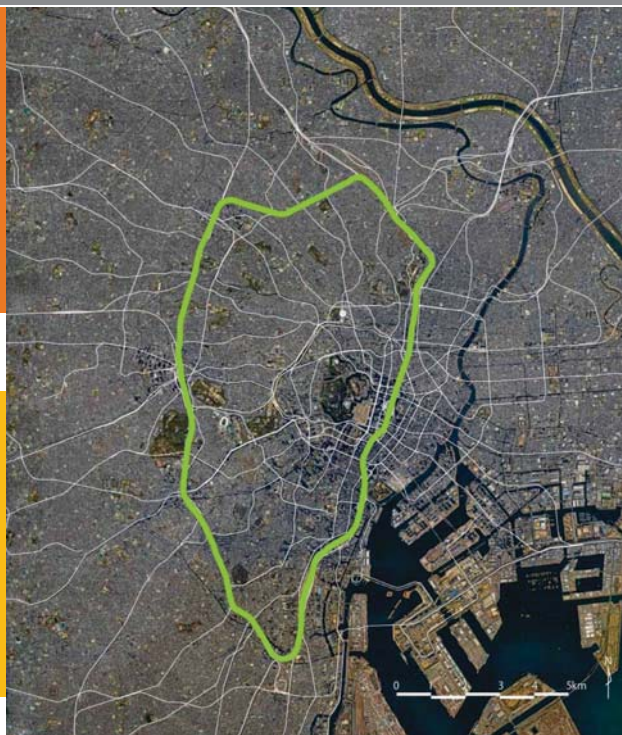


Reducing commuting time by compacting living areas

Well-being of Communities



Revitalizing urban neighborhood communities

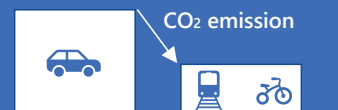


Well-being of Networks



Easing commuting train congestion by promoting proximity of work and residence

Well-being of Environments



Reducing environmental impact of travel through public transportation and bicycles

3-2. EXPO2025 – Future Life Expo/Future City –



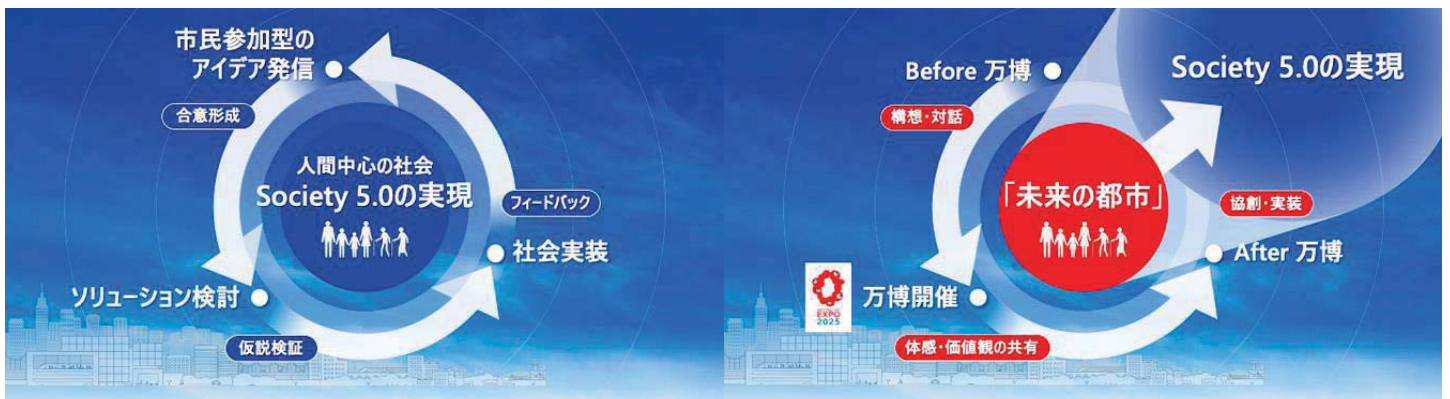
By involving citizens in the creation of new social systems, we aim to realize a **participatory society** in which each individual plays a central role by promoting social implementation and obtaining feedback on the values that new ideas bring to society.

Orientation toward a **new participatory society**



3-3. Hitachi's thoughts on "Future Life Expo/Cities of the Future"

Through the fusion of cyber and physical, we aim to realize a society in which citizens' ideas are actually implemented, a society in which each and every person plays a central role.



3-4. H-UTokyo Lab : Hitachi and U-Tokyo Joint Research



Creation of a vision for the realization of Society 5.0 (super-smart society)
 Dissemination of models for solving social issues toward post COVID-19

Habitat Innovation Project

Theme: Community Development

Realization of a Data-driven People-centric Smart City

Energy Projects

Topic: Energy Systems

Establishment of a total energy system to realize a net-zero society

© Hitachi, Ltd. 2024. All rights reserved. 57

Summary



- 1

Society 5.0 and CPS

A human-centered society in which economic development and the resolution of social issues are compatible with each other through a highly integrated system of cyberspace and physical space.
- 2

6 Key factors for Smart City

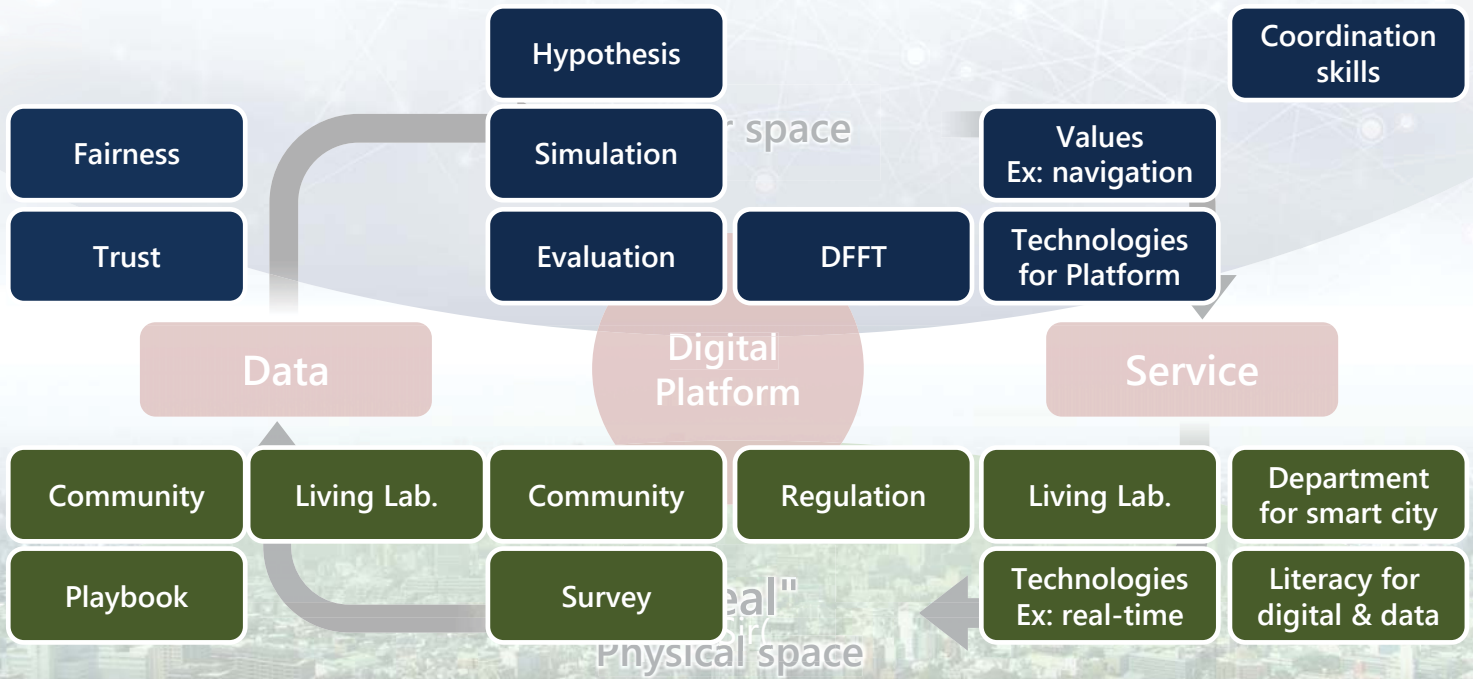
Key factors: Social Acceptance, Customer Participation, QoL Evaluation, Data Governance, Data Infrastructure Ecosystem and Human Resource development. Continuous trials with multi stakeholders are important to validate and update.
- 3

Next step and for future – new world

By involving citizens in the creation of new social systems, we aim to realize a participatory society in which each individual plays a central role by promoting social implementation and obtaining feedback on the values that new ideas bring to society.

© Hitachi, Ltd. 2024. All rights reserved. 58

Summary – Digital connects and opens...but activities in real space are the key.



Session 2:
Sensing and Analysis
(Chair: Tomoya Kitani)

Design and Implementation of Smartphone Beacons in Stay Estimation System using BLE

Kota Togawa[†], and Katsuhiko Kaji[†]

[†]Graduate School of Business Administration and Computer Science, Aichi Institute of Technology
{b24716bb, kaji}@aitech.ac.jp

Abstract - People spend 88.9% of their day indoors and are mainly active indoors in physically separated spaces such as their own rooms, laboratories, and conference rooms. Therefore, room-level location information, rather than highly accurate location information, is also valuable. We have proposed a stay estimation system that receives signals from BLE beacons carried by each user and estimates the room location using receivers installed in the environment. However, conventional methods using only physical beacons have problems such as battery replacement, time-consuming initial setup, and users moving from room to room without a physical beacon. In this study, we implement smartphone beacons that have high tracking performance, do not require initial setup or battery replacement, and consumes little battery power. Smartphones are often carried around at all times, so they are considered to be highly trackable.

Keywords: Room-level indoor localization, BLE beacon, Smartphone application

1 INTRODUCTION

Since many indoor activities take place within the confines of a room, room-level indoor location information is valuable. People spend the majority of their time indoors, as reported in the study that people spend 88.9% of their day indoors. Indoors, they are mainly active in physically separated spaces such as their own rooms, laboratories, and conference rooms. Therefore, not only highly accurate location information but also room-level location information is valuable. For example, there are commercial packages available that use room-level location information to manage attendance [1], monitor congestion, and reserve meeting rooms[2].

A method using radio signals is available for room-level location estimation (hereafter referred to as “room-level localization”). Among them, the method using BLE signals use dedicated terminals that transmit BLE signals (hereinafter referred to as “physical beacons”) and a receiver that receives BLE signals. The receiver receives BLE advertising packets from nearby physical beacons. Room-level estimation using the identifier contained in the beacon to determine which physical beacon the receiver received the BLE signals from. There are two main types of room-level localization methods using BLE signals. The first one is to install physical beacons in each room and the users carry the receiver. The second one is to install receivers in

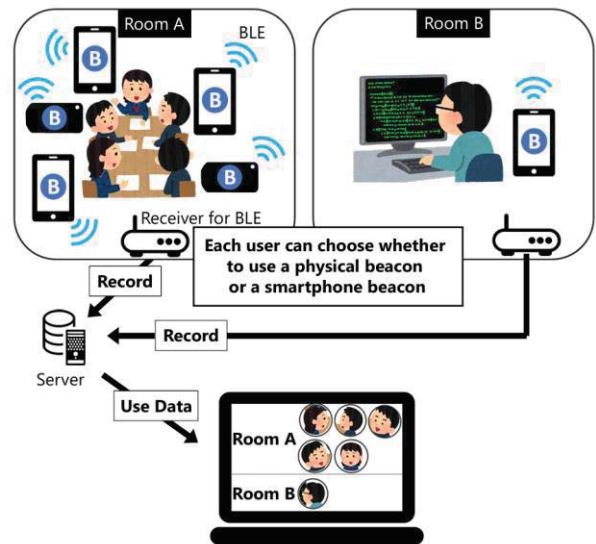


Figure 1: Schematic diagram of the stay estimate system to be realized in this study. The user can choose whether to use a smartphone beacon or a physical beacon.

the rooms and the users carry the physical beacons. We have proposed the system called “StayWatch” that manage the stay information of our laboratory members using the latter method because of its low burden on the users, and we are operating it in our laboratory.

There are also multiple problems with the method where the users carry a physical beacon. Since each user carries one physical beacon, the number of beacons needed to be distributed must match the number of users in the community where the beacons are to be deployed. The financial cost is high because many physical beacons are required. In addition, it takes a lot of time and effort to setup physical beacons. When performing room estimation, individuals are identified using UUIDs advertised from physical beacons. When installing in a community, it is necessary to manually enter and set the UUID of the physical beacon for each user. Physical beacons may be placed in rooms and the user may move between rooms, resulting in poor tracking.

The purpose of this research is to realize BLE beacons with high tracking performance and low financial and human costs. As an approach, we implement an application that substitutes the role of physical beacon with a smartphone (hereinafter referred to as “smartphone beacon”). Smartphone beacons are expected to offer better tracking performance than physical beacons, as many people carry

	Smartphone Beacon	Physical Beacon
User Burden	Install the application on the user's own smartphone.	Battery needs replacement (About once a year)
Administrator Burden	None	Initial setting of UUID, etc.
Financial Cost	Low (Use user's smartphone)	High (Thousands of yen per beacon)
User Restrictions	No restrictions on app installation	Available to everyone

Figure 2: Features of smartphone beacons and physical beacons

their smartphones with them at all times. An overview of this application is shown in Figure1. As shown in Figure2, smartphone beacons and physical beacons have their own advantages and disadvantages. Therefore, the system should be configured so that each user can choose whether to use a smartphone beacon or a physical beacon.

2 RELATED RESEARCH

There are room-level localization methods that use real-time radio wave information, such as Wi-Fi and BLE. Some Wi-Fi-based methods detect and use packets sent by smartphones to locate Wi-Fi base stations in the surrounding area [3]. Users do not need to install special applications on their device. However, in recent years, MAC addresses of smartphones are often randomized to improve privacy, in which case this method cannot be used. The fingerprinting method [4] is a method that enables highly accurate location estimation using radio waves. The fingerprinting method has high administrative costs. Room-level localization does not require as much accuracy as the fingerprint method. Therefore, the proximity method, which estimates location at the area level, is often used for room-level localization.

BLE allows for easier adjustment of the location and number of base stations than Wi-Fi. Wi-Fi base stations are installed to improve the wireless communication environment. Therefore, it is difficult to change or increase the number of Wi-Fi base station locations to improve the accuracy of room-level localization. In contrast, physical beacons that transmit BLE signals are installed specifically for room-level localization, making it easier to place them in optimal locations and to adjust the number of beacons.

There are two main methods for room-level localization using proximity to a BLE beacon terminal: one is to install a physical beacon in rooms and the other is to have users carry physical beacons. In the method that installs physical beacons in rooms, physical beacons are first installed in rooms. Each physical beacon should advertise a unique ID. Users run the application on their smartphones to receive signals from the physical beacons. When the user enters the room, the smartphone receives the signals from the physical beacon. The unique ID contained in the signals from the physical beacon is sent to the server to determine which smartphone is being used and which room the physical beacon is in. In this way, it is possible to estimate who is in which room. Physical beacons do not require any cables; they can be simply fixed to a wall or ceiling. Users do not

need to carry any special equipment. The disadvantage is that it is necessary to install a specific application on the smartphone. In addition, the smartphone must continuously receive BLE signals from the surrounding area and communicate with the server, which places a heavy burden on the smartphone and consumes a large amount of battery power.

In the method where each user carries a physical beacon[5], receivers are first installed in each room. The administrator distributes small physical beacons to users, which are configured to advertise a unique ID. When a user enters a room, the receiver in that room receives signals from the physical beacon the user is carrying. The unique ID contained in the signals from the physical beacon and the unique ID of the receiver are sent to the server. The server checks whose physical beacon it is and which room the receiver is in. In this way, it can estimate who is in which room. Users only need to carry the distributed physical beacon in their wallets or bags. The disadvantage is that the financial cost is high because physical beacons are required for each user. In addition, the initial configuration of physical beacons is time-consuming because it is done manually for each beacon using a configuration application.

There is a study that uses BLE signals emitted from smartphones to estimate the level of congestion in a room[6]. For this purpose, we use COVID-19 contact tracing application (hereafter referred to as COCOA) [7], which was provided by the Ministry of Health, Labour and Welfare. Smartphones with COCOA installed transmit BLE signals periodically. This BLE signal is received by a receiver installed in the room. The number of smartphones in the surrounding area is determined based on the information about how many BLE signals the receiver receives from them. In this way, the number of people in the room is estimated. Users only need to install COCOA on their smartphones and do not need to install any additional applications for the stay estimate system. Users who do not wish to install COCOA can be provided with physical beacons to carry instead. However, COCOA cannot identify individuals because the contents of advertised packets are encrypted and updated periodically. Therefore, COCOA cannot be incorporated into a stay estimation system.

The acquired stay information can be used in various situations, contributing to the realization of a smart city[8]. For example, it can be used to estimate the room occupancy rates for indoor disaster rescue activities and power management. Another study has been conducted that promotes laboratory visits by predicting and presenting future stay information based on each laboratory member's past stay data and calendar schedules [9].

3 SMARTPHONE BEACONS IN STAY ESTIMATION SYSTEMS

In this chapter, we discuss the requirements specification and implementation of the smartphone beacon proposed in this study. Users choose whether to use a physical beacon or a smartphone beacon and carry the chosen device. Receivers in the room receive BLE signals transmitted from the

beacon device and send them to the server to record stay information.

3.1 Examination of Requirement Specifications

Based on the background of this research presented in Chapter 1 and related research presented in Chapter 2, the requirements for a smart beacon are as follows.

- (R1) Each user can choose whether to use a smartphone beacon or a physical beacon
- (R2) Small battery consumption of the smartphone
- (R3) Small initial setup effort
- (R4) Users do not need to open the application except when changing settings

R1: In order to allow the user to choose whether to use the smartphone beacon or the physical beacon, the smartphone beacon is made to behave in the same way as the physical beacon. It is appropriate for users to be able to choose between smartphone beacons and physical beacons, as each has its own advantages and disadvantages. Physical beacons periodically advertise packets containing a UUID to identify the user. Smartphone beacons should behave in the same way. This would eliminate the need for major modifications to the existing stay management system.

R2: In order to suppress battery consumption low, the application only transmits BLE signals during normal operation. It is necessary to make it a low battery consuming application, because smartphone beacon operates constantly. Periodic transmission of BLE signals can continue for several months to several years with a coin battery. Thus, transmitting BLE signals is low power consumption. Therefore, as in the solution of R1, battery consumption can be suppressed by having the smartphone beacon operate in the same way as the physical beacon. The network communication with a smart phone consumes a large amount of battery power. Therefore, the smartphone beacon does not require network communication except when it is set up, and its normal operation is limited to transmitting BLE signals, which consume low battery power.

R3: To reduce the time and effort required for initial setup, the application automatically performs the setup. A community administrator performs the initial configuration of the physical beacon using a configuration application and manually enters the UUID and transmission frequency. The initial setup takes more than one minute per beacon, and about one in every 20 beacons is incorrectly configured. Additionally, while the initial setup of physical beacons is done by the administrator, it is not desirable for the administrator to directly operate and configure the smartphone beacon, which is installed on each user's smartphone. We believe that even if the user performs the initial setup, it takes a long time for a person who is familiar with the initial setup to make a mistake, and that an unfamiliar user will take more time and make a mistake. The smartphone beacon automati-

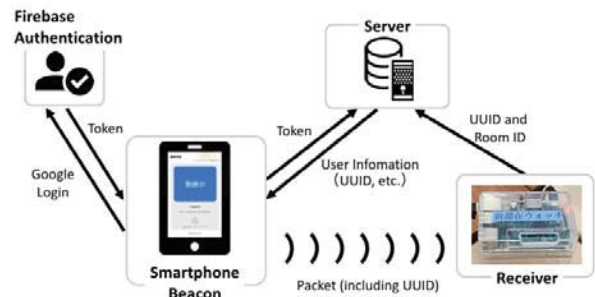


Figure 3: Initial setup flow

cally performs the initial setup, to reduce this burden on users.

R4: To prevent users from forgetting to operate the application, they should not need to open it except during initial setup. Applications that need to be opened periodically require active user action. In the case of such an application, the application may not run due to forgotten operations or troublesome operations. Therefore, the user does not need to actively open the smartphone beacon except for the initial setup or when changing settings. Therefore, smartphone beacons should continue operating unless users intentionally stop it.

3.2 How to set up outgoing packets

Figure 3 shows the flow of automated initial setup to satisfy R3. The beacons use OAuth authentication with a Google account to get the user's UUID. OAuth authentication prevents unauthorized acquisition of the user's UUID. At the first startup, the user logs in with their Google account according to the application screen. A token is issued upon successful authentication with Firebase Authentication. The smartphone beacon sends the token to the server that manages the user information. The server uses the information in the token to verify whether the user has authenticated with Firebase Authentication and whether the user is registered. If the server confirms that the token has been authenticated and the user has been registered, it sends the user information associated with the e-mail address to the smartphone beacon. The beacons use the user information got from the server to perform the initial setup and advertise a packet containing the UUID. In this way, the smartphone beacon satisfies R3.

3.3 Implementation in Android

To satisfy R1, smartphone beacons for Android advertise the UUID got from the server as it is included in the packet. Smartphone beacons transmit BLE signals using android.bluetooth.le package provided by Android. The smartphone beacons use this to set the UUID to be advertised, the frequency of transmission, and the transmission strength, and transmit BLE signals. When the application is not open (hereafter referred to as "background"), the device transmits at the lowest frequency

(ADVERTISE_MODE_LOW_POWER) for any of the settings. In order to accommodate large rooms, the transmission strength is set to the strongest setting ADVERTISE TX POWER HIGH. In this way, the smartphone beacon satisfies R1.

To satisfy R4, the smartphone beacon must also run in the background. For this purpose, the smartphone beacons use a foreground service that allows applications to run in the background. However, it is not possible to satisfy R4, because this alone may stop the transmission of BLE signals.

To satisfy R4, when the smartphone is unable to continue transmitting BLE signals, it will automatically resume when it is ready to transmit. There are three scenarios where BLE signals transmission in Android becomes uncontrollable: when the Bluetooth function is disabled, when the smartphone is turned off and the foreground service is terminated, or when the foreground service is forcibly terminated by Android.

When the smartphone becomes unable to use the Bluetooth function, the smartphone automatically resumes transmitting BLE signals when it becomes able to do so. We use the BluetoothAdapter.ACTION_STATE_CHANGED intent broadcasted by Android when the Bluetooth status of the smartphone is changed. When the foreground service detects this intent, it executes a process to start transmitting BLE signals.

When the smartphone is turned off and the foreground service is stopped, BLE signal transmission automatically resumes when the smartphone is turned on. We use the ACTION_LOCKED_BOOT_COMPLETED intent broadcast by Android when the smartphone is turned on. When the beacon detects this intent, it starts the foreground service and begins transmitting BLE signals.

When Android forces the application to stop, the BLE transmission of the smartphone beacon is automatically resumed when the resource is reused. To do this, set the return value of the method that is executed when the foreground service is started to START_STICKY. When Android kills the foreground service, it will be automatically restarted as soon as the resources become available again. In this way, the smartphone beacon satisfies R4.

3.4 Implementation in iOS

iOS does not allow the user’s UUID to be included in the packet as it is in the background. When an iOS application is in the background, the UUID enters the overflow area and is converted to a different value and included in the packet. The value that changes according to the UUID in the packet actually advertised by the smartphone beacon is a 152-bit hexadecimal number. Of the 152 bits, 24 bits are fixed values and 128 bits are values that change according to the set UUID. The initial value of the value that changes according to the UUID is 0 for all 128 bits.

The smartphone beacon for iOS uses the table to set multiple UUIDs to represent the values it wants advertised in the packet. When a single UUID is set, one bit out of 128 bits changes from 0 to 1. When multiple UUIDs are set, all the bit positions corresponding to each UUID will change. We created a correspondence table (Figure 4) to express the

Changing bit positions	UUID to be set
0	00000000-0000-0000-0000-000000000039
1	00000000-0000-0000-0000-000000000072
2	00000000-0000-0000-0000-00000000002b
126	00000000-0000-0000-0000-000000000037
127	00000000-0000-0000-0000-00000000007c

Figure 4: Table of correspondence between bit positions to be changed and UUIDs to be set

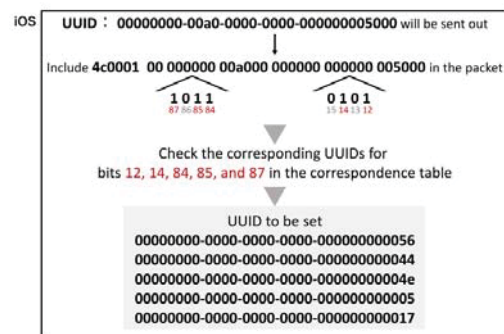


Figure 5: Example of setting multiple UUIDs with correspondence table in iOS

value to be advertised in the packet using the above properties. An example of setting multiple UUIDs using the correspondence table is shown in Figure 5. The smartphone beacon wants to advertise 00000000-00a0-0000-0000-0000-00000000005000. In this case, the value advertised should be 4c0001 for the fixed value part and 000000000000000000a0000000000000000000000000000000000005000 for the value that changes according to the UUID. To achieve this, because bits 12, 14, 84, 85, and 87 be changed from 0 to 1, the UUID corresponding to each bit position should be checked in the corresponding table. When all the corresponding UUIDs are set and advertised, the smartphone beacon advertises the following value. 4c0001000000000000000000000000000000000000a000000000 0005000

This allows the UUID to be advertised from an iOS like a physical beacon. In this way, the smartphone beacon satisfies R1.

We made the smartphone beacon to automatically resume transmission when it becomes possible to transmit when it is unable to continue transmitting BLE signals to satisfy requirement 4. There are two scenarios where BLE signal transmission cannot be continued in iOS: when the smartphone is turned off and when the application is forced to terminate by the system.

We used the save and restore state feature to address these issues. The feature allows the system to take over the background process and start the application in the background when necessary, even if the application is stopped. This allows the system to request that BLE-related tasks be performed in place of the application when it is stopped. In this way, the smartphone beacon satisfies R4.

4 CONCLUSION

In this paper, we discuss the problem of using only physical beacons in a stay management system using BLE, and implement a smartphone beacon that enables smartphones to behave like a physical beacon. The method using only physical beacons has problems such as high financial cost, time and effort for initial setup, and forgetting to replace batteries. We have developed a smartphone beacon with high tracking performance that does not require initial setup or battery replacement.

Current the smartphone beacons are always traceable by unintended third parties because the UUIDs included in the BLE advertising packets are fixed and unchanging for each user. As long as the user's UUID is known, the owner of the beacon terminal can determine when he or she was around the receiver by installing a receiver or other device without permission. We believe that a solution to this problem requires a mechanism that allows the stay estimation system to identify whose BLE beacon terminal the signal came from, but prevents third parties from determining the owner of the BLE beacon terminal.

REFERENCES

- [1] ACCESS CO., LTD. Linkit Kintai. <https://linkit.access-company.com/kintai/>. (reference 2024-1-25).
- [2] UCHIDA YOKO CO., LTD. SmartOfficeNavigator. <https://www.uchida.co.jp/it/products/smartofficenavigator/>. (reference 2024-1-25).
- [3] Musa, A.B.M., Eriksson, J. Tracking Unmodified Smartphones Using Wi-fi Monitors. *Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems*, pp. 281–294, 2012.
- [4] Ramsey, F., Robert, H. Location Fingerprinting With Bluetooth Low Energy Beacons. *IEEE Journal on Selected Areas in Communications*, Vol. 33, No. 11, pp. 2418–2428, 2015.
- [5] Barsocchi, P., Crivello, A., Girolami, M., Mavilia, F., Palumbo, F. Occupancy detection by multi-power bluetooth low energy beaconing. *2017 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pp. 1–6, 2017.
- [6] Akira, T., Yutaka, A., Shigeru, T., Shigemi, I. Examination of Automatic Parameter Adjustment in Hybrid Congestion Measurement by WiFi and BLE. *DPS Technical Report*, Vol. 187, No. 16, pp. 1–8, 2021.
- [7] Labour Ministry of Health and Welfare. COVID-19 Contact-Confirming Application (COCOA). https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/cocoa_00138.html, 2020. (reference 2024-1-25).
- [8] Cabinet Office. Smart city. https://www8.cao.go.jp/cstp/society5_0/smartcity/. (reference 2024-5-6).
- [9] Yuto, T., Taku, F., Takashi, Y. Docoitter: A Presence Display System Capable of Predicting Future In-the-room Information. *IPSJ Journal*, Vol. 54, No. 9, pp. 2265–2275, 2013.

Anomaly Detection for Weather Sensors Based on Relationship Among Multiple Sensors

Kiyoto Kimura[†], Takuya Yoshihiro[‡]

[†]Graduate School of Systems Engineering, Wakayama University, Japan

[‡]Faculty of Systems Engineering, Wakayama University, Japan

[†]e-mail: s256078@wakayama-u.ac.jp

[‡]e-mail: tac@wakayama-u.ac.jp

Abstract - Measurements of weather sensors are important indicators for decision-making in disaster warnings, everyday life, industries, and agriculture, etc. Therefore, ensuring the reliability of weather sensor measurements is crucial. To maintain reliability, it is necessary to promptly detect anomalies by applying anomaly detection methods to the sensor measurements. Conventional methods model the probability distribution of a sensor's measurement values based on past data and apply it to current measurements to detect anomalies. However, this method can suffer from biases in the probability distributions due to factors such as imbalanced datasets, which can degrade the accuracy of anomaly detection. On the other hand, with the recent proliferation of the Internet of Things (IoT), weather sensor data are being collected via wireless communication and utilized in various applications. Therefore, we propose an anomaly detection method using multiple weather sensors. Specifically, we generate a deep learning-based prediction model using data from the target sensor and surrounding weather sensors to perform anomaly detection. By using data from multiple weather sensors, we aim to reduce the impact of biases in individual weather sensor data and improve the accuracy of anomaly detection. In the evaluation, we compare the accuracy of anomaly detection between the proposed method and conventional methods using AMeDAS weather data and artificially created anomaly data. The results show that the proposed method improves the accuracy of anomaly detection.

Keywords: IoT, anomaly detection, weather sensor, deep learning

1 INTRODUCTION

Measurements of weather sensors are important indicators for decision-making in disaster warnings, everyday life, industries, and agriculture, etc. With the recent proliferation of the Internet of Things (IoT), weather sensor data can be collected via wireless communication and utilized in various applications. In this situation, ensuring the reliability of weather sensor data is crucial for many applications. At meteorological observation stations operated by the Japan Meteorological Agency, expensive sensors are installed for each weather measurement item, and large management costs and efforts are spent to maintain their reliability [1][2]. However, this management strategy requires unacceptably high costs as the number of sensors increases, unsuitable for an IoT-populated society.

Consequently, to make the most of IoT technologies, a highly accurate and inexpensive way of surveillance to maintain the reliability of cheaper IoT sensors is mandatory. Especially in inexpensive IoT devices, anomaly detection is the key technique to maintain the reliability of the sensor sets [2]. Traditionally, for anomaly detection of sensors, we often make a distribution of the past measurements of a sensor and check if the current measurements of the sensor are outliers or not [3]. However, this traditional method inevitably involves a problem that the outliers hidden in dynamic time-series data are hard to find because a static distribution is hardly obtained [4]. Thus, recently, deep-learning-based anomaly detection has been studied, in which the current measurement value is compared with the predicted value from the past measurements data to check if the predicted value is outlier or not [5]. However, in this case, we suffer from some kind of biases possibly contaminated in the past measurement data set (called the imbalanced data), which may reduce the anomaly detection performance [6].

One of the ideas to address this issue is to utilize measurements of other sensors located relatively near the target sensor. Because sensors located nearby have a correlation in their measurements with each other, prediction of measurements of a sensor from its nearby sensors will be possible, and thus fault detection will also be possible. Here, note that the individual sensors are considered to have their trend of biases. Thus, integrating prediction results from multiple sensors for detecting the anomaly of a target sensor will be more accurate because biases of different sensors are averaged and reduced.

Based on the idea above, we in this study propose an anomaly detection method that integrates multiple prediction models learned from the past measurements of surrounding weather sensors. Since the multiple prediction models learn different features from different data sets, they are considered to have different biases in their prediction results. Therefore, when predicting measurements of a target sensor, the biases are averaged and canceled, which leads to high-accuracy anomaly detection that allows for detecting rare anomaly events. We evaluated the proposed method using AMeDAS weather sensor data and a data set artificially created from them. The results show that the proposed method improves the accuracy of detecting anomalies created in a weather station using measurements of surrounding weather station data.

This paper consists of 4 sections. Section 2 describes the proposed method. Section 3 describes the evaluation of the proposed method using AMeDAS weather sensor data and ar-

tificially created anomaly data. Finally, Section 4 summarizes this study.

2 PROPOSED METHOD

2.1 Problem Setting

The purpose of this study is to detect long-term anomalies, i.e., sensor failure, where the target weather sensor outputs inaccurate values over a certain period of time, due to factors such as sensor equipment failure or the surrounding environmental conditions. Weather sensor anomalies can occur for various reasons, not only sensor malfunctions but also influences from plants and animals or human errors. Taking temperature sensors as an example, According to the documents from the Japan Meteorological Agency [1][2], instances of temperature sensor anomalies include sensor degradation, ventilation fan failures, ventilation issues caused by spider webs or plant vines, short-term temperature spikes due to field burning, aircraft blasts, or heat from heavy machinery exhausts. These anomalies manifest in various forms in the measured value, such as significantly higher temperature readings during the day or temporary spike-like abnormal values. Among these anomalies, those that persist for long periods are particularly impactful. If such anomalies continue for long periods without being noticed by the administrators, they can affect various applications that utilize the weather for a long time, greatly diminishing the reliability of the sensors. Therefore, it is crucial to constantly monitor the weather sensor measurements, promptly identify abnormal values, and respond accordingly.

In recent years, the proliferation of the Internet of Things (IoT) has increased the number of examples where multiple sensors are connected to networks for operation. Various inexpensive weather sensors are also available, many of which can collect data through wireless communication. By using these devices, if multiple sensors monitor each other's measurements and analyze the data collected through the network in real-time to detect anomalies, it would allow for automatic anomaly detection without requiring administrators to calibrate each sensor periodically. This study proposes a method for automatic detection of long-term anomalies, which are difficult to detect using conventional anomaly detection methods that rely solely on past measurements of the sensor in question, by leveraging the cooperation of multiple sensors.

This study assumes a situation where multiple weather sensors are installed within a certain range. Types of weather sensors include temperature, precipitation, humidity, atmospheric pressure, wind direction, and wind speed. There is a relatively strong correlation among the measurements of these sensors. Using methods such as deep learning, it is possible to predict mutual measurements with a certain degree of accuracy among the same type of sensors. Also, concerning different types of sensors, their measurements are related to each other as a physical phenomenon and can therefore be an important factor for prediction. By utilizing these correlations, any long-term anomaly in one of the sensors can be detected without delay. Specifically, when checking anomalies in a particular sensor (hereafter referred to as the target sen-

sor), multiple prediction models can be generated to predict the target sensor's measurements from the surrounding sensors. The target sensor's measurements are always predicted using these prediction models. If the target sensor measures abnormal values over a long period of time, the predicted values are expected to have a significant difference from the real measurements of the target sensor. By capturing this phenomenon, long-term anomalies in the target sensor can be detected.

2.2 Overview of Proposed method

In this study, we propose an anomaly detection method using multiple prediction models based on the relationships between multiple weather sensors. Using the measurements of the target sensor as the prediction target, multiple prediction models with different inputs are generated using the measurements of the past target sensor and those of multiple weather sensors installed in the vicinity. Each prediction model performs outlier detection. To detect long-term anomalies, we focus on the mean absolute value of the prediction errors of each prediction model over a certain period of time rather than capturing the instantaneous magnitude of the prediction error. Subsequently, the anomaly detection results from each prediction model are integrated to produce the final anomaly detection result. Since each prediction model predicts the target sensor's measurements based on different characteristics of different input data, the timings that large prediction errors appear are expected to vary. Nonetheless, if multiple prediction models simultaneously detect anomalies, it will not be a false detection, i.e., the target sensor would actually be experiencing an anomaly.

2.3 Anomaly Detection Using a Prediction Model

A prediction model performs outlier detection. Outlier detection is performed by calculating PMAE (Period Mean Absolute Error) and modeling it as a probability distribution. PMAE is the mean absolute value of the prediction errors occurring within a certain time interval. Let f be the prediction model that predicts the measurements of the target sensor. For the measurements d_t of the target sensor at time t , let y_t be the predicted value output by the prediction model f . The prediction error e_t is obtained by Equation (1), and PMAE $E(t)$ is defined by Equation (2):

$$e_t = y_t - d_t, \quad (1)$$

$$E(t) = \frac{\sum_{k=t}^{t+d} |e_k|}{d}, \quad (2)$$

where d is a parameter that determines the time interval. For example, if the prediction frequency of model f is every 10 minutes and parameter d is set to $d = 6$, PMAE $E(t)$ is calculated on an hourly basis. By using the absolute value of the prediction error e_t to calculate PMAE $E(t)$, we evaluate the magnitude of errors by preventing offsetting when positive values and negative values are mixed within the period.

Since the PMAE $E(t)$ is the mean absolute value of the prediction errors, it is distributed within the range $[0, \infty]$. In outlier detection, when a threshold T is set, PMAE $E(t)$ within the range $(T, \infty]$ are considered anomalies. The threshold T is set using a percentile method [7]. The percentile method is an effective threshold setting technique when the dataset doesn't follow a normal distribution. In this study, for example, if the threshold is set at the 95th percentile, the threshold T is set such that 95% of the data falls within the range $[0, T]$, and the remaining 5% of the data falls within the range $(T, \infty]$.

2.4 Anomaly Detection by Integrating Prediction Models

The final anomaly detection result for the target sensor using the proposed method is determined by integrating the anomaly detection results obtained from multiple prediction models. When there are N prediction models, each performs anomaly detection using the method explained in Section 2.3, resulting in N anomaly detection results. The final anomaly detection result for the target sensor is considered an anomaly if at least n out of the N anomaly detection results indicate an anomaly, where n is a parameter that determines the condition for integration. The parameter n is determined by the majority rule. By using the majority rule, it is believed that people can obtain reliable results in real environments. Additionally, even if one of the prediction models makes an incorrect decision, the final result can be accurate.

3 EVALUATION

3.1 Evaluation Methods

To confirm that the proposed method improves the accuracy of anomaly detection, the following two folds of evaluations are performed.

- (1) Verifying independence of individual prediction models.
- (2) Evaluating the accuracy of the proposed method.

In evaluation (1), we verified that PMAE $E(t)$ calculated by each prediction model is independent of each other. If the PMAE of each model is independent of each other, it means that the biases included in each model are not correlated, and that the probability that multiple prediction models detect an anomaly simultaneously is very low. Thus, this will be one of the evidence that our idea of integrating multiple prediction models is valid. In evaluation (2), we evaluated the accuracy of anomaly detection using the proposed method. We compared the performance of the proposed method with a basic outlier detection method explained in Sec. 2.3 using each prediction model learned from a nearby weather station's data set. We used three indexes to evaluate the accuracy of the proposed method: Precision, Recall, and F-measure (called also as F1-score).

The parameters of the proposed method were set as follows. The parameter d which determines the range of PMAE $E(t)$ was set to $d = 6$. Here, since the prediction frequency

of the prediction model used in this study is every 10 minutes, PMAE is the mean of the absolute errors for each an hour. The threshold T was set using the percentile method, with the 95th percentile as the threshold. A dataset used to set the threshold was different from a dataset used to evaluate the accuracy of anomaly detection.

3.2 Dataset

3.2.1 Dataset Details

In this study, we evaluated the proposed method using weather sensor data from AMeDAS, obtained by the Japan Meteorological Agency [8]. We used three years of data from 2019 to 2021; weather sensor data were AMeDAS 1-minute values observed every minute. The types of weather sensors were temperature, precipitation, wind speed, and wind direction. For training the prediction models, we used data from 2019, while for evaluating the proposed method, we used data from 2020 and 2021.

The AMeDAS data contains missing values. As a preprocess of missing values, we complemented them with the mean of the values before and after the missing value if the missing values were not consecutive. If the missing values were consecutive, we deleted the values for the corresponding times. To compress the data, we transformed the AMeDAS 1-minute values to 10-minute values. Specifically, the sensor data of 1-minute values were compressed for every ten samples, and the mean value was used as the 10-minute values.

The time and wind direction data, which are input to the prediction model, were transformed using the sine and cosine functions to account for their periodicity. Wind direction data is observed as 36-direction values, ranging from 1 to 36 (0 being no wind). The time data was obtained at ten-minute intervals, with 00:00 set as zero and incremented by one every 10 minutes to reach 143 at 23:50. The transformation using the sine and cosine functions was performed on these values x using the following Equations (3), (4):

$$z_{sin} = \sin\left(2\pi \frac{x}{X}\right), \quad (3)$$

$$z_{cos} = \cos\left(2\pi \frac{x}{X}\right), \quad (4)$$

where z_{sin} is the sine transformed value, z_{cos} is the cosine transformed value, and X is the number of possible values of x , which is 36 for wind direction values and 144 for time values.

To eliminate bias among the features of the training data and improve prediction accuracy, we normalized the values of all features to be within the range $[0, 1]$. The normalized value z_i was calculated using the following Equation (5):

$$z_i = \frac{x_i - \min(x)}{\max(x) - \min(x)}, \quad (5)$$

where x_i is the value before normalization, and $\min(x)$ and $\max(x)$ are the minimum value and the maximum value in the training data, respectively.



Figure 1: Weather observation station locations
(Source: Google Maps)

3.2.2 Target Weather Observation Station

In this study, we evaluated our method using weather sensor data from ten weather observation stations as shown in Fig. 1: Wakayama, Katsuragi, Kawabe, Kankujima, Kumatori, Koyasan, Sakai, Shimizu, Tomogashima, and Ryujin. From the weather observation station in Wakayama, we obtained temperature sensor data [$^{\circ}\text{C}$], while from the other nine stations, we obtained four types of weather sensor data: temperature [$^{\circ}\text{C}$], precipitation [mm], wind speed [m/s], and wind direction (36 directions). In this study, the target sensor for anomaly detection was the temperature sensor in Wakayama. Multiple prediction models were generated using weather sensor data and time data from all ten locations and used for anomaly detection.

3.2.3 Generation of Anomaly Data

In this study, we artificially generated the anomaly data used to evaluate our proposed method. We assumed a scenario where a malfunction of the ventilation fan causes the daytime temperature to be observed higher than normal. Because of the ventilation fan failures, heat tends to accumulate inside the ventilation tube more than normal, and the temperature is expected to be observed higher than normal during the daytime because of sunshine. Therefore, we transformed daytime temperature data from the normal data to generate the anomaly data. Specifically, we added a constant value b [$^{\circ}\text{C}$] to the temperature sensor data of Wakayama from 10:00 to 16:00 to generate the anomaly values. The reason for using the period from 10:00 to 16:00 is that the effect of failures is assumed to be fully reflected because enough time has elapsed since sunrise. We confirmed the magnitude of anomalies that could be detected with various values for the constant value b

Table 1: Training settings

Item	Values
Optimizer	Adam (0.001)
Epoch	Early Stopping (10)
Error function	MSE (Mean Squared Error)
Batch size	128
Language	Python version 3.9.13
Library	PyTorch version 1.13.1

[$^{\circ}\text{C}$].

From the data from 2020 and 2021 used to evaluate the proposed method, the data from 10:00 to 16:00 was extracted, the data from 2020 was used as the dataset for setting the threshold of outlier detection explained Sec. 2.3, and the data from 2021 was used as the dataset for evaluating the accuracy of anomaly detection. The number of samples for both normal and anomaly data was 1586 samples.

3.3 Detail for Prediction Models

3.3.1 Configuration of Prediction Models

We used LSTM (Long Short-Term Memory) to generate the prediction models. Each step of the LSTM layer inputs 10-minute values, and at the final step, the outputs of the LSTM layer are input to the fully connected layer, which outputs the prediction value of the target sensor. The LSTM layer consists of 144 time steps. Therefore, the prediction model is input to 24-hours data. The LSTM layer consists of a single layer, and the number of units in the hidden layer is set to 10.

3.3.2 Training Settings

The settings for training are shown in Table 1. The initial value of the learning rate was set to 0.001 and adjusted using Adam. To prevent overfitting, the early stopping technique was adopted, where training is stopped if accuracy does not improve in 10 epochs. The general stochastic gradient descent (SGD) method was used to learn networks, and the general backpropagation technique was applied to calculate the gradient of the loss function.

3.3.3 Construction of Prediction Models

The names of the generated prediction models and the input features are shown in Table 2. The Wakayama model only inputs the target sensor data (the temperature sensor in Wakayama). The other nine prediction models were constructed using weather sensor data from a single observation station and time data. Particularly, for temperature sensor data prediction, it is considered effective to input temperature sensor data from nearby stations, and time is also considered a value highly related to temperature.

If anomaly values are included in the input, accuracy significantly decreases, so the input data of the Wakayama model is the data from 24 hours to 48 hours before the prediction time, and it is assumed that there are no anomaly data in the

Table 2: Names and input of Each Prediction Model

Name	Input
Wakayama model	temperature (in Wakayama observation station)
Katsuragi model	temperature, precipitation, wind speed, wind direction, time (in Katsuragi observation station)
Kawabe model	temperature, precipitation, wind speed, wind direction, time (in Kawabe observation station)
Kankujima model	temperature, precipitation, wind speed, wind direction, time (in Kankujima observation station)
Kumatori model	temperature, precipitation, wind speed, wind direction, time (in Kumatori observation station)
Kouyasan model	temperature, precipitation, wind speed, wind direction, time (in Kouyasan observation station)
Sakai model	temperature, precipitation, wind speed, wind direction, time (in Sakai observation station)
Shimizu model	temperature, precipitation, wind speed, wind direction, time (in Shimizu observation station)
Tomogashima model	temperature, precipitation, wind speed, wind direction, time (in Tomogashima observation station)
Ryujin model	temperature, precipitation, wind speed, wind direction, time (in Ryujin observation station)

input of the prediction model in this study. The other nine prediction models used data from the prediction time to 24 hours before as their input data.

3.4 Results

3.4.1 Verification of independence of Prediction Models

To confirm the independence of each prediction model, we calculated the correlation coefficients of the PMAE $E(t)$ between every two prediction models. The results are shown in Fig. 2. According to Fig. 2, there are no cases for which the correlation coefficients exceeded 0.7, and there is no strongly correlated combination. Because the mean of the 45 correlation coefficients calculated from the ten prediction models was 0.356, the correlations were not strong. Therefore, the PMAE $E(t)$ calculated by each prediction model is independent of each other.

3.4.2 Anomaly Detection Accuracy of the Proposed Methods

Figure 3 shows the anomaly detection accuracy of the proposed method and the conventional methods. The conventional methods adopt the basic outlier detection method described in Sec. 2.3 using Wakayama model. The Wakayama model's input was solely the temperature sensor measurements in the Wakayama observation station. Figure 3 (a) shows the precision, Fig. 3 (b) the recall, and Fig. 3 (c) the F-measure, respectively. The horizontal axis is the constant value b [$^{\circ}\text{C}$] added when generating the anomaly values. Parameter n , which determines the condition for integration, was tested for five values: 6, 7, 8, 9, and 10. According to Fig. 3, the proposed method had higher precision and recall than the conventional methods for all anomaly values b . Consequently, the proposed method had a higher F-measure than the conventional method. These figures show that the proposed method improves anomaly detection accuracy.

Next, we examine the effect of integrating multiple models. We selected three models (Kumatori model, Kawabe model, and Tomogashima model) with the highest accuracy. The anomaly detection accuracy of the proposed method integrating these three models, and each prediction model, are

compared in Fig. 4. The evaluation index was the F-measure. The horizontal axis was the constant value b [$^{\circ}\text{C}$] added when generating the anomaly values. According to Fig. 4, the proposed method had a higher F-measure than the three prediction models. We conclude that integrating multiple prediction models with the proposed method achieved more accurate anomaly detection than each prediction model with outlier detection.

3.4.3 Effect of the Number of Prediction Models

We confirmed the effect in the accuracy of the number of integrated prediction models. The results are shown in Fig. 5. Figure 5 shows four cases where the number of models was changed to 3, 5, 7, and 9. The order of prediction models in descending order of accuracy, is Kumatori model, Kawabe model, Tomogashima model, Sakai model, Katsuragi model, Shimizu model, Kouyasan model, Kankujima model, and Ryujin model. When integrating k -models in Fig. 5, models were selected in the descending order of accuracy. According to Fig. 5, the anomaly detection integrating three models had the highest accuracy, and the accuracy slightly decreased with the number of models increased. This indicates that increasing the number of models does not contribute to improving accuracy. Therefore, the proposed method can achieve high accuracy by integrating a small number of high-accuracy models.

4 CONCLUSION

In this study, we proposed an anomaly detection method for weather sensors by integrating multiple prediction models based on the relationship between multiple sensors. By integrating multiple prediction models, we aim to decrease the effect of biases contained in each model's input data and improve the accuracy of anomaly detection.

We evaluated the proposed methods using AMeDAS data and anomaly data artificially generated from them. We compared the performance by comparing the proposed method with the conventional method. The conventional method is the outlier detection using a single prediction model. From the results, we first showed that the correlation in prediction

	Wakayama model	Katsuragi model	Kawabe model	Kankujima model	Kumatori model	Kouyasan model	Sakai model	Simizu model	Tomogashima model	Ryujin model
Wakayama model	1	0.282	0.134	0.340	0.206	0.149	0.280	0.122	0.266	0.279
Katsuragi model		1	0.403	0.431	0.588	0.565	0.599	0.590	0.199	0.416
Kawabe model			1	0.285	0.373	0.449	0.341	0.555	0.155	0.530
Kankujima model				1	0.622	0.274	0.597	0.269	0.288	0.256
Kumatori model					1	0.404	0.692	0.427	0.199	0.304
Kouyasan model						1	0.395	0.622	0.031	0.440
Sakai model							1	0.390	0.203	0.385
Simizu model								1	0.0513	0.520
Tomogashima model									1	0.131
Ryujin model										1

Figure 2: Correlation coefficients of PMAE $E(t)$ of two prediction models

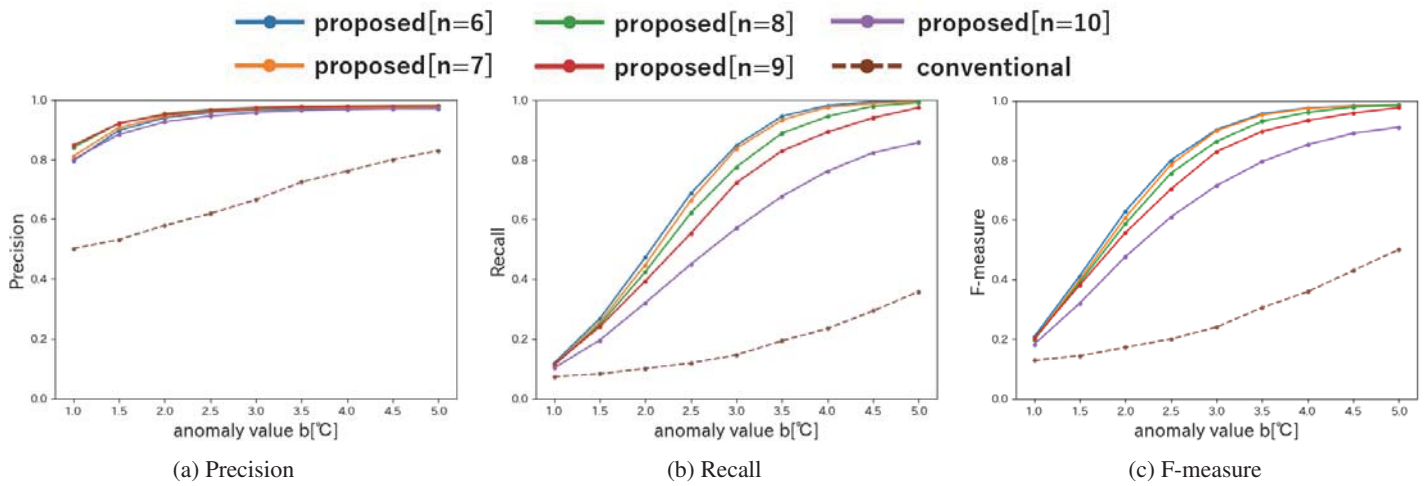


Figure 3: Accuracy of anomaly detection

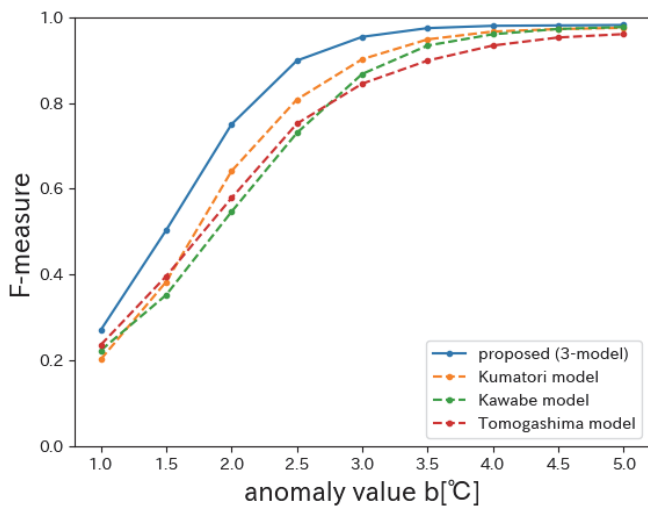


Figure 4: Accuracy of anomaly detection (the proposed method integrating the three most accurate prediction models)

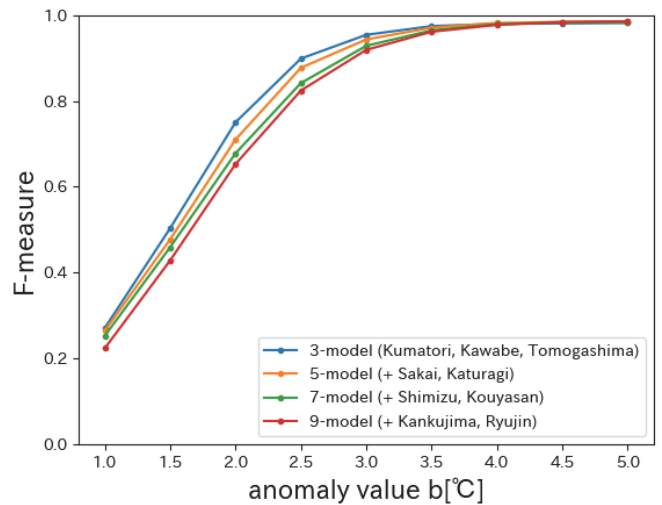


Figure 5: F-measure of the proposed method when changing the number of integrated prediction models

error among each prediction model is not high, meaning that the biases included in each model are independent of one another. Second, we showed that the anomaly detection performance improves when multiple prediction models (relatively independent in bias) are integrated. From this result, we show that our idea to integrate multiple prediction models is effective. Finally, we examined the anomaly detection performance in terms of the number of integrated models. As a result, we found that integrating many models is not always good, but integrating a small number of high-accuracy models is a good strategy.

In future work, it is necessary to confirm the properties of the proposed method. The results of comparing various numbers of models indicate that integrating a small number of models with high accuracy improves the anomaly detection accuracy, rather than merely increasing the number of models. However, the optimal number and quality of models to achieve the highest accuracy remain unknown. One of our future researches should focus on determining the optimal combination of models to maximize the accuracy of the proposed method.

<https://www.jma.go.jp/jma/en/Activities/amedas/amedas.html>

REFERENCES

- [1] Japan Meteorological Agency.: "Kishou kansoku guide-book", Accessed February 21, 2024.
https://www.jma.go.jp/jma/kishou/know/kansoku_guide/guidebook.pdf
- [2] Japan Meteorological Agency, kansoku bu keikaku ka joho kanri shitsu.: "Kansoku deta no hinshitsu kanri kyoka to kodoka no genjo - kion ni kansuru jido hinshitsu kanri kino, taiwateki hinshitsu kanri tool, kansokusho un'yo kiroku joho no gyomu e no riyo -", Sokko jiho, vol. 82, pp. 55-79, 2015. Accessed February 21, 2024.
<https://www.jma.go.jp/jma/kishou/books/sokkou/82/vol182p055.pdf>
- [3] Meek, D. W., and J. L. Hatfield. "Data quality checking for single station meteorological databases." *Agricultural and Forest Meteorology* 69.1-2 (1994): 85-109.
- [4] C. C. Aggarwal, "An introduction to outlier analysis, " in *Outlier Analysis* Springer, pp.1-40, 2016.
- [5] M. Munir, S. A. Siddiqui, A. Dengel and S. Ahmed, "DeepAnT: A Deep Learning Approach for Unsupervised Anomaly Detection in Time Series," in *IEEE Access*, vol. 7, pp. 1991-2005, 2019, doi: 10.1109/ACCESS.2018.2886457.
- [6] Fernández, A., García, S., Galar, M., Prati, R. C., Krawczyk, B., and Herrera, F.: *Learning from imbalanced data sets*, 2018, Vol.10, Cham: Springer.
- [7] FreeCodeCamp.: "How to Detect Outliers in Machine Learning – 4 Methods for Outlier Detection", Accessed May 2, 2024.
<https://www.freecodecamp.org/news/how-to-detect-outliers-in-machine-learning/>
- [8] Japan Meteorological Agency: "AMeDAS - Japan Meteorological Agency", Accessed February 21, 2024.

A Study on Accuracy Management Methods for IoT Sensors Considering Spatio-temporal Correlations

TAIKI YAMADA[†], TAKUYA YOSHIHIRO[‡]

[†]Graduate School of Systems Engineering, Wakayama University, Japan

[‡]Faculty of Systems Engineering, Wakayama University, Japan

[†]s256281@wakayama-u.ac.jp

[‡]tac@wakayama-u.ac.jp

Abstract - With the proliferation of IoT technology, numerous applications utilizing sensor data have emerged. In these applications, the quality management of sensors, i.e., the maintenance and management of measurement accuracy, becomes critical. A systematic error known as bias, which can gradually fluctuate over time, is a common cause of reduced measurement accuracy in sensors. As bias can increase significantly over time, it is essential to perform regular sensor calibration, such as annually, to correct for bias and maintain accuracy. Calibration is typically performed manually for each sensor, but as the number of sensors increase, it becomes challenging to calibrate all sensors manually. Therefore, an automatic calibration method has been proposed in environments with dense sensor deployments, utilizing the proximity of neighboring sensors to estimate correction values for each sensor. However, existing automatic calibration methods focus solely on estimating optimal correction values, and the accuracy of bias estimation remains unclear. This study proposes a new automatic calibration method that aims to sufficiently minimize the error in all sensors. The proposed method uses a hierarchical Bayesian model to estimate the probability distribution of each sensor's error and periodically calibrates selected sensors based on this distribution, thus maintaining and managing the overall accuracy of the sensors.

Keywords: auto calibration, error estimate, bayesian hierarchical model, MCMC method

1 INTRODUCTION

The Internet of Things (IoT) is important in many emerging applications such as smart agriculture, smart cities, etc [1][2]. In many cases, these applications require deploying many sensors in a wide range and use low-cost sensors. However, if those sensors are used for a long time, their readings start to deviate from the true value. Thus, using applications for a long time requires maintaining the accuracy of sensor measurements.

Sensor measurements involve random noise and systematic error called bias. Sensor measurements deviate from the true value because of these errors. The sensor bias is slowly changing over time. As the bias grows large, measurement accuracy gradually decreases. Therefore, we are required to maintain the accuracy by calibrating sensors, i.e., by making their bias zero, periodically. The example of bias is shown in Fig. 1. Generally, bias changes slowly both in a positive and

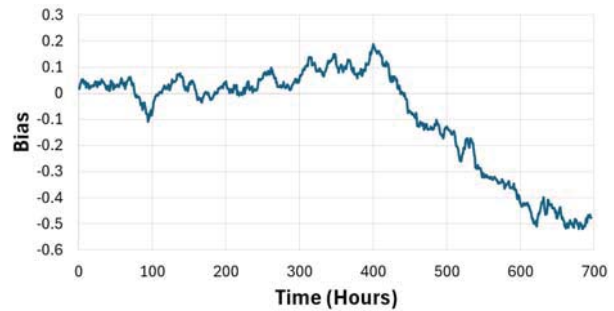


Figure 1: Example of the Bias

negative direction. Therefore, bias takes on a small value in the beginning while it takes on a large value as time passes.

Traditionally, calibration is performed manually. Manual calibration compares the sensor readings including bias with a highly accurate sensor reading, and corrects the sensor based on the comparison. However, manual calibration is impractical when the number of sensors is large because it is extremely time consuming and costly. Hence, several automatic calibration methods have been proposed to calibrate many sensors simultaneously [3]. Automatic calibration methods assume that sensors are densely deployed and the neighboring sensors' measurements are correlated. Under this assumption, automatic calibration estimates each sensor's bias by using neighboring sensor measurements, and corrects each sensor measurement based on the estimated bias. Automatic calibration maintains measurement accuracy by correcting measurement values, instead of calibrating the sensors themselves. Therefore, automatic calibration enables us to save time and cost in comparison with the manual calibration.

Although automatic calibration seems useful for maintaining sensors, existing methods have a serious drawback. If we calibrate sensors with automatic calibration for a long time, each sensor's bias grows larger. This results in lower accuracy in bias estimation, leading to inaccurate calibration. In addition, the existing automatic calibration methods estimate the most likely bias value without estimating how accurate it is. This means that the accuracy of the calibrated results is unknown and cannot be guaranteed. This is practically fatal when we want to keep the sensors' accuracy for a long time.

In this study, we propose an automatic calibration method to maintain the accuracy of sensor measurements within a guaranteed range for a long time. By modeling the bias based on the Bayesian Hierarchical Model and applying the Markov Chain Monte Carlo (MCMC) method, we estimate the proba-

bility distribution of sensors' biases, enabling us to know how accurate the estimated biases are. Additionally, we designed a calibration procedure that combines manual and automatic calibrations. By manually calibrating a part of sensors with large biases, we can reduce the magnitude of sensor biases, enabling us to improve the accuracy of bias estimation. As a result, we designed a novel method that enable us to maintain the accuracy of many sensors for a long time with an acceptable cost.

This paper is an extended version of [4]. We mainly extended the bias estimation model of [4] by introducing a time-domain relationship.

This paper is organized as follows. In Section 2, we present related work. In Section 3, we describe the proposed method. In Section 4, we summarize this study.

2 RELATED WORK

Automatic calibration aims to manage the accuracy of sensors by correcting biases based on the estimated bias values, provided that there is a correlation between neighboring sensors when sensors are densely deployed. Because automatic calibration does not need manual sensor calibration and just corrects the measurement values, it is one of the promising techniques in the IoT era.

Becnel et al. proposed an automatic calibration method under the assumption that there is an accurate sensor [5]. This method constructs a graph that defines edges between neighboring sensors, and calibrates the sensors in turn from the neighboring sensors of the accurate one. Bychkovskiy et al. proposed a method that uses a calibration graph that represents the correlation relationship of sensors [6]. Balzano et al. proposed a method that models measurements, true values, and errors (called gain and offset) using a linear model and estimates gains and offsets of sensors with the least square metrics [7]. Sun et al. proposed a multipoint correction method that uses a multivariable regression modeling and estimates the measurement error based on the least square criterion [8]. These methods estimate the accurate sensor values at a specific time using the correlation of neighboring sensors, meaning that they do not consider the time-series correlation and thus their performance is limited.

Automatic calibration methods that consider clusters of the sensors to improve the accuracy of error estimation are proposed. Moon et al. proposed a method that creates clusters of sensors with similar measurement tendencies and estimates error values by applying multiple regression analysis [9]. Takruri et al. proposed a method that estimates the error value by subtracting the actual measurement of the sensor from the average of the sensors in the cluster [10]. However, the situation in which these assumptions of cluster are satisfied is limited.

There are automatic calibration methods that consider the correlation of time-series. Takruri et al. proposed an automatic calibration method that uses Support Vector Regression (SVR) and Kalman filter [11]. This method uses SVR to estimate the true values at the locations of each sensor by giving the measurements of the neighboring sensors as input data. After estimating the true values, they use the Kalman filter

to estimate the error value of each sensor from the estimated true value and the actual measurement of the sensor. Then, they calibrate the sensors based on the estimated results. Kumar et al. proposed a method that uses Kriging and Kalman filter [12]. This method uses spatial estimation called Kriging to estimate the true values at the locations of each sensor by the measurements of the neighboring sensors and uses the Kalman filter to estimate the error value of each sensor from the estimated value and the actual measurement of the sensor. WU et al. proposed a method that uses a Constrained Extreme Learning Machine (CELM) and Kalman filter [13]. This method divides many sensors deployed randomly into multiple clusters and estimates the measurement of the sensor by the measurements of other sensors in the same cluster with the assumption that the measurements of the sensors in the same cluster tend to measure similar values over time. Afterward, they use the Kalman filter to estimate the error value of each sensor from the estimated value and the actual measurement of the sensor. These methods can estimate the error value considering temporal as well as spatial correlations by applying the Kalman filter recursively.

Instead of Kalman filter, some studies use Particle filter to consider temporal correlation. Fujino proposed a method that uses Gaussian process regression (GPR) [14] and a Particle filter [15][16]. This method estimates the true values at the locations of each sensor from the neighboring sensors' measurements by using GPR. Then, this method estimates the error value from the estimated value and the actual measurement of the sensor by using the Particle filter.

As above, recent auto-calibration methods are based on Kalman or Particle filters. Both Kalman and Particle filters estimate unknown values from a statistical computation based on the state-space model, in which they typically compute future values from the past measurements. In contrast, our method adopts MCMC methods to estimate the statistical distribution of unknown values based on our space-time model within the framework of Bayesian hierarchical modeling, which enables us to simultaneously estimate the statistical distribution of all variables corresponding to any time instant. Different from Kalman or Particle filters, our method performs the optimal estimation using the whole data period, i.e., not only from the past measurements but also from the future. Due to the better optimization and the output of statistical distribution, our method is considered to output better estimation.

In addition, our method is novel in that it can guarantee the accuracy of sensors continuously for a long time, instead of just estimating biases as is done by all the existing methods. First, existing methods just estimate the most likely value of bias and do not output the probability distribution. Accordingly, we cannot get to know how accurate the estimation value is, and so guaranteeing accuracy in the auto-calibration process is not possible. Second, most of the existing auto-calibration methods does not incorporate manual calibration of sensors. Thus, the bias of each sensor grows larger as time passes, which makes error in estimating bias larger. By incorporating manual calibration in the operation steps of the proposed method, we enable to keep the sensor biases small enough for a long time. To the best of our knowledge, our

method is the first method that can maintain the bias of sensors within small values.

3 PROPOSED METHOD

3.1 Overview of the Proposed Method

In this study, we propose an automatic calibration approach to maintain many sensors' accuracy for a long time. We use a Bayesian hierarchical model and the MCMC method to estimate the amount of bias as a probability distribution. Then, we apply an affine transformation to the estimated results to obtain a more accurate probability distribution. Afterward, we specify the sensors that need calibration based on the probability distribution and correct those sensors. We repeat this process periodically. Consequently, we can maintain many sensors' accuracy in a long term.

3.2 Proposed Model

We use an error estimation model based on the Bayesian hierarchical model to estimate the bias. We assume that the temperature sensors are densely deployed in a flat area and calibrate the sensors periodically. Let S be a set of the sensors. The measured value of a sensor $s \in S$ at time instant t , the ground truth value of temperature at the location of the sensor s , and the sensor s ' bias are given by

$$y_s^t = x_s^t + b_s^t + e_1, \quad (1)$$

where y_s^t , x_s^t and b_s^t are the measured value, the true value and the bias of the sensor s at time instant t respectively, and e_1 is the measurement noise following a normal probability distribution. We estimate the true value and the bias at each time point from the measured values at those time points based on Eq. (1). However, in that case, accurately estimating the bias is difficult because of lacking relation among each measurement. Thus, our model assumes that the ground truth temperature values and biases have spatio-temporal correlations to estimate the highly accurate probability distributions.

First, we consider temporal correlations of bias. We model the relationship of the biases at time instant t and $t - 1$ as follows.

$$b_s^t = b_s^{t-1} + d, \quad (2)$$

where d is the difference between the biases at time instant t and $t - 1$ and is a small value following a normal probability distribution. Eq. (2) enables us to constraint the value that bias can take.

Additionally, we consider spatio-temporal correlations of the true value. The true value changes over time when using temperature sensors for measurement environment. Thus, we model the true value as follows.

$$x_s^t = x_s^{t-1} + r_t + e_2, \quad (3)$$

where r_t is the inclination of the temperature change at time t and e_2 is the random noise following the normal distribution. The inclination of temperature change means the amount of change in the true value over a unit of time which depends on time in a day. For example, the inclination of temperature

change r_t is 0.5 if x_s^{t-1} , the true value at time instant $t - 1$, is 20.0, and x_s^t , the true value at time instant t , is 20.5. Use of r_t enables us to consider the spatio-temporal effects among true values in time series.

We also define x_s^t assuming that the true values of sensors located in the neighborhood are highly correlated. We use the ICAR model to consider the spatial correlation. The ICAR model is applied to the true value at the location of sensor s as follows.

$$x_s^t \sim \mathcal{N}\left(\frac{\sum_{j \in N_s} w_j x_j^t}{\sum_{j \in N_s} w_j}, \frac{\tau^2}{\sum_{j \in N_s} w_j}\right), \quad (4)$$

where N_s is the set of neighboring sensors of sensor s , w_j is a weight of the neighboring sensor j and τ is variance. Also, \mathcal{N} is the normal distribution. The expectation of x_s^t is given as the weighted average of the neighboring sensors and thus the true values are correlated among the neighboring sensors.

3.3 Sensor Accuracy Management Method

In the proposed method, we estimate the amount of bias as a probability distribution and manually calibrate a part of the sensors based on the estimated results. Subsequently, we apply an affine transformation to the estimated results to obtain a more accurate probability distribution with a correct scale. Finally, we calibrate the sensors based on the estimated bias with a correct scale. We show the procedure of our accuracy management method in the following.

1. We estimate the posterior probability of all parameters that are the bias b_s^t , the true value x_s^t , the value of temperature change r_t , the difference d , and the noise e_1 and e_2 by applying the MCMC method to the proposed model.
2. We choose two sensors that have the maximum and the minimum bias based on the posterior probability obtained in Step 1, and calibrate those sensors manually. Then, we get to know the true biases of those two sensors.
3. We apply an affine transformation to the posterior probability obtained in Step 1, based on the true biases obtained in Step 2.
4. We additionally calibrate the sensors if the 90% confidence interval of the posterior probability obtained as a result of the affine transformation exceeds the acceptable bias range $[-0.6, 0.6]$.

In Step 1, We estimate the posterior probability of all parameters such as the bias b_s^t and the true value x_s^t to apply the MCMC method to the proposed model. We regard the value of the highest posterior probability as the estimated value of each parameter.

Here, we have a problem that the estimated posterior probability loses the scale information because the value has a high degree of freedom. Fig. 2 shows an example of the relationship between the true bias and the estimated bias when scale information is lost. The horizontal axis represents the true

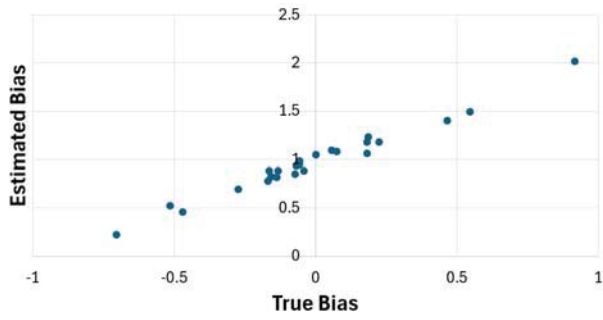


Figure 2: Example of the Relationship between True Bias and Uncorrected Bias Estimation

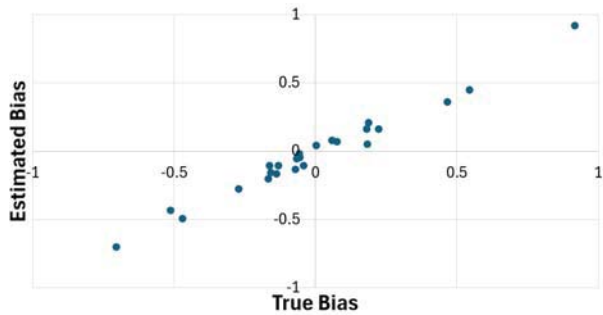


Figure 3: Example of the Relationship between True Bias and Corrected Bias Estimation

bias and the vertical axis represents the estimated bias. In the case of Fig. 2, the true bias is in interval $[-1, 1]$, but the estimated bias is in interval $[0, 2.5]$. When scale information is lost, we can not estimate the bias appropriately.

Thus, we need to correct the posterior probability to obtain a more accurate posterior probability. In this study, we apply an affine transformation which performs scaling and translation to the posterior probability. Before applying the affine transformation, in Step 2, we calibrate two sensors manually that have the maximum and the minimum bias based on the posterior probability in Step 1. Then, we know the true bias of the two calibrated sensors. In Step 3, we correct the posterior probability by using the true bias. When \hat{b}_s represents the corrected bias estimation and b_s represents the uncorrected bias estimation, the uncorrected bias estimation is corrected by Eq. (5).

$$\hat{b}_s = p \times b_s + q, \quad (5)$$

where p and q are parameters representing the degree of scaling and translation. We can determine the optimal values for p and q by using the true bias that we know in Step 2. Then, we can correct all sensors that are not calibrated manually. In this way, we correct the estimated bias. The corrected relationship between the corrected bias estimation and the true bias is shown in Fig. 3. The uncorrected bias estimation in Fig. 2 is in the interval $[0, 2.5]$, while the corrected bias estimation in Fig. 3 is in the interval $[-1, 1]$ which is the same interval of true bias. By the affine transformation described above, We can estimate the bias more accurately.

Finally, in Step 4, We calibrate other sensors additionally based on the corrected bias estimation. By defining the range of acceptable bias values as the interval $[-0.6, 0.6]$, we choose

the sensors that need calibration. If the 90% confidence interval of the corrected posterior probability exceeds the acceptable range $[-0.6, 0.6]$, We additionally calibrate the sensor.

By performing these four steps at regular intervals, such as annually or monthly, we can maintain the bias of all sensors within the acceptable range with a sufficiently high probability. Also, comparing the proposed method to the traditional method which calibrates all sensors manually, the proposed method can maintain the measurement accuracy of sensors within a certain range with less labor and cost because we calibrate only the sensors that exceed the acceptable range based on the confidence interval.

4 CONCLUSION

In this study, we propose the automatic calibration method that can maintain the bias of sensors within small values for a long time. We estimate the probability distribution of bias by applying the MCMC method to our proposed space-time model within the framework of the Bayesian hierarchical modeling. Then, we additionally calibrate the sensors that need calibration due to large biases based on the estimated probability distribution. Note that we involve manual calibration in the procedures of the proposed method, which enables us to maintain the bias small enough for a long time.

The estimation accuracy of the proposed method is evaluated by comparing it with the methods that exclude key techniques from the proposed method. As a result, we showed that we can estimate bias most accurately and can maintain bias smallest by the proposed method. This means that the key techniques proposed in this paper actually worked, and we achieved a practical automatic calibration method that maintains the bias of sensors within a certain range for a long time by considering the temporal correlation of bias and incorporating manual calibration.

In this study, we assume that sensors are deployed in a mesh pattern. As a future issue, it will be meaningful to apply the proposed method in the case that sensors are deployed in a random coordination. We expect that we can achieve this goal by determining weight in Eq. (4) appropriately depending on the distance between the neighboring sensors.

REFERENCES

- [1] Friha, Othmane, et al. "Internet of things for the future of smart agriculture: A comprehensive survey of emerging technologies." *IEEE/CAA Journal of Automatica Sinica* 8.4 (2021): 718-752.
- [2] Talari, Saber, et al. "A review of smart cities based on the internet of things concept." *Energies* 10.4 (2017): 421.
- [3] Barcelo-Ordinas, Jose M., et al. "Self-calibration methods for Uncontrolled Environments in Sensor Networks: A Reference Survey." *AdHocNetworks*(2019): 142-159.
- [4] T. Yoshihiro, "A Two-step Automatic Calibration Method for Sensor Accuracy Management." 2023 19th International Conference on Intelligent Environments (IE), IEEE, 2023.

- [5] Becnel, Thomas, et al. "A recursive approach to partially blind calibration of a pollution sensor network." 2019 IEEE International Conference on Embedded Software and Systems (ICESS). IEEE, 2019.
- [6] Bychkovskiy, Vladimir, et al. "A collaborative approach to in-place sensor calibration." Information processing in sensor networks. Springer, Berlin, Heidelberg, 2003.
- [7] Balzano, Laura, and Robert Nowak. "Blind calibration of sensor networks." Proceedings of the 6th international conference on Information processing in sensor networks, 2007.
- [8] Sun, Jian, Kai Zhao, and Tao Jiang. "A multipoint correction method for environmental temperature changes in airborne double-antenna microwave radiometers." *Sensors* 14.5, 2014: 7820-7830.
- [9] Moon, Nam-Won, and Yong-Hoon Kim. "Optimized thermal compensation method using clustering and drifted response stability for total power radiometer calibration." *IEEE Sensors Journal* 17.5, 2016: 1269-1276.
- [10] Takruri, Maen and Challa, Subhash, "Drift aware wireless sensor networks," in Proc. of the 10th intl. conference on information fusion, (Quebec City, Canada), 2007
- [11] Takruri, Maen, et al. "Online drift correction in wireless sensor networks using spatio-temporal modeling." 2008 11th International Conference on Information Fusion. IEEE, 2008.
- [12] Kumar, Dheeraj, Sutharshan Rajasegarar, and Marimuthu Palaniswami. "Automatic sensor drift detection and correction using spatial kriging and kalman filtering." 2013 IEEE International Conference on Distributed Computing in Sensor Systems. IEEE, 2013.
- [13] Wu, Jiawen, and Guanghui Li. "Drift calibration using constrained extreme learning machine and Kalman filter in clustered wireless sensor networks." *IEEE Access* 8, 2019: 13078-13085.
- [14] C.E. Rasmussen and C.K.I. "Williams Gaussian Process for Machine Learning." MIT Press, 2006.
- [15] T. Fujino and S. Honda, "An automatic calibration method for networked thermal sensors." SICE Annual Conference 2011, pp. 1528–1532, 2011.
- [16] T. Fujino and S. Honda, "Automatic Calibration of Sensing Systems for Distributed Physical Fields." *SICE Journal of Control, Measurement, and System Integration*, 6:3, 221-229, DOI: 10.9746/jcmsi.6.221, 2013
- [17] The Agro-Meteorological Grid Square Data, Naro, <https://amu.rd.naro.go.jp/wiki_open/>(accessed 2024-07-08).
- [18] Ohno, H., K. Sasaki, G. Ohara, K. Nakazono, "Development of grid square air temperature and precipitation data compiled from observed, forecasted, and climatic normal data. *Climate in Biosphere*", 16, 71-79 (in Japanese with English title), 2016.
- [19] Sasaki, K., M. Nishimori, M. Nemoto, "User's manual for The Agro-Meteorological Grid Square Data, NARO Ver.5". NARO, 77pp (in Japanese), 2022.
- [20] Ministry of Land, Infrastructure, Transport and Tourism: Geospatial Information Authority of Japan, <<https://www.gsi.go.jp/top.html/>>(accessed 2024-07-08).

A Bicycle Caution Spot Detection Architecture using Brainwave and Driving Data

Hayato Tomisu[†], Naoto Kai[‡], and Tomoki Yoshihisa[†]

[†]Graduate School of Data Science, Shiga University, Japan

[‡]University Library, Osaka University, Japan

s6023130@st.shiga-u.ac.jp, kai.n.lib@osaka-u.ac.jp, yoshihisa@biwako.shiga-u.ac.jp

Abstract - The population of sport cyclists has been on the rise. Cyclists often encounter caution spots, which pose significant risks. While experienced cyclists may sometimes avoid accidents with quick identification, overconfidence can lead them to accidents that could have been prevented. Recognizing these caution spots can help reduce overconfidence and promote safer bicycling. This paper presents a novel architecture for caution spot detection using electroencephalography (EEG), driving video, and driving data. The proposed system includes a future EEG prediction dynamic model. Performance evaluation was conducted by comparing models with and without future EEG predictions. We found that models with future EEG predictions have higher accuracy compared to those without, indicating closer detection distances and the ability to predict caution spots. In addition, the architecture achieved 0.8171 in accuracy, 0.2041 in precision, 0.3030 in recall, and 0.2439 in f1-score to detect caution spots.

Keywords: electroencephalography, model updating, wearable device, multimodal, image analysis

1 INTRODUCTION

The use of bicycles as personal mobility is well known globally. Furthermore, bicycles are used not only for transportation, but also for sports. As the population of sport cyclists increases, the number of beginners also increases, necessitating the establishment of strict traffic rules. Ensuring safety in cycling is deeply influenced by societal rules and individual awareness. In Japan, for example, there has been a shift towards stricter regulations and encouraging cycling on roadways rather than pavements, a trend emphasized by stricter enforcement in 2022 [1]. Beginner cyclists, often unfamiliar with cycling on roadways, may encounter unexpected caution spots, leading to slips, falls, and accidents. Especially in the case of sport cycling, the driver is moving at high speed and must make an instantaneous judgment of danger.

On the other hand, experienced cyclists can quickly identify and avoid caution spots, adjusting their riding accordingly. However, amateur cyclists, who may be overconfident in their abilities, are at a higher risk of serious accidents. In fact, Dong et al. found that distraction and overconfidence among bicyclists were associated with the occurrence of serious injury crashes [2]. This situation is not unique to cycling but is also observed in driving [3].

Self-assessment of their abilities is essential for safe cycling and effective traffic education. When focusing on the

actions of an individual cyclist, road and environmental conditions are perceived differently based on their experiences, making it important to address these variations. Monitoring heart rate and facial expressions have been employed to address these subjective aspects. Research using the R-R Interval (RRI) has shown that a smaller RRI indicates higher stress levels [4][5]. However, it is also reduced by physical exertion, complicating its use in sport cycling. In addition, facial expression analysis is limited by the use of helmets and eyewear, making it less applicable for sports cycling[6]. Therefore, new methods that use data from any style of cyclist and situation are needed for caution spot detection.

This paper¹ aims to detect caution spots for sports cycling using electroencephalography (EEG), driving video, and driving data. To achieve this objective, we propose a novel architecture that includes a future prediction component for addressing time-series issues in EEG-based caution spot detection. This paper contributes to the recognition of caution spots after cycling and provides an opportunity for self-assessment. The rest of this paper is organized as follows. Section 2 presents a review of the latest literature related to this paper. In Section 3, we introduce the focused issues and the proposed architecture for caution spot detection using EEG, driving video, and driving data. In Section 4, we explain the implementation of the proposed architecture. In Section 5, we describe experiments conducted to evaluate the proposed method's effectiveness and present the results. Section 6 discusses the results of the performance evaluation. Finally, Section 7 concludes this paper.

2 RELATED WORK

This section reviews several related issues, including EEG analysis for traffic safety and EEG usage for cycling, respectively.

2.1 EEG Analysis for Traffic Safety

Most research on EEG for traffic safety has focused on car drivers rather than cyclists. Saman et al. indicated that many studies have concentrated on detecting, monitoring, and forecasting driver fatigue, investigating cognitive responses and brain dynamics during navigation, obstacle avoidance, and external distractions [7]. Le et al. demonstrated that driver fatigue is a significant cause of traffic accidents, emphasizing the growing interest in fatigue detection systems [8]. Turker et al. reviewed eight fatigue

detection systems and proposed a robust method using single-channel EEG [9].

Some studies have explored EEG changes during emergencies. Li et al. conducted driving simulator experiments with 19 men and 19 women, measuring EEG during pedestrian collision avoidance [10]. They found that specific EEG activities could occur simultaneously in urgent and safety critical situations, suggesting the need for personalized danger assessment based on EEG. Xinran et al. identified that EEG in the alpha/delta, alpha/theta, beta/delta, and beta/theta frequency bands were higher during collisions compared to successful avoidance, with critical decision periods occurring two seconds before braking [11].

2.2 EEG Usage for Cycling

The use of EEG during cycling has distinct characteristics. Szu-Yu et al. investigated how increasing the average maximum heart rate ratio during continuous cycling affects brain activity, revealing that EEG power showed similar increases across all electrodes [12]. Enders et al. also found that EEG power significantly increased with fatigue during exercise [13]. These studies conducted measurements indoors using conductive gel. Siddharth et al. demonstrated that dry EEG systems could be used in outdoor environments with high movement, such as cycling [14]. Zink et al. showed that mobile EEG could reliably acquire data during outdoor cycling, though electrical noise from the rear and baseline periods posed issues [15]. These considerations emphasize the differences between using EEG in cars and bicycles.

Regarding the application of EEG for hazard recognition, Harish et al. focused on the safety of urban cyclists involved in logistics, using EEG, eye-tracking data, and biometric data to identify areas of distraction [16][17]. Chu et al. examined six different cycling activity patterns [18]. They extracted EEG features using discrete wavelet transform, revealing that EEG patterns were related to various cycling activities, especially during multitasking, such as phone use.

These studies primarily address hazard recognition in city cycling, leaving a gap in understanding the unique hazards associated with sports cycling.

3 PROPOSED METHOD

In this section, we first introduce focused important issues on sport cycling. Then, we discuss the proposed bicycle caution spot detection architecture with feature fusion based on integrated EEG, driving video, and driving data.

3.1 Focused Issues

To enhance cyclist safety, it is crucial to detect and avoid caution spots. While some of these spots can be identified through visual cues or image analysis, others are less visible and rely more on psychological or intuitive factors. These specific spots cannot be detected solely through images, necessitating the exploration of alternative detection methods. This study employs EEG to detect caution spots that may not be immediately apparent to cyclists. A multimodal

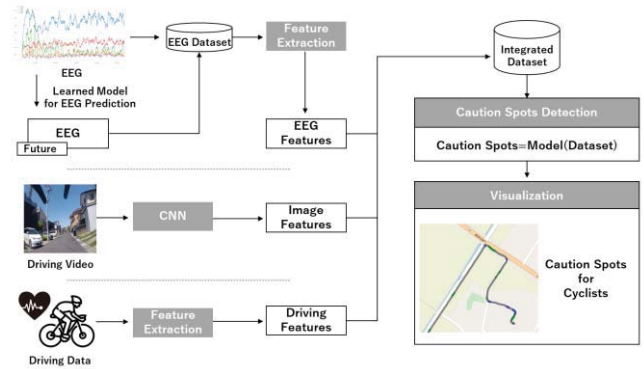


Figure 1: Overview of the proposed architecture

approach incorporating EEG, driving video, and driving data is used to improve prediction performance.

3.2 Overview of The Proposed Architecture

Figure 1 shows the overview of the proposed architecture. In this architecture, features obtained from EEG are used in combination with other data to make caution spot detection decisions. One of the novel proposals considers this and uses the future prediction of the EEG. In addition, feature extraction of images using CNN reveals common features of driving conditions without explicit objects. After that, common driving data is also combined to build an integrated dataset. Finally, using the integrated dataset, caution spot detection is performed.

We will explain the details in the following.

3.3 Future EEG Prediction

Predicting future EEG is important to detect of caution spots. EEG is known to exhibit a lag in response following cognitive recognition. Future predictions can help address this delay. On the other hand, sudden judgments are made reflexively, causing EEG to change faster at caution spots than the conscious response. This consideration is essential in constructing prediction tasks.

A common method for detecting abnormalities of time-series data is the difference from a rolling average. However, using a rolling average that reflects a small window can be overly influenced by peculiar behaviors. Increasing the window size results in an average that reflects only overall movement trends. Although this method can detect reflexive signals, it is unsuitable for detecting foresight signals to notice caution spots. Figure 2 shows the comparison of the pattern of significant signal differences calculated by rolling average or future prediction in simulated EEG.

The upper section of the figure shows the simulated EEG data with an anomaly compared to rolling averages calculated using small (5) and large (50) window sizes. Significant signal differences indicate reflexive actions and are easily understood in regions where the rolling average deviates substantially from the actual data. The lower section of the figure presents the same EEG data with a 2 second future prediction model. The prediction model forecasts caution spots by indicating significant differences

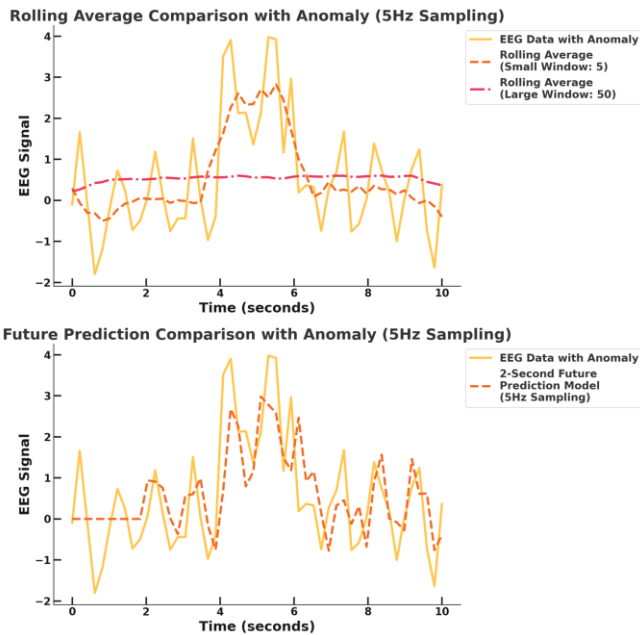


Figure 2: Comparison of the pattern of significant signal differences calculated by rolling average or future prediction in simulated EEG

ahead of time, allowing for proactive measures rather than reflexive reactions. This comparison found the advantage of future prediction models in identifying caution spots by forecasting than the rolling average method. By analyzing these differences, we aim to highlight features of caution spots that are not immediately apparent and contrast those that require sudden, reflexive judgments. The dataset in this figure is simulated and manually created for illustrative purposes and does not represent actual data or calculations.

3.4 Application Example of the Proposed Method

Figure 3 shows the application example of the proposed method. This application example allows users to reflect on their riding result after cycling. Existing private sector methods only provide information on performance, such as speed and distance. By using the proposed method, it is possible to identify individual recognized caution spots. In addition, object-detected caution spots that are not recognized as dangerous by the user are shown simultaneously. Since it is difficult to pinpoint caution spots perfectly, a threshold value for a specific indicator is used to make judgments. This enables cyclists to self-assess their ability to recognize caution spots

4 IMPLEMENTATION

In this section, we discuss the detailed mechanism and implementation of the proposed architecture. In implementation, the experiments were conducted using a computational environment with the following specifications. The system was equipped with an AMD Ryzen7 2700X CPU and an NVIDIA RTX 2070 GPU, supported by 48 GB of RAM. The operating system used

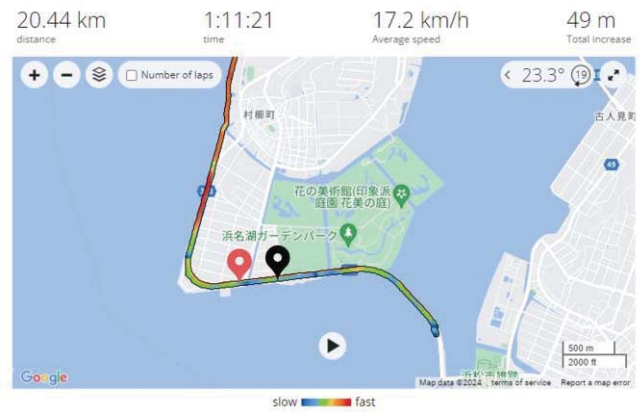


Figure 3: Application example: Cycling result of performance and caution spots' location

Figure 3: Application example: Cycling result of performance and caution spots' location

was Windows 10, version 22H2. For the software environment, Python 3.10.11 was used as the primary programming language, with key libraries and tools including Pandas (version 2.2.2), Numpy (version 1.24.3), scikit-learn (version 1.2.2), and PyTorch (version 2.3.1). To leverage GPU acceleration, the CUDA Toolkit version 12.1.105 was installed and configured.

4.1 Obtaining Data from Sensors

Samples were obtained from three locations: Yasu-shi in Shiga, and Hachioji-shi in Tokyo from one participant. Each experimental route included various environments such as residential areas, arterial roads, and cycle tourism roads. In the experiment, data was obtained at 5 minute intervals. In total, ten datasets were obtained for the experiment. Table 1 indicates the datasets' obtained location and usage in this paper. All datasets were obtained during the day from 11:00 am to 3:00 pm, in good weather conditions.

We used FocusCalm with the GoodBrain application to capture EEG. This application could obtain the timestamp and raw EEG. Also, driving data were obtained from Garmin Forerunner 255 as a cycling computer. We obtained heart rate, enhanced speed, enhanced altitude, latitude, and longitude. In addition, driving videos were captured by a GoPro HERO5 Black.

All the devices were connected to the rooted Pixel 4 and synchronized time by NTP just before the experiment. Figures 4 show the mounting position of devices.

4.2 Training Model Using Open EEG Dataset for Prediction

We used the "OPM-MEG, SQUID-MEG, and EEG (OSE)" dataset provided by Okito et al. to train a time-series prediction model for EEG. This dataset included EEG recordings while moving their bodies in a stable situation. The dataset records EEG from 64 channels. However, our datasets only obtained EEG from 3 channels. To

Table 1: The Datasets' Obtained Location and Usage

No.	Location	Usage
1	Yasu-shi	CNN Training
2	Yasu-shi	CNN Training
3	Yasu-shi	CNN Training
4	Hachioji-shi	Detection Model Training
5	Hachioji-shi	Detection Model Training and Validation



Figure 4: Sensor mounting positions. Left: Camera and smartphone mounted on the bicycle's center. Right: Focus-culm (EEG capture device) worn under the helmet.

accommodate this limitation, we extracted data from channels 61, 62, and 63, which had the same electrode positions as our device. The average value of these channels was used as each band's power.

The trained model aimed to predict EEG 2 seconds into the future. The 2 second prediction was made here to account for the lag between reflexive motor movements, perceptions, and the EEG [11]. We employed a LightGBM[19], using specific parameters: {n_estimators=100, random=42}. The performance of the trained model was evaluated using the Mean Squared Error (MSE) and Mean Average Error (MAE). Table 2 shows each band's prediction performance evaluation and metrics.

After training the model, it predicted the future EEG. In this process, the predicted data on timestamps with ranges not in the existing dataset were cut off.

4.3 Caution Spots Annotation

We annotated the caution spots on the routes traveled to train the classifier model. Participants evaluated the annotations by viewing driving videos taken within 10 minutes and rating them as binary values. The frequency of annotations was 10 [Hz]. Caution spots here were not bound by rules, such as the presence of a specific object, but were annotated based on when the users themselves felt the route was dangerous.

4.4 Feature Extraction

Feature extraction was performed on the respective data obtained from each device according to the following method.

4.4.1 EEG

We used segmented EEG into Delta, Theta, Alpha, Beta, and Gamma bands. The ratios between the Delta, Theta, Alpha, Beta, and Gamma bands, such as the Theta/Beta Ratio (TBR), were also calculated. The 5 and 25 rolling

Table 2: LightGBM Models Performance for Each Band

Band	MSE	MAE	SD
Delta	5.708×10^{-12}	1.897×10^{-6}	2.176×10^{-6}
Theta	2.804×10^{-12}	1.367×10^{-6}	1.562×10^{-6}
Alpha	1.535×10^{-12}	9.742×10^{-7}	1.844×10^{-6}
Beta	2.654×10^{-12}	1.226×10^{-6}	2.205×10^{-6}
Gamma	7.853×10^{-13}	6.900×10^{-7}	1.213×10^{-6}

averages from each EEG band and rate were calculated. This calculation was also adopted to EEG predicted after 2 seconds. After that, we calculated the difference between the original EEG and the predicted EEG and between the original EEG and 5 rolling averages.

4.4.2 Driving Video

The acquired driving video obtained prediction probability by customized MobilenetV3[20]. Figure 5 shows a schematic diagram of the architecture used in this study. EEG and driving data are numerical and have a small data volume, but driving video has a large data volume. This implementation adopted an architecture that can be used even with small computing resources on prediction tasks. This realized feature extraction to be performed without spending a great deal of time on image processing. The model was trained 10 Epochs using a dataset obtained from the area of Yasu-shi, which included 15 minutes of video and annotation information. We employed Google Colab Pro's A100 runtime to train architecture. As a result of the log of training, we employed Epoch 10 trained architecture for feature extraction.

4.4.3 Driving Data

We computed heart rate change, RRI, 5 rolling average of heart rate, these standard deviations, and altitude gain from the driving data. The cycle computer used in this paper obtains data at 1 [Hz], and an RRI was obtained by dividing 60 seconds by the calculated heart rate.

4.5 Multimodal Data Integration

Each data was compared with the timestamp of the EEG and merged based on the data with the closest timestamp. Therefore, the sampling rate of all data was fixed at 5 [Hz] in this process. In addition, driving data obtained less frequently than EEG was the same value transcribed in multiple records. The data in the first 2 seconds were not used in this experiment to employ the data holding the predicted EEG 2 seconds ahead.

4.6 Training Detection Model

Before the main experiment, a preliminary test was conducted to predict caution spots by combining EEG and driving data. We used Random Forest, Support Vector Machine, and Logistic Regression in this preliminary study. Among these, Random Forest demonstrated the best performance. Therefore, we adopted Random Forest as the classifier model for this implementation.

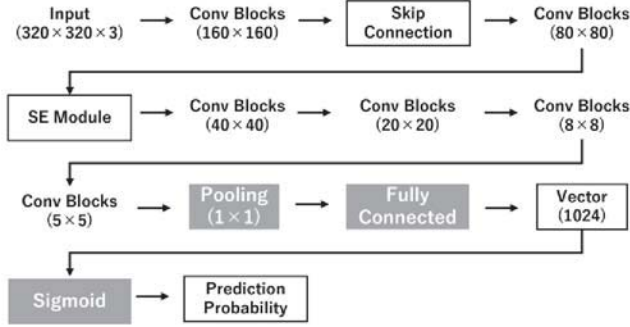


Figure 5: The implemented schematic diagram of the customized MobilenetV3 architecture

We used two 5 minute datasets, totaling 10 minutes of data obtained in Hachioji-shi for initial training. The initial model was trained using the following parameters: {n estimators = 10, max features = "sqrt", max depth = 10, min samples split = 2, min samples leaf = 4, random state=42}. We also implemented time-series-aware cross-validation to validate the model. Cross-validation was performed in three parts, and the model was selected using f1-score as the evaluation index.

5 PERFORMANCE EVALUATION

In this section, we explain the procedure and result of performance evaluation. The objective of this evaluation was to compare the prediction performance with and without predicted future EEG.

5.1 Procedure

We created two integrated datasets to evaluate the effect of including future prediction data: one including predicted future EEG and one without it. We used two 5 minute datasets, totaling 8 minutes of data obtained in Hachioji-shi for training Random Forest, and 2 minutes of data for validation of this model. The performance of each trained model was evaluated using a confusion matrix, accuracy, precision, recall, and f1-score.

5.2 Result

The confusion matrices for both models are presented in Figure 6. The proposed method shows a strong ability to correctly classify normal states, with 267 true negatives. However, it also had a notable number of false negatives, 23 instances where caution states were misclassified as normal. In contrast, the model without future prediction had a lower true negative count, 75, suggesting a less accurate classification of normal states. It also had a high number of false positives, 221, indicating many normal states were incorrectly classified as caution. However, it performed better in detecting caution states with fewer false negatives, only 9.

The performance evaluation result of the model using a dataset with predicted future EEG indicated 0.8171 in accuracy, 0.2041 in precision, 0.3030 in recall, and 0.2439 in f1-

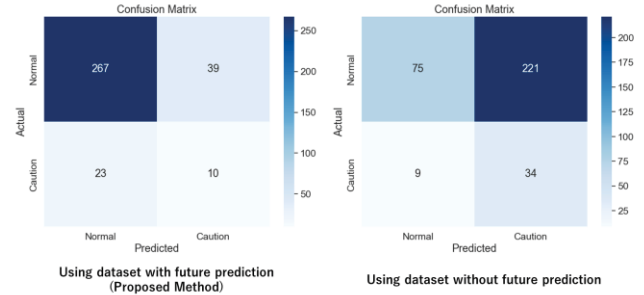


Figure 6: Confusion matrices comparing model performance with and without predicted future data during validation

score. On the other hand, the performance evaluation result of the model using a dataset without predicted future EEG indicated 0.3215 in accuracy, 0.1333 in precision, 0.7907 in recall, and 0.2282 in f1-score.

6 DISCUSSION

In the evaluation, we clarified the characteristics of the predictions made by the proposed method. Training with a dataset with predicted future EEG indicated high accuracy but had low precision and recall. The particularly low recall indicates the possibility of missing many positive cases. On the other hand, training with a dataset without predicted future EEG indicated lower accuracy but very high recall. However, the low precision in this case means that there were many false positives. This difference suggests that the proposed method can offer significantly different insights when reviewing cycling activities to understand potential caution spots.

The limitation of this study is that this experiment was analyzed from data acquired by a single person. In the future, similar experiments can be conducted with multiple users to compare the differences in prediction results or train a generic model using a dataset obtained from multiple users. Also, this study does not encompass different risks that may arise during peak traffic hours, off-peak hours, or in adverse weather conditions such as rain. These factors can significantly affect the risk levels and the corresponding caution spots for cyclists. Conducting experiments during these varied conditions could provide a more comprehensive understanding of the potential risks and improve the robustness of the proposed method.

7 CONCLUSION

This paper introduced a novel architecture to detect caution spots for sports cycling using EEG, driving video, and driving data. We found that models with future EEG predictions have higher accuracy compared to those without, indicating closer detection distances and the ability to predict caution spots. The best validation results using the proposed architecture were 0.8171 in accuracy, 0.2041 in precision, 0.3030 in recall, and 0.2439 in f1-score, demonstrating the effectiveness of the future prediction. However, there are some limitations; for example, this experiment was analyzed from data acquired by a single person. Future research will focus on developing and

implementing an architecture adopted by multiple users and locations. It will contribute to recognizing caution spots after cycling and make an opportunity for self-assessment.

Acknowledgement

The data using for EEG prediction model was obtained from the OSE dataset (<https://vbmeg.atr.jp/nictitaku209/>).

REFERENCES

- [1] Cabinet Office, Government of Japan, “Promotion of Safe Bicycle Use,” Cabinet Office. https://www8.cao.go.jp/koutu/taisaku/bicycle/bicycle_r04.html (accessed May 29, 2024).
- [2] X. Dong, D. Zhang, C. Wang, and T. Zhang, “Analysis of factors influencing the degree of accidental injury of bicycle riders considering data heterogeneity and imbalance,” *PLoS One*, vol. 19, no. 5, p. e0301293, May 2024.
- [3] R. W. Wohleber and G. Matthews, “Multiple facets of overconfidence: Implications for driving safety,” *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 43, pp. 265-278, ISSN 1369-8478, Nov. 2016.
- [4] K. Tatsuno, “An Experimental Study on Level of Stress for Bicycle Facilities by Using Vital Reaction According to Driving Tasks,” Master dissertation, Graduate School of Engineering, Osaka City University, Osaka, Osaka, Japan, 2019.
- [5] M. Taniyama, M. Arakida, R. Yamashita, R. Hashimoto Koichi, S. Okubo, and I. Kai, “Structure of health characteristics that affect welfare recipients’ willingness to participate in workfare programs,” *Journal of Japan Academy of Nursing Science*, vol. 38, pp. 263–273, Feb. 2019.
- [6] R. Yamaguchi, P. Siriaraya, T. Yoshihisa, S. Shimojo, and Y. Kawai, “A Detection System for Comfortable Locations Based on Facial Expression Analysis While Riding Bicycles,” in *Companion Proceedings of the ACM Web Conference 2023*, in *WWW ’23 Companion*. New York, NY, USA: Association for Computing Machinery, Apr. 2023, pp. 306–309.
- [7] S. Sargolzaei, S. Hatchett, M. Okashita, and A. Sargolzaei, “Decoding Driving Neural Dynamics: An Exploratory Survey of EEG Applications in Unraveling Cognitive Processes in Virtual Reality Environments,” in *SoutheastCon 2024*, IEEE, pp. 830–836, Mar. 2024.
- [8] L. He, L. Zhang, X. Lin, and Y. Qin, “A novel deep-learning model based on τ -shaped convolutional network (τ Net) with long short-term memory (LSTM) for physiological fatigue detection from EEG and EOG signals,” *Medical & Biological Engineering & Computing*, vol. 62, no. 6, pp. 1781–1793, June 2024.
- [9] T. Tuncer, S. Dogan, and A. Subasi, “EEG-based driving fatigue detection using multilevel feature extraction and iterative hybrid feature selection,” *Biomed. Signal Process. Control*, vol. 68, p. 102591, July 2021.
- [10] X. Li, L. Yang, and X. Yan, “An exploratory study of drivers’ EEG response during emergent collision avoidance,” *Journal of Safety Research*, vol. 82, pp. 241–250, Sep. 2022.
- [11] X. Zhang, X. Yan, J. Stylli, and M. L. Platt, “Exploring the effects of EEG signals on collision cases happening in the process of young drivers’ braking,” *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 80, pp. 381–398, July 2021.
- [12] S. Y. Lin, C. W. Jao, P. S. Wang, and Y. T. Wu, “Analysis of Electroencephalography Alteration During Sustained Cycling Exercise Using Power Spectrum and Fuzzy Entropy,” *International Journal of Fuzzy Systems*, vol. 19, no. 2, pp. 580–590, Apr. 2017.
- [13] H. Enders, F. Cortese, C. Maurer, J. Baltich, A. B. Protzner, and B. M. Nigg, “Changes in cortical activity measured with EEG during a high-intensity cycling exercise,” *Journal of Neurophysiology*, vol. 115, no. 1, pp. 379–388, Jan. 2016.
- [14] S. Kohli and A. J. Casson, “Towards out-of-the-lab EEG in uncontrolled environments: Feasibility study of dry EEG recordings during exercise bike riding,” in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE, pp. 1025–1028, Aug. 2015.
- [15] R. Zink, B. Hunyadi, S. Van Huffel, and M. D. Vos, “Mobile EEG on the bike: disentangling attentional and physical contributions to auditory attention tasks,” *Journal of Neural Engineering*, vol. 13, no. 4, 046017, Aug. 2016.
- [16] H. Moturu, J. Schering, and J. M. Gómez, “Visualization of Attention Distraction Points obtained by Modelling and Analysis of an Intelligent Bicycle Assistant System,” *Procedia Computer Science*, vol. 219, pp. 987–998, Jan. 2023.
- [17] H. Moturu, J. M. Gómez, and J. Schering, “A New Approach to Determine Cycling KPIs in the Context of Behavioral Data,” in *INFORMATIK 2022*, Gesellschaft für Informatik, Bonn, pp. 763–770, Sep. 2022.
- [18] W. L. Chu, C. J. Lin, and P. W. Hsiao, “Using Encephalography to Classify Mind States Associated with Cycling Activities,” *Sensors and Materials*, vol. 32, no. 1, pp. 387-406, 2020.
- [19] G. Ke et al., “LightGBM: A highly efficient Gradient Boosting Decision Tree,” in *Proceedings of Advances in Neural Information Processing Systems 30 (NIPS 2017)*, Long Beach, CA, USA, pp. 3146–3154, Dec. 2017.
- [20] A. Howard et al., “Searching for MobileNetV3,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision 2019 (ICCV)*, Seoul, Korea, pp. 1314–1324, Nov. 2019.

A Social Sensor Simulation Platform for Sensor Evaluation and Tuning

Kei Hiroi[†], Akihito Kohiga[‡], Sho Fukaya, and Yoichi Shinoda[‡]

[†]Disaster Prevention Research Institute, Kyoto University, Japan

[‡]Japan Advanced Institute of Science and Technology,
Japan Suwa University of Science, Japan
hiroii@dimisis.dpri.kyoto-u.ac.jp

Abstract - Social Sensing Data (SSD) posted by residents of disaster-stricken areas during a disaster holds great potential for flood damage assessment. It is particularly promising for use in areas where installation, operation, and maintenance of observation equipment are difficult or where such equipment is in short supply. We are developing a social sensor tuning function to effectively utilize SSD for identifying safe evacuation routes and timing during urban flood events. In this paper, as part of the social sensor tuning function, we examine methods for determining activity levels that indicate the density of SSD necessary for flood assessment, aiming to effectively utilize SSD. This determination method consists of four main processes. The function analyzes the relationship between SSD density and data assimilation accuracy, introducing the concept of activity levels to enable effective data collection and high-precision flood assessment. Simulations using actual flood cases demonstrate that increasing SSD density under specific conditions leads to improved accuracy, verifying the effectiveness of the proposed function. We also examine several challenges for practical application. This research provides important insights for pre-verification of flood assessment utilizing SSD.

Keywords: flood estimation, state-space model, temporal-spatial analysis, data assimilation

1 INTRODUCTION

There are concerns that flood risks are intensifying on a global scale. The Fifth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC) states that global warming is gradually progressing, and it is likely that the frequency and intensity of rainfall will change accordingly [1]. Many regions worldwide are experiencing an increase in the frequency and intensity of heavy rain and flooding [2]-[3]. Among flood damage countermeasures, observing rainfall, rivers, and flooding, and understanding the changing situations of rainfall and rivers and their influences enables us to determine what actions should be taken and to take effective measures to prevent or mitigate damage. Many previous studies have attempted to estimate flood risks using regional vulnerability. For example, [4] estimates urban flood risk with a detailed spatial resolution of about 2 meters. [5]-[6] conduct research to estimate index-based flood risk using theoretical hydraulic engineering models. Furthermore, a Chinese case in [7]

examines risk perception during the 1997 Red River flood situation. Active research is being conducted to accurately analyze risks and lead to mitigation behaviors by presenting risks to people in affected areas and raising awareness of individual flood risks [8]-[9].

However, these previous studies are not temporal estimation methods but static estimation approaches for calculating maximum water levels. Static estimation results are risk estimates where risk values may change due to rainfall fluctuations. Considering evacuation behavior, dynamic risk estimation is necessary because flood situations change very rapidly due to complex flooding phenomena in urban areas caused by water overflowing from small rivers and waterways, and rainwater that cannot be completely drained. Therefore, to understand risk with high temporal resolution as a guide for evacuation behavior, it is necessary to calculate high spatio-temporal flood levels that fluctuate according to rainfall conditions.

Our research goal is to achieve high-accuracy flood assessment by integrating static estimation approaches for calculating maximum water levels and data for understanding dynamic risks. This compensates flood analysis simulations, which are static estimation approaches, using dynamic data such as time-series data obtained from observation devices and social sensing data (SSD). SSD is effective in assessing flood levels even in places where it is difficult to install, operate, and manage observation equipment. Although there have been many studies on flood damage detection using SSD, its effectiveness has not been clearly established, and the amount and content of collected data vary depending on flood damage cases. However, the greatest advantage of SSD is that it has the potential to obtain data from diverse locations without the need to install monitoring devices. We are developing a compensation method using monitoring data and SSD to take advantage of this benefit. We believe that if this compensation method is realized, it will contribute to improving the accuracy of flood assessment and enable a system for pre-verification of information services based on flood assessment.

For compensation, we use the state-space model (SSM) used in our previous research. By applying SSM to flood analysis simulation results with the addition of monitoring data and SSD, we calculate the flooding situation for the entire affected area. What we are developing this time is social sensor tuning that enables effective utilization of SSD. While SSD has many advantages, it also has the disadvantage that necessary data may not always be obtained. However, given the recent advancements in

information services, it is possible to request necessary data from users. Therefore, we consider a function to evaluate and tune how much data is needed to

meet the required accuracy by setting an activity level based on the relationship between SSD density and flood estimation accuracy. Activity level refers to the density of collected SSD considering the time series. Social sensor tuning first analyzes the relationship between SSD density and flood estimation accuracy to define the activity level. Next, considering temporal elements, simulations are performed with different SSD densities and posting times to quantitatively evaluate the impact on accuracy. Finally, based on these analysis results, the quantity and quality of SSD needed to meet the required accuracy are determined, and parameters for effective data collection are set. If we can set the necessary activity level, we will be able to determine optimal SSD collection parameters. In this paper, we investigate whether tuning of SSD is possible in flood assessment and examine how to set activity levels to calculate the necessary amount of SSD.

2 RELATED WORK

2.1 Flood Monitoring

Traditional river sensor technology has succeeded in detecting signs of disaster in large-scale rivers [10], [11]. While these sensors have the advantage of stable monitoring, they also come with significant limitations in terms of installation. Specifically, they require very large equipment, incur high installation costs of several million dollars, and demand complicated preconfiguration. However, if these installation limitations could be improved, it would enable the possibility of installing a large number of sensors, leading to more reliable detection through the implementation of higher-resolution monitoring sensors. In the field of flood prediction, advanced forecasting methods such as hydrological techniques [12] and artificial neural networks [13], [14] have been proposed. These methods have resulted in highly precise predictions of rising river levels from the perspective of large-scale river analysis. However, these conventional methods have a significant limitation: they cannot predict the flooding of smaller rivers and waterways. The main reason for this limitation lies in the elements required for complex water flow prediction. Forecasting floods in smaller rivers and waterways necessitates analyzing the complicated relationships among multiple confluent rivers while also factoring in the impact of rainfall dynamics. Due to the intricate interplay of these factors, traditional methods are currently insufficient for adequate prediction.

2.2 Risk Estimation with Higher Spatial Resolution

Various studies have already been conducted with the aim of generating information about dangerous areas. Case studies such as [15]-[16] have focused on research presenting

risks on maps. For instance, Sinnakaudan et al. [15] developed an ArcView GIS extension, providing an efficient and interactive spatial decision support tool for flood risk analysis. This extension can analyze water surface profiles of the Pari River and create associated flood maps within ArcView GIS. Another example of GIS-based flood risk assessment is the study by Lyu et al. [16] on the vulnerability of the Guangzhou metro

system. Using the flood event that occurred in Guangzhou on May 10, 2016, they revealed the vulnerability of several metro stations. These studies demonstrate the importance of risk assessment focused on specific areas or facilities.

Meanwhile, research proposing more detailed modeling methods has also been conducted. Studies such as [4],[17]-,[18] have proposed modeling techniques using strictly collected input data. Particularly noteworthy is the research by Ernst et al. [4], which presented a microscale flood risk analysis procedure with a high resolution of 2-meter grid. This method is based on detailed 2D inundation modeling and high-resolution topographic and landuse databases. However, these detailed risk estimation methods also face challenges. They require detailed data measurements, such as laser altimeter data, making it impractical to measure such high-precision data in all areas. This highlights the difficulty in balancing high-precision risk assessment with wide-ranging applicability.

2.3 Flood Detection through Social Networking Services

Social networking services are gaining attention as a new method for gathering information about flooded areas. According to research by Kim et al. [19], social networking is the fourth most popular source for accessing emergency information. They applied social network analysis to the 2016 Louisiana flood, converting emergency SSD into useful knowledge. The aim of this study was to support the development of social media operation strategies for disaster mitigation plans, exploring the interaction patterns between online users and disaster responses.

Development of disaster monitoring systems utilizing SSD is also progressing. Sufi et al. [20] designed a system to monitor disaster-related social media feeds using AI and NLP-based sentiment analysis. This system is reported to have an average accuracy of 0.05 and can indicate potential disaster locations with an average accuracy of 0.93. Additionally, Teodorescu [21] proposed an SSD analysis method for disaster prediction and relief and mitigation measures. This method aims to establish correlations between disaster characteristics and SSD responses.

However, while studies using SSD have been applied to many flood damage cases, there are also some challenges. SSD are not always posted as researchers expect, and the high accuracy reported in these studies may not always be achieved. This suggests that when utilizing SSD, it is necessary to consider its limitations and uncertainties.

2.4 Issues and Approaches

Understanding road conditions in urban areas is crucial for identifying safe evacuation routes. Currently, flood damage assessment relies on two primary methods: numerical simulations (e.g., flood analysis) and monitoring using low-resolution ground observation data (precipitation and river water levels). Numerical simulations are based on differential equations of urban flood flow for given precipitation amounts. These simulations calculate maximum flood levels for specific detailed areas (e.g., 10m grids). Based on these calculations, authorities publish information about areas expected to be at risk during heavy rainfall. However, these analyses use idealized models with fixed parameters such as precipitation amount, runoff coefficient, and drainage outflow conditions. Consequently, in urban areas with complex rainfall distributions and land use patterns, there can be significant discrepancies between analytical results and actual flood levels. This discrepancy can lead to road flooding occurring before warnings or evacuation information are issued, resulting in potential damage.

On the other hand, monitoring sets danger level thresholds at specific locations where road underpasses or river breaches are of concern. This method includes situation monitoring to detect flooding based on observational data. While this approach can easily assess actual damage, it is limited by the number of observation points. SSD (Social Sensing Data) is expected to overcome these monitoring limitations. As mentioned in the previous section, the importance of SSD in flood damage detection has been recognized for some time and has been applied to many flood damage cases. However, there is a fundamental problem with using SSD for water damage detection. SSD is not necessarily posted in all cases. While it may function effectively in floods with a high number of posts, it is likely to be less accurate than reported in floods with fewer posts. Particularly in severely damaged areas, it can be difficult to post while ensuring safety, and communication problems may prevent posting altogether. We are convinced that these issues are obstacles to the effective use of SSD in flood damage detection.

We are developing a system to estimate flooding through data assimilation to identify safe evacuation routes and timing during urban flood events. The distinctive feature of our system is its ability to utilize Flood Monitoring, risk estimation, and SSD collectively, compensating for each other's shortcomings to enable high-precision assessments. Previously, we developed a method to complement flood analysis simulations using monitoring data, proposing a way to estimate the expansion process of flooding. We demonstrated its effectiveness in time-series flood assessments [22]. Furthermore, we extended this approach to investigate how SSD could contribute to flood assessments, examining the effectiveness of SSD in terms of the number of posts, their content, and posting timing [23]. Our findings revealed that while SSD requires careful handling due to potential errors, its utilization can indeed contribute to flood assessments. At this stage, our experiments assumed that SSD could be collected at speci-

fied densities. However, considering real-world scenarios, we identified the need to account for variations in the density of collected SSD as a challenge for future research.

In this paper, we advance the development of our system to the next stage. We tackle the challenge of developing a "social sensor tuning" function for our social sensing simulator that generates SSD. SSD has the potential to gather information from a wide area without the need for installing sensor devices. However, the density of this data can be unevenly distributed. If SSD is not posted from areas where we need to assess flood conditions, flood assessments using SSD may not function as we expect. Therefore, understanding the accuracy of flood assessments in areas where we need situational awareness is crucial for effective flood assessment. If we can determine the accuracy of flood assessments for detailed areas, it may become possible to encourage data collection in those specific regions. (Of course, we recognize the need for safety considerations. We do not recommend posting SSD from dangerous areas.)

In this paper, we explore a tuning model for social sensor tuning to accurately calculate activity levels. Activity level refers to the density of collected SSD, considering the time series. By incorporating social sensor tuning that accurately calculates activity levels into our system, we anticipate the following benefits:

- Social sensor tuning, when functioning effectively, will enable advance planning for flood assessments. For instance, our system could request SSD collection about an hour before expected rainfall. This approach offers the advantage of obtaining timely flood assessment values.
- This development allows us to adapt the concept of semi-active social sensing. Most social sensors are passive in nature, meaning we must find ways to effectively utilize data without knowing who will write what. However, when social sensor tuning functions well, social sensors can adopt a semi-active nature. We can stimulate sensors in advance by providing requests for specific types of data at designated times.
- This shift towards semi-active social sensing represents a significant advancement in our approach to flood assessment. It allows for more targeted and timely data collection, potentially improving the accuracy and relevance of our assessments. By proactively engaging with social sensors, we can better guide the data collection process to meet specific assessment needs.

3 PROPOSAL OF SENSOR EVALUATION AND TUNING USING SOCIAL SENSOR SIMULATOR

3.1 Overview

This research proposes a system to estimate flooding through data assimilation, aimed at identifying safe evacuation routes and timing during urban flood events. We present a flood estimation system based on data assimilation. This system seeks to improve the accuracy of flood estimation by assimilating data using observational time-series data and SSD. Furthermore, we aim to enhance accuracy by encouraging system users to provide information that compensates for the lack of observation points. Our system leverages the advantages of both observational methods, which can provide accurate time-series data, and SSD, which can easily provide large amounts of data based on the number of posted messages. This dual approach allows for a more comprehensive and accurate flood estimation. By combining the precision of traditional observational data with the volume and breadth of SSD, we can create a more robust and responsive system for urban flood management. The integration of user-provided information further enhances the system's adaptability and coverage, potentially filling gaps in areas where traditional observations may be limited or unavailable.

Phase 1: SSD Validation requires verification of the effectiveness of SSD in improving the accuracy of flood assessments. While some previous studies have shown SSD to be effective in disaster damage assessment, these are empirical validations based on data collected during specific disasters. We cannot rule out the possibility that highly effective data were simply collected by chance. There has been little verification of what types of SSD posted during disasters are effective or ineffective. Therefore, a scientific validation of the relationship between SSD and analysis is necessary. Addressing this point, we investigated in Reference [23] whether SSD containing various errors is effective in improving the accuracy of data assimilation and flood estimation. This systematic approach to validating SSD is crucial for establishing its reliability in flood assessment. By scientifically examining the effectiveness of different types of SSD, including those with errors, we can develop more robust methods for integrating this data into our flood estimation models. This research contributes to a more nuanced understanding of how SSD can be effectively utilized in disaster management systems.

Phase 2: Flood Assessment and Promotion Requirement Validation is a more detailed investigation compared to the validation conducted in Phase 1. This phase involves examining the flooding conditions of various past flood events and the SSD posted during these events. Concurrently, we simulate SSD based on the relationships between past SSD and physical measurements of events, derived from the learning process. Furthermore, our social sensing simulator generates numerical flooding data under disaster occurrence scenarios and produces SSD based on the measured data. Throughout Phase 2, the effectiveness of SSD in flood

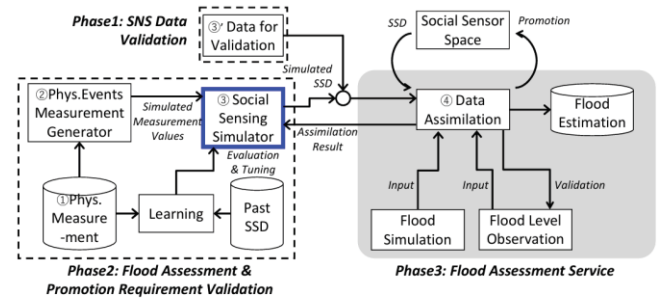


Figure 1: Proposed System for Flood Assessment

assessment is validated for as many flood disasters as possible. Additionally, the system develops dissemination requirements to determine what kind of SSD is necessary to improve the accuracy of the assessment. This comprehensive approach allows us to refine our understanding of SSD's role in flood assessment and to develop more targeted strategies for data collection and utilization. By simulating various scenarios and analyzing a wide range of historical data, we can identify patterns and correlations that might not be apparent in single-event studies. This process enhances our ability to predict the most valuable types of SSD for different flood situations, potentially leading to more efficient and effective flood management strategies.

Phase 3: Flood Assessment is illustrated on the right side of Figure 1. In Phase 2, we accumulate knowledge about SSD that is useful for flood assessment. Our system then uses this knowledge to disseminate desirable data to the SSD space. In Phase 3, we conduct high-precision flood assessments using the collected data. This assessment includes estimating flood levels through data assimilation. In the data assimilation process, the estimated water levels analyzed by flood simulations are compensated using a state-space model with observational data and SSD. The results of data assimilation are compared with observational data and SSD. In flood affected areas where low accuracy is detected, SSD is acquired through commercial services and other means. These results can also be used to improve the data assimilation process. We are planning to develop a dissemination function for the SSD space that encourages SSD posting in areas with insufficient data. By providing SSD from the SSD space, we can update the data assimilation results and improve accuracy. This comprehensive approach integrates simulation, real-time data collection, and adaptive improvement strategies. By continuously refining our models and encouraging targeted data collection, we can create a more responsive and accurate flood assessment system. This system has the potential to significantly enhance flood management and evacuation planning in urban areas.

3.2 Intended Use of the Proposed System

We envision two primary uses for our system. The first is the pre-verification of information services based on flood assessment. The need for providing flood situation information is increasingly crucial. However, even if service

developers wish to develop and offer such services, there are currently no means to verify their usefulness. This is because disasters occur infrequently, and it is not yet clear how people are affected by various flood situations and what kind of SSD they would post. We intend to use our system for such pre-verification purposes. In pre-verification scenarios, our system generates SSD from people in the midst of flooding and provides it as verification data. This approach allows service developers to test and refine their information services under realistic conditions without waiting for actual flood events. By simulating various flood scenarios and the corresponding SSD, our system can provide valuable insights into how different flood situations might impact people and what kind of information they might share. This capability can significantly enhance the development and refinement of flood-related information services, potentially leading to more effective disaster response and management tools.

Phase 2 in Figure 1 illustrates our expected flow of SSD generation for pre-verification. Past SSD collected from previous floods is compared with actual flood damage (①Physical measurement). The ②Physical events measurement generator uses these results to recreate the damage that occurred during the flood as simulated measurement values. Based on these simulated measurements, the ③Social Sensing Simulator generates simulated SSD. Using the simulated SSD collected in this phase, we perform ④data assimilation, combining flood analysis simulations and observational data. This process calculates under what conditions SSD should be collected to improve flood estimation accuracy. The social sensor tuning within the ③Social Sensing Simulator determines whether the accuracy of data assimilation meets the precision required by the information service. If not, the ②Physical events measurement generator and ③Social Sensing Simulator generate additional SSD for that area. In this paper, we develop and evaluate our system for use in pre-verification scenarios. This approach allows for a comprehensive testing and refinement process, enabling the creation of more accurate and reliable flood assessment tools before their deployment in real-world situations.

The second intended use of our system is for flood assessment itself. Although this is not the focus of this paper, it represents our ultimate goal. We aim to identify safe evacuation routes and timing during urban flood events, accurately estimate flooding through data assimilation, and provide this information to the public. The pre-verification method described above can be applied to such a service. Now, let us explain the operational flow of this information provision service when implemented:

In the second use case, the operation of ①Physical measurement and ②Physical events measurement generator is straightforward. ①Physical measurement collects actually posted SSD. This data is then passed to the ②Physical events measurement generator, which removes data containing obvious errors or false reports. The processed SSD is sent through the ③Social Sensing Simulator to ④Data Assimilation, where the assimilation process is executed. Based on the results, the social sensor tuning within the ③Social Sensing Simulator determines

whether the estimation accuracy meets the precision required by local authorities and residents. If the accuracy is insufficient, a request is sent from the ③Social Sensing Simulator to this information service (referred to as Social Sensor Space in Figure 1). The request promotes the areas, quantity, and types of SSD needed. SSD posted in response to the promotion undergoes ④Data Assimilation processing again. By repeating this process or promoting in advance, we aim to utilize SSD as semi-active social sensing, enabling high-precision estimation by timely obtaining necessary data. This approach represents a novel method of integrating real-time social data into flood assessment systems. By actively promoting data collection in specific areas and of specific types, the system can adapt to changing conditions and maintain high accuracy even in complex urban flood scenarios.

3.3 Main Objective of This Paper

Among the three phases necessary for developing our system, this paper focuses on the ③Social Sensing Simulator in Phase 2: Flood Assessment Promotion Requirement Validation (blue box in Figure 1). The purpose of this paper is to develop and test the performance of the social sensor tuning function, a feature of the ③Social Sensing Simulator in Phase 2. Our social sensor tuning sets an activity level indicating the density of SSD that has spatial bias in data collection areas, and calculates the accuracy of the assessment. While SSD is known to be useful for flood assessment using data assimilation, there are issues with how this data is collected. SSD may not always be available from necessary areas, and its accuracy varies by location. By setting an activity level to represent data density, we can achieve good results even in cases of low density by increasing the volume of data generation or collection. This approach addresses a crucial challenge in utilizing SSD for flood assessment: the uneven distribution and varying reliability of data across different areas. By quantifying the activity level, our system can adaptively adjust its data collection and generation strategies to maintain high assessment accuracy even in areas with sparse data. Moreover, this method enables a more nuanced understanding of data quality and coverage, allowing for targeted improvements in areas where data is lacking or unreliable. This capability is essential for creating a robust and adaptable flood assessment system that can perform well across diverse urban environments and flood scenarios.

Our proposed simulation is based on the use of observational data and Social Sensing Data (SSD). Previous research [22] has shown that observational data significantly contributes to the accuracy of flood level compensation. However, there is an upper limit to the number of observation devices due to installation requirements. Therefore, effective utilization of SSD is considered crucial for improving accuracy when devising a system. As mentioned in 2.4, there are numerous studies on flood damage detection using SSD, but most of these analyze flood damage where large amounts of data are available. However, the ability to collect large amounts of SSD depends on the characteristics of the flooded area.

Consequently, it is not always possible to collect SSD that can be used for flood analysis in all flood disasters. Additionally, SSD collected during actual floods varies in terms of information representation and posting timing. Furthermore, the accuracy of much of this data cannot be verified. Thus, to sustainably utilize SSD for disaster management, it is necessary to verify its effectiveness in various cases. This approach acknowledges the challenges and limitations of SSD while recognizing its potential value in flood assessment. By systematically evaluating SSD's effectiveness across different scenarios, we can develop more robust methods for integrating this data into flood management systems. This comprehensive validation process is essential for creating reliable and adaptable flood assessment tools that can perform well even in situations where data quality and quantity may vary significantly.

In Reference [23], we investigated the conditions under which SSD can contribute to flood estimation and how it can improve the accuracy of flood analysis. Specifically, we examined whether variations in information representation and posting timing affect flood assessment, and to what extent these influences are acceptable. Our next consideration is to incorporate a function into our system that evaluates these impacts. This involves measuring the accuracy impact due to SSD bias and determining whether to request additional SSD. This requires further refinement of the approach in Reference [23]. The challenge lies in the fact that the demand for flood assessment accuracy cannot be uniformly defined. To estimate all flood-risk areas with high precision would require requesting high-density SSD from all areas, which is unrealistic in practice. To effectively utilize limited resources, it's more practical to determine which areas need high-precision estimates and what level of accuracy is required, allowing for more targeted requests that are more likely to yield SSD collection. The difficulty in our flood assessment approach lies in dynamically determining which areas need high-precision estimates and what level of accuracy is required. To address this challenge, we consider how to dynamically set activity levels. Achieving this would enable us to operate a stimulation mechanism aimed at obtaining specific activity levels in targeted areas. This dynamic approach represents a significant advancement in flood assessment methodology. By adapting to changing conditions and prioritizing data collection based on real-time needs, we can create a more efficient and effective system for flood management. This method not only optimizes resource utilization but also enhances the system's ability to provide critical information where and when it's most needed during flood events.

4 SOCIAL SENSOR TUNING

The basic concept of social sensor tuning is as follows: This function evaluates the accuracy of compensated flood levels and determines the necessary activity levels. Figure 2 illustrates the data assimilation process where social sensor tuning is applied. Data assimilation is performed using a SSM. Monitoring devices are already installed throughout the city. In data assimilation, time-series data obtained from these monitoring devices is first used to compensate flood

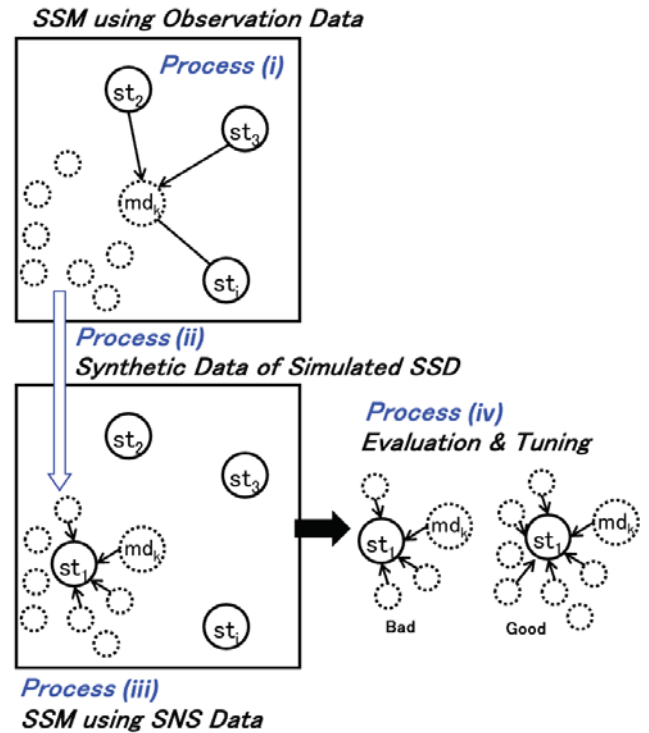


Figure 2: Social Sensor Tuning Processing Flow

analysis simulation results. However, this compensation is insufficient due to the limited number of installed monitoring devices. Data assimilation using SSD attempts to achieve further high accuracy. Social sensor tuning receives the results of this data assimilation. In pre-verification scenarios, social sensor tuning has knowledge of the compensate flood levels. Based on these levels, it calculates the accuracy of data assimilation. If the accuracy is insufficient, it returns an evaluation result to the information service, requesting the areas and quantities of additional SSD needed. This data assimilation uses the spatio-temporal SSM proposed in our previous studies [22] and [23]. The types of SSD and included errors discussed in this paper are explained in [23]. Note that the spatio-temporal state-space model in this paper is applied without using the waterway and sewer data used in the previous study [22], as these data are generally limited in availability. From here, we will explain the processing flow of our social sensor tuning. This approach represents a sophisticated method for dynamically adjusting flood assessment accuracy based on real-time data and predefined accuracy requirements. By integrating both fixed monitoring devices and variable SSD, the system can adapt to changing flood conditions and data availability, potentially leading to more accurate and timely flood assessments in urban areas.

4.1 Process(i) Data Assimilation using Monitoring Data

Initial Process (i): Flood Level Estimation In the initial process (i), we use a SSM to compensate the results of flood analysis simulations with time-series data collected from observation points, estimating flood levels that closely

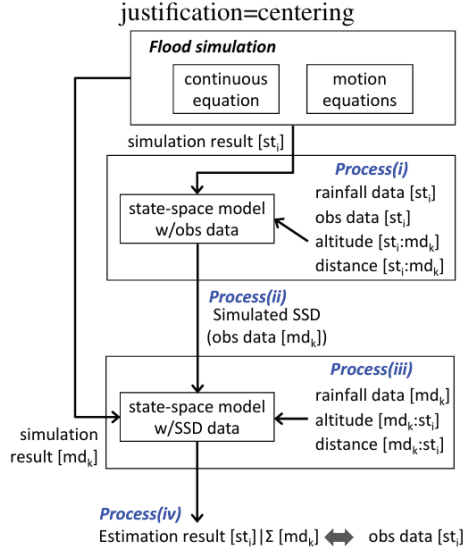


Figure 3: SNS Data Process Flow ($st_{1,\dots,i}$:observation location, md_k :SNS data posting location)

approximate reality. The subsequent process flow is illustrated in Figure 3. This section provides an overview of the SSM used in our data assimilation. The fundamental flood analysis is based on conventional simulations using a surface flooding model. This method calculates the runoff at each grid position by expressing the flood flow through continuity and momentum equations. Our approach integrates advanced statistical techniques with traditional hydraulic modeling to enhance the accuracy of flood level predictions. By assimilating real time observational data into the simulation results, we can dynamically adjust the model to better reflect actual conditions. This process is crucial for capturing the complex and rapidly changing nature of urban flooding. The use of SSM allows for a more nuanced understanding of flood dynamics, accounting for uncertainties and variabilities that static models might miss. This methodology forms the foundation for our subsequent social sensor tuning processes, enabling a more adaptive and responsive flood assessment system.

A continuous equation is defined as follows.

$$\frac{\partial h}{\partial t} + \frac{\partial M}{\partial x} + \frac{\partial N}{\partial y} = 0 \quad (1)$$

The motion equations are given as follows:

$$\frac{\partial M}{\partial t} + \frac{\partial UM}{\partial x} + \frac{\partial VM}{\partial y} + gh \frac{\partial H}{\partial x} + \frac{1}{\rho} \tau_x(b) = 0 \quad (2)$$

$$\frac{\partial N}{\partial t} + \frac{\partial UN}{\partial x} + \frac{\partial VN}{\partial y} + gh \frac{\partial H}{\partial y} + \frac{1}{\rho} \tau_y(b) = 0 \quad (3)$$

Each parameter is defined as t : time, H : water level, h : flood level, U : flow velocity (X direction), V : flow velocity (Y direction), g : gravity acceleration, ρ : water density, M : flux (X direction), and N : flux (Y direction) ($M = uh$, $N = vh$).

Here, the shear force in the x direction $\tau_x(b)$ and the shear force in the y direction $\tau_y(b)$ are defined as follows.

$$\tau_x(b) = \frac{\rho g n^2 \bar{U} \sqrt{U^2 + V^2}}{h^{\frac{1}{3}}} \quad (4)$$

$$\tau_y(b) = \frac{\rho g n^2 \bar{V} \sqrt{U^2 + V^2}}{h^{\frac{1}{3}}} \quad (5)$$

The roughness coefficient n (the resistance value of river water to touch obstacles) can be expressed as follows, considering the influence of a building.

$$n^2 = n_0^2 + 0.020 \times \frac{\theta}{100 - \theta} \times h^{\frac{4}{3}} \quad (6)$$

(n :bottom roughness coefficient, n_0 :composition equivalent roughness coefficient, and θ :building occupancy rate)

Equation (1)-(3) calculates flood level h for each grid, accounting for the runoff from the inside of the sewer line to the ground surface and the flooding to the ground surface due to rainwater. For the above equations, the inflow into each grid represents the flux into each grid from adjacent grids and the effect of buildings on the inflow in each grid.

We define D as the two-dimensional space corresponding to the region of interest and divide D into m grids of d meters each. Let $s_i \in D$ denote the location coordinates of each grid (s_i is denoted by i). Using equation (1)-(3), we calculate $h_t(i)$ for the flood level of each grid at time t .

Then, using the state-space model, we estimate the flood level of grid s_k from the observations $y_t^{(i)}$ collected at the observation location at time t . Grid s_k ($k = 1, 2, 3, \dots, m$) is the location indicated by the SNS data. The state-space model is represented by two types of observation equations; the flood analysis simulation result $h_t(i)$ at grid s_i , and the difference between the flood analysis simulation and the observed value at the observation location. This state-space model is defined by the equations (7)(8)(9).

$$y_t^{(i)} = S_t r_t^{(i)} + G_t^{(i)} x_t^{(i)} + e_t^{(i)} \quad (7)$$

$$r_t^{(i)} = r_{t-1}^{(i)} + v_t^{(i)} \quad (8)$$

$$x_t^{(i)} = x_{t-1}^{(i)} + u_t^{(i)} \quad (9)$$

The $r_t^{(i)}$ denotes the state at time t and $v_t^{(i)}$ denotes noise. The term $G_t^{(i)} x_t^{(i)}$ represents the total inflow/outflow, and $x_t^{(i)}$ is the difference between the flood analysis simulation results and the observed values. The $u_t^{(i)}$ denotes the noise at time t . $G_t^{(i)}$ is the adjacency matrix indicating the spatial component.

4.2 Process (ii) Synthetic Data of Simulated SSD

The next process we need to explain is how we create simulated SSD for use in data assimilation. This simulated SSD can be considered synthetic data related to disaster damage. This is because it essentially involves artificially generating information about flood damage. For pre-verification purposes, we only have access to flood analysis

simulation re-sults and limited monitoring data. From these restricted data sources, we must generate synthetic data that realistically rep-resents potential SSD during a flood event.

Flood analysis simulation results often differ from actual flood levels. Consequently, the only truly accurate data we have access to is from monitoring devices. However, these monitoring points are limited in their spatial distribution. To address this limitation, we create simulated SSD from flood analysis simulation results compensated using monitoring data. Yet, we recognize that compensation based solely on monitor-ing data is insufficient for creating fully appropriate simulated SSD. Therefore, we generate synthetic data anticipating a dis-crepancy between the compensated values using monitoring data and actual flood levels. We assume this difference prob-abilistically falls within the interquartile range and randomly add values within this range to the compensated values. It is important to note that these are merely projected SSD val-ues. However, we do not require high accuracy from SSD. Our goal is to develop a method that produces high-precision flood estimation outputs even when using SSD with lower ac-curacy or containing errors. This approach acknowledges the inherent uncertainties in flood modeling and social sensing data while leveraging the strengths of both. By incorporat-ing this controlled uncertainty into our synthetic data, we can create a more robust and realistic testing environment for our flood assessment system. The key lies in developing data as-similation and social sensor tuning methods that can effectively filter out noise and errors in SSD while still extracting valuable information. This strategy allows us to harness the potential of widespread, albeit imperfect, social sensing data to enhance the accuracy of flood assessments in urban areas.

4.3 Process(iii) Data Assimilation using Simulated SSD

In Process (iii), we use a State-Space Model (SSM) to compensate the results of flood analysis simulations with Simulated Social Sensing Data (SSD) around observation points. This step aims to estimate flood levels that closely approximate reality. The SSD used in this process is the synthetic data generated in Process (ii). The data assimilation technique employed here uses the same equations as the SSM used in Process (i). However, there's a crucial distinction: while Process (i) applied SSM to time-series data, Process (iii) applies it to data at a specific time t .

4.4 Process(iv) Evaluation & Tuning by Social Sensor Simulator

Process (iv) involves social sensor tuning, where we evaluate the accuracy of data assimilation performed in Process (iii). In this paper, we investigate the feasibility of such evaluations by focusing on the relationship between activity levels and compensation accuracy. Activity levels represent the density of SSD in target areas over time. In pre-verification scenarios, data assimilation is repeatedly performed with varying SSD activity levels. Social sensor tuning calculates the required activity level to meet accuracy demands in areas where compensation precision falls short,

considering the relationship between activity levels and compensation accuracy. We examine the feasibility of this verification by comparing flood levels compensated through data assimilation with "correct" levels estimated from monitoring data. It's important to note that simulated SSD for specific locations is not continuous time-series data. Therefore, there is only one time point t for simulated SSD, and the state-space model in Process (iii) is applied only spatially. Consequently, more SSD may sometimes be necessary. In pre-verification, increasing the density of synthetic data is easily achievable. However, when applying this system to actual flood events, we must request SSD from people in affected areas. As the scale of flooding increases, providing SSD becomes more challenging. Therefore, determining acceptable activity levels during pre-verification becomes a crucial process. This step helps us understand

the minimum data requirements for accurate flood assessment while considering the practical limitations of data collection during real disaster events. By systematically evaluating the relationship between activity levels and assessment accuracy, we can optimize system performance, balance the reality of data availability with precision needs during flood events, and enable effective and responsible utilization of SSD in actual flood situations.

5 CONCLUSION

This study proposed a new flood assessment system utilizing Social Sensing Data (SSD) to identify safe evacuation routes and timing during urban flood events. We focused on developing and evaluating the core component of the system: the social sensor tuning function. This function analyzes the relationship between SSD density and data assimilation accuracy, setting optimal activity levels to enable effective data collection and high-precision flood assessment.

Our system consists of four main processes. Process (i) performs initial flood estimation using data assimilation with observational data. Process (ii) generates simulated SSD, which is then used in Process (iii) for further data assimilation. Finally, Process (iv) evaluates results and adjusts parameters through the social sensor tuning function. This sequence of processes enables high-precision flood assessment by integrating conventional observational data with SSD.

A key feature of the social sensor tuning function is its method of utilizing SSD that considers temporal elements. This allows for the evaluation of SSD importance based on posting time, enabling more appropriate data assimilation. By introducing the concept of activity levels, we were able to quantitatively analyze how changes in SSD density in specific areas and time periods affect accuracy. Through simulations using actual flood cases, we demonstrated that increasing SSD density under certain conditions leads to improved accuracy, verifying the effectiveness of our proposed system. This research has demonstrated the potential of an SSD- based flood assessment system centered on the social sensor

tuning function, providing important insights for pre-verification. The proposed four-stage process enables more comprehensive flood assessment by effectively integrating conventional observational data with SSD. These findings are expected to contribute to the development of more reliable flood prediction and assessment systems, significantly contributing to disaster prevention and mitigation in urban areas. Moving forward, it is necessary to conduct simulations considering a wider variety of flood scenarios and regional characteristics

to further improve and validate the proposed system for practical application.

ACKNOWLEDGEMENT

This work was supported by JST FOREST Program, Grant Number JPMJFR226Z, and by National Institute of Information and Communications Technology (NICT), Japan. (Grant Number: JPJ012368C08201)

REFERENCES

- [1] “Intergovernmental Panel on Climate Change Fifth Assessment Report (AR5)”, Retrieved April 17, 2018, from <https://www.ipcc.ch/report/ar5/>.
- [2] P. Milly, D. Christopher, T. R. Wetherald, A. K. Dunne, T. L. Delworth, “Increasing Risk of Great Floods in a Changing Climate, *Nature*, Nature Publishing Group”, Vol.415, No.6871, page514, (2002).
- [3] Y. Hirabayashi, R. Mahendran, S. Koirala, L. Konoshima, D. Yamazaki, S. Watanabe, H. Kim, S. Kanae, “Global Flood Risk under Climate Change”, *Nature Climate Change*, Nature Publishing Group, Vol.3, No.9, page816, (2013).
- [4] J. Ernst, J. B. Dewals, S. Detrembleur, P. Archambeau, S. Erpicum, M. Pirotton, “Micro-scale Flood Risk Analysis based on Detailed 2D Hydraulic Modelling and High Resolution Geographic Data”, *Natural Hazards*, Springer, Vol.55, No.2, pp.181–209, (2010).
- [5] A. S. M. Saudi, I. S. D. Ridzuan, A. Balakrishnan, A. Azid, D. M. A. Shukor, Z. I. Rizman, “New Flood Risk Index in Tropical Area Generated by using SPC Technique”, *Journal of Fundamental and Applied Sciences*, Vol.9, No.4S, pp.828–850, (2017).
- [6] S. F. Silva, M. Martinho, R. Capitão, T. Reis, C. J. Fortes, J. C. Ferreira, “An Index-based Method for Coastal-flood Risk Assessment in Low-lying Areas (Costa de Caparica, Portugal)”, *Ocean & Coastal Management*, Elsevier, Vol.144, pp.90–104, (2017).
- [7] D. H. Burn, “Perceptions of Flood Risk: A Case Study of the Red River Flood of 1997”, *Water Resources Research*, Wiley Online Library, Vol.35, No.11, pp.3451–3458, (1999).
- [8] J. W. Hall, I. C. Meadowcroft, P. B. Sayers, M. E. Bramley, “Integrated Flood Risk Management in England and Wales”, *Natural Hazards Review*, American Society of Civil Engineers, Vol.4, No.3, pp.126–135, (2003).
- [9] R. Miceli, I. Sotgiu, M. Settanni, “Disaster Preparedness and Perception of Flood Risk: A Study in an Alpine Valley in Italy”, *Journal of Environmental Psychology*, Elsevier, Vol.28, No.2, pp.164–173, (2008).
- [10] L. Pulvirentia, M. Chinib, N. Pierdicca, L. Guerrieroc, P. Ferrazzolic, “Flood Monitoring using Multi-temporal COSMO-SkyMed Data: Image Segmentation and Signature Interpretation”, *Remote Sensing of Environment*, Vol.115, No.4, pp.990–1002, (2011).
- [11] E. A. Basha, S. Ravela, D. Rus, “Model-based Monitoring for Early Warning Flood Detection”, In *Proceedings of the 6th ACM conference on Embedded network sensor systems (SenSys)*, pp.295–308, (2008).
- [12] A. Elshorbagy, G. Corzo, S. Srinivasulu, D. Solomatine, “Experimental Investigation of the Predictive Capabilities of Data Driven Modeling Techniques in Hydrology - part 2: Application”, *Hydrology and Earth System Sciences*, Vol.14, pp.1943–1961, (2010).
- [13] A. Rafieinasab, A. Norouzi, S. Kim, H. Habibi, B. Nazari, D. Seo, H. Lee, B. Cosgrove, Z. Cui, “Toward High-resolution Flash Flood Prediction in Large Urban Areas – Analysis of Sensitivity to Spatiotemporal Resolution of Rainfall Input and Hydrologic Modeling”, *Journal of Hydrology*, Vol.531, part 2, pp.370–388, (2015).
- [14] F. A. Ruslan, A. M. Samad, Z. M. Zain, R. Adnan, “Flood Prediction using NARX Neural Network and EKF Prediction Technique: A Comparative Study”, In *Proceeding of the IEEE 3rd International Conference on System Engineering and Technology (ICSET)*, pp.203–208, (2013).
- [15] S. K. Sinnakaudan, A. A. Ghani, M. S. S. Ahmad, N. A. Zakaria, “Flood Risk Mapping for Pari River Incorporating Sediment Transport”, *Environmental Modelling & Software*, Elsevier, Vol.18, No.2, pp.119–130, (2003).
- [16] H. M. Lyu, W. J. Sun, S. L. Shen, A. Arulrajah, “Flood Risk Assessment in Metro Systems of Megacities using a GIS-based Modeling Approach”, *Science of the Total Environment*, Elsevier, Vol.626, pp.1012–1025, (2018).
- [17] N. M. Hunter, P. D. Bates, M. S. Horritt, M. D. Wilson, “Simple Spatially-distributed Models for Predicting Flood Inundation: A Review”, *Geomorphology*, Elsevier, Vol.90, No.3–4, pp.208–225, (2007).
- [18] M. S. Horritt, P. D. Bates, “Evaluation of 1D and 2D

- Numerical Models for Predicting River Flood Inundation”, *Journal of Hydrology*, Elsevier, Vol.268, No.1-4, pp.87–99, (2002).
- [19] J. Kim, M. Hastak, “Social Network Analysis: Characteristics of Online Social Networks after a Disaster”, *International Journal of Information Management*, Vol.38, No.1, pp.86–96, (2018).
- [20] F. K. Sufi, I. Khalil, “Automated Disaster Monitoring From Social Media Posts Using AI-Based Location Intelligence and Sentiment Analysis”, *IEEE Transactions on Computational Social Systems*, pp.1–11, (2022).
- [21] H. N. Teodorescu, “Emergency-Related, Social Network Time Series: Description and Analysis”, *Time Series Analysis and Forecasting*, Springer International Publishing, pp.205–215, (2016).
- [22] K. Hiroi, D. Murakami, K. Kurata, T. Tashiro, Y. Shinoda, “A Proposal of Data Assimilation Approach for Flood Level Estimation and Evaluation with Urban Flood Disasters”, *Journal of Information Processing: Consumer Devices and Systems*, Vol.10, No.3, pp.55– 64, (2020) (In Japanese).
- [23] K. Hiroi, A., Kohiga, Y., Shinoda, “A Study on Effectiveness of SNS Data in Flood Estimation”, *International Journal of Informatics Society (IJIS)*, Vol.15, No.3, pp.133–144, (2024) .

Session 3:
Systems and Applications2
(Chair: Katsuhiko Kaji)

Individual Information Presentation for Autonomous Mobility

Yuichi Yaguchi*, Takumi Yotsuji*, Kouta Imoto*,
Kazuki Tamegai*, Seiji Komiya*, and Toshihiro Wakita*

*Graduate School of Engineering, Kanagawa Institute of Technology
s2384003@cco.kanagawa-it.ac.jp

Abstract - Autonomous mobility is now driving not only outdoors, but also indoors, such as in restaurants, commercial facilities, and other places. A common method to make people aware of approaching mobility is to use an alarm sound. However, there is a problem where if the volume of the warning sound is too soft or small, pedestrians may not notice it, and if it is too loud, it can be heard by people outside the roadway. In this paper, an information presentation system is developed that individually notifies pedestrians, who may be at risk of collision and are presumed to be unaware of the presence of mobility, about the approach. The system uses image recognition to recognize the direction of the pedestrian's face, and if the pedestrian is likely to be unaware of the vehicle's presence, the system uses parametric speakers to individually notify the pedestrian of the approaching vehicle.

Keywords: Autonomous Mobility, Human Interface

1 INTRODUCTION

1.1 Research Background

In recent years, the implementation of autonomous mobility has been advancing in restaurants, commercial facilities, and other places. Autonomous mobility is being used for automating luggage transportation and patrolling duties within facilities. These mobility systems operate in environments where pedestrians are present. They need to reach their destination without colliding with pedestrians or obstacles. Pedestrians can avoid collisions by predicting each other's movements through cues like eye contact. However, mobility units face the risk of collisions if they cannot communicate properly with pedestrians or if pedestrians are unaware of their approach. Therefore, it is necessary to explore methods to achieve smooth and cooperative movement when autonomous mobility navigates situations where pedestrians may not be aware of their presence. This discussion considers two examples.

In the first example, consider the scenario of driving a car. Major roads on which cars travel are typically equipped with traffic signals and crosswalks, and there are clear traffic rules based on road signs and traffic laws. This enables cars to navigate smoothly and safely with other vehicles and pedestrians. In the second example, consider a cyclist approaching a pedestrian who is looking at their smartphone while walking. To alert the pedestrian of their approach, the cyclist can ring a bell or maintain a safe distance to avoid a collision.

As seen in these examples, a smooth traffic environment is achieved when humans drive cars or ride bicycles by making judgments based on their surroundings. In contrast, when autonomous mobility travels on residential roads or sidewalks, the routes and traffic rules are not as clearly defined as they are for cars [1]. Therefore, "cooperative movement," where the mobility predicts and responds to the movements of others, becomes essential.

Thus, if the mobility itself can predict the state of pedestrians and take appropriate actions, it can achieve cooperative navigation that also considers the pedestrians' sense of security.

1.2 Acoustic Vehicle Alerting System

When autonomous vehicles or automated guided vehicles (AGVs) in factories approach pedestrians, they often use warning sounds to notify nearby people of their approach for safety reasons (Fig. 1). Recently, the number of incidents involving pedestrians walking while looking at their smartphones, known as "Texting while walking" has increased. According to data from the Tokyo Fire Department, 35.4% of emergency transports due to accidents involve collisions, and 75.9% of these collisions occur on roads or in traffic facilities [2].

The dangers of texting while walking include "narrowed field of vision", "inability to hear surrounding sounds" and "being defenseless" [3]. If pedestrians are unaware of an approaching mobility, they might suddenly change their path, leading to collisions. Therefore, using proximity warnings helps reduce the risk of such collisions. In automobiles, the installation of vehicle approach notification systems to alert pedestrians of an approaching vehicle by sound has been mandated by law [4]. For autonomous mobility, which are expected to operate on sidewalks and residential roads, it is essential to ensure that pedestrians can reliably recognize their approach to maintain safe and secure operations. It is anticipated that similar considerations will advance autonomous mobility in the future.

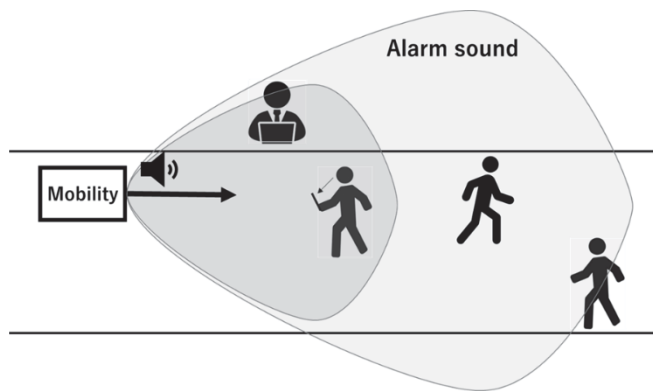


Fig. 1 Presenting information to the surroundings using a proximity alarm sound

However, audible alerts can be tolerated in open spaces like roads or factories where noise levels are generally higher. In quieter environments like office settings, frequent mobility vehicle movements with loud alarms might be perceived as disturbing or annoying by individuals who shouldn't normally hear them. To address this issue, adjusting the volume of the alarm or using softer tones can reduce the disturbance. Nevertheless, this approach may lead to challenges such as individuals who need to hear the alarm not being able to do so or becoming less attentive to approaching vehicles.

1.3 Research question and hypothesis

The classification of pedestrians is shown in Fig. 2. Individuals on the path where the mobility vehicle travels are termed "traffic participants," while those outside the path are termed "non-traffic participants." Among traffic participants, some may be aware of the approaching vehicle while others may not. Factors contributing to lack of awareness include pedestrians walking in front of the vehicle or being distracted by smartphones. While it's desirable to issue audible alerts to ensure safety for unaware traffic participants, it's also preferable for non-traffic participants not to hear the sound to maintain tranquility.

To convey information about the approach of vehicles and mobility to pedestrians, research on warning sounds is being conducted. One example is the study by Pavlo et al., which demonstrated a strong correlation between the loudness of external vehicle sounds and the deterrent effect on road crossing [5].

Research on approach warning sounds often focuses on the impact of the sound's timbre on traffic participants and its perceptibility. However, studies considering the impact on non-traffic participants are lacking. As a result, approach warnings might provide information to non-traffic participants as well, which can be perceived as noisy. Additionally, traffic participants might not receive sufficient information, making it difficult to notice the warnings.

To solve these problems, this study introduces a method termed "individual information presentation" aimed at delivering information to pedestrians individually, thereby addressing the limitations of proximity warnings. Specifically, the method employs image recognition to identify pedestrian states and subsequently delivers tailored

information to them using parametric speakers. This approach aims not only to prioritize safe arrival at the destination but also to consider pedestrians' sense of security. The proposed method was integrated into a mobility system, and its effectiveness was validated through sensory evaluation experiments.

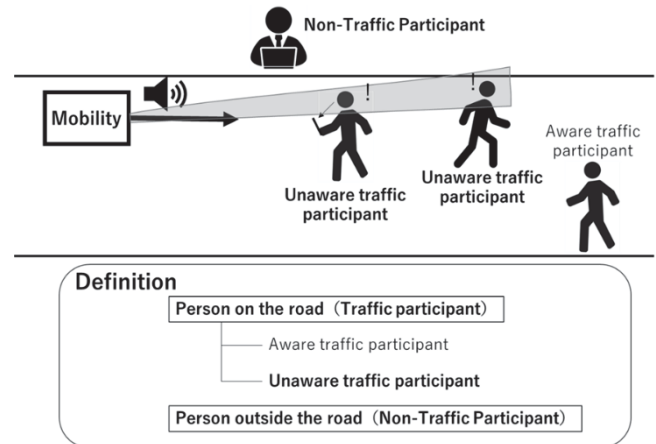


Fig. 2 Pedestrian Classification

2 METHOD

2.1 Proposed Algorithm

To issue proximity warnings individually to unaware traffic participants, the system estimates the position and condition of pedestrians. If it determines that a pedestrian is unaware, it activates the warning sound only in such cases. The procedure unfolds as follows:

- Step1: Utilizing a camera mounted on the mobility, pedestrian position and state are estimated through image recognition.
- Step2: Identifying nearby pedestrians who are deemed unaware, the system controls the direction of parametric speakers.
- Step3: Based on recognition results and pedestrian positions, audible warnings are emitted.

An overview of the proposed information presentation system is depicted in Fig. 3.

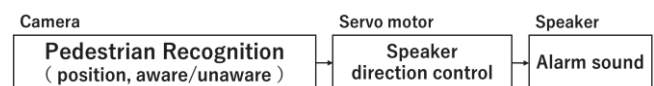


Fig. 3 Overview of Information Presentation System

2.2 Pedestrian location and condition detection

The position and state of pedestrians were determined using the YOLO [6] image recognition software. The pedestrian's position was calculated by projecting the coordinates of the bottom of the bounding box (bbox) representing the "person" class obtained through image recognition (Fig. 4). This

method allows for determining the pedestrian's position without depending on their height. To evaluate the detection accuracy, a coordinate axis was set up as shown in Fig. 5, with vertical distances of 7m and horizontal distances of 1.4m. The accuracy of pedestrian position estimation was investigated by comparing the calculated distances and angles with the actual values at each coordinate set in Fig. 5. One male subject (23 years old) was sequentially positioned upright at each coordinate point, and the measured distances and angles were compared with the calculated values. The subject wore long-sleeved shirts and pants matching the attire worn during the evaluation experiments, and the measurements were conducted under two conditions: indoor and outdoor. RMS (Root Mean Square) errors were calculated from a total of 210 data points collected at 21 measurement points. The results are presented in Table 1 and Table 2.



Fig. 4 Method of obtaining pedestrian coordinates

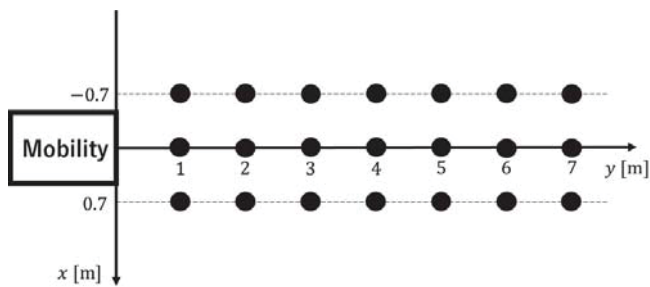


Fig. 5 Location where detection accuracy is evaluated

Table 1 Pedestrian location estimation results (indoor)

Distance error [m]	0.461
Angle error [deg]	5.75

Table 2 Pedestrian location estimation results (outdoor)

Distance error [m]	0.414
Angle error [deg]	4.61

The state of pedestrians was assumed to be such that they were not aware of the mobility vehicle's presence if they were facing away from the direction of the vehicle or if their faces were looking downwards. Therefore, images of pedestrians in various states were captured at multiple locations within Kanagawa Institute of Technology. These images were labeled into four categories: "person", "forward-facing (look)", "backward-facing (no_look)" and "downward-facing (smartphone_ped)" as shown in Fig. 6. A total of 7548 images were used, and the training process took 9 hours and 51 minutes, with 8000 epochs conducted for learning. Table 3 shows the accuracy of the recognition results.

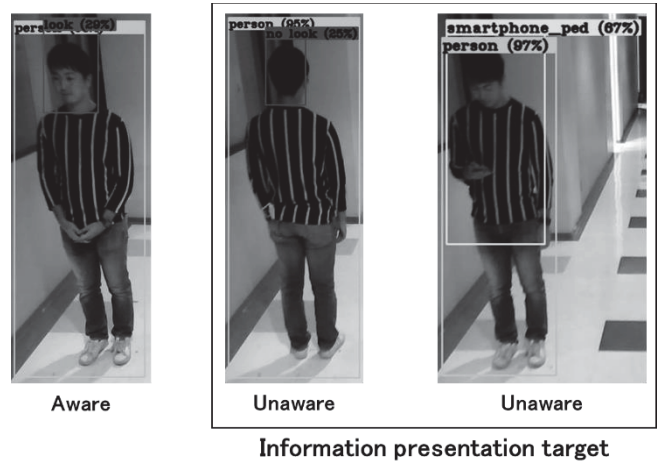


Fig. 6 Pedestrian Condition Detection

Table 3 Correctness of recognition results (Accuracy)

person	93.4%
look	75.4%
no look	64.7%
smartphone ped	69.9%

There is a need to improve the recognition rates for "look", "no_look" and "smartphone_ped" in the future.

2.3 Parametric speaker

To transmit warnings to individual pedestrians, a highly directional parametric speaker was used. The speaker's orientation was controlled by a servomotor to align with the pedestrian's direction. The parametric speaker, which uses ultrasound, is a single-directional audible speaker with a narrower audible range compared to conventional speakers. For this purpose, a parametric speaker manufactured by TriState was utilized (Fig. 7). This speaker generates a 40 kHz ultrasonic carrier wave from each of its 50 ultrasonic transducers, which is frequency modulated with the target sound and emitted. These ultrasonic waves cause nonlinear distortion in the air, allowing the target sound to be heard with sharp directivity [7]. This technology has been applied in previous research to systems providing hands-free environments [8].

To evaluate the characteristics of the parametric speaker used in this study, the variation in perceived sound based on

positioning was investigated. Three male subjects (average age: 22.7 years, ranging from 22 to 23 years) were positioned upright at various measurement locations (Fig. 8) in a quiet environment (background noise: 32dBA). They were then asked to respond to a questionnaire evaluating the "audibility of the alert sound." Responses were rated on a 5-point scale: "1. hardly audible", "2. a little audible", "3. audible", "4. well audible", "5. very audible". The alert sound, a sine wave with a fundamental frequency of 880Hz, was continuously emitted from the parametric speaker. The experimental results, depicted in Fig. 9, highlight areas where the mean score of the audibility evaluation exceeded 3.0 (indicating "audible"), marked with shading. Despite the angle error in image recognition being 5.75 degrees, the evaluations were consistently in the "audible" range. This suggests that controlling the orientation of the speaker towards pedestrians and transmitting alerts is indeed feasible.

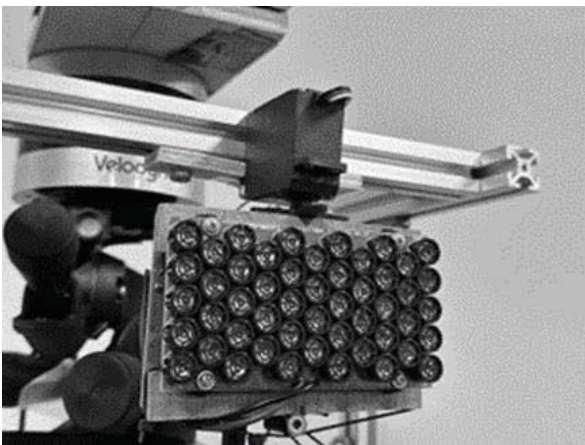


Fig. 7 Parametric speaker

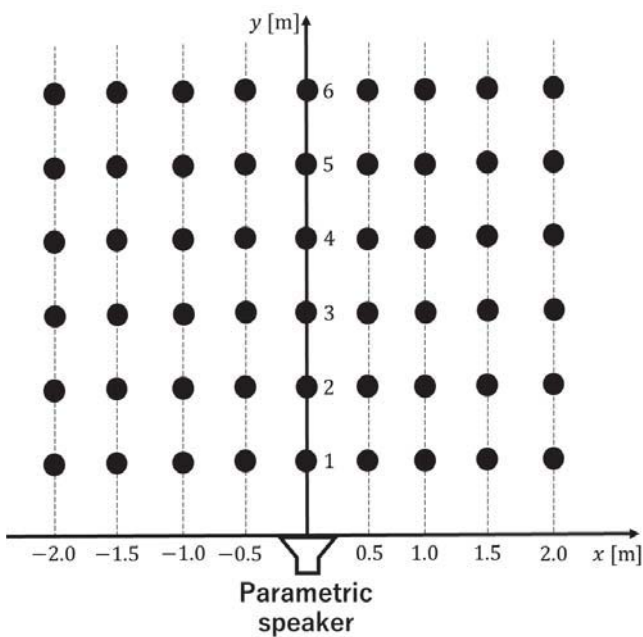


Fig. 8 Location of the "Hearing the alarm" rating.

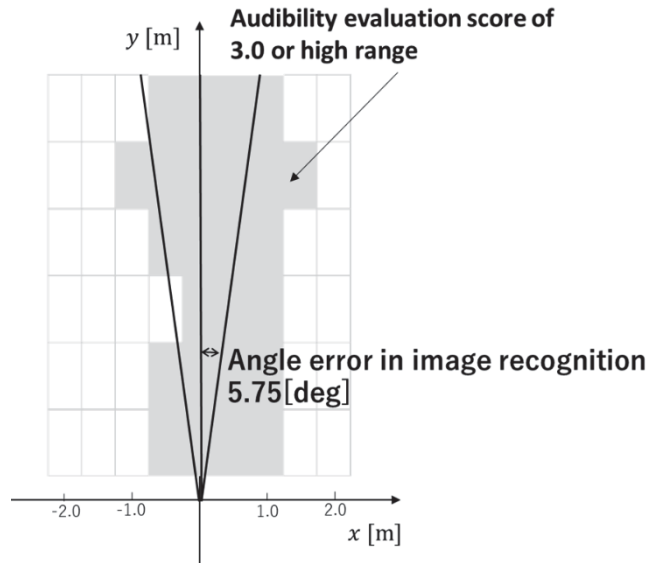


Fig. 9 Experimental Results Average

3 EXPERIMENT

3.1 Experimental device

An evaluation experiment was conducted to assess the effectiveness of the proposed method. The autonomous mobility device used was the WHILL Model CR (hereafter referred to as the mobility). The mobility is equipped with an omnidirectional laser sensor and a camera for obstacle detection and image recognition, a control computer, and an information transmission system. The control computer uses a Jetson Xavier NX for image recognition and a Shuttle for autonomous movement. The information transmission system employs a parametric speaker manufactured by TriState for sounding warnings and a SONY wireless portable speaker SRS-XB43 for comparison as a general speaker.

3.2 Experiment conditions

A subjective evaluation experiment was conducted with an information presentation system installed in a mobility. The experiment was conducted on three male subjects (average age: 22.7 years, ranging from 22 to 23 years). The experiment was conducted under the approval of the Human Ethics Review Committee of Kanagawa Institute of Technology. This experiment was conducted with the assumption of an office environment. The location was the same as where the sound pressure measurements of the parametric speaker were conducted. A course was set up as depicted in Fig. 10, and the mobility was driven at a speed of 0.5m/s with the subjects on board. Pedestrians were instructed to stand upright facing the direction of the mobility, while non-traffic participants were seated perpendicular to the mobility vehicle in chairs, with instructions to perform tasks freely, such as using smartphones. Path planning utilized the Dynamic Window Approach [9] (hereafter referred to as DWA), a method

originally proposed by D. Fox for static obstacle avoidance. This approach generates candidate paths considering vehicle constraints of the mobility and factors like destination arrival time. It evaluates each candidate path and selects the one with the highest evaluation score to control the mobility.

The mobility navigated from the starting point towards a goal located 13 meters ahead, using DWA to avoid pedestrians encountered along the way.

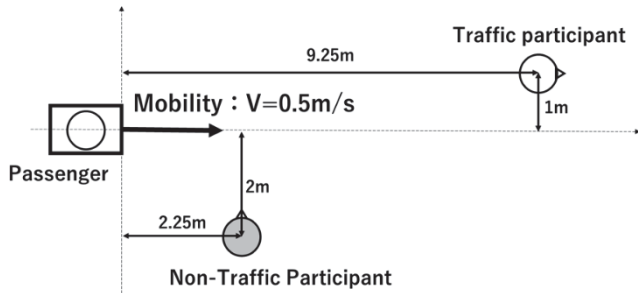


Fig. 10 Experimental Layout

The participants undertook two trials for each of the three roles: as traffic participants, as non-traffic participants, and as passengers. After each trial, they completed an evaluation questionnaire. Two experimental conditions were employed:

1. Conventional Method: In this method, an alarm sound (alarm sound A: consisting of "C4", "E4" and "G4" combined for 1 second each, with a cycle of 5.5 seconds) was continuously emitted from a conventional speaker.
2. Method 1: In this method, the same alarm sound (alarm sound A) was emitted repeatedly from the parametric speaker.

The evaluation criteria included "Audibility of the alarm sound", "Perception of being alerted", "Direction from which the mobility vehicle approached" and "Sense of distance when the mobility vehicle approached". Participants were asked to rate each criterion on a 5-point scale. The evaluation questionnaire is depicted in Fig. 11.

Please tell me about the experiment. (Information transmission)

1. How did you hear the warning sound?
 1. hardly audible
 2. a little audible
 3. audible
 4. well audible
 5. very audible
2. Did you feel that you were alerted?
 1. didn't feel it
 2. didn't feel it that much
 3. couldn't say either
 4. felt it a little
 5. felt it
3. Direction of approaching mobility vehicles * answered only by traffic participants
 1. very difficult to understand
 2. difficult to understand
 3. normal
 4. easy to understand
 5. very easy to understand
4. Distance from approaching mobility * answered only by traffic participants
 1. very difficult to understand
 2. difficult to understand
 3. normal
 4. easy to understand
 5. very easy to understand

Fig. 11 Evaluation Questionnaire

3.3 Experimental results

The trajectory of the mobility vehicle is illustrated in Fig. 12, while the subjective evaluation results of the participants are presented in Fig. 13, Fig. 14, and Fig. 15. The error bars in the evaluation result graphs represent the standard deviation σ of the evaluation scores. Additionally, significant differences are denoted by "*" for a significance level of 5% and "***" for a significance level of 1%.

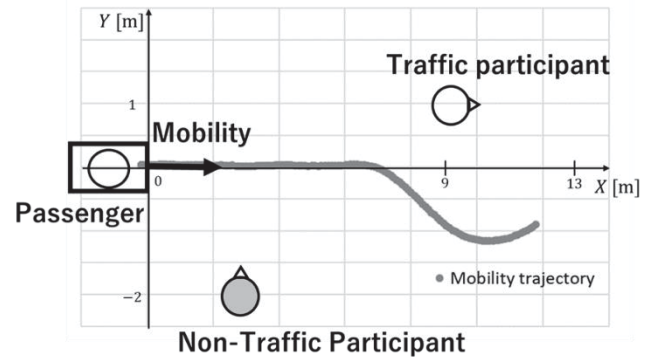


Fig. 12 Trajectory

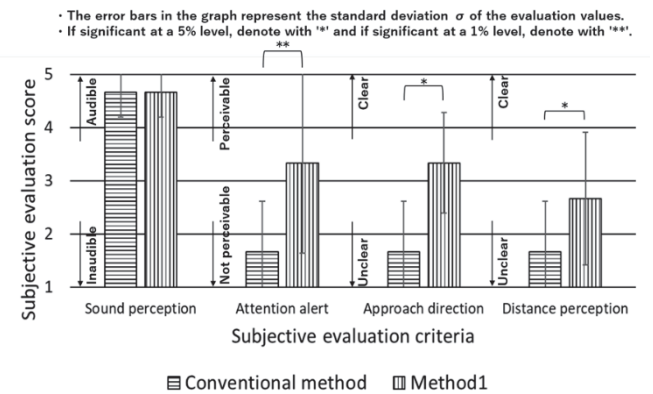


Fig. 13 Comparison results of information audibility and clarity (Traffic participant)

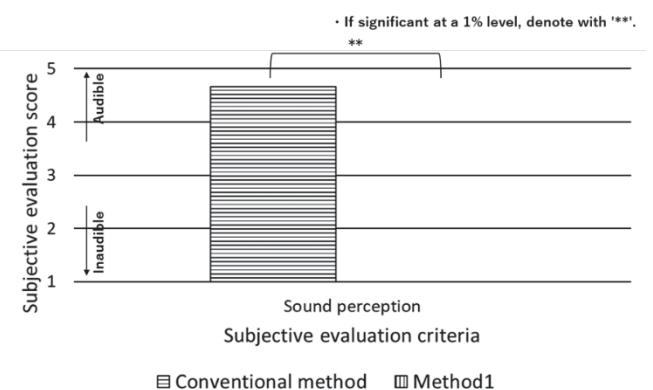


Fig. 14 Comparison results of information audibility (Non-Traffic participant)

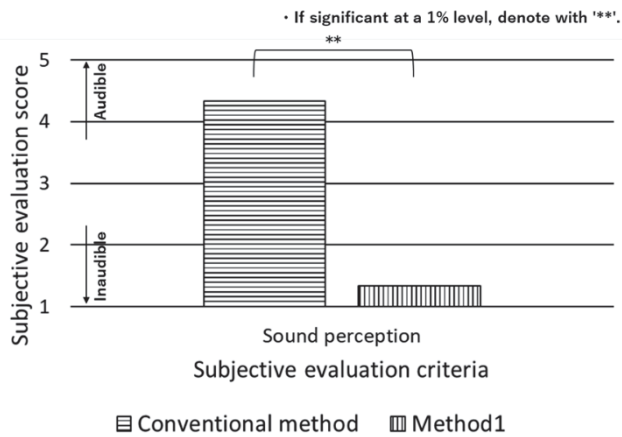


Fig. 15 Comparison results of information audibility (Passengers)

For the traffic participants, it was evident that both the conventional method and Method 1 of using warning sounds were rated as "4. well audible." Furthermore, compared to the conventional method, Method 1 elicited a significantly stronger sense of alertness. However, the direction and sense of distance during approach were rated as "3. normal," with the sense of distance particularly falling below "3."

For non-traffic participants, Method 1 was rated as "1. hardly audible," achieving the intended outcome.

4 DISCUSSION

4.1 Issues with conventional proximity alarms and information presentation systems

The first challenge of traditional proximity warnings is that "unrelated non-traffic participants also hear the warning sound, resulting in noise disturbance." To address this issue, the use of parametric speakers allows for transmitting information only to traffic participants, reducing noise for non-traffic participants. However, the second challenge of "not noticing the approach" remains even when using parametric speakers, showing limited effectiveness.

Through this study, a new consideration emerged regarding the preliminary experiments conducted in an open lobby with glass windows and concrete pillars. It was observed that when using parametric speakers, alarm sounds were audible not only from the intended direction but also from other directions. This occurrence is attributed to the characteristic of parametric speakers where sound travels directly and reflects off surrounding walls. Further investigation is needed to address this issue.

4.2 Future considerations for subject subjective evaluation

According to feedback from traffic participants, "The approach of the mobility was noticed, but it was unclear whether it would overtake from the left or the right." This

suggests that the characteristics of the parametric speaker influenced the ability to determine which side the mobility would pass on, making it difficult to assess. On the other hand, non-traffic participants mentioned, "The alarm sound was not heard at all, so when focused on something else, the approach would go unnoticed." This indicates that using this method could make it possible to avoid conveying the mobility's approach information.

5 CONCLUSION

This study developed an information presentation system to propose a method for individually alerting pedestrians who are deemed to be at risk of collision and may not be aware of the approaching vehicle. Furthermore, the effectiveness of this method was confirmed through sensory evaluation experiments. Implementing this system into mobility would enable cooperative autonomous mobility operations with pedestrians.

REFERENCES

- [1] Ministry of Economy, Trade and Industry, https://www.meti.go.jp/shingikai/mono_info_service/jidosoko_robot/pdf/007_05_00.pdf
- [2] Tokyo Fire Department, <https://www.tfd.metro.tokyo.lg.jp/lfe/topics/nichijou/mobile.html>
- [3] MEGURO CITY, https://www.city.meguro.tokyo.jp/dobokukanri/bousai/ianzen/koutsuanzen/kiken_nagarasumaho.html
- [4] Ministry of Land, Infrastructure, Transport and Tourism, <https://www.mlit.go.jp/common/001148213.pdf>
- [5] Pavlo Bazilinskyy et al., Exterior sounds for electric and automated vehicles: Loud is effective, *Applied Acoustics*, Vol.214, 109673 (2023).
- [6] J.Redmon et al., "You Only Look Once: Unified, Real-Time Object Detection", *IEEE Conference on Computer Vision and Pattern Recognition* (2016).
- [7] Hideyuki Nomura, Parametric Speaker Principles and Applications, *The Journal of the Acoustical Society of Japan*, Vol.76, No.11, pp.648-653 (2020).
- [8] Kazuhiro Kondo, Development of a pedestrian tracking, autonomously moving, super-directional loudspeaker system, *The Telecommunications Advancement Foundation*, No.31 (2016).
- [9] D. Fox, W. Burgard, and S. Thrun, "The dynamic window approach to collision avoidance," *IEEE Robot. Autom. Mag.*, Vol. 4, No. 1, pp. 23-33 (1997).

Title: Individual Information Presentation for Autonomous Mobility

Author: Yuichi Yaguchi, Takumi Yotsuji, Kouta Imoto, Kazuki Tamegai, Seiji Komiya, Toshihiro Wakita

Corresponding Author: Yuichi Yaguchi

Tel: 046-291-3091

Fax: 046-242-6806

E-mail: s2384003@cco.kanagawa-it.ac.jp

A Method for Automated Viewpoint Adjustment to Eliminate Blind Spots in Drone Remote Operation Using Virtual TPV

Reo AKAMINE*, Noriaki TAKENOUE**, Yuichi TOKUNAGA*

*Graduate School of Engineering, Kanazawa Institute of Technology, Japan
c6301767@st.kanazawa-it.ac.jp, y.tokunaga@neptune.kanazawa-it.ac.jp

**GSEC Inc., Japan
takenoue@gsec.biz

Abstract - We proposed a Third Person View (TPV) system for remote drone control that enabled pilots to easily perceive their surroundings within a virtual space. In the case of fixed virtual TPV, blind spots are created by obstacles and their own aircraft. Blind spots are dangerous because if blind spots continue to occur, obstacles in the blind spots cannot be identified. The remote pilot's manipulation of the virtual TPV's viewpoint to eliminate blind spots complicates the remote operation of the drone and may lead to accidents. Therefore, manually operating the virtual TPV viewpoint by the remote pilot affects the safety of remotely controlling the drone.

In this study, we developed an automatic virtual TPV viewpoint adjustment method to eliminate blind spots occurring during virtual TPV flight of a drone. This approach managed blind spot duration using grid units and determined the viewpoint's movement path, addressing blind spots that a fixed TPV method could not resolve.

In this paper, we present an algorithm for searching for virtual TPV viewpoints to eliminate blind spots and outline an evaluation plan. The plan is to use three scenarios to evaluate the search for viewpoints and the reduction of blind spots, and to demonstrate the effectiveness of the algorithm by adjusting its parameters.

Keywords: Blind Spots, Drone, Third Person View, Remote Operation

1 INTRODUCTION

Currently, the workload per worker is increasing in Japan's logistics industry due to an increase in small-lot deliveries and strict delivery time requirements from customers. As a result, the working environment has deteriorated, and a labor shortage has become an issue [1]. A similar phenomenon is occurring abroad [2][3]. As a solution, the U.S. has taken measures to increase the number of workers. Specifically, a special corporation called Driver Solutions has been established [4]. Driver Solutions assists inexperienced drivers to obtain a commercial driver's license. Europe has also launched the Women in Transport network, a platform aimed at increasing the number of female employees [5]. In Japan, the installation of delivery boxes as a mechanism to reduce re-delivery and the review of working conditions to improve salaries and benefits are being implemented. Aside from these measures to increase the number of workers, the use of drones for delivery services is attracting significant attention.

Currently, various services utilizing drones are being proposed [6]. For example, there are delivery services, agricultural applications, disaster relief, and critical infrastructure monitoring [7]. Among these, drone delivery is expected to play a significant role in the last mile. The last mile refers to the segment from the final logistics hub to the end user. Overseas, Amazon has introduced Amazon Prime Air, a drone delivery service [8]. In various countries, the development of regulations regarding drone operations is advancing. The Federal Aviation Administration in the United States permits aircraft registration on dedicated sites and allows flights below 400 feet [9]. Similar progress is being made in other countries.

In Japan, the Civil Aeronautics Law was amended in December 2022, allowing for flights beyond visual line of sight in populated areas, making drone operations in urban settings possible [10]. Currently, companies and local governments are collaborating on demonstration experiments for drone delivery of pharmaceuticals and lightweight packages, as well as for flights beyond visual line of sight. However, the presence of complex buildings and utility poles in urban areas poses obstacles that complicate drone operation. In 2023, the number of reported accidents reached 447, of which 89 were classified as serious incidents [11]. Reported accidents include contact with power lines and collisions with buildings and vehicles. Abroad, numerous incidents of contact with people have been reported, raising the possibility of similar issues arising in Japan in the future.

To address this issue, we propose a virtual Third Person View (TPV) system for remotely piloting drones [12]. The system allows users to understand the spatial relationship between the piloted drone and surrounding obstacles through VR imagery. Our research assumes the presence of a pilot. On the other hand, research and demonstrations of drone autopilot systems not requiring a pilot are increasing. However, accepting the risks of unmanned drones is not easy for the populace [13]. Aircraft with unmanned systems tend to be less willing to accept the risks of new technologies such as drones while there are parties involved who would take responsibility in the event of a serious accident [14]. Safe aviation requires not only minimizing human error, but also taking advantage of human concepts such as flexibility, goal awareness, and common sense.

As a prerequisite for human remote operation, the necessary information must be provided as visual information. In this context, a virtual TPV can provide the information satisfactorily. However, in cases where the

virtual TPV viewpoint is fixed, the shadow of the actual aircraft creates blind spots that affect hazard perception. Therefore, for safe remote operation, it is necessary to present these blind spots so that the operator can understand them effectively. In this study, we propose providing information on blind spots by moving the viewpoint. In this paper, we describe an overview of the algorithm for searching for a virtual TPV viewpoint to eliminate blind spots that have occurred, and a plan for evaluating its effectiveness.

2 RELATED WORK

2.1 Elimination of Blind Spots

Blind spots are a major threat to the safe operation of mobility. For example, automated vehicles cannot detect hazards in blind spots [15]. In addition, fatal accidents have occurred with construction heavy equipment due to blind spots [16]. If the blind spots are temporary, blind spots can be eliminated immediately, but if blind spots persist for a long time and are difficult to recognize, accidents may occur. This is called the collision course phenomenon and is considered a dangerous phenomenon in all forms of mobility, including cars, airplanes, and ships [17][18].

Figure 1 shows blind spot detection technology. Blind spot detection technology is broadly classified into detection by sensors and visual recognition through visualization. As an example of the former, ultrasonic sensors are used to monitor other vehicles that cannot be seen through side mirrors [19]. In this study, potential risks are defined in three levels, and LEDs corresponding to each risk are activated based on distance data obtained from ultrasonic sensors. On the other hand, to prevent accidents between large vehicles such as trucks and pedestrians hiding in blind spots, Bluetooth Low Energy and a received signal strength indicator are used to warn both drivers and pedestrians of the danger [20]. In this study, attachable LED lights and vibration motors are utilized as means of warning each other of their presence. This method of detecting blind spots relies on sensors. Therefore, a sensor failure or malfunction could affect the results. For this reason, recognizing blind spot situations by visualizing blind spot areas is also important.

There are two methods of visualization: drone utilization and virtual space utilization. As an example of drone utilization, Inoue et al. developed BirdViewAR, a new remote drone operation system, as a spatial recognition assistant for remote pilots in drone operations [21]. BirdViewAR always reveals the surroundings of the drone being maneuvered by controlling a tracking drone in response to the speed and direction of the maneuvered drone. The tracking drone also provides a third-person view to assist the remote pilot in controlling the drone at all times. The screen provided has an additional AR overlay to enhance spatial awareness of both the piloted drone and its surroundings. In this study, three patterns were evaluated: First Person View (FPV) only, BirdView without AR, and BirdViewAR with AR, with BirdViewAR showing better results in both spatial

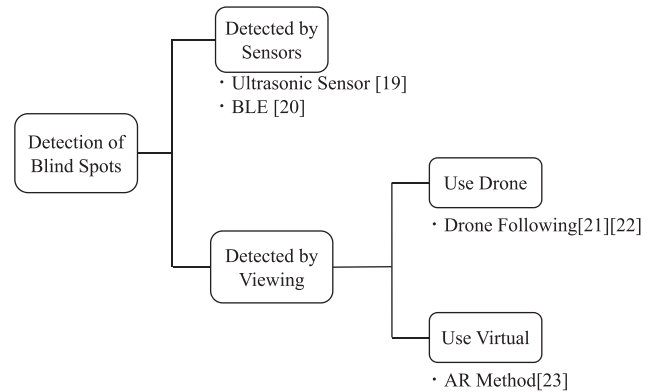


Figure 1: Blind Spot Detection Method

recognition and maneuvering performance. However, there are still issues in ensuring the safety of tracking drones providing TPV images. In a similar study, Temma et al. also used a tracking drone to provide a TPV to enhance the pilot's understanding of the drone's surroundings but described safety issues with the tracking drone [22]. The use of drones as a means of visualization has not solved the safety issue. Therefore, a method utilizing virtual space has been proposed.

Takeuchi et al. used pre-mapped three-dimensional spatial information to make obstacles causing blind spots transparent for the purpose of narrow space exploration within blind spots area[23]. Additionally, they proposed two methods for recognizing obstacles around the drone at the transparency destination and compared them with conventional methods. In the evaluation experiment, subjects wearing headsets piloted drones behind walls using an AR-based TPV. As a result, the AR-based method improved drone maneuverability compared to conventional methods. However, by transmitting obstacles causing blind spots, the operator cannot determine the distance relationship between the obstacle and the drone.

2.2 Automatic Viewpoint Movement

Controlling the viewpoint to check the relationship between the drone and obstacles from different angles can increase the workload for the operator. Therefore, technologies that automatically move viewpoints are needed. For example, autonomous cameras that adapt to the surgeon's movements using machine learning [24], and methods that present a viewpoint adapted to the operator to assist in the remote operation of a robot [25], are available. These methods are suitable for presenting a viewpoint in environments where the operator's work pattern is fixed. However, they are not adaptable to cases with a high degree of freedom, such as drone flight.

Thomason et al. automatically detect dangerous obstacles and make the pilot aware of the positional relationship between the drone and the obstacle by moving the viewpoint in the VR environment [26]. In this study, when the virtual drone detects an obstacle, the camera position changes according to the situation. If the drone is about to collide with an obstacle, the viewpoint moves to look between the drone

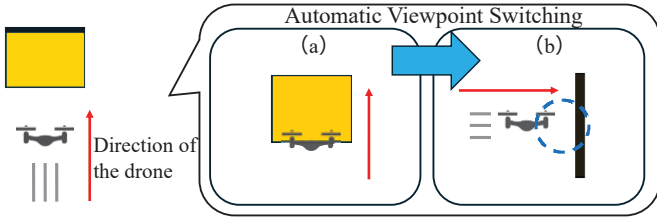


Figure 2: Automatic Viewpoint Switching

- (a): The drone is about to collide with the wall ahead in its direction of travel.
 (b): The viewpoint moves to look halfway between the drone and the obstacle.

and the obstacle (Fig. 2). Otherwise, the control viewpoint is placed 3 meters behind the virtual drone and is automatically moved to a position that maximizes visibility depending on the situation. The subjects wore headsets and flew the drone using only VR video to compare the effectiveness of FPV and TPV viewpoints. The results showed the effectiveness of automatic viewpoint movement during maneuvering, as it reduced the number of obstacles collisions more than FPV or TPV. However, although this method is designed for obstacle avoidance, obstacles that are always present are likely to be handled safely based on the pilot's experience. For sudden obstacles, however, there is a risk that the system may misjudge the obstacle. For this reason, relying on the machine to make obstacle judgments is not an appropriate solution.

As described above, there are many related studies on the use of virtual space for blind spots detection and automatic viewpoint adjustment. However, even if temporary blind spots can be eliminated, the problem of eliminating blind spots that are difficult to recognize, such as the collision course phenomenon, which lasts for a long time and is unnoticed by anyone, has not been solved. In this study, we propose a virtual TPV viewpoint control to solve this problem.

3 PROPOSED METHOD

3.1 Architecture for Automatic Viewpoint Shift

In this proposal, as a method to eliminate the blind spots that occur, we detect the areas in the virtual space where blind spots appear and seek a viewpoint that minimizes the duration of these blind spots. Specifically, the space is divided into 3D grids of a defined size, and the duration of blind spots is managed on a grid-by-grid basis. The sum of the durations across all grids is used as the blind spot cost, and the virtual TPV viewpoint with the lowest blind spot cost is chosen as the next virtual TPV viewpoint. However, since a large viewpoint shift can disturb the pilot's situational awareness, we aim to select a virtual TPV viewpoint with both a low blind spot cost and minimal shift to prevent such disturbances.

Figure 3 shows the process flow from virtual space generation to viewpoint shift. In this study, information on blind spot costs is managed through grid-based management of blind spots, and blind spot detection is performed based on this information. To find a new virtual TPV viewpoint, the

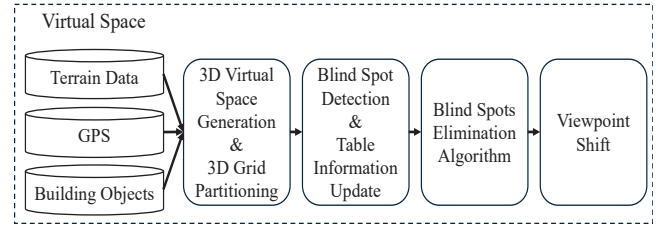


Figure 3: Architecture for Automatic Viewpoints

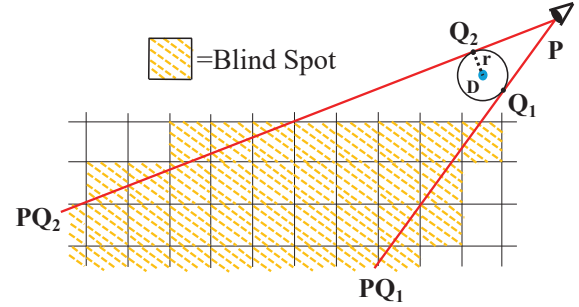


Figure 4: Image of Blind Spot Detection

Table 1: Management Table Information

Management Information	Type	Description
Position	Tuple	Grid Coordinate
Building	Bool	Presence of Building
Drone	Bool	Presence of Drone
BlindParameter	Float	Elapsed Time of Blind Spot

system searches among viewpoint candidates for the next position that can reduce the accumulated blind spot costs most effectively. The system then shifts from the previous virtual TPV viewpoint position to the new position.

3.2 3D Grid Partitioning

3D grids are generated by dividing a pre-generated 3D virtual space. 3D grids are used as a method for managing blind spots. Each grid generates a table that can manage the location information of the corresponding building, the presence or absence of a virtual TPV viewpoint or drone, and the blind spot cost (Table 1). At this time, the position coordinates stored in the table are calculated relative to the absolute coordinates of the virtual space. In addition, the grid size can be set freely. However, the grid size must be set appropriately because it affects the amount of viewpoint movement and the time taken to search blind spots.

Let i be the total set of divided grids, and j be the area where the virtual TPV viewpoint can detect blind spots. Area j is called the view area and is set to a size centered on the drone coordinates. In addition, there is a set k of grids that have been judged to be blind spots in the view area. The relation between the total set i , the view area j , and the set k of blind spots is $k \in j \in i$.

3.3 Blind Spot Detection

In this study, blind spot detection caused by drones in virtual space was conducted. The drone is represented as a sphere with radius r , centered at the drone's position D . Blind spot detection is performed within the view area j .

As shown in Fig. 4, grids touching or inside the straight lines PQ_1 , PQ_2 which are the boundary lines passing through virtual TPV viewpoint P and the surface coordinates of the sphere, Q_1 and Q_2 , were determined to be blind spots. Here, for the purpose of eliminating all hazards, grids touching the boundary lines are also classified as blind spots. In cases where the viewpoint is closer to the drone, the blind spot area tends to increase, while in cases where the distance is greater, the blind spot area decreases.

Grids identified as blind spots will have an increased blind spot cost value in the grid management table. The cost value is weighted, with a weight $w(Distance)$ assigned based on the distance between the virtual TPV viewpoint and each grid. To prioritize the elimination of blind spots in grids closer to the virtual TPV viewpoint, a higher weight is assigned to these grids. As the distance from the virtual TPV viewpoint increases, the priority for blind spot elimination decreases.

Additionally, let a_k represent the total sum of grids within the view area j that have been identified as blind spots. This value is defined as follows. The term $w(Distance(P, x))$ represents the weight based on the distance from viewpoint P to grid x .

The calculated total of blind spot determinations, a_k , is considered the blind spot cost and is used to derive a new virtual TPV position aimed at resolving the existing blind spots.

$$a_k = \sum_{x \in j} m_k(x) \cdot w(Distance(P, x)) \quad (1)$$

The function m_k indicates whether grid x belongs to the set of blind spots k , and it is defined as follows.

$$m_k(x) = \begin{cases} 1 & \text{if } x \in k \\ 0 & \text{if } x \notin k \end{cases} \quad (2)$$

3.4 Blind Spots Elimination Algorithm

To eliminate the detected blind spots, it is necessary to move the current virtual TPV viewpoint to a new position. Currently, the set L is defined as the set of adjacent grids that include the current virtual TPV viewpoint, and the new virtual TPV viewpoint is selected from this set.

From the set L , we search for a new virtual TPV viewpoint that can eliminate the largest number of blind spots. As a search method, we perform blind spot judgment within the new view area j' from the viewpoint candidate P_l . At this time, the set of blind spots determined in the view area is kP_l .

Next, the complement set $\overline{kP_l}$ is obtained by excluding the blind spot set kP_l from the entire set i of grids.

$$\overline{kP_l} = i - kP_l, l \in L \quad (3)$$

Equation (3) shows the part excluded from the view area j' , that is, the set that is not a blind spot. Next, the common set between the blind spot set k at the previous virtual TPV viewpoint P and the complement set $\overline{kP_l}$ is determined. This is called S_{P_l} .

$$S_{P_l} = k \cap \overline{kP_l} \quad (4)$$

Equation (4) represents the set that was determined to be a blind spot at the previous virtual TPV viewpoint P but is not determined to be a blind spot at the new virtual TPV viewpoint. The total cost of this set is called the blind spot elimination cost $C(P_l)$, which is the cost of the blind spot that occurred at the previous virtual TPV viewpoint P .

The maximum blind spot elimination cost $maxC(P_l)$ becomes the new virtual TPV viewpoint P_{app} . Additionally, $maxC(P_l)$ is evaluated by the cost function.

$$C(P_l) = \sum S_{P_l} \quad (5)$$

$$maxC(P_l) > 10 \cdot n \quad (6)$$

$$P_{opt} = argmaxC(P_l) \quad (7)$$

The variable n in Equation (6) represents the number of steps taken to explore the new virtual TPV viewpoint P_{app} . A higher number of steps indicates that the new virtual TPV viewpoint P_{app} moves farther away from the previous virtual TPV viewpoint P , resulting in a larger movement distance. This is a cost function designed to prevent such excessive movement.

Finally, if the grids that were determined to be blind spots at the previous virtual TPV viewpoint P remain at the new virtual TPV viewpoint P_{app} , these grids will be used as parameters to determine the new virtual TPV viewpoint P_{app} . This process yields the total blind spot cost a'_k . Additionally, if the time step t is greater than or equal to 1, the cumulative blind spot cost $X(i, t)$ will be incremented by this value. Let the equation for calculating the total blind spot cost be designated as (8) and let the equation for calculating the cumulative blind spot cost be designated as (9).

$$a'_k = a_k - argmaxC(P_l) \quad (8)$$

$$X_{i,t+1} = X_{i,t} + a'_k, t \geq 1 \quad (9)$$

The new virtual TPV viewpoint P_{app} obtained by equation (7) is moved from the previous virtual TPV viewpoint P to provide the pilot with a new virtual TPV viewpoint.

3.5 Viewpoint Movement

The method for moving between the new virtual TPV viewpoint and the previous virtual TPV viewpoint uses Slerp to ensure a smooth transition between the viewpoints. Slerp utilizes a technique called quaternion, which efficiently handles rotations in three-dimensional space, to perform interpolation on the surface of a sphere [27]. This allows for the creation of a natural and smooth trajectory between the

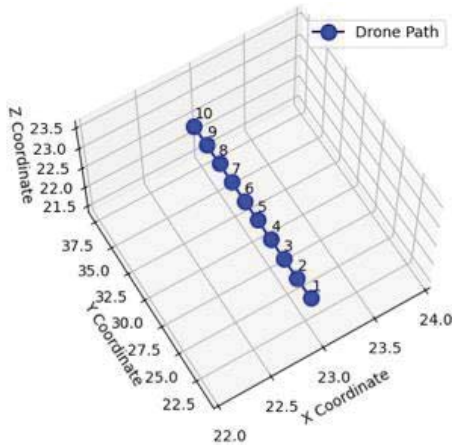


Figure 5: Trajectory of Constant Velocity Drone

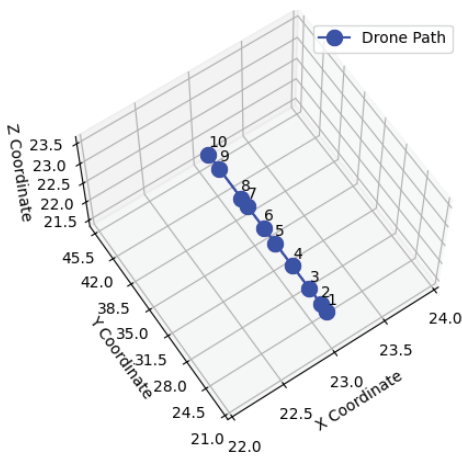


Figure 6: Trajectory of Speed Change Drone

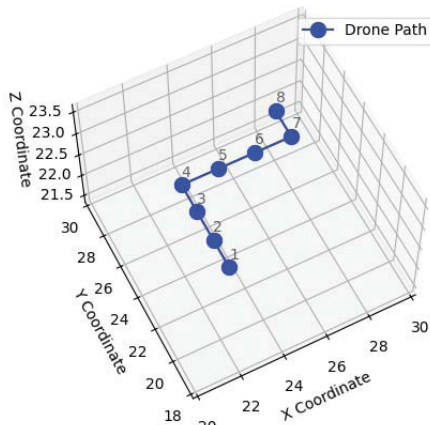


Figure 7: Trajectory of a Direction-Changing Drone

previous virtual TPV viewpoint and the new virtual TPV viewpoint.

These algorithms are used to find new virtual TPV viewpoints and eliminate blind spots.

4 EVALUATION PLAN

To validate the effectiveness of the proposed algorithm, the exploration of new virtual TPV viewpoints and their evaluation will be conducted based on trajectories derived

from three different scenarios. The trajectories to be evaluated are as follows:

(A) In the case where the drone moves in a straight line at a constant speed (Fig. 5)

(B) In the case where the drone moves in a straight line while changing its speed (Fig. 6)

(C) In the case where the drone's direction of travel changes suddenly (Fig. 7)

In each scenario, the effectiveness of the algorithm in accurately deriving the new virtual TPV viewpoint will be evaluated. Based on the obtained results, adjustments will be made to parameters such as the view area of the virtual TPV viewpoint, the size of the 3D grids, and the weights used in the algorithm. The goal is to enable the pilot to operate the drone more safely and accurately. Additionally, the safety of the virtual TPV viewpoint derived from the readjusted algorithm will be evaluated, and the practicality of providing viewpoints through TPV footage for drone operations will be examined.

5 CONCLUSION

This study proposes that eliminating blind spots occurring during drone deliveries in urban areas can address safety concerns in drone operations, providing a solution to the labor shortage problem in the logistics industry. The research suggests that blind spot information can be conveyed through viewpoint movement, enhancing situational awareness for the drone operator.

While related studies have employed virtual spaces for detecting surrounding blind spots and enabling free viewpoint movement, they have not yet addressed the challenge of eliminating blind spots that are difficult to recognize and persist over extended periods. To solve this problem, this research developed a viewpoint exploration algorithm that identifies new virtual TPV viewpoints. The algorithm operates based on a management table that tracks blind spot costs and object information within a 3D grid in the virtual space.

To validate the effectiveness of the proposed algorithm, three different courses will be prepared, and the exploration and evaluation of new virtual TPV viewpoints will be carried out. Based on the evaluation results, adjustments will be made to the algorithm's parameters, such as viewpoint area, grid size, and weights, with the goal of enabling safer and more accurate drone operation.

REFERENCES

- [1] METI • MLIT • MAFF, Current status of logistics in Japan and status of initiatives(in Japanese), https://www.meti.go.jp/shingikai/mono_info_service/sustainable_logistics/pdf/001_02_00.pdf (9 June 2024).
- [2] C. Ji-Hyland, D. Allen, "What Do Professional Drivers Think about Their Profession? An Examination of Factors Contributing to the Driver Shortage", International Journal of Logistics Research and Applications, Vol. 25, No. 3, pp.231-246 (2022).
- [3] J. de Winter, T. Driessen, D. Dodou, A. Cannoo, "Exploring the Challenges Faced by Dutch Truck

- Drivers in The Era of Technological Advancement”, *Frontiers in public health*, Vol. 12, DOI: <https://doi.org/10.3389/fpubh.2024.1352979> (2024).
- [4] Driver Solutions, Driver Solutions, <https://greatcdltraining.com/> (9 June 2024).
- [5] Vic Gov, Women in Transport Network, <https://www.vic.gov.au/women-transport-network> (9 June 2024).
- [6] L. Kapustina, N. Izakova, E. Makovkina, M. Khmelkov, “The global drone market: main development trends”, *SHS Web of Conferences*, Vol. 129, p. 11004 (2021).
- [7] N. Elmeseiry, N. Alshaer, T. Ismail, “A Detailed Survey and Future Directions of Unmanned Aerial Vehicles (UAVs) with Potential Applications”, *Aerospace*, Vol. 8, No. 12, p. 363 (2021).
- [8] H. Chen, Z. Hu, S. Solak, “Improved delivery policies for future drone-based delivery systems”, *European Journal of Operational Research*, Vol. 294, No. 3, pp. 1181-1201 (2021).
- [9] Federal Aviation Administration, FAA DroneZone Access – Home, <https://faadronezone-access.faa.gov/#> (9 June 2024).
- [10] MLIT, Unmanned Aerial Vehicle Level 4 Flight Portal Site, <https://www.mlit.go.jp/koku/level4/> (9 June 2024) (in Japanese).
- [11] Unmanned Aircraft Safety Division, JCAB, MLIT, Accidents and Serious Incidents, <https://www.mlit.go.jp/common/001623401.pdf> (18 July 2024) (in Japanese).
- [12] R. Akamine, N. Takenoue, Y. Tokunaga, “Proposal of Third Person View for Safe Drone Remote Control Operation”, *Proceedings of the 85th National Convention of IPSJ*, Vol. 2023, No. 1, pp. 231-232 (2023) (in Japanese).
- [13] J. Keebler, E. H. Lazzara, K. Wilson, E. L. Blickensderfer, *HUMAN FACTORS IN AVIATION AND AEROSPACE*, ELSEVIER, (2022).
- [14] P. Slovic, *The perception of risk*, Routledge, (2000).
- [15] J. Lu, Z. Peng, R. Shi, S. Yang, L. Gao, B. Zhang, R. Wang, Z. Pang, Q. Zhang, Y. Cao, “A quantitative blind area risks assessment method for safe driving assistance”, *Journal of Systems Architecture*, Vol. 150, DOI: 10.1016/j.sysarc.2024.103121 (2024).
- [16] J. W. Hinze, J. Teizer, “Visibility-related fatalities related to construction equipment”, *Safety science*, Vol. 49, No. 5, pp. 709-718 (2011).
- [17] P. Kevin, “Constant Bearing, Decreasing Range”, *US Naval Institute Proceedings*, Vol. 122, No. 12 (1996).
- [18] JAF Mate Online, Accidents that occur despite good prospects! The dangers of the Collision Course Phenomenon, https://jafmate.jp/car/traffic_topics_20230117.html (10, June 2024) (in Japanese).
- [19] Z. Adnan, M. Z. Hassan, N. Ab Wahab, S. M. Najib, N. S. Nasir, “Vehicle Blind Spot Monitoring Phenomenon using Ultrasonic Sensor”, *International Journal of Emerging Trends in Engineering Research*, Vol. 8, No. 8, (2020).
- [20] N. De Raeve, M. De Schepper, J. Verhaevert, P. Van Torre, H. Rogier, “A bluetooth-low-energy-based detection and warning system for vulnerable road users in the blind spot of vehicles”, *Sensors*, Vol. 20, No. 9, p. 2727 (2020).
- [21] M. Inoue, K. Takashima, K. Fujita, Y. Kitamura, “BirdViewAR: Surroundings-aware Remote Drone Piloting Using an Augmented Third-person Perspective”, *Proceedings of the 2023 CHI Conference on Human Factors in Computing System*, Vol. 84, No. 31, pp. 1-19 (2023).
- [22] R. Temma, K. Takashima, K. Fujita, K. Sueda, Y. Kitamura, “Enhancing Drone Interface Using Spatially Coupled Two Perspectives”, *IPSJ Journal*, Vol. 61, No. 8, pp. 1319-1332 (2020) (in Japanese).
- [23] K. Takeuchi, R. Teng, K. Sato, “Implementation and Evaluation of a Drone-based AR Visualization Method for Narrow Space Surveillance”, *IPSJ Journal*, Vol. 64, No. 2, pp. 614-625 (2023) (in Japanese).
- [24] M. Wagner, A. Bihlmaier, H. G. Kenngott, P. Mietkowski, P. M. Scheickl, S. Bodenstedt, A. Schiepe-Tiska, J. Vetter, F. Nickel, S. Speidel, H. Worm, F. Mathis-Ullrich, B. P. Muller-Stich, “A learning robot for cognitive camera control in minimally invasive surgery, *Surgical Endoscopy*”, Vol. 35, No. 9, pp. 5365-5374 (2021).
- [25] D. Rakita, B. Mutlu, M. Gleicher, “An Autonomous Dynamic Camera Method for Effective Remote Teleoperation”, *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 325-333 (2018).
- [26] J. Thomason, P. Ratsamee, J. Orlosky, K. Kiyokawa, T. Mashita, Y. Uranishi, H. Takemura, “A comparison of adaptive view techniques for exploratory 3D drone teleoperation”, *ACM Transactions on Interactive Intelligent Systems (TiiS)*, Vol. 9, No. 2-3, pp. 1-19 (2019).
- [27] K. Shoemake, “Animating rotation with quaternion curves”, *Proceedings of the 12th annual conference on Computer graphics and interactive techniques*, Vol. 19, No. 3, pp. 245-254 (1985).

An Efficient Route Planning Method for 3D Modeling of Utility Poles

Koji YAMAGISHI*, Yuichi TOKUNAGA*

*Graduate School of Engineering, Kanazawa Institute of Technology, Japan
c6301756@st.kanazawa-it.ac.jp
y.tokunaga@neptune.kanazawa-it.ac.jp

Abstract - The project aims to reduce the cost of remote inspection of utility poles by utilizing 3D models generated from image capture through drone aerial shooting. The 3D model allows inspectors to have a free perspective without having to go to the site. Thus, time efficiency is improved, and utility pole inspections are more efficient. On the other hand, more efficient travel routes are needed to shoot many utility poles at once.

In this paper, we propose an algorithm to generate the shooting points and mobility travel routes that can both reduce the shooting time and generate a highly accurate 3D model. Specifically, take multiple shots while moving from one shooting point to another in a discrete pattern. Thereby creating a flight route that does not require extra detours that cause travel speed reduction. The above efficient flight route reduces the time required to capture the images needed to generate the 3D model. Search for the mobile travel route with the shortest total shooting time via multiple shooting points.

In addition to the above proposed algorithm, this paper introduces the evaluation plan. This plan shows the proposed algorithm and the conventional method of comparing the path with the least number of shots. The plan is to show the effectiveness of the proposed algorithm based on the compatibility evaluation of accuracy-by-accuracy evaluation index and efficiency by aerial photography time.

Keywords: 3D model, Drone, Utility pole, SfM, MVS

1 INTRODUCTION

In recent years, many social infrastructure facilities have been aging as 50 years have passed since their construction because the useful life of equipment is 50 years [1].

As an example, the United States has been developing its infrastructure on a large scale since the New Deal policies of the 1930s [2]. Fifty years later in the 1980s the U.S. faced a fundamental problem of aging infrastructure that impacted various aspects of the economy and daily life [3]. Japan's social infrastructure was built intensively during the period of rapid economic growth around 1955-1970. Fifty years later Japan already has many aging facilities, and their number is expected to increase at an accelerating rate in the future [4].

Maintenance is essential to maintain these facilities. Maintenance includes inspection and repair. Inspection is the process of selecting equipment to be repaired. Efficient and reliable inspections can improve overall maintenance efficiency. Many efficient inspection methods for large facilities have been studied. However, few studies have focused on inspection methods for utility poles which are

small structures. Utility poles contain significantly more equipment compared to large facilities. The total number of utility poles in 2016 was 35.78 million and increased by approximately 70,000 utility poles per year. Of these, 67% are electric poles for power transmission and 33% are telegraph poles for telecommunications. These are called utility poles. Time and cost as well as human efficiency must be emphasized to inspect all utility poles.

Until now, utility pole inspections have had no choice but to rely on visually inspections that are characterized by high reliability and accuracy. Yet, visual inspection requires the inspector to visit all poles, and the condition is difficult to ascertain depending on geographical factors and height above the ground. Consequently, visually inspecting all utility poles is inefficient regarding time and money as well as workforce.

Inspection methods based on images are being considered as a new method to increase efficiency. A camera mounted on the roof of the vehicle allows remote inspectors to review images taken while driving along the road [5]. But the issue lies in the fact that image capture is limited to the roadside which results in the creation of blind spots. In addition, the number of images required to check the entire object in an image-based inspection is large because the images are required to be changed depending on the area to be checked, a factor that lowers inspection efficiency.

On the other hand, inspection methods using 3D models are being considered for large facilities. Zhang Tion et al. describe a 3D reconstruction method for multi-vision-based inspection for pipelines [6]. An efficient and accurate 3D reconstruction method for industrial pipeline inspection using photogrammetry was presented. The five photogrammetric methods used in this study were stereovision, photometric stereo, CAD-based photogrammetry, motion-based photogrammetry, and shading-based photogrammetry.

In addition, the use of drones is being considered for acquiring data used to generate 3D models. Miyake et al. used a drone to take aerial shots of the slope and fill on the road and the bridge to create a 3D model that represents structures and slopes and fill on the road [7]. This allowed us to identify any signs based on the difference between the current and past 3D models.

Realizing manpower saving in utility poles as well as in other facilities is necessary. The number of construction workers including infrastructure inspectors has been declining since 1997 with 5 million in 2015 which is 73% of the peak and the shortage of inspectors will continue [8]. Thus, like large facilities drones and 3D models are expected to be utilized however conducting efficient inspections of utility poles remains important.

Therefore, this study examines a method for generating high-precision 3D models of utility poles using aerial images taken by drones with the aim of improving efficiency.

Creating a 3D model of an entire utility pole using aerial photography by a drone allows inspections to be conducted from any viewpoint without needing to go to the site. As a result, inspections without relying on manpower become possible and efficiency improves.

This paper proposes an algorithm for generating aerial shooting points and mobility paths that can both reduce aerial shooting time and generate highly accurate 3D models.

Chapter 2 introduces 3D models and related studies on the use of 3D models in inspections and the generation of 3D models by aerial photography, Chapter 3 describes the proposed framework based on hypotheses derived from these studies, and Chapter 4 describes the evaluation. Finally, Chapter 5 presents the conclusions.

2 RELATED WORK

The prior art on drone 3D model generation methods will be presented and the requirements for 3D modeling of utility poles will be clarified.

The 3D model consists of point cloud data that is obtained by laser scanning or photogrammetry methods. Laser scanning is often used with LiDAR and 3D laser scanners. The time-of-flight method, which measures the time of travel due to laser emission and reflection, is used to obtain the spatial coordinates of an object's surface. Thereby, spatial coordinates can be accurately measured within an error range of 1 to 10 mm [9]. However, laser scanners are expensive, and their use may be limited by certain circumstances that can distort measurements such as laser penetration and diffuse reflection.

The photogrammetric method acquires point cloud data by reconstructing 2D images with overlapping intervals into a 3D point cloud. Point cloud data is acquired using techniques such as Structure from Motion (SfM) and Multi-View Stereo (MVS). Structure from Motion (SfM) obtains spatial coordinates by superimposing multiple photos and calculating distances based on key points in the photos. Multi-View Stereo (MVS) refers to the process of recovering a 3D shape from the camera's captured position. Photogrammetry methods are challenged by low accuracy.

But photogrammetry methods are attracting attention as an alternative to LiDAR because of their low cost and ease of acquiring point clouds [10]. Therefore, various mobility-based imaging methods are being considered to acquire data efficiently in terms of time and money as well as manpower.

Drones and other unmanned aircraft are used for data acquisition. Sungsik Yoon et al. proposed an unmanned aerial vehicle-based missing area detection and damage location method based on 3D image coordinates as an alternative to visual inspection of piers [11]. Sungsik's methodology consists of three phases. In Phase 1, coordinate transformation is performed by the position and attitude information of the unmanned aircraft and camera, and the distance information between the camera and the target surface is used to obtain the coordinates of the center point of each acquired image. In Phase 2, the camera's focal length

and working distance are used to calculate the size of the field of view for each acquired image. In the last phase, Phase 3, information about the size of the field of view of all images computed in the previous phases is used to identify missing portions of the region of interest. This makes it possible to locate damage detected at the individual image level to the entire inspection area.

The results showed that the missing areas and damaged locations could be identified with an accuracy of 10 cm error while taking advantage of the efficiency provided by unmanned aerial vehicles. However, because utility poles are small-scale facilities the 10 cm error in utility poles is significant and cannot be ignored. A means to eliminate this error and generate a high-quality 3D model is needed.

Moritani et al. proposed a method for estimating camera positions for additional shooting based on the results of low-quality region extraction as a guideline for optimal shooting planning for efficient and high-quality 3D model generation by SfM-MVS [12]. The paper focuses on the fact that it is currently difficult to estimate in advance how many images should be taken from which location to generate a high-quality 3D model to generate a high-precision 3D model. In addition, we also noted that if too many images are taken, the MVS process to generate a highly accurate model takes an enormous amount of time. The related study uses the procedure shown in Fig. 1 to predict the quality of the high-density model generated by the MVS process. The prediction relies solely on the 3D tie points and camera poses estimated by the SfM process which can be computed in a short time. The study also estimates the optimal additional image capture positions to improve the quality of regions predicted as low-quality. The goal is to estimate the optimal additional image capturing locations that can improve the low quality and the quality of the predicted regions.

The specific procedure is to calculate candidate additional shooting positions where orthographic imaging of the low-quality area is possible from the center of gravity position of the low-quality area, and then to obtain an index of candidate additional shooting positions. The relationship between the calculated candidate additional shooting positions and the shooting positions that best cover the low-quality area among the already-captured cameras is expressed as a certain index. The above two indices are weighed and added together. This procedure is performed for all candidate additional shooting

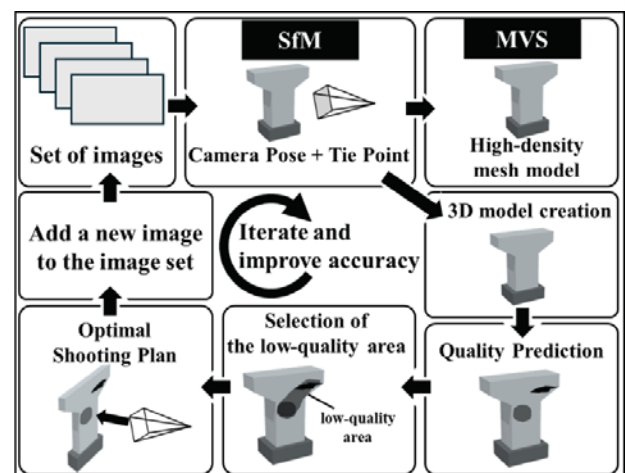


Figure 1: SfM-MVS Process and Proposed Process.

positions, and the candidate additional shooting position corresponding to the index with the lowest value is selected as the optimal additional shooting position. Repeating this method allows efficient determination of additional shooting positions.

We compared the accuracy of the 3D model with three groups of images: 30 randomly selected images, 72 randomly selected images, and 72 images selected by the proposed method from a set of 171 original images taken densely around the outer circumference of the piers. As a result, the average error of the 30 randomly selected images was 12.3 mm, and that of the 72 randomly selected images was 6.1 mm. In contrast, the proposed method of 72 images resulted in an error of 2.2mm, indicating the effectiveness of the proposed method for 3D model accuracy. However, while high accuracy can be obtained, the number of sheets is large. Therefore, reducing the number of shots while maintaining the accuracy of the 3D model is necessary.

Yamazaki et al. proposed an optimal acquisition planning method for 3D reconstruction using SfM (Structure from Motion) / MVS (Multi View Stereo) for construction sites [13]. This study is based on the "model-based methods" method that uses a simple 3D model based on geospatial information to plan optimal shooting locations and routes in advance [14]. The specific process of this related study is to first homogenize the model by polygon meshing the simplified model. Next, from the voxels that are candidates for shooting positions the shooting positions that minimize the number of shots required for modeling are extracted. Thereby, minimizing the number of shots to capture all polygons reduces the cost of shooting.

As a result, the number of shooting position postures was reduced from 1,790 to 83 for the target building. Yet, there is no consideration of time efficiency in case multiple shooting positions are far apart, and operational issues remain because the path setting was not considered.

Acquiring image data of a 3D model requires taking images from the appropriate position with fewer images to avoid generating missing parts. In addition, the use of a drone is effective for efficient image capture.

Thus, this study examines a route for aerial shooting of utility poles that is both efficient and accurate using a drone. Specifically, an aerial shooting course is determined based on the shooting position that minimizes the number of shots, while performing supplementary shooting. Scattered shooting positions require supplementary shooting based on the estimation of additional positions so aerial shooting can be done without degrading the accuracy of the 3D model.

3 PROPOSAL OF THE FRAMEWORK

3.1 Summary Proposal

To achieve both accuracy and time efficiency, this study creates a flight path plan based on the optimal aerial shooting position that minimizes the number of cameras.

Furthermore, to perform drone aerial shooting without compromising temporal efficiency a path is created that allows continuous flight at a constant speed. We follow the obtained route and apply the complementary shooting

position candidate extraction method of Moritani et al. We perform interpolated shooting to take aerial shots without degrading the accuracy of the 3D model by capturing images necessary for 3D model generation. The goal is to find the path with the shortest total aerial photography time using multiple aerial photography locations.

3.2 Flight Route Planning

The target is subdivided into smaller areas using a square mesh by referring to the method of Yamazaki et al. Determine the coordinates at which more of this square mesh can be taken at one time. The above can be expressed as a mathematical formula to optimize so that the binary variable x in the following Equation (1) is minimized. Each component of x corresponds to each candidate camera with x_i extracted being 1 and x_i not extracted being 0.

$$\min \sum_{\forall i \in I} x_i \quad \text{where } x_i \in \{0,1\} \quad (1)$$

Extracting the optimal shooting position for each polygon is important as well; during the SfM process each polygon must be shot from at least two viewpoints to triangulate its distance. Therefore, the constraint in Equation (1) is that each polygon is shot at least r times as shown in Equation (2).

$$\text{subject to } \sum_{\forall i \in I} P_{ji} x_i \geq r, \forall j \in J \quad (2)$$

In each polygon, the binocular parallax of around 30 degrees is represented by $A = [A_1 \dots A_j]$. A_j is the symmetric matrix and A_{jkm} is defined as 1 if the binocular parallax between camera k and camera m with respect to the center of gravity of polygon j is around 30 degrees and 0 otherwise and Equation (3) is used as the constraint condition.

$$\text{subject to } \sum_{\forall i \in I} A_{jni} x_i \geq (r - 1), \forall j \in J, \forall n \in I \quad (3)$$

One advantage of drones is the ability to change the direction of travel without changing the direction of the aircraft though sudden movements such as turning around force deceleration. Therefore, a constant-speed flight is possible by setting a waypoint that includes an appropriate drone turn path. Two constraint conditions are set to make the above possible.

The first is the upper limit of the bank angle relative to the direction of travel. Specifically, the distance L from the drone to the destination and the steering angle ψ_{MAX} per fixed distance l of the drone are set. Thus, the maximum bank angle φ_{MAX} from the drone point to the destination is determined.

$$\varphi_{MAX} = \frac{l}{L} \times \psi_{MAX} \quad (4)$$

The second constraint is the setting of the turning radius R . The turning radius R can be derived from Equation (5). V is the speed of the drone and θ is the upper limit of the bank angle derived in Equation (4). 11.26 is the gravitational acceleration converted to feet value.

$$R = \frac{V^2}{11.26 \times \tan \theta} \quad (5)$$

Thereby revealing the drone's trajectory in case turning at the maximum bank angle. The intersection of this trajectory and the vector from the current location to the destination is determined. If the distance to the intersection is less than or

equal to the distance to the destination the destination can be reached within the maximum bank angle and the destination is used as the waypoint. If the distance to the intersection is greater than the distance to the destination the destination cannot be reached even at the maximum bank angle. In this case, the waypoint is the intersection of the normal of the vector from the current location to the destination through the destination and the locus at the maximum bank angle.

The trajectory and waypoints calculated above are used to set the route. To adapt to these conditions the start and goal points need to be updated.

In this study, the start point was set as the drone location. The drone's movement vector was determined by connecting the coordinates of the start point and the drone point before it rotated 180 degrees around the start point. The goal point was defined as the next point after the optimal aerial shooting position that the drone had been aiming for until then.

If the vector between the drone's current location and the next optimal aerial position is \vec{a}_i and the inverse vector between the drone's current location and the previous waypoint is \vec{b}_i Equations (6) and (7) represent the angle formed between \vec{a}_i and the optimal aerial shooting position P_i .

$$\cos \theta = \frac{\vec{a}_i \cdot \vec{b}_i}{|\vec{a}_i| |\vec{b}_i|} \quad (6)$$

$$P_i = \vec{a}_i \times \cos(\theta - \varphi_{MAX}) \quad (7)$$

Connecting the trajectories between the waypoints calculated in these ways creates a final route plan for the drone to fly without slowing down.

The above settings make it possible to create a flight path that does not slow down the drone.

3.3 Calculation of aerial shooting points

This section calculates the aerial shooting points on the calculated route. The issue in shooting from the waypoint calculated in the previous section is the distance from the utility poles compared to the optimal aerial shooting position of Yamazaki et al. The accuracy of the 3D model depends on the resolution of the image. Therefore, the further the shooting position is from the object the less accurate the 3D model becomes. To address this issue the algorithm of Moritani et al. is used to calculate the optimal additional shooting position. Specifically, additional shooting is performed to ensure efficient shooting without compromising the quality of the 3D model.

Moritani et al. use two indices to find additional shooting points: The first is an evaluation index for orthogonal shooting. The second is an evaluation index of baseline ratio. The baseline ratio is the ratio of the distance between two images and the distance from the image to the target. These two indices are used to calculate the optimal additional shooting position.

The evaluation index for positive shooting is calculated using (8). c_i is the multiple representative points between waypoints obtained in the previous section. In this paper, 10 were set between each waypoint to be candidates for c_i . p is the center of gravity of each mesh in case the poles are meshed. n is the normal vector of the poles at the center of gravity of each mesh. From the above variables, we obtain an

evaluation index of directly opposite shooting from each representative point in the pathway to the center of gravity of each mesh.

$$NBV_{frontarity}^i = 1 - \left(n \cdot \frac{c_i - p}{\|c_i - p\|} \right) \quad (8)$$

The baseline ratio evaluation index is calculated by the following procedure.

BH^i is a value indicating whether the angle formed by the additional shooting position candidate c_i and the reference image position c_r centered at the center of gravity p of each mesh is within the specified angle. In this paper, it is set to 45 degrees the same value as the standard angle of view. Although the reference image position c_r originally corresponds to the optimal aerial shooting position, there are cases where it is not via the optimal aerial shooting position. In such a case, the candidate additional shooting position located at the nearest coordinate from the optimal aerial shooting position is used as the reference image position c_r . c_{ri} is the midpoint of c_r and c_i . 0.536 uses Moritani's empirically recommended angle obtained from between images.

$$BH^i = \left| \frac{\|c_i - c_r\|}{\|c_{ri} - p\|} - 0.536 \right| \quad (9)$$

$$NBV_{BH}^i = \frac{BH^i - BH_{max}}{BH_{min} - BH_{max}} \quad (10)$$

The baseline ratio obtained in Equation (9) is used in Equations (8) and (10) to calculate the evaluation index for each additional shooting position candidate c_i . w is the weight, and in this paper, $w_f = w_{BH} = 1/2$ in order to use both evaluation indices equally, referring to the method of Moritani et al. From the above, the candidate additional shooting position $c_i \in C$ with the lowest value is the optimal additional shooting position c_{NBV} .

$$c_{NBV} = \underset{c_i \in C}{\operatorname{argmin}} (w_f NBV_{frontarity}^i + w_{BH} NBV_{BH}^i) \quad (11)$$

4 EVALUATION PLAN

This study evaluates the compatibility of temporal efficiency and 3D model accuracy. The 3D model accuracy is evaluated using the accuracy prediction index of the 3D model used by Moritani et al. In addition, the minimum shooting position calculated by the optimal shooting planning method of Yamazaki et al. is used for comparison. Yamazaki et al. assume that the camera stops and shoots from the minimum shooting position. In this evaluation the aircraft is assumed to stop at the minimum shooting position for a certain period and to fly in a straight line between the minimum shooting positions while accelerating and decelerating.

4.1 Evaluation Method

The poles evaluate in this study are 16 m high and 340 mm in diameter using as reference poles with a height of 16 m, a terminal diameter of 240 mm and a former diameter of 453 mm. Since this evaluation is a basic evaluation there are assumed to be no appendages or surrounding structures. In

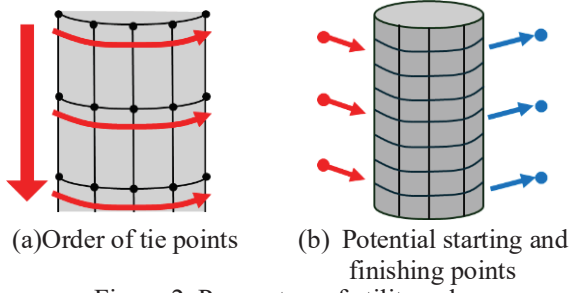


Figure 2: Parameters of utility poles

In addition, the utility pole is divided into 8 sections in the circumferential direction and 8 sections in the height direction making a total of 64 voxels. Therefore, the tie points are the vertices of each voxel, resulting in a total of 72 candidate tie points. The accuracy of the tie point candidates is evaluated. The tie points are numbered from the top of the poles as shown in Fig. 2(a). Based on the above conditions the degree of shape degradation at all tie points is evaluated using two patterns: the conventional method and the proposed method. The speed of the drone in this evaluation is variable, but the acceleration is unified to 10 km/h². This value is determined since the maximum acceleration of a typical aircraft is approximately 10 km/h². The candidate starting and finishing points of the drone are the top and bottom of the utility poles, as shown in Fig. 2(b).

4.2 Accuracy Evaluation Methods

The method of Moritani et al. is used as the evaluation value to assess accuracy. Moritani et al. use the following six indices. The evaluation method predicts the quality of a high-density model based on the tie points that constitute the vertices of the 3D model and the camera pose of the captured image.

① Reliability F_R (Fig. 3(a))

Evaluate the number of images with visible tie points.

$$F_R(i) = |V_i| \quad (12)$$

② Average of mesh area F_a (Fig.3(b))

The average surface mesh area $A_{i,n}$ of the n th approximate object connected to the tie point i on the surface of the approximate object is denoted by $F_a(i)$ and evaluated using Equation (13).

$$F_a(i) = \frac{1}{N_{A^i}} \sum_{n=0}^{N_{A^i}} A_n^i \quad (13)$$

③ Average of mesh edges F_e (Fig.3(c))

On the approximate object surface model, the edge lengths e^i connected to tie points i are calculated and their average value F_e is evaluated using Equation (14). where F_e is the number of edges connected to tie point i .

$$F_e = \frac{1}{N_{e^i}} \sum_{n=0}^{N_{e^i}} e_n^i \quad (14)$$

④ Average Baseline F_b (Fig.3(d))

Calculate the distance between the centers of projection c_j and c_k of the two visible cameras j and k from tie point i and evaluate it using Equation (15) with F_b as the average value for all camera pairs.

$$F_b = \frac{1}{N_{c^i} c_2} \sum_{j=0}^{N_{c^i}} \sum_{k=j+1}^{N_{c^i}} \|c_j - c_k\| \quad (15)$$

⑤ Average Height F_h (Fig. 3(e))

The distance from point c_j, c_k , which bisects the baseline of the two cameras j, k visible from tie point i , to tie point i is calculated and evaluated using Equation (16) with F_h as the average value for all camera pairs.

$$F_h = \frac{1}{N_{c^i} c_2} \sum_{j=0}^{N_{c^i}} \sum_{k=j+1}^{N_{c^i}} \|p_i - c'_{jk}\| \quad (16)$$

⑥ Base line ratio F_{bh}

The ratio of the average baseline length to the average subject distance at tie point i is F_{bh} and is evaluated using Equation (17).

$$F_{bh} = \frac{F_b(i)}{F_h(i)} \quad (17)$$

Each quality prediction index $F_X(i)$ ($X \in R, a, e, bh$) derived in Equations (12)-(17) is normalized using Equation (18) below and converted to $E_X(i)$ ($\in [0,1]$).

$$E_X(i) = 1 - L(F_X - \mu_X, \sigma_X) \quad X \in [R, a, e, bh] \quad (18)$$

However, μ_X is the mean of indicator $F_X(i)$, σ_X is the standard deviation, and $L(F_X - \mu_X, \sigma_X) = 1/(1 + \exp(\frac{-2(x-\mu)}{\sigma}))$ is the standard deviation.

$$E_{deg}(i) = (E_R + E_a + E_e + E_{bh})/4 \quad (19)$$

Finally, the average of the six energies $E_X(i)$ values is used to evaluate the degree of shape degradation, $E_{deg}(i)$, using Equation (19). $E_{deg}(i)$ is a value between 0 and 1. The higher the value, the greater the shape degradation and the less accurate the 3D model.

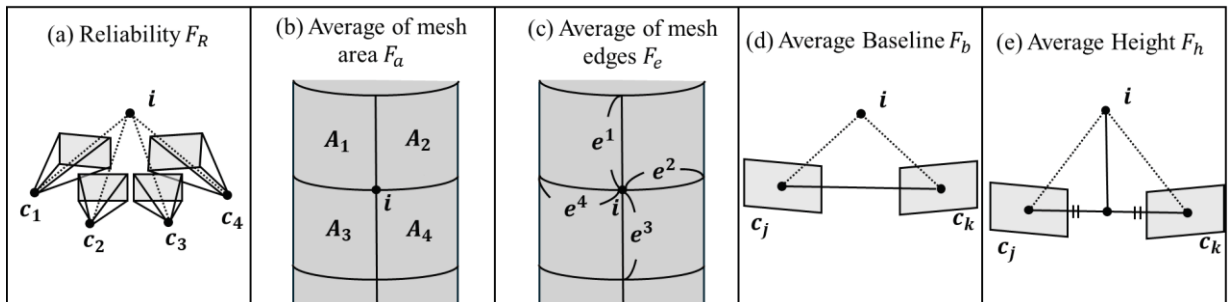


Figure 3: Summary Chart of Quality Prediction Indicators

4.3 Methods of efficiency evaluation

The efficiency evaluation in this study uses the aerial shooting time per utility pole. The procedure for calculating the specific aerial shooting time varies depending on the shooting method. In the conventional method, the drone must stop at each shooting point, so acceleration and deceleration are performed between each shooting point according to the acceleration of the drone. From the above, the aerial shooting time t between each shooting point in the conventional method is $t = 2 \times \sqrt{L/a}$, where a is the acceleration.

In contrast, the proposed method in this study requires the aircraft to fly at a constant speed v without decelerating. Therefore, the shooting time t between each shooting point in the proposed method is $t = L/v$.

5 CONCLUSION

This paper aims to reduce the cost of remote inspection of utility poles by utilizing 3D models generated from aerial drone images. 3D models allow inspectors to inspect from any viewpoint without having to go to the site. Thus, time efficiency is improved, and utility poles can be inspected more efficiently. On the other hand, more efficient travel routes are needed to photograph and 3D model utility poles that exist over a wide area.

In this paper, we propose an algorithm for generating shooting points and moving travel paths that can both reduce shooting time and generate highly accurate 3D models. Specifically, multiple shots are taken without discretely moving the index shooting points. This created a flight route that does not require extra detours that cause travel speed reduction. The above method reduces the time required to take the images needed to generate the 3D model.

In addition to the above proposed algorithm, an evaluation plan was introduced in this paper. This plan shows a method for comparing the proposed algorithm with the conventional method of finding the route with the least number of shots. In the future, we will show the effectiveness of the proposed algorithm based on the compatibility evaluation of accuracy-by-accuracy evaluation index and efficiency by aerial shooting time.

REFERENCES

- [1] Jamesine Rogers Gibson, Built to Last Challenges and Opportunities for Climate-Smart Infrastructure in California, Union of Concerned Scientists, pp.6 (2017).
- [2] Fact Sheet: The President's Plan to Make America a Magnet for Jobs by Investing in Manufacturing, <https://obamawhitehouse.archives.gov/the-press-office/2013/02/13/fact-sheet-president-s-plan-make-america-magnet-jobs-investing-manufactu> (23 July 2024)
- [3] Highway History, In Memory of Ronald Reagan, <https://www.fhwa.dot.gov/infrastructure/reagan.cfm> (23 July 2024)
- [4] Ministry of Land, Infrastructure, Transport and Tourism, Changes in utility pole phoenixes, https://www.mlit.go.jp/road/road/traffic/chicyuka/chi_13_03.html (23 July 2024)
- [5] Tomohiro Mizoguchi, Evaluation of Classification Performance of Pole-Like Objects from MMS Images Using Convolutional Neural Network and Image Super Resolution, International Journal of Automation Technology, Vol.12, No.3, pp.369-375 (2018).
- [6] Zhang Tian, Liu Jianhua, A 3D reconstruction method for pipeline inspection based on multi-vision, Measurement, Vol.98, No.5, pp.35-48 (2017).
- [7] Hiroyuki Miyake, Makoto Fujiu, Shoichiro Nakayama, Junichi Takayama, EVALUATION OF INFRASTRUCTURE CONDITIONS BY 3D MODEL USING DRONE, 4th International Conference on Road and Rail Infrastructure, Vol.9, No.3, pp.597-602 (2016).
- [8] Ministry of Land, Infrastructure, Transport and Tourism, Securing infrastructure development leaders, improving on-site productivity, introducing new technologies, etc., <https://www.mlit.go.jp/hakusyo/mlit/h27/hakusho/h28/html/n1321000.html>(23 July 2024)
- [9] Joel B. Sankey, Sujith Ravi, Cynthia S. A. Wallace, Robert H. Webb, Travis E. Huxman, Quantifying soil surface change in degraded drylands: Shrub encroachment and effects of fire and vegetation removal in a desert grassland, Journal of Geophysical Research: Bio geosciences: Volume 117, Issue G2(2012)
- [10] Daeyoon Moon, Suwan Chung, Soonwook Kwon, Jongwon Seo, Joonghwan Shin, Comparison and utilization of point cloud generated from photogrammetry and laser scanning: 3D world model for smart heavy equipment planning, Automation in Construction, Volume 98, Pages 322-331(2019)
- [11] Sungsik Yoon, Gi-Hun Gwon, Jinhwan Lee, Hyung-Jo Jung, Three-Dimensional Image Coordinate-Based Missing ROI Area Detection and Damage Localization for Bridge Visual Inspection Using Unmanned Aerial Vehicles, Structural Health Monitoring, Vol.20, No.4, pp.1462-1475 (2020).
- [12] Ryota Moritani, Satoshi Kanai, Hiroaki Date, Yasuhito Niina, Ryohei Honma, View planning for high-quality and efficient as-is model reconstruction by SfM-MVS (6th report) — Estimation of the additional camera poses based on the extracted low-quality regions —, 2020 Journal of the Japan Society of Precision Engineering, No.22, pp.43-44 (2020).
- [13] Kento Yamazaki, Kohei Okahara, Akira Minesawa, Proposal of an optimal imaging planning method for 3D reconstruction using geospatial information (in Japanese), IPSJ Symposium series DICOMO Multimedia, Distributed, Cooperative, and Mobile Symposium, (DICOMO 2023), pp.611- 616 (2023).
- [14] Hiroshige DAN, Naoki INAZU, Taira OZAKI, Satoshi KUBOTA, and Yoshihiro YASUMURO, OPTIMAL OF UAV PATH PLANNING FOR AERIAL SHOOTING SURVEY OF OUTDOOR STRUCTURES, Japanese Journal of JSCE F3, Vol.74, No.2, pp.159-166 (2020).

Investigation on Access Performance of Table Partitioning in PostgreSQL

Yoshiharu Kikuchi*, Sora Terao*, Takuro Sakiyama*, and Mikiko Sode*

*National Institute of Technology (KOSEN), Niihama College, Japan
e1212018@niihama.kosen-ac.jp, m.sode@niihama-nct.ac.jp

Abstract Companies are becoming increasingly internationalized. Many companies have factories located all over the world. For such companies, how to manage factories in distant locations and achieve the best performance is an important issue. In recent years, factories have been increasingly adopting the IoT, and methods have been proposed for collecting sensor data in the area where the headquarters is located and using that data to manage the factory. In such cases, it is becoming increasingly important to know how to store the data being sent in real time in a single database in real time while analyzing it. This paper reports the results of our study on a database to be used in a method for collecting sensor data in real time from factories in multiple regions and analyzing that data in real time.

Keywords: Database, Multi-core, Parallel, Factory, IoT.

1 INTRODUCTION

Companies are becoming increasingly internationalized. Many companies have factories located all over the world (Figure 1). For such companies, how to manage factories in distant locations and achieve the best performance is an important issue. In recent years, factories have been increasingly adopting the IoT, and methods have been proposed for collecting sensor data in the area where the headquarters is located and using that data to manage the factory. There are many different types of equipment in factories. In recent years, the number of factories that manage human movements is on the rise. This means that a large amount of data is generated in real time. In such cases, it is becoming increasingly important to know how to store the data being sent in real time in a single database in real time while analyzing it [1, 2].

Existing relational databases (RDBMSs) are built on the premise of a small number of cores and limited memory capacity, and the basic architectural concept has not changed for a long time, making it difficult to demonstrate performance on new hardware architectures (many-core, large memory). For example, PostgreSQL is known to have poor scalability from 8 to 16 cores compared to 4 to 8 cores, so it is often used on 8-core machine configurations from a cost-performance perspective [3].

PostgreSQL supports multiple processes, so by splitting one process into multiple independent processes and executing them in parallel, it is possible to utilize multiple cores and shorten the processing time. However, since a high degree of parallelism means a large amount of processing per unit time, the load on the storage naturally increases, and the processing performance becomes saturated due to storage bottlenecks.

In this paper, we investigate the performance of table partitions in PostgreSQL. Partitioning is a function that divides and stores data in a table in a database. Reducing I/O by partitioning tables by specifying filtering conditions in SQL, you can narrow down the range of access to only specific partitions. We also investigate the difference in I/O access between SSDs and HDDs.

Tsurugi [4] is designed to match new hardware architectures (many-core, large capacity memory). Tsurugi is known for its high processing performance, as it has better write performance than traditional RDBs and scales in-memory/many-core (especially according to the number of cores). This paper describes the results of our investigation into methods for utilizing relational database Tsurugi in factory applications. And we report on our investigation into a method for collecting sensor data from factories in multiple regions in real time and analyzing that data in real time.

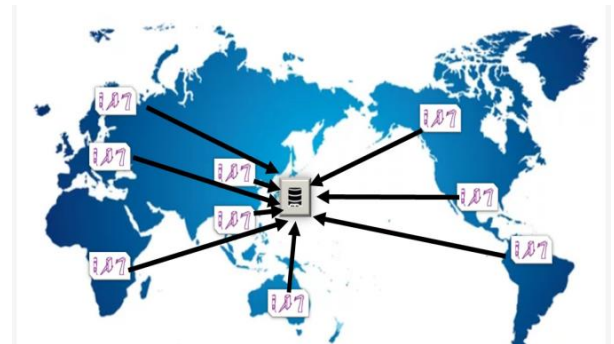


Figure 1: A system that collects data from factories around the world.

2 PROPOSED DATA COLLECTION SYSTEM FOR FACTORIES

When a machine makes an anomalous noise, it is often necessary to take measures such as stopping the factory lines. As a result, losses are often very large. Therefore, it is necessary to detect abnormalities early and take measures. Only experienced engineers can naturally detect the slightest anomalous machine noises, and up until now, maintenance and inspection of equipment at factory production sites has relied on the experience and intuition of veteran engineers. In recent years, factories are becoming increasingly smart factories. This technology detects failures and signs of failure in mechanical equipment without relying on human intuition.

In the field of anomalous sound detection, it had been mainstream to use voice data directly for machine learning

to detect anomalous sounds. In the recent years, there is a fact that AI in the field of image processing is evolving. To utilize AI's image recognition ability, we have been proposed the method that converts sounds in a factory into images using wavelet transformations and uses these images for input of machine learning[5-8]. There, in addition to various data such as temperature, humidity, and luminosity, we collect sound data converted into images.

3 TABLE PARTITIONS IN POSTGRESQL

3.1 Benefits of PostgreSQL table partitions

Existing PostgreSQL implements three types of partitioning features:

- I. Range Partition: Divides tables based on a specified range or interval of key values.
- II. List Partition: Divides tables based on predefined specific key values.
- III. Hash Partition: Divides tables based on the hash values derived from specified key values.

Each partitioning method contributes to improving access performance through distributed I/O processing and enhances maintainability. By specifying filtering conditions in SQL, you can narrow down the range of access to only specific partitions. In addition, frequently used parts of a partition are more likely to be cached in memory, which reduces disk I/O and improves access performance. In addition, split tables are placed in different table spaces on physical disks. This allows reading and writing to separate physical disks in parallel, distributing disk I/O and improving processing performance.

3.2 PostgreSQL Experimental Results

We explain the access performance evaluation experiments for HDD and SSD. We conducted repeated data registration from multiple PCs to a single DB server to evaluate access efficiency. On the DB server side, we used both regular tables and tables partitioned by the List Partition method. To investigate the difference in I/O performance, we compared scenarios where the database storage was allocated on HDDs and SSDs. On the client side, we simulated data writes from multiple IoT devices by repeatedly registering data in short cycles using multiple processes on a single PC.

The system architecture is explained below. There is one DB server, and storage is allocated to HDD or SSD. Table or list partition table (divided by client host name) is used. Also, there are three clients, each with eight processes writing to the server at the same time. The clients were implemented using PHP. We developed a program that continuously sends SQL requests using the `pg_query()` function while maintaining a connection to the DB server, with a random wait time (2 to 5 ms) between each request, for a total of 500,000 iterations. By recording the client's timestamp at the request time and the server's timestamp at the data registration time, we measured the latency in the data registration process. This program was executed on

three clients with different network environments, with eight concurrent processes running on each PC.

PostgreSQL was used as the DB system, and we compared cases with and without partitioning. To assess the impact of storage access speed, we compared scenarios where the database storage was allocated on HDDs and SSDs.

The results of the comparison experiment between HDD and SSD are shown in Figure 2 and Table 1. We compared the time difference between the client and server timestamps when the database storage was allocated on HDDs versus SSDs, under conditions without partitioning. The difference between the client timestamp and the server timestamp showed no significant variance between the HDD and SSD configurations. This time difference is primarily attributed to network latency and does not appear to be dependent on the load of the DB server.

However, there was a large difference in the intervals between the server timestamps of the data registered on the DB server, as shown in Figure 3 and Table 2, 3. While the experiment took over 2 hours and 30 minutes to complete using the HDD, it was completed in less than 1 hour and 30 minutes with the SSD. Since neither the client settings nor the program was altered, this difference can be attributed to the time required to execute the `pg_query()` function. The `pg_query()` function sends an SQL request to the DB server, waits for the execution result, and then completes. Assuming that network latency is consistent, the time required for this function can be regarded as the time taken by the DB server to process the registration.

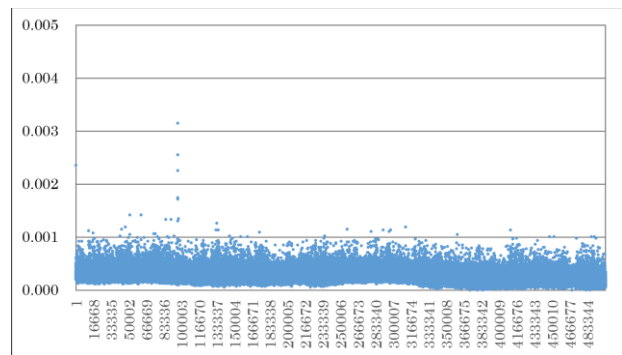


Figure 2(a): DB storage on HDDs.

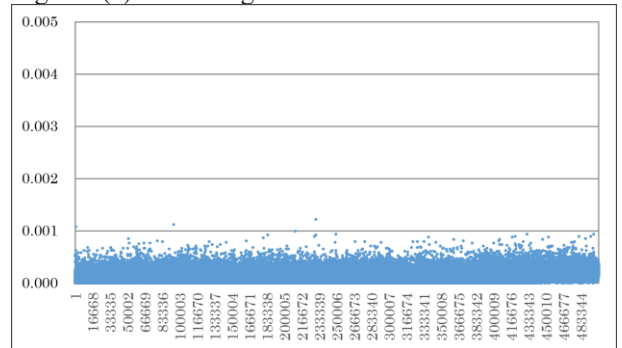


Figure 2(b):DB storage on SSDs

Table 1. Average time difference

HDD	SSD
0.254ms	0.191ms

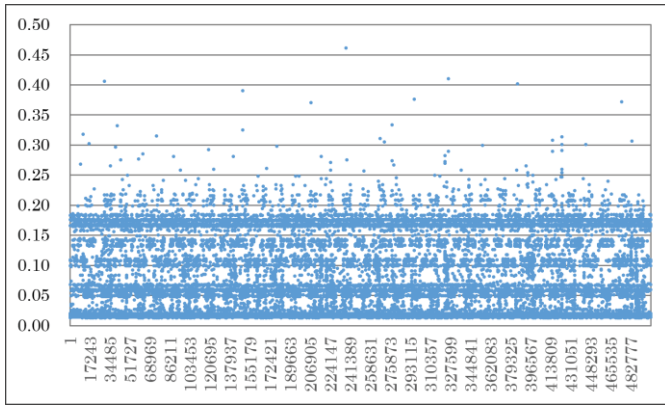


Figure 3(a): Intervals between server timestamps (on HDD)

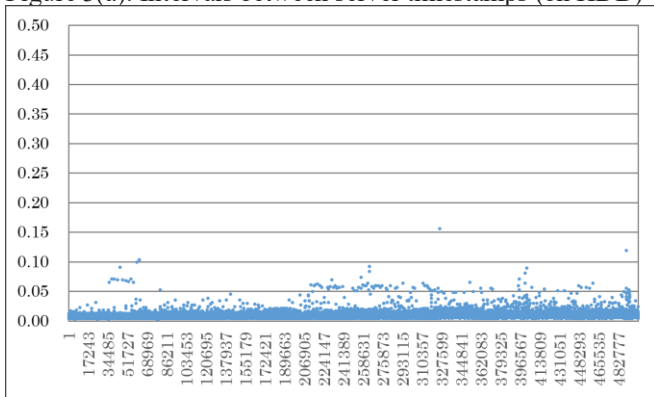


Figure 3(b): Intervals between server timestamps (on SSD)

Table 2 Average intervals between server timestamps

HDD	SSD
20.190ms	9.765ms

Table 3 Time required to complete the experiment

HDD	SSD
2:48:25	1:21:22

The effect of table partitions was investigated using a machine equipped with an Intel Xeon 16-core CPU, 132GB of memory, and 1TB of SSD MVME storage. PostgreSQL ver14 was installed on Ubuntu. The IoT table structure is: device_id, device_name, factory_id, factory_name, temperature, humidity, pressure, illuminance, rssi, battery, dat timestamp with time zone and there are 1 million records registered. We partitioned the table by factory_id. The results are shown in Table 4. It was confirmed that selecting factory_id without partitioning took 215.803 ms, and 125.410 ms with partitioning, which is roughly half the processing time. On the other hand, selecting device_id from a database partitioned by factory_id took 214.996 ms, which confirmed that it is meaningless not to partition depending on the purpose. The fixed table partition approach is not suitable for IoT.

Table 4 Effect of table partitioning

Select method	Processing time
factory_id	125.410ms
device_id	214.996ms

4 LARGE-SCALE DATA PROCESSING USING MULTI-CORE DATABASES

Existing relational databases(RDBMSs) are built on the premise of a small number of cores and limited memory capacity, and the basic architectural concept has not changed for a long time, making it difficult to demonstrate performance on new hardware architectures (many-core, large memory). We decided to choose the relational database Tsurugi [4] to solve these problems. Tsurugi is designed to match new hardware architectures (many-core, large capacity memory). Tsurugi is known for its high processing performance, as it has better write performance than traditional RDBs and scales in-memory/many-core (especially according to the number of cores).

We looked into Tsurugi as a factory database. It has built-in mechanisms for multi-core processing, such as automatic partitioning, so we evaluated its IoT features.

5 EXPERIMENT RESULTS: RELATIONSHIP BETWEEN WI-FI AND ETHERNET

5.1 Method of Compression Experiments

We investigated whether there was a difference between data collection via Wi-Fi and via wired LAN. The images are shown in Figures 4 and 5. Assuming that the Raspberry Pi is an IoT terminal, we will install the client software and write data to the database.

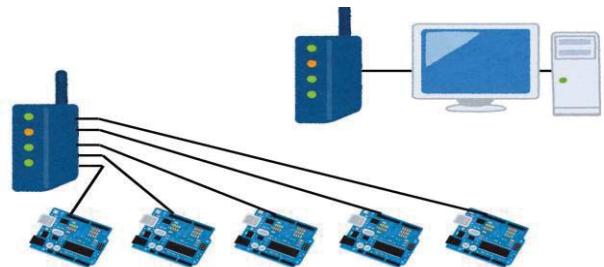


Figure 4: Data collection via wired LAN



Figure 5: Data collection via Wi-Fi

5.2 Compression of Wi-Fi AND Ethernet

We set up the server machine so that Tsurugi uses 3 CPUs, wrote data from the IoT device using the Insert command, and measured the time it took to write. The measurement was done by acquiring time information before executing the Insert command, then collecting time information again after the Insert command was executed and the completion signal arrived from the server, and calculating the difference. The results are shown in Figures 6 and 7. Since there is no difference in processing time, it is determined that there is no difference due to the route.

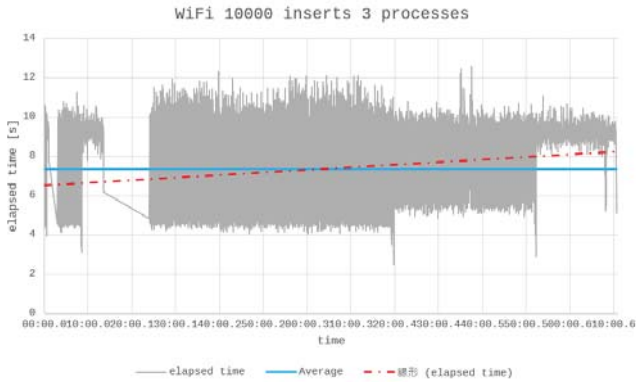


Figure 6: Results over Wi-Fi: 10,000 Insert commands executed on 3 CPUs on the server.

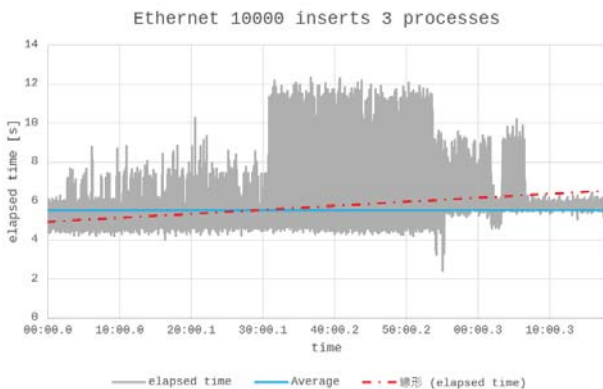


Figure 7: Results over Wired LAN: 10,000 Insert commands executed on 3 CPUs on the server.

6 EXPERIMENTAL RESULTS: RELATIONSHIP BETWEEN DATA VOLUME AND CPU

We evaluated the difference in processing time depending on the amount of data. Figures 8 and 9 show the results of sending 100 and 1,000 pieces of data, with three CPUs specified on the server. It can be seen that as the amount of data increases, the processing time increases, but the increase is small.

Next, the difference in processing time depending on the number of CPUs used is shown in Figures 10, 11, and 12. These are the results of executing the Insert command 1,000 times. We can see that the processing time tends to decrease

as the number of CPUs increases. Figure 13 shows the CPU operation status of the server when data is inserted.

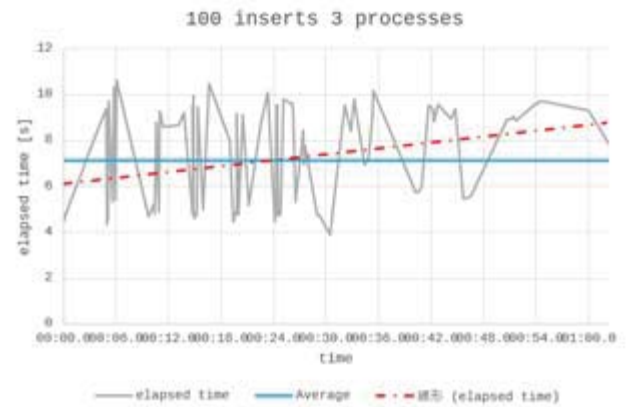


Figure 8: Result of 100 Inserts with Tsurugi 3 CPU.

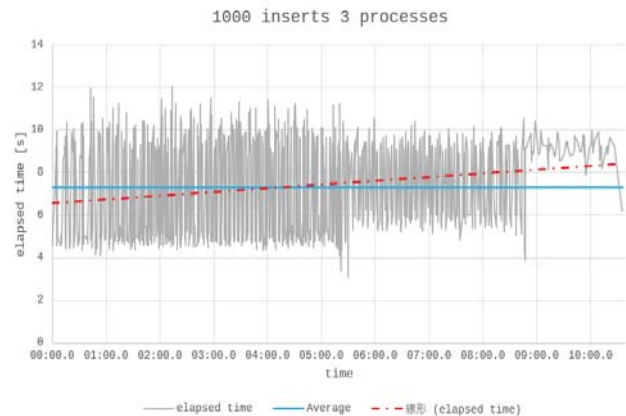


Figure 9: Result of 1000 Inserts with Tsurugi 3 CPU.

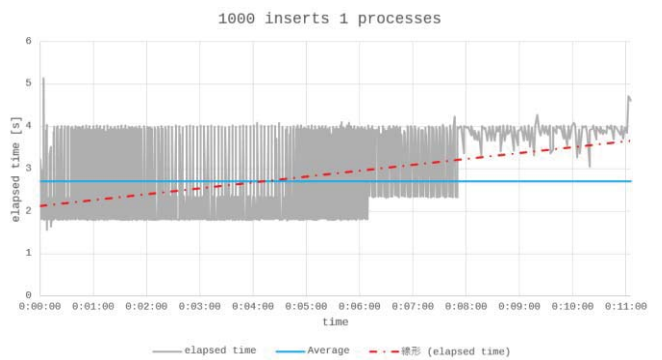


Figure 10: Result of 1000 Inserts with Tsurugi 1 CPU.

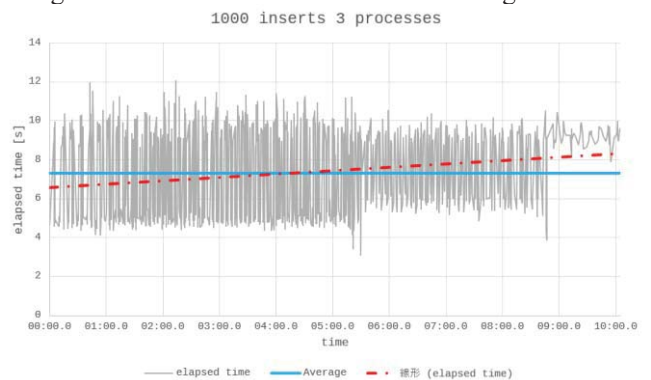


Figure 11 Result of 1000 Inserts with Tsurugi 3 CPU.

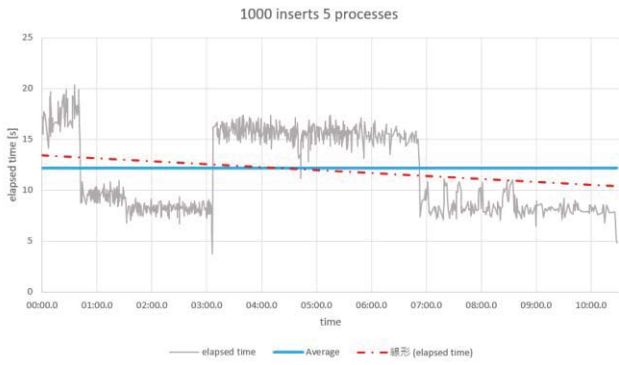


Figure 12: Result of 1000 Inserts with Tsurugi 5 CPU.

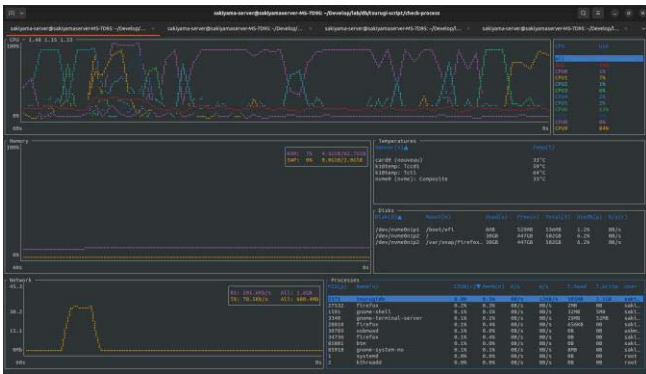


Figure 13: Server CPU usage status.

7 SUMMARY

As companies become more internationalized, an increasing number of companies have factories all over the world. For such companies, how to manage factories in remote locations and ensure their best performance is an important issue. In recent years, factories have become more IoT-enabled, and methods have been proposed for collecting sensor data from the headquarters area and using that data to manage the factory. In this case, it is important to accumulate and analyze the data sent in real time in a single database in real time. In this paper, we evaluated PostgreSQL data partitions and Tsurugi's multi-core as a method and study for collecting sensor data from factories in multiple regions in real time. Having confirmed the effect of multi-core, we plan to continue evaluating Tsurugi and verify its functions for IoT.

REFERENCES

[1] S. Park and J. -H. Huh, "A Study on Big Data Collecting and Utilizing Smart Factory Based Grid Networking Big Data Using Apache Kafka," in *IEEE Access*, vol. 11, pp. 96131-96142, 2023, doi: 10.1109/ACCESS.2023.3305586.

[2] Qingxin Xia, Joseph Korpela, Yasuo Namioka, Takuya Maekawa, "Robust Unsupervised Factory Activity Recognition with Body-worn Accelerometer Using Temporal Structure of Multiple Sensor Data Motifs," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, Volume 4,

Issue 3 Article No.: 97, Pages 1 – 30 <https://doi.org/10.1145/3411836>.

[3] Kenji Yoneda, Yoshifumi Kumano, "Evaluation and Improvement of the CPU Scalability for PostgreSQL" *UNISYS TECHNOLOGY REVIEW* Vol.94, NOV. 2007, <https://www.biprogy.com/pdf/9404.pdf>.

[4] Toshi Kanbayashi, Suguru Arakawa, Masato Hishida, Shinichi Umekin, Akira Kawaguchi, Takashi Horikawa, Takayuki Tanabe, Ryoji Kurosawa and Kohei Okada, "Next-generation high-speed open source RDB Tsurugi" *Nikkei BP*, October 5, 2023.

[5] Masaya Ueda, Daisuke Tanaka, Mikiko Sode, "A study on anomalous sound detection in factories for early failure detection using wavelet transform," *IEEE International Conference on Consumer Electronics – Taiwan, 2024 (IEEE ICCE-TW)*, July 9-11, 2024.

[6] Keisuke Kikuchi, Mikiko Sode, "Anomalous Sound Analysis Using Wavelet Transform," *The 7th NIT-NUU Bilateral Academic Conference 2024*, July 14-15, 2024.

[7] Hiroki Yamamoto, Mikiko Sode, "Anomaly Analysis Using Wavelet Transform," *The 7th NIT-NUU Bilateral Academic Conference 2024*, July 14-15, 2024.

[8] Syuuno, Mikiko Sode, "," *The 7th NIT-NUU Bilateral Academic Conference 2024*, July 14-15, 2024.

[9] S. Dominico, E. C. de Almeida, J. A. Meira and M. A. Z. Alves, "An Elastic Multi-Core Allocation Mechanism for Database Systems," *2018 IEEE 34th International Conference on Data Engineering (ICDE)*, Paris, France, 2018, pp. 473-484, doi: 10.1109/ICDE.2018.00050.

Session 4:
Multimedia and Communication
(Chair: Kozo Okano)

Audio Signal Compression in Surround Environments Applying Wavelet Transform

Miyuki Shirai*, Yuhi Shuno*, Hiroki Yamamoto*, Sho Ishikawa*, Mikiko Sode*

*National Institute of Technology (KOSEN), Niihama College, Japan
{ m.shirai, m.sode }@niihama-nct.ac.jp

Abstract When a machine makes an anomalous noise, it is often necessary to take measures such as stopping the factory lines. Thus, we have been making a system to detect machine failure using sounds. The feature of the proposed system is that it converts voice data into an image using wavelet transform, and then uses the image as input to determine anomalies using machine learning. The important thing in this system is the size of the audio data. We want to compress audio data and make it easier to transmit data. In this paper, audio signal compression using wavelet transform is discussed. We will consider reducing the size of audio data without removing abnormal sounds contained in the audio data. The features of abnormal sound on the time-frequency plane by applying some different conversion methods are compared each other.

Keywords: Wavelet, Features extract with sound processing.

1 INTRODUCTION

When an abnormality occurs in a machine operating in a factory, it is necessary to stop the factory line in order to deal with the problem. However, this reduces the factory's operating hours, and if the abnormality goes unnoticed for a long period of time, even if the factory line is stopped and measures are taken, it may result in significant losses for the company. Therefore, it is necessary to detect abnormalities or signs related to abnormalities and respond to them quickly. Factory equipment always makes some kind of noise, and if an abnormality occurs, the machines may make sounds that are different from normal sounds. Only experienced engineers can naturally detect the slightest anomalous machine noises, and up until now, maintenance and inspection of equipment at factory production sites has relied on the experience and intuition of veteran engineers. In recent years, due to the influence of generational change in companies, there has been an issue regarding the succession of maintenance work techniques at factories and other work sites. Thus, factories are becoming increasingly smart factories. This technology detects failures and signs of failure in mechanical equipment without relying on human intuition. In factories and other places, the presence or absence of faults or failures is often based on sounds observed at specific locations, and a system that observes the sounds at relevant locations in a noisy environment and notifies of abnormalities is useful, so various methods have been proposed[1].

We have proposed a method in which audio data of inside a factory is processed to image data using wavelet transformation, its characteristics are clarified using some processes, and then machine learning is applied [2-4]. By grasping the characteristics from the time variation of the

frequency characteristics, the accuracy of judging normality or abnormality could be improved.

Figure 1 shows the overall figure of the system we are developing. In the first stage, machine learning is performed using only normal sounds. First, we apply wavelet transformation to each sound, visualize each sound data. Second, feature extraction processes are applied to the image data. This becomes the input data for machine learning. The wavelet transform uses a basis that shortens the time width at high frequencies and widens the time width at low frequencies, thus providing local frequency information and efficient time-frequency analysis. Time domain is essential for environmental sounds and the operating sounds of factory machinery. The AI learned from this input data is used to analyze the voice data collected in real time, and if an event occurs in which abnormal voice is observed, an alert is sent to the administrator (by email, etc.).

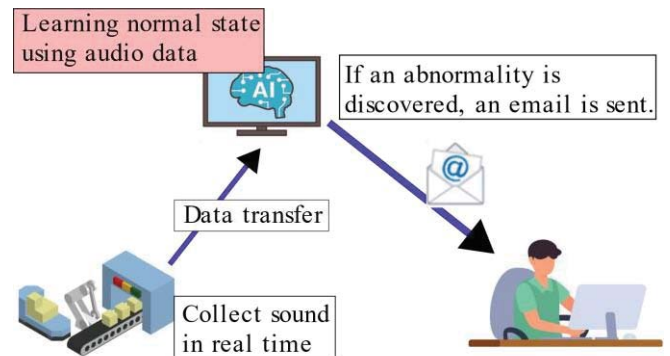


Figure 1: Proposed remote monitoring type system configuration for anomalous sound detection.

An important point in our system is how to collect and transfer audio data. There are multiple machines in a factory and multiple factories, but it is not realistic to install a server for each machine and each factory for the collected data. It is desirable to transfer the collected data in real time and consolidate it on one server.

In such a situation, the size of the voice data is important. The smaller the size of the voice data, the easier it is to collect and transfer. However, this requires that the compressed voice data can be analyzed to determine whether it is a normal sound or an abnormal sound. This will make it possible in the future to build a system that stores collected data on the cloud, as shown in Figure 2, and allows for real-time understanding of the situation from remote locations.

In this paper, we examined the compression method of audio data for anomalous sound detection system in factories for early failure detection using machine learning, thus we

will report. In this study, we hypothesize a system that observes and monitors sound, compresses and analyzes the sound using wavelet transform, and aims to verify the possibility and usefulness of applying this to sound source preservation and anomaly detection.

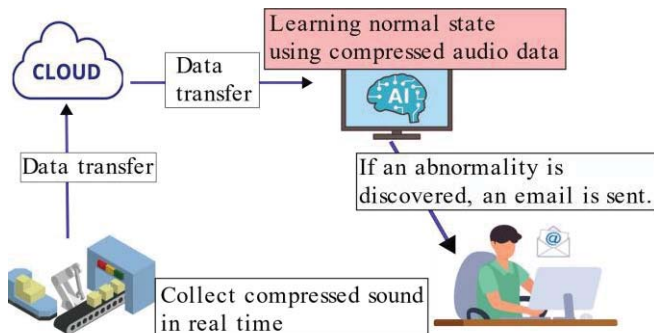


Figure 2: Proposed remote monitoring type system configuration for anomalous sound detection in the future.

2 WAVELET TRANSFORM

There are two types of wavelet transform. One is continuous wavelet transform(abbreviated as CWT in the following text and figures) and the other is discrete wavelet transform(abbreviated as DWT in the following text and figures).

The wavelet transform is known as a method for analyzing both time and frequency. The wavelet transform is a method that calculates accurate frequency information by increasing the time width in the low-frequency region and calculates accurate time information by decreasing the time width in the high-frequency region. In the wavelet transform, the original waveform is expressed as an appropriate waveform $\Psi(t)$. This $\Psi(t)$ is called the mother wavelet, and the one that is appropriate for the waveform to be analyzed is selected appropriately. The mother wavelets used in this study are Morse, Molet, and Bump for the CWT, and SYM4 and db DB4 for the DWT[5].

The CWT has applications such as detecting abnormal signals, and the DWT is used as a standard for image compression, and its applications are being actively discussed.

A typical analysis method used for audio signals is the Fourier transform, which includes frequency information but loses time domain. In contrast, the wavelet transform preserves time information, making it possible to analyze even sudden signal fluctuations. It is also possible to perform flexible operations such as varying the time interval according to the frequency domain, making it possible to perform dynamic analysis according to the situation.

SYM4 wavelet and db4 wavelet are particularly suitable for analysis of biological signals, such as detection of QRS complexes in ECG (ECG). The QRS complex is the most prominent feature of the electrocardiogram and reflects the depreciation of the left and right ventricles[6].

The waveform that informs the abnormalities of the machine often has a signal where the amplitude of the waveform increases rapidly. It is similar to the waveform of

the electrocardiogram, and we thought that SYM4 and DB4 suitable for electrocardiograms may be effective.

3 EXPERIMENT RESULTS: COMPRESSION OF AUDIO SIGNALS USING DISCRETE WAVELET TRANSFORM

3.1 Method of Compression Experiments

We conducted an experiment to confirm the effect of compression using the DWT. The items we confirmed were:

1. Whether compression can be used to determine whether something is normal or abnormal

2. The amount of reduction in file size due to compression

The audio data used in the experiment was DCASE 2022 Challenge Task 2[7-9]. The original data to be compressed was a 10-second uncompressed file with a sampling frequency of 16,000 [Hz]. We used the sound of a fan as a model for factory sounds. We also used data in which abnormal sounds were mixed into the sound of the fan for comparison. To compress the audio, we used Matlab's Wavelet Toolbox in MATLAB R2024a.

3.2 Compression of Normal and abnormal data

First, we analyzed the normal sound of a fan. We performed a DWT using SYM4 as the mother wavelet. Next, we applied a cut using the same threshold at each of the four frequency decomposition levels, and samples with values below the threshold were set to zero. Figure 3 shows the chronological waveform diagram of the signal reconstructed based on the coefficient distribution of each decomposition level after cutting. The horizontal axis is the sampling number along the time series, and the vertical axis is the signal amplitude.

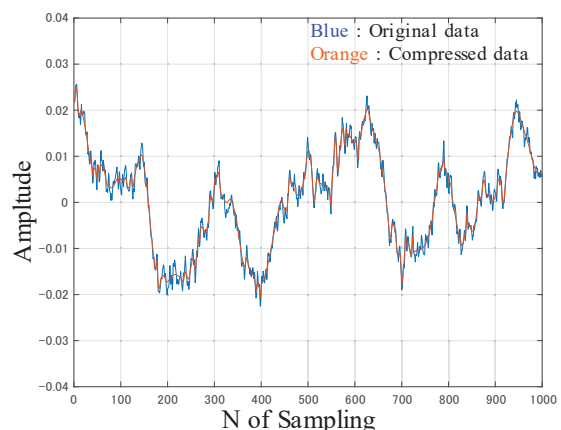


Figure 3: Comparison of Waveform diagram of Normal data using SYM4-Wavelet.

The reconstructed data was output as an audio file (wav file), and the audio file was converted to a flac file and compressed.

During the DWT, the ratio of the number of samples that did not become 0 when cut by the threshold to the total number of samples was defined as the "retention rate H" and calculated. It was found that the data could be classified into those with high and low retention rates. Figure 3 shown above corresponds to the group with high retention rate ($H = 0.857$). In contrast, Figure 4 shows a time series waveform diagram of data with low retention rate ($H = 0.570$).

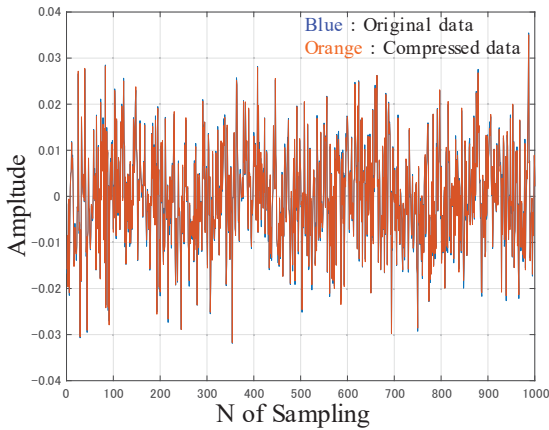


Figure 4: Comparison of Waveform diagram of data with low retention rate ($H = 0.570$) of Normal data using SYM4-Wavelet.

Figure 4 shows that the waveform oscillates violently with a very short period. When looking at the waveform diagram of the amplitude for each frequency decomposition level when DWT is performed for this data, the waveform oscillates with a constant amplitude at all decomposition levels. The data size after compression of such data was almost the same as the data size after compression without DWT. It is considered that data like this, which does not change with respect to the time domain and does not change the weight of each decomposition level, has no effect on compression by DWT.

Therefore, in the subsequent analysis, the retention rate was calculated when DWT was performed, and if it was 3σ smaller than the average value of the group of data with high retention rate ($H = 0.857 \pm 0.006$), it was not used in the analysis. For the data that was not used, we visually confirmed that all of it was inappropriate waveforms as explained above.

For the 13 pieces of normal sound data used in the analysis, the average retention rate was $H = 0.857 \pm 0.006$. In contrast, for the data that contained abnormal sounds, 13 pieces of data were similarly extracted from the group with a high retention rate, and the average retention rate was calculated, which was $H = 0.863 \pm 0.006$. The difference between these two values was 0.69σ , which is not significant.

Figure 5 shows the time domain waveform of the reconstructed signal after cutting the data containing abnormal sounds using DWT. It is similar to the waveform of normal sounds in Figure 3, and it can be seen that the amplitude of the signal waveform is gradually lost in both cases. However, when these signals were saved and played back and listened to by humans, no difference was perceived from the original voice data. The reason why the retention rate is higher when abnormal sounds are included is thought

to be because abnormal sounds expand the fluctuations in the frequency band.

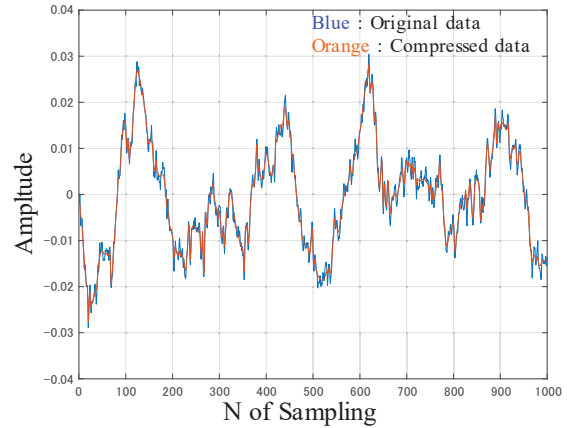


Figure 5: Comparison of Waveform diagram of Abnormal data using SYM4-Wavelet.

Next, to evaluate the amount of data file reduction achieved by running DWT, we took the ratio of the data size of each sound file when it was converted to a flac file before running DWT to the data size when it was converted to a flac file after running DWT. We defined this as the Compression Rate.

The results were $R = 0.744 \pm 0.031$ for normal sound data, and $R = 0.826 \pm 0.029$ for data containing abnormal sounds. This shows that by running DWT, file sizes are reduced by around 70% to 80%. In addition, the difference in compression ratio between normal sounds and abnormal sounds was 1.93σ , showing that data containing abnormal sounds was significantly exceeded.

This indicates the possibility of distinguishing between normal and abnormal sounds by evaluating the file size after file compression following DWT.

3.3 Dependence on Different types of Mother Wavelet

A similar analysis was performed using db4 to verify the impact of different types of mother wavelets. The results are shown in Table 1. The threshold value used when performing DWT was the same as that used in SYM4.

Table 1: Comparison of Compression rate R for other MotherWavelet

Mother wavelet	Compression rate R	
	NormalData	AbnormalData
SYM4	0.744 ± 0.031	0.826 ± 0.029
DB4	0.746 ± 0.029	0.826 ± 0.029

No significant difference in compression ratio was observed due to the difference in mother wavelets, SYM4 and DB4.

Figure 6 shows the chronological waveform diagram of the signal reconstructed based on the coefficient distribution of each decomposition level after performing DWT using DB4 as the mother wavelet of Normal data.

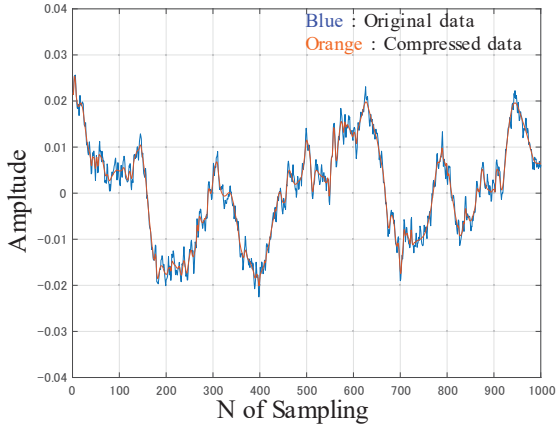


Figure 6: Comparison of Waveform diagram of Normal data using DB4-Wavelet.

And Figure 7 shows about Abnormal data in same analysis. In both case, there were no difference of reconstructed signal wave between Figure 3 and Figure 6 or Figure 5 and Figure 7 by human eye.

However, this may change if a different mother wavelet that matches the sound source is used. In order to compress data, it is necessary to have a lower Compression rate and not have humans feel uncomfortable when playing back the compressed sound, so it is necessary to find a mother wavelet that is useful for the target sound source.

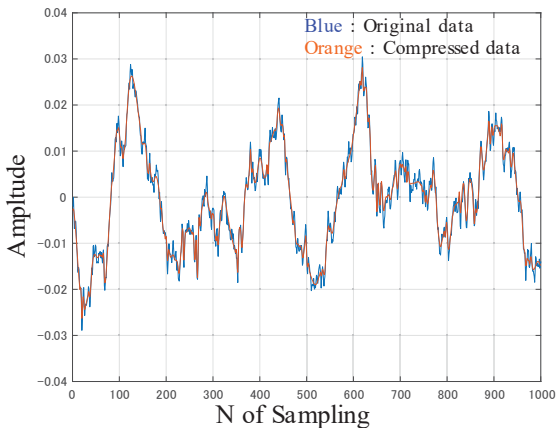


Figure 7: Comparison of Waveform diagram of Abnormal data using DB4-Wavelet.

4 EXPERIMENTAL RESULTS: CONTINUOUS WAVELET TRANSFORM OF COMPRESSED VOICE DATA

4.1 Comparison of Raw data and Compressed data

Next, a continuous wavelet transform was performed on both the normal and abnormal sounds in the compressed voice data to create a scalogram that represents the signal strength on the time and frequency plane.

Figure 8 shows the scalogram of Normal raw data where Morse was used as Mother Wavelet. Figure 9 shows the

scalogram on Compressed data using the DWT of SYM4 Wavelet.

From this, it was found that by first compressing using the discrete wavelet transform, the components corresponding to the white noise that are common to both were suppressed, resulting in a distribution that makes it easier to determine the characteristics.

The abnormal data were analyzed in a similar manner. The results were shows in figure 10 and 11. Abnormal data has a high retention rate, but it is observed as a more characteristic figure.

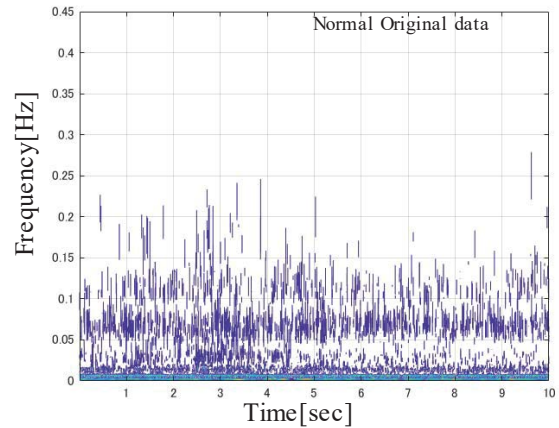


Figure 8: Scalogram of Normal Original Data

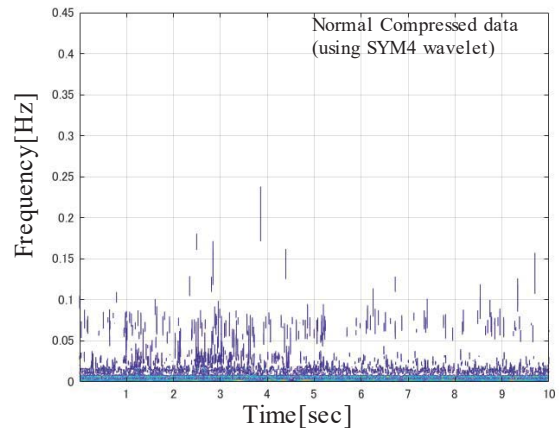


Figure 9: Scalogram of Compressed Normal Data (compressed using sym4 Wavelet)

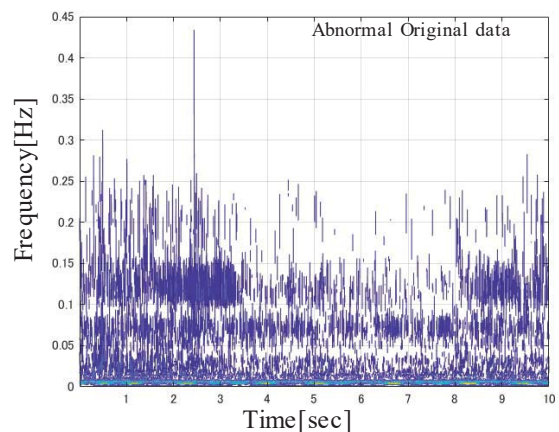


Figure 10: Scalogram of Abnormal Data

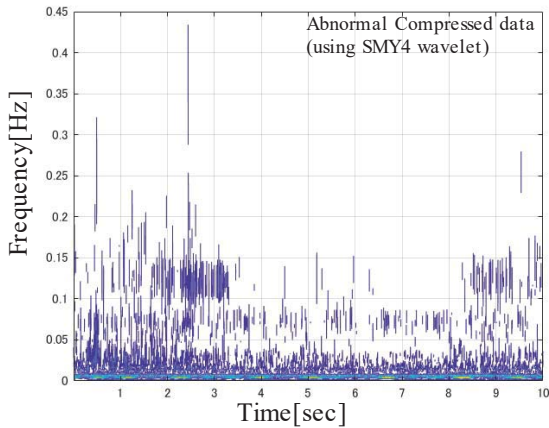


Figure 11: Scalogram of Compressed Abnormal Data (compressed using SYM4 Wavelet)

4.2 Dependence on MotherWavelet

To investigate the dependence on MotherWavelet, a similar analysis was performed for two types of MotherWavelet, amor(Morlet) and Bump, in addition to the basic Morse.

Figure 9 shows the case where the basic Morse Wavelet were used as mother wavelets for Normal data analysis, figure 12 shows amor(Morlet) and figure 13 shows Bump Wavelet.

From Figure 12, this scalogram was more widely distributed than in the case of Morse and Bump. So, Morlet was not suitable to handle this data.

From this, it was found that for continuous wavelets, differences in the scalogram appear depending on the mother wavelet.

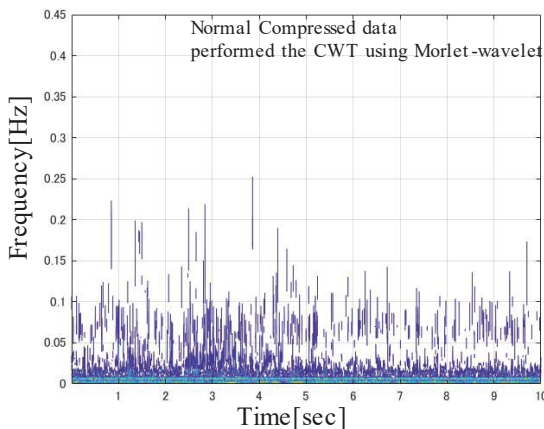


Figure 12: Scalogram of Compressed Normal Data for Morlet-Wavelet (compressed using SYM4 Wavelet).

The abnormal data were analyzed in a similar manner. The results were shown in figure 11, 14 and 15.

For Abnormal data, the distribution characteristics were completely different. By comparing Figure 11 and Figure 14, as expected, Morlet(amor) was found to be inappropriate for extracting features from this data.

On the other hand, in Figure 15, the scalogram using Bump-Mother Wavelet, it was found only in a few places, and was

very distinctive. Figure 16 shows the scalogram of uncompressed data transformed by continuous wavelet transform using Bump. We can see that Bump is a wavelet that is quite suitable for extracting features from this Abnormal data.

This can also be derived by comparing Figure 14 (Morlet-Wavelet) and Figure 15 (Bump-Wavelet).

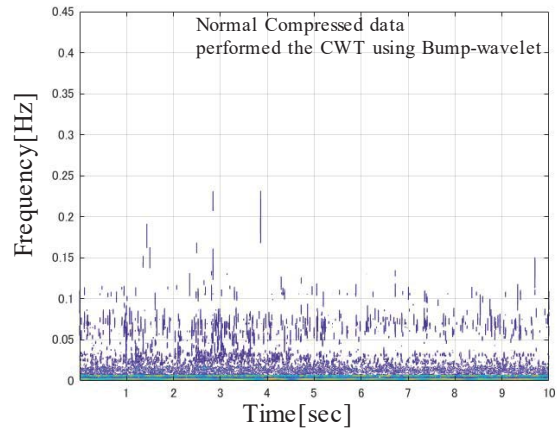


Figure 13: Scalogram of Compressed Normal Data for Bump-Mother Wavelet (compressed using sym4 Wavelet).

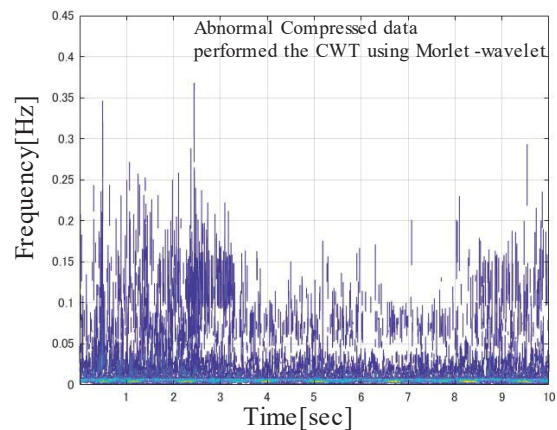


Figure 14: Scalogram of Compressed Abnormal Data for Morlet-Mother Wavelet (compressed using sym4 Wavelet)

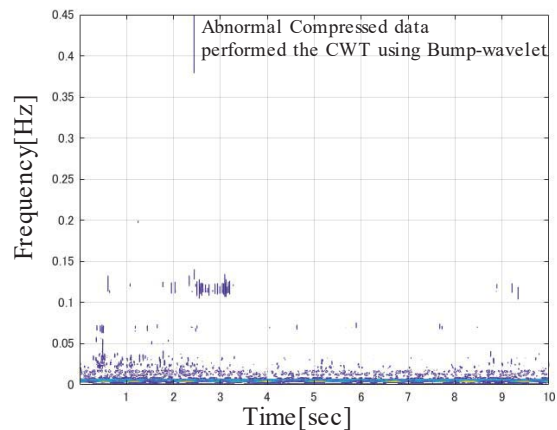


Figure 15: Scalogram of Compressed Abnormal Data for Bump-Wavelet (compressed using sym4 Wavelet).

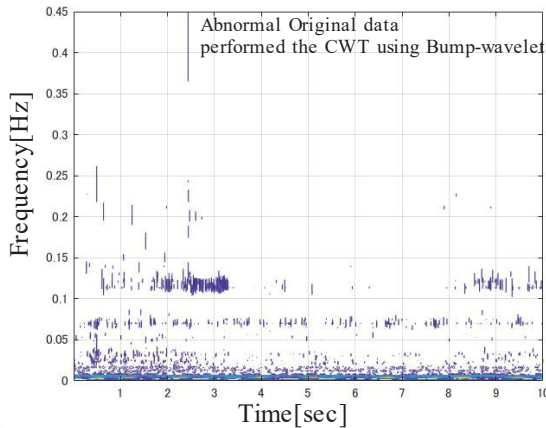


Figure 16: Scalogram of Abnormal Original Data for Bump-Wavelet(Uncompressed data).

So, it is useful to first identify the wavelet that most clearly shows the characteristics of the target sound source.

Since the range of the distribution of the scalogram is the very characteristics of the sound source, it is possible to use these distribution maps as data for machine learning to build a system that detects and judges abnormalities on behalf of humans.

These results confirmed that the proposed method makes it possible to visually determine whether data is normal or abnormal even after compression, and that the file size can be reduced to 70% of original data.

5 SUMMARY

In this study, data compression was performed using discrete wavelet transform with SYM4 and DB4 mother wavelets. As a result, it was found that using wavelet transform is useful as a method for compressing data within a range that does not sound strange to the human ear and then storing the data. In addition, by performing continuous wavelet transform on compressed data, we were able to obtain characteristic scalograms in some cases. In the future, it is possible to use the scalogram of the compressed audio signal as a simple two-dimensional image and use machine learning to determine the situation.

REFERENCES

- [1] Emad Awada, Aws Al-Qaisi, Eyad Radwan, Mutasim Nour, "Motor fault detection using sound signature and wavelet transform," *International Journal of Power Electronics and Drive Systems (IJPEDS)*, vol.13, pp247-255(2022).
- [2] Masaya Ueda, Daisuke Tanaka, Mikiko Sode, "A study on anomalous sound detection in factories for early failure detection using wavelet transform," *IEEE International Conference on Consumer Electronics – Taiwan, 2024 (IEEE ICCE-TW)*, July 9-11, 2024(2024).
- [3] Keisuke Kikuchi, Mikiko Sode, "Anomalous Sound Analysis Using Wavelet Transform," *The 7th NIT-NUU Bilateral Academic Conference 2024*, July 14-15, 2024(2024).
- [4] Hiroki Yamamoto, Mikiko Sode, "Anomaly Analysis Using Wavelet Transform," *The 7th NIT-NUU Bilateral Academic Conference 2024*, July 14-15, 2024(2024).
- [5] Choose a Wavelet, <https://jp.mathworks.com/help/wavelet/gs/choose-a-wavelet.html>(last visited: 07.20.2024)
- [6] Wavelet analysis of biological signals, <https://jp.mathworks.com/help/wavelet/ug/wavelet-analysis-of-physiologic-signals.html>(last visited: 07.23.2024)
- [7] Kota Dohi, Keisuke Imoto, Noboru Harada, Daisuke Niizumi, Yuma Koizumi, Tomoya Nishida, Harsh Purohit, Takashi Endo, Masaaki Yamamoto, Yohei Kawaguchi, Description and Discussion on DCASE 2022 Challenge Task 2: Unsupervised Anomalous Sound Detection for Machine Condition Monitoring Applying Domain Generalization Techniques. In *arXiv e-prints: 2206.05876(2022)*.
- [8] Kota Dohi, Tomoya Nishida, Harsh Purohit, Ryo Tanabe, Takashi Endo, Masaaki Yamamoto, Yuki Nikaido, and Yohei Kawaguchi. MIMII DG: sound dataset for malfunctioning industrial machine investigation and inspection for domain generalization task. In *arXiv e-prints: 2205.13879(2022)*.
- [9] Noboru Harada, Daisuke Niizumi, Daiki Takeuchi, Yasunori Ohishi, Masahiro Yasuda, and Shoichiro Saito. ToyADMOS2: another dataset of miniature-machine operating sounds for anomalous sound detection under domain shift conditions. In *Proceedings of the 6th Detection and Classification of Acoustic Scenes and Events 2021 Workshop (DCASE2021)*, pp.1–5. Barcelona, Spain, November 2021(2021).

A Study on Effectiveness of Emphasis Methods for MR Stamps in 360-degree Internet Live Broadcasting

Yoshia Saito*

*Faculty of Software and Information Science, Iwate Prefectural University, Japan
y-saito@iwate-pu.ac.jp

Abstract - We have proposed Mixed Reality (MR) stamps for supporting the broadcaster in 360-degree internet live broadcasting. In the previous study, the MR stamps that were sent from the viewers could be displayed and fixed on the real space, which enabled the broadcaster to check a holographic stamp on the real space through an MR device and communicate smoothly with the viewers. However, all MR stamps had the same appearance, making it impossible to identify the important ones. In this paper, we study the effectiveness of emphasis methods for the MR stamps implementing a prototype system that can emphasize the MR stamps by color, size, sound, and display time.

Keywords: 360-degree internet live broadcasting, Mixed Reality

1 INTRODUCTION

Affordable omnidirectional cameras have made it easy to deal with 360-degree video and many 360-degree videos can be watched via online video-sharing services. Recently, anyone can easily try 360-degree internet live broadcasting that combines internet live broadcasting with 360-degree videos. In 360-degree internet live broadcasting, a broadcaster can provide a 360-degree video to viewers in real time using an omnidirectional camera. Each viewer can be free to change their POV (Point Of View) which means viewing direction of the 360-degree video. The viewers can also communicate with the broadcaster using text chat. However, 360-degree live Internet broadcasts make communication between the broadcaster and viewers difficult. This is because the broadcaster does not grasp which direction the viewers are looking, and the viewers' POV indicates viewers' gaze is important for communication [1][2]. In this case, only text chat is not a sufficient means of communication.

We have proposed Mixed Reality (MR) stamps to enhance the means of communication, supplementing viewers' POV information [3]. The MR stamps can be displayed and fixed on the real space, which enables the broadcaster to check a holographic stamp on the real space through an MR device to help the broadcaster understand what the viewers are talking about and find the object. From the evaluation experiments, the MR stamps could reduce the average time until the broadcaster found the object that the viewers were talking about. However, there were only two viewers in the experiments and multiple MR stamps were not displayed at the same time. In actual usage scenarios, various viewers may simultaneously display multiple MR stamps to the broadcaster. Then, the broadcaster would not



Figure 1: An example of the MR stamps

know which stamp to look at. This problem could be solved by prioritizing stamps and making it easier for broadcasters to find high-priority stamps.

In this paper, we assume that stamps are prioritized by the amount of Super Chat such as YouTube, and study the emphasis methods to make the MR stamps more noticeable to broadcasters. We implemented a prototype system that can emphasize the MR stamps by color, size, sound, and display time and evaluated the effectiveness of the emphasis methods.

2 PREVIOUS WORK

In the previous work, we proposed MR stamps to realize smooth communication between a broadcaster and viewers [3]. Communication errors can be reduced by introducing MR technology. The MR stamps display viewers' requests for specific objects and locations in real space to the broadcaster in the form of holographic stamps through the MR device. Examples of the viewers' requests include "Look" to indicate that they want the broadcaster to look at a specific object, and "Go" to indicate that they want the broadcaster to move to a specific location. Figure 1 shows an example of the MR stamps which means a viewer requests the broadcaster to look at a book on the desk. The MR stamps assist communication between the broadcaster and the viewers. We found that the MR stamps could reduce the average time to find the target object and the subjective easiness of finding the target object.

The MR stamps of the previous work were intended for use in small-scale 360-degree live Internet broadcasts of a few dozen viewers, and we did not consider the case of multiple MR stamps appearing simultaneously. In practice, however, multiple MR stamps are likely to be displayed simultaneously by multiple viewers. In this case, the problem arises that the broadcaster does not know which

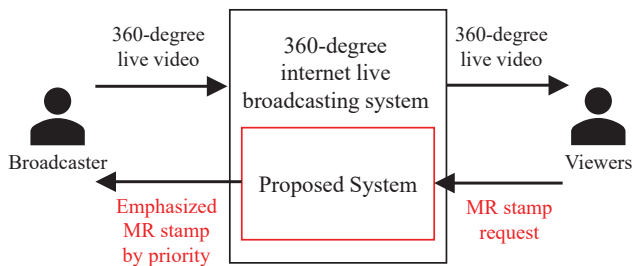


Figure 2: The model of the proposed system

stamp to pay attention to. Therefore, a solution to this problem requires some methods that emphasize the high-priority MR stamps that the broadcaster should pay attention to.

3 PROPOSED SYSTEM

We propose a system for emphasizing MR stamps in 360-degree internet live broadcasting so that the broadcaster can pay attention to the high-priority MR stamps. Figure 2 shows the model of the proposed system. Each MR stamp has a priority calculated based on some information. The high-priority MR stamps are emphasized so that the broadcaster can easily notice them. On the other hand, low-priority MR stamps are displayed moderately or not displayed at all so as not to overburden the broadcaster.

There are two issues to realize the proposed system. The first issue is what information should be used as an indicator to determine priority. There are various ways to determine priority, depending on what kind of MR stamp the broadcaster wants to be noticed. For example, if the same MR stamp is sent to the same location by multiple viewers, it may be a good idea to raise the priority level since various viewers often request the same MR stamp at the location. Alternatively, viewers' contributions to the broadcast could be reflected in the priority of MR stamps. In this case, priority can also be calculated using information such as whether the viewer is a subscribing member of the broadcast and the number of times the viewer has watched the broadcast so far. In terms of contributions to the broadcast, existing live broadcasting services often introduce a mechanism to emphasize chat messages by paying money as in the case of YouTube's Super Chat [4]. The method of calculating priority by money paid is practical because it is a mechanism that is already used in existing live broadcasting services. Therefore, in this paper, we assume that the priority is determined by the money paid by the viewers for the MR stamps and the effect of the prioritization method is out of scope.

The second issue is how to emphasize the MR stamps. It is necessary to find out how to emphasize MR stamps to grab broadcasters' attention more. There are several ways to emphasize the MR stamps. In UI/UX design, factors such as color, size, sound, and display time affect the user experience [5-8]. Moreover, we need to investigate how to emphasize the different priority levels. Therefore, the purpose of this paper is to clarify through experiments what

kind of emphasis method is suitable and how to vary the emphasis methods according to the priority level.

4 IMPLEMENTATION

We implemented a prototype system for the proposed system to evaluate to verify the effectiveness of the emphasis methods.

4.1 System Architecture

The prototype system was implemented by enhancing the system of our previous study [3]. The red words and square shows new implementation in this study and the other parts are diverted from the previous study. A broadcaster can start 360-degree internet live broadcasting using the client for broadcaster on a web browser. The 360-degree internet live broadcasting server distributes it to viewers.

The viewers can watch the 360-degree live video in spherical format and change the POV as they want while commenting. The viewers are also able to use the Look and Go stamps by selecting the kind of stamp and clicking on the video. When using a stamp, the viewers can specify the amount they are willing to pay for the stamp. The stamp/comment server receives the comments and the prioritized stamps which are based on the amount paid, and sends them to all clients for viewers and an MR device of the broadcaster. The broadcaster uses a Microsoft HoloLens 2 as the MR device. In the HoloLens2, the MR stamp application is running. The application presents emphasized MR stamps and comments to the broadcaster. The broadcaster can check the MR stamps in real space through the HoloLens 2.

The MR stamp application receives 2D position information of the stamp and needs to transform it to 3D position information for the MR stamp. To realize the coordinate transformation, we use a Raycast function in Unity, which irradiates a 3D ray from an origin point to a target direction and detects intersecting collisions. The origin point is the coordinates of the omnidirectional camera. The target direction can be given by the coordinates of the stamp. The HoloLens 2 has a function of spatial mapping. The spatial mapping provides a detailed representation of real space surfaces in the environment around HoloLens 2. The raycast detects the intersecting collisions with the real space surfaces and returns the 3D coordinates. The MR stamp is displayed on the 3D coordinates.

The comments from the viewers are displayed on a comment window. The comment window tracks the broadcaster's sight.

4.2 Emphasis Methods

The priority of MR Stamps changes with the amount of payment, based on the mechanism of YouTube's Super Chat, in which the chat displayed changes depending on the price range. The prototype system also changes the MR stamps displayed depending on the price range. In the proposed

Table 1: emphasis methods of MR stamps by payment amount in the prototype system

Payment amount (yen)	Color	Size	Sound	Display time (sec)
0	White	Small	No sound	5
1 – 499	Blue	Small	Low-tone sound	5
500 – 999	Blue	Middle	Low-tone sound	10
1000 – 4999	Yellow	Middle	High-tone sound	10
5000 – 9999	Yellow	Large	High-tone sound	15
10000 – 19999	Red	Large	Long high-tone sound	15
20000 –	Red	Extra-large	Long high-tone sound	20

system, the price range is divided into seven levels. When the price of money is higher, the priority level is higher. In the prototype system, the actual method of money exchange is out of scope.

The lowest price of 0 yen was set as the standard display with no emphasis, and the remaining six levels were set as MR stamps with emphasis. Four elements were used for the emphasis: color, size, sound, and time, which can change the display of MR stamps. Table 1 shows emphasis methods of MR stamps by payment amount in the prototype system. The colors are blue, yellow, and red, in that order, based on the colors used in YouTube’s Super Chat. The size is set so that the higher the price, the larger the size. The sound is set in the order of no sound, low-tone, high-tone, and long high-tone sound, so that the higher the price, the easier it is for the broadcaster to hear. The display time is set so that the higher the price, the longer the time is displayed by 5 seconds.

5 EXPERIMENT

Using the implemented prototype system, an experiment was conducted to see which emphasis method was more suitable and whether changing the emphasis levels according to priority level would be effective.

5.1 Environment for the Experiment

The experiment was conducted 7 times with one broadcaster and one viewer. The reason for experimenting 7 times was to evaluate the emphasis methods for the seven priority levels. The equipment used in the experiment was one notebook PC for the broadcaster, one notebook PC for the viewer, one notebook PC for the server, and a RICOH THETA V of the omnidirectional camera and a HoloLens 2 of the MR device.

The viewer sent an MR stamp on an object in the room where the broadcaster was present. The broadcaster looked around the objects in the room where the MR stamp was sent and raised his/her hand when could find the MR stamp. Figure 3 shows the layout of the objects in the broadcaster’s room. The objects were placed three in each direction on the front, left, and right from the broadcaster.

The details of the MR stamps sent by the viewers in the experiments were as follows. There were a total of 14 MR

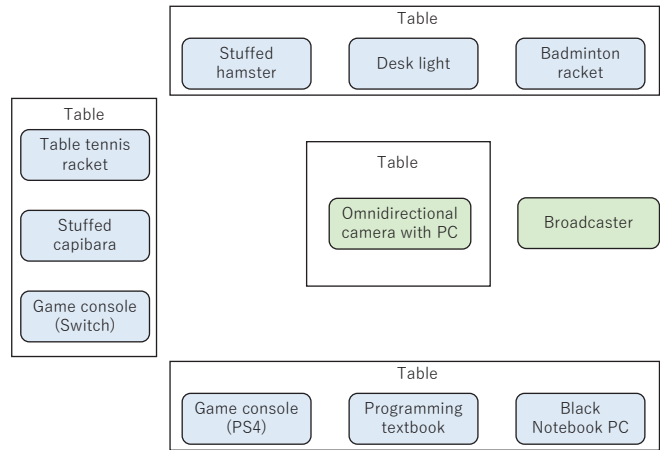


Figure 3: The layout of the objects

stamps sent by viewers in each experiment. These 14 stamps were Look and Go stamps, each of which was assigned a priority level of 7. In order to verify whether there was a difference in the discovery time depending on the payment, in each experiment, the type of MR stamp and the objects to be stamped were all fixed, and only the amount of stamp was changed by the broadcaster. In this way, we were able to measure the discovery time for the same object at seven different levels.

There were two evaluation criteria in the experiment. The first evaluation criterion is the measurement time for the discovery when the MR stamp is displayed. This is the time from when the MR stamp is sent to when the broadcaster raises his/her hand. A stopwatch was used to measure the time. If the broadcaster did not find the MR stamp after the first transmission, a second and a third MR stamp with the same conditions were sent. The second evaluation criterion was a subjective evaluation of the ease of finding MR stamps. The subjective evaluation method was a questionnaire, given only to the broadcaster after the experiment was finished. The questionnaire consisted of a five-point scale in terms of whether the elements used for the emphasis affected the ease of discovery. The closer the rating was to 5, the greater the impact on discoverability. The questionnaire also asked the broadcaster to rate the positive aspects of the MR stamp as well as areas for improvement, using free-text answers.

5.2 Results

Table 2 and Table 3 show the discovery time depending on the payment for the Look and Go stamp. The results in Table 2 show that one broadcaster failed to find the stamp two times in the case of 100 yen, taking almost 30 seconds due to the retry. The results in Table 3 show that two broadcasters failed to find the stamp two times in the case of 0 yen, taking almost 20 seconds due to the retry. The broadcaster had to look around to find the MR stamps. In the case of 0 yen and 100 yen, because the MR stamps were displayed for only 5 seconds, the broadcaster could not find the MR stamps before they disappeared. In some cases, even when the payment was 500 yen or more, it took more than 5

Table 2: An experimental result for the discovery time on the Look stamp

	¥0	¥100	¥500	¥1,000	¥5,000	¥10,000	¥20,000
Broadcaster A	4.75	2.12	1.83	2.82	2.60	1.96	1.88
Broadcaster B	3.81	3.15	2.55	4.56	2.87	2.96	4.33
Broadcaster C	4.38	6.79	3.97	5.33	7.03	3.74	3.46
Broadcaster D	2.20	2.25	1.79	1.48	2.76	1.36	2.03
Broadcaster E	3.47	5.23	3.33	3.19	3.11	6.17	3.01
Broadcaster F	2.63	2.06	2.68	1.97	2.14	1.87	1.37
Broadcaster G	3.95	27.64	2.41	5.60	2.99	5.83	6.17
Average time	3.60	7.03	2.65	3.56	3.36	3.41	3.18
SD	0.91	9.26	0.78	1.62	1.65	1.93	1.67

Table 3: An experimental result for the discovery time on the Go stamp

	¥0	¥100	¥500	¥1,000	¥5,000	¥10,000	¥20,000
Broadcaster A	1.88	3.04	2.55	2.01	1.82	2.02	1.77
Broadcaster B	3.67	2.57	2.89	3.07	3.52	3.90	3.44
Broadcaster C	18.71	6.10	5.00	7.87	5.42	5.98	5.06
Broadcaster D	1.91	2.17	2.29	2.31	1.66	2.28	1.87
Broadcaster E	3.84	3.40	3.06	4.20	2.01	4.30	2.89
Broadcaster F	2.65	2.31	1.64	1.83	2.32	2.44	3.94
Broadcaster G	18.80	4.33	6.62	7.39	5.59	8.25	6.23
Average time	7.35	3.42	3.44	4.10	3.19	4.17	3.60
SD	7.83	1.39	1.75	2.54	1.69	2.28	1.63

Table 4: Questionnaire on whether the elements used for the emphasis affected the ease of discovery

	A	B	C	D	E	F	G	Average Score
Color	1	4	3	5	2	5	5	3.6
Size	3	5	4	5	4	3	5	4.1
Sound	1	2	4	5	2	4	4	3.1
Display time	3	4	4	3	4	1	4	3.3

seconds to find the MR stamp, indicating that a retry might have been necessary if the display time had been 5 seconds. The display time of the MR stamp was found to be an important factor for the stamp to be noticed by the broadcaster.

Table 4 shows the result of the questionnaire on whether the elements used for the emphasis affected the ease of discovery. The ratings of each broadcaster showed that only the size factor was rated no lower than 3 out of 5. The average scores also showed that size was the most highly rated of the elements used for the emphasis. The broadcasters felt that size had the greatest impact on findability. On the other hand, the average scores for the color, sound, and display time were above 3, but there were no significant differences, suggesting that they were not perceived as having much impact on findability.

To summarize the results of the experiment, the emphasis method that had the greatest impact on the discovery time of the MR stamps was the length of the display time. The broadcaster's subjective perception was that the size of the MR stamp had the greatest impact on the ease of discovery. Therefore, it was found that by varying the size and display time factors of the MR stamps, the MR stamps with higher priority could be given more emphasis.

6 CONCLUSION

In this paper, we proposed a method to emphasize MR stamps according to priority. Assuming that the priority level is determined by the amount of money paid to send the MR stamp, we conducted experiments to investigate the effective factors in emphasizing MR stamps. A prototype system was developed in which the color, size, sound, and display time of MR stamps changed in seven levels depending on the priority level. Experimental results with the prototype system showed that display time had the greatest impact on MR stamp discovery time, while size had the greatest impact on the broadcaster's subjective evaluation.

For future work, the experimental method needs to be improved. In our experiment, we tested the effect of the emphasis methods in an environment where only one MR stamp was displayed at the same time, but the effect should also be tested when multiple MR stamps are displayed. Moreover, it is necessary to verify the upper limit of the number of stamps to be displayed when multiple MR stamps are displayed. Furthermore, we would like to improve the usability of MR stamps.

REFERENCES

- [1] Roel Vertegaal: The GAZE groupware system: mediating joint attention in multiparty communication and collaboration, Proc. of SIGCHI conference on Human Factors in Computing Systems, pp.294-301 (1999).
- [2] David M. Grayson and Andrew F. Monk: Are you looking at me? Eye contact and desktop video conferencing, ACM Transactions on Computer-Human Interaction (TOCHI), Vol. 10, Issue 3, pp.221-243 (2003).
- [3] Yoshia Saito and Kei Sato, Proposal of a Broadcaster Support Method using MR Stamp in 360-degree Internet Live Broadcasting, International Journal of Informatics Society (IJIS), Vol. 15, No. 3, pp. 115-122 (2024).
- [4] YouTube: Buy a Super Chat or Super Sticker (Online), available from <<https://support.google.com/youtube/answer/9178363#zippy=>> (accessed 2024-7-1).
- [5] Mario Silic: Understanding Colour Impact on Warning Messages: Evidence from US and India, CHI EA'16, pp. 2954-2960 (2016).
- [6] Valeeya Tadayin, Nuttanont Hongwarittorn, and Lumpapun Punchoojit: The Effect of Icon Size and Grid Size on Smartphone Menu Selection, ICSEB'18, pp. 64-68 (2018).
- [7] Lotus Zhang, Jingyao Shao, Augustina Ao Liu, Lucy Jiang, Abigale Stangl, Adam Fourney, Meredith Ringel Morris, and Leah Findlater: Exploring Interactive Sound Design for Auditory Websites, CHI'22, pp. 1-16 (2022).
- [8] Jakob Nielsen: Enhancing the Explanatory Power of Usability Heuristics, CHI'94, pp. 152-158 (1994).

“It’s Still My Turn!”: Visualizing Speech Intervals for Enhanced Conversational Satisfaction.

Atsuto Kurokochi[†], Yo Kuwamiya[‡], Anri Nakano[†] and Minoru Kobayashi[†]

[†]Faculty of Interdisciplinary Mathematical Sciences, Meiji University, Japan

[‡]Graduate School of Advanced Mathematical Sciences, Meiji University, Japan

atsuto.kurokochi@koblabs.org

Abstract - Smooth turn-taking during conversation is important for enhanced conversational satisfaction. However, unintended turn-taking for the speaker can occur. This turn-taking interrupt the speaker’s speech and cause discomfort. In addition, it also leads to incomplete transmission of information. From the listener’s perspective, they cannot fully understand the speaker’s intentions and information. This research considers that the cause of these unintended turn-taking is the difficulty in consistently interpreting the “Speech Intervals” that occur during conversation: “Interval in speech” and “Interval after speech”. The purpose of this research is to enhance conversational satisfaction by unifying the interpretation of these “Speech Intervals” between the speaker and listener, thereby preventing such unintended turn-taking. To achieve the purpose, this paper proposes two methods to visualize speech intervals: 1) The Ball Method and 2) The Rise-up Button Method. We investigated the effect of these methods of visualizing speech intervals during conversation.

Keywords: Speech Intervals, Turn-taking, Communication Support Systems

1 INTRODUCTION

1.1 Background

Smooth speakers turn-taking during a conversation is important for improving conversational satisfaction. However, there are cases where turn-taking occur at unintended timings for the speaker. These unintended turn-taking can cause discomfort by interrupting the speaker’s speech, as well as causing it to end incompletely. Furthermore, the intention behind the speech may not be adequately conveyed to the listener. These unintended turn-taking are problematic for both the speaker and listener and can reduce conversational satisfaction. In this research, we considered that one of the factors contributing to unintended turn-taking is the gaps that occur during speech. We express the gaps to “Speech Intervals”¹.

1.2 Definition of Speech Intervals

We identify two types of speech intervals: “Interval in speech” and “Interval after speech”. Figure 1 shows the spe-

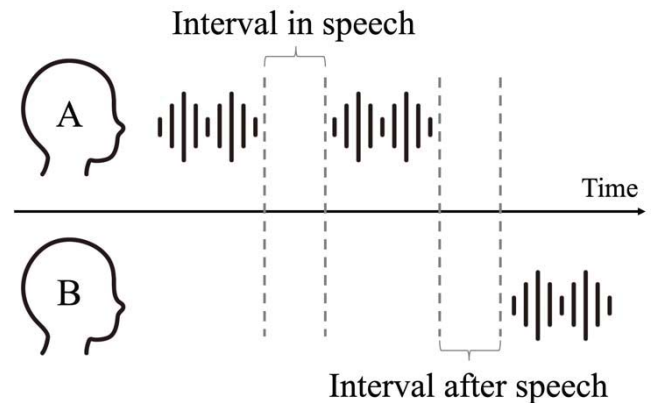


Figure 1 : Speech Intervals

ech intervals. A and B are conversation participants. The waveform portions represent times when they are vocalizing, and the dotted line segments represent speech intervals when they are not vocalizing. “Interval in speech” refer to the segments where the speaker temporarily pauses or interrupts their speech due to reasons such as breathing or thinking. In Figure 1, the interval on the left enclosed by the dotted line is “Interval in speech,” after which the same speaker continues speech. On the other hand, the interval on the right in Figure 1 represents “Interval after speech” which occurs when a speaker has completely finished their speech. Ideally, turn-taking should occur during this interval. However, if the speaker and listener have different recognitions of this interval, unintended turn-taking may occur. For example, if the listener judges it as “Interval after speech” while the speaker intends it as “Interval in speech” and continues speaking, a turn-taking may occur. This turn-taking is unintended for the original speaker.

Sacks [1] proposed a set of speaker transition rules designed to minimize overlap (when two or more people are speaking at the same time) and gaps (when no one is speaking, similar to intervals) in conversation. According to these rules, if a gap occurs during a conversation, one will start speaking. Additionally, Stivers et al. [2] reported that among speakers of 10 languages including English, Italian, and Chinese, Japanese speakers responded with the shortest gaps on average. In other words, Japanese people tend to start speaking when a gap occurs during conversation.

¹ This work is supported by JSPS KAKENHI Grant Numbers 22H03635, 23K24891.

1.3 Research Subject

The objective of this research is to prevent unintended turn-taking and improve conversational satisfaction. In this research, conversational satisfaction is defined as follows:

1. The speaker's intention is conveyed.
2. The speaker does not feel discomfort.
3. There is no complaint with the conversation content and conclusions.

To achieve the objective, this research proposes a method to visualize speech intervals during conversation. Specifically, it visualizes whether a speech interval is an "Interval in speech" or an "Interval after speech," to prevent turn-taking from the listener during the speech. We implemented two prototypes of visualizing speech intervals methods: the Ball Method and the Rise-up Button Method. We conducted an experiment comparing these two proposed methods with a condition without using a visualization method. Figure 2 shows the proposed methods and the experimental state. The experiment yielded the following results:

- Visualizing speech intervals does not improve conversational satisfaction.
- Depending on visualization methods, conversational satisfaction may decrease.

These results suggest that visualizing speech intervals to prevent unintended turn-taking does not improve conversational satisfaction, and in some cases, the visualization method can decrease satisfaction. Based on these results, this paper discusses the effect of visualization of speech intervals on conversations, discusses the possibility that visualization of speech intervals could be effective under certain conditions and adequacy of visualizing methods.



Figure 2 : The state of experiment

2 RELATED WORKS

In an area related to this research, Beattie [3] studied interruptions during speech. When simultaneous speech occurs, the original speaker's speech gets interrupted. He stated that due to this interruption, the current speech may end up incomplete. Additionally, he mentioned the possibility of speech interruptions occurring even without simultaneous speech. He speculated that such interruptions occur at locations called "encoded pauses [4]". Ferguson [5] referred to these interruptions as "silent interruptions".

In addition, Beattie analyzed four types of interruptions from group conversation video data: three involving speech overlap (simple interruption, overlap, butting-in interruption) and one without overlap (silent interruption). The results showed that on average 34.3% of speaker transitions involved some kind of interruption. Silent interruptions accounted for 10% of all interruptions in those conversations. In contrast, overlaps occurred more than four times as frequently as silent interruptions. While overlap has been studied extensively [6], and Beattie regarded it as the most common type of interruption, he noted that "Nevertheless, almost 10% of all interruptions are silent interruptions." ([3], p.24) Therefore, if speech interruptions are considered a type of unintended turn-taking, preventing their occurrence has significance.

3 METHOD

3.1 Method Details

In this research, we propose a method to visualize speech intervals during conversations. In considering this method, we defined the following two requirements:

Requirement1. Speech continuation must be clear.

Requirement2. Speech completion must be clear.

To visualize speech intervals, there is a need to express the intention to continue speaking (desire to continue the speech) and the intention to end speaking (no desire to continue the speech). Therefore, we defined these two requirements. We propose the following two methods that meet these requirements:

- The Ball Method
- The Rise-up Button Method

Figure 3 shows the Ball Method and Figure 4 shows the Rise-up Button Method. For the Ball Method, we used a red rubber ball with a diameter of about 13 cm for easy visibility. For the Button Method, we used a device [7] which pressing a red button at hand causes a plate with a red circle drawn on it to stand up.



Figure 3: The Ball Method



Figure 4 : The Rise-up Button Method

3.2 Method Points

There is a difference between these methods - whether the method is shared or independent for everyone. This difference is considered to change the resistance to speech. In the Ball Method, the physical action of taking the ball is required to speak. It can impose a large physical load. On the other hand, the Button Method only requires the relatively low-load action of pressing a button at hand. This difference may influence the speech. Additionally, there is a difference in how pauses after speech completion are expressed. In the Ball Method, after completing speech, the ball is returned and the next speaker takes it, visualizing the speaker transition flow during the pause after completion. The role of gaze in speaker turns has also been reported in previous studies [8], and the ball serves the role of guiding the gaze during turn-taking. On the other hand, in the case of the Buttons, since each system is independent, the time required for speaker turns is expected to be shorter than in the ball condition, but it can also be described as discontinuous. The effect of these differences on the experiment is also investigated.

4 EXPERIMENT

4.1 Experiment Details

The participants were 12 students aged 20-24 years old (seven males, five females). The experiment was conducted in a university seminar room that accommodates 24 people. Four desks were arranged in the center of the room, and the participants were seated facing each other. A camera was set up to record the experiment. Additionally, one desk was placed on each side, with a computer displaying instruction slides and a timer on the desk at the back from the camera's view, and a mobile phone for audio recording on the desk in the front. The state of experiment is shown in Figure 2.

The experiment was conducted with groups of four people, repeated three times each. In each experiment, the participants had a 10-minute conversation. The topic was "If the group members were to have 3 meals together tomorrow, what dishes would you want to eat?" with the sub-topics of breakfast (topic 1), lunch (topic 2), and dinner (topic 3), and the participants had to reach a conclusion. For the three repetitions, the two methods proposed in the previous chapter were applied, along with a condition without any method, for a total of three conditions.

4.2 Instruction Contents

Regarding the three conditions shown above, they were defined as follows:

- Condition 1: Control condition (None)
- Condition 2: Ball condition
- Condition 3: Button condition

The experiment was conducted considering the potential influence of the order of conditions. For the first group, the conditions were in the order of Control condition, Ball condition, Button condition. For the second group, the order was Button condition, Control condition, Ball condition. For the third group, the order was Ball condition, Button condition, Control condition. For Ball condition and Button condition, the following rules were set:

Ball Condition:

- Place the ball at the central location among the four participants.
- When speaking, hold the ball while doing so.
- After finishing your speech, return the ball to its original position or pass it to another person.

Button Condition 3:

- Each participant has the button device shown in Figure 2 in front of them.
- When speaking, press the button to make the circular plate stand up.
- After finishing your speech, return the circular plate being down.
- While a participant's circular plate is standing up, the other participants do not speak.

In the experiments for Ball Conditions and Button Condition, the above respective rules were instructed, and the participants followed them during their conversations.

4.3 Evaluation

The evaluation was conducted through a questionnaire. After each condition, a paper questionnaire was distributed (response time: 10 minutes), and participants were asked to respond. The evaluation items consisted of 10 questions encompassed within the following three major categories, as well as an open-ended comments section and a post-experiment questionnaire, for a total of 12 indicators. For the 10 questions mentioned, a 7-point Likert scale was used. The contents of the questions are shown in Table 1 to Table 3.

Table 1: Your own participation in the conversation

Q1	Do you think you were able to actively express your opinions?
Q2	Do you satisfy with your own comments?
Q3	Do you think you were able to concentrate on the discussion?
Q4	Do you think you were able to communicate your intentions to the other participants?

Table 2: Impressions of other participants

Q5	Do you think you understood the intentions behind the other participants' comments?
Q6	Do you think the other participants were able to communicate their intentions to each other?
Q7	Did you feel positively towards the other participants?

Table 3: Impressions of other participants

Q8	Do you think your group was able to have an active conversation?
Q9	Do you satisfy with the process your group went through to reach a conclusion?
Q10	Do you satisfy with the conclusion your group reached?

In the post-experiment questionnaire, participants' attributes were surveyed, and they were asked to rate their closeness with each other group member on a scale of 1-5 (1 : not close ~ 5 : very close)

5 RESULT

The results will be explained in two types: the results obtained from the questionnaire and the results obtained from the audio recordings. Due to equipment trouble during the experiment, results from the questionnaire were obtained for three groups, while results from the video were obtained for groups 1 and 2.

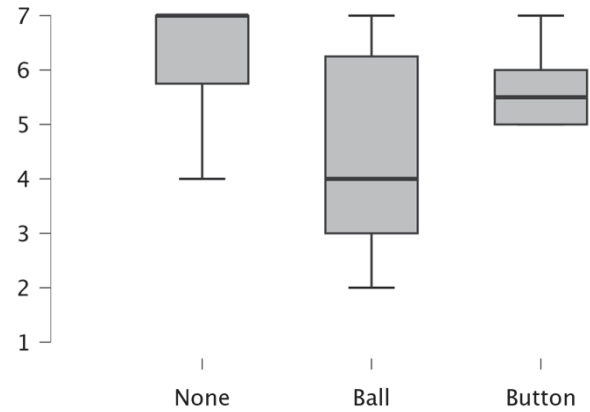


Figure 5: The result of Q.1

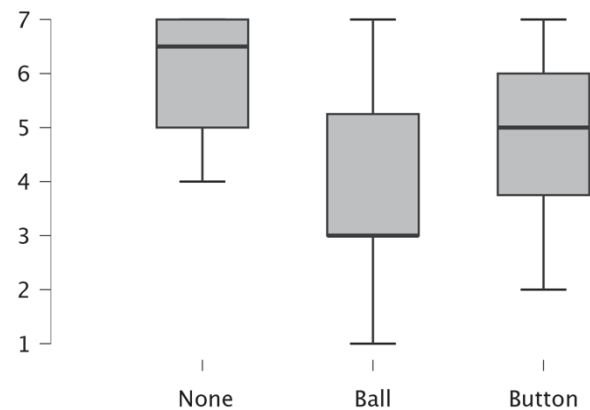


Figure 6: The result of Q.3

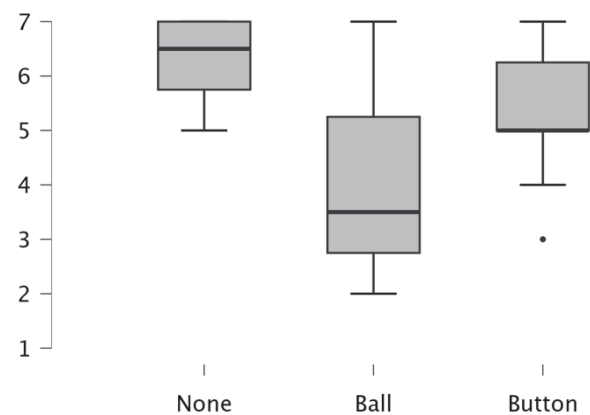


Figure 7: The result of Q.4

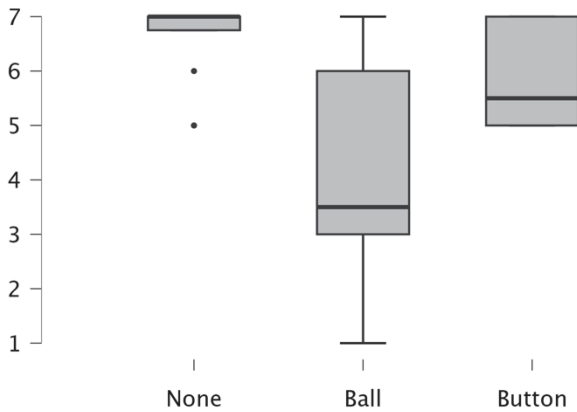


Figure 8: The result of Q.8

5.1 Survey Result

The Friedman test was used for the analysis, and Bonferroni's method was used for the post-hoc multiple comparisons. The significance level was set at $\alpha=0.05$. Significant differences were found in the Friedman test for Q1, Q3, Q4, and Q8. For the question items with significant differences, multiple comparisons using Bonferroni's method were performed. For Q1, there was a significant difference between Control condition and Ball condition ($p=0.024$). For Q3, there was a significant difference between Control condition and Ball condition ($p=0.002$). For Q4, there were significant differences between Control condition and Ball condition ($p<0.001$), and between Button condition and Ball condition ($p=0.022$). For Q8, there were significant differences between Control condition and Ball condition ($p<0.001$), and between Button condition and Ball condition ($p=0.015$). Figures 5-8 show the results for the question items with significant differences. For the other questions (2, 5, 6, 7, 9, 10), there were no significant differences between any conditions. In particular, there were no significant differences between Control condition and Button condition for all 10 questions.

Table 4: Silence time

	None	Ball	Button
Group 1	166.2 s	322.35 s	340.54 s
Group2	200.79 s	239.68 s	274.83 s

Table 5: The number of turn-taking

	None	Ball	Button
Group 1	271	93	80
Group2	189	50	63

Table 6: The number of unintended turn-taking

	No condition	Ball condition	Button condition
Group 1	19	2	1
Group2	8	0	0

5.2 Audio Result

Tables 4-6 show the results of the speech analysis. The audio analysis was conducted using ELAN [9]. From the audio, the durations of silence time and the number of turn-taking were obtained. The duration of silence time is the total time where there was no vocalization for 0.1 seconds or more during the conversation. For group 1, the silence durations for Control condition, Ball condition, and Button condition were 166.2 seconds, 322.35 seconds, and 340.54 seconds respectively. For group 2, they were 200.79 seconds, 239.68 seconds, and 274.83 seconds respectively. As for the number of turn-taking, for group 1 in Control condition, Ball condition, and Button condition, there were 271, 93, and 80 instances respectively. For group 2, there were 189, 50, and 63 instances respectively. Additionally, based on inferring from the speech content, the numbers of unintended turn-taking judged to have occurred were 19, 2, and 1 for group 1 in Control condition, Ball condition, and Button condition respectively. For group 2, they were 8, 0, and 0 respectively. Due to the small amount of data, statistical testing could not be performed on the audio analysis results.

6 CONSIDERATION

Since the proposed methods showed a decreasing trend in the targeted turn-taking for both Ball condition and Button condition compared to Control condition, we can consider that the proposed methods were able to prevent unintended turn-taking for the speaker. However, the survey results showed that conversational satisfaction levels did not increase in either Ball condition or Button condition compared to Control condition. Therefore, we can infer that visualizing speech intervals and preventing unintended turn-taking for the speaker does not have an effect of improving conversational satisfaction. Additionally, since Ball condition showed significantly lower scores for four of 10 questions compared to Control condition and two questions compared to Button condition, we can say that the visualization speech intervals method may decrease satisfaction levels in some cases.

7 DISCUSSION

7.1 Potential for Visualizing Speech Intervals

The experimental results suggest that visualizing speech intervals does not have an effect of improving conversation satisfaction. However, we think that this result cannot be

generalized to all situations. Here are some comments from participants in the condition 1 experiment:

- I felt happy being able to talk without thinking about using the system.
- In the Ball condition and Button condition, unnecessary talk was not possible.

These comments suggest that some people felt Control condition was better in this experiment. However, these comments can only be judged as better in the context of this experimental setting. In this experiment, all participants belonged to the same laboratory in university and regularly interacted with each other, so the intimacy levels of groups 1-3 were relatively high at 3.42, 4.08, and 3.67 (rounded to two decimal places), indicating the groups were composed of relatively close members. Additionally, the conversation content was similar to daily conversation. In other words, the conversations in this experiment were "casual conversations among close individuals with low importance." Therefore, talking without thinking or engaging in unnecessary speech did not decrease conversational satisfaction. However, we speculate that the effect of such speech may differ depending on the relationships among members and the importance of the conversation. Consequently, visualizing speech intervals may have the potential to improve satisfaction in conversations under specific conditions. We suggest the following two conditions:

- Conversations among participants with low intimacy
- Highly important conversations

Going forward, the effect of visualizing speech Intervals on conversational satisfaction under such specific conditions will be investigated. Additionally, appropriate visualization methods for those situations will also be considered.

7.2 Adequacy of Methods

In the experiment, we visualized speech intervals using the two methods of the Ball Method and the Rise-up Button Method, and intentionally limited speech to prevent unintended turn-taking for the speaker. In this section, we will discuss the appropriateness of these methods.

Both methods were able to prevent the targeted turn-taking. Additionally, in the conditions using the methods, there was an increasing trend in silent time and a decreasing trend in the number of turn-taking compared to Control condition. However, the survey results showed that while satisfaction levels significantly decreased in Ball condition, there was no significant difference in Button condition. This may be attributed to the characteristics of the systems. In the comments from participants in Ball condition, seven people expressed negative opinions about the method. Furthermore, in the comments obtained from participants in Button condition, one said, "There was only one ball, but each person had their own button, so it was easy to press." As mentioned in Section 3.2, this is a clear difference between the Ball Method and the Button Method. Because of this, we infer those participants felt resistance to speaking in the Ball

Method. From these results, since sharing one object to visualize speech intervals seemed to place a load on participants, the Ball Method is not appropriate.

For Button condition, there was no significant decrease in the survey results compared to Control condition. Additionally, the comments did not clearly suggest any negative effects from using the method. Based on this, we consider the Button Method to be suitable for visualizing speech intervals. However, one comment stated, "I had the impression that it was somewhat difficult to understand the timing of speech from the person sitting right next to me." This does not imply that the method itself made it difficult to understand the timing of speech, but rather that it was difficult to understand compared to the two people in front. If this method was not used, there would be fewer cues to infer the timing of speech, which would make it even more difficult to understand. For this reason, we do not consider this comment to suggest a negative effect of the method. However, this is an area for improvement in the method, and the details of the method need to be reconsidered.

Another method that should be compared to the visualizing speech intervals methods used in this experiment is the automatic visualization methods. The results suggested that operating a system could place a load on the speaker. Therefore, automating the process could enable visualization without burdening the speaker. As a future prospect, we will reconsider the Button Method and explore methods to automatically visualization system and compare them to propose a more suitable method for visualizing speech intervals.

7.3 Limitation of the Experiment

All participants in this experiment were Japanese. However, the length of speech intervals taken by speakers varies across different languages [2]. If the length of speech intervals differs, it can be inferred that the effect of visualizing those intervals would also differ. Therefore, the insights gained from this experiment have limitations and cannot be realistically generalized to all languages.

8 CONCLUSION

In this research, we focused on the gaps during speech to prevent unintended turn-taking for the speaker. We considered two types of speech intervals: "Interval in speech" and "Interval after speech". By visualizing these intervals and limiting speech, we aimed to prevent unintended turn-taking for the speaker and improve conversational satisfaction. To achieve the purpose, we proposed the Ball Method and the Rise-up Button Method to visualize speech intervals. The experimental results showed that the proposed methods did not improve satisfaction levels, and in some cases, the methods even decreased satisfaction. Comments from participants suggested that in this experiment, conversations in Control conditions had higher satisfaction, but also indicated that results could change depending on factors such as the intimacy between conversation participants and the importance of the conversation. As a future prospect, we will consider investigating situations under which visualizing speech intervals would be effective. Additionally, we

will reconsider the Button Method and explore methods to automatically visualization method of speech intervals, and investigate the effects that these methods have on conversational satisfaction.

REFERENCES

- [1] Sacks, H., Schegloff, E. A., and Jefferson, G. A., A simplest systematics for the organization of turn-taking for conversation, *Language*, 50, pp.696-735 (1974)
- [2] Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Henieman, T., Hoymann, G., Rossano, F., Ruitter, J. P., Yoon, K. E., and Levinson, C. S., Universals and cultural variation in turn-taking in conversation, *Proc. Natl. Acad. Sci. U. S. A.*, 106, pp.10587-10592 (2009)
- [3] Beattie, G. W., Interruption in conversational interaction, and its relation to the sex and status of the interactants, *Linguistics*, 19, pp.15-36 (1981)
- [4] Beattie, G. W., Planning units in spontaneous speech: some evidence from hesitation in speech and speaker gaze direction in conversation, *Linguistics*, 17, pp.61-78 (1979)
- [5] Ferguson, N., Interruptions: speaker-switch nonfluency in spontaneous conversation, *Linguistics*, Edinburgh University, Department of Linguistics, Work in progress, 9, pp.1-19 (1976)
- [6] Schegloff, E. A., Overlapping talk and the organization of turn-taking for conversation, *Language in Society*, 29, pp.1-63 (2000)
- [7] Sasagawa Co., Ltd., <https://www.sasagawa-brand.co.jp/>, last accessed 2024/6/11
- [8] Kendrick, K. H., Holler, J., and Levinson, C. S., Turn-taking in human face-to-face interaction is multimodal: gaze direction and manual gestures aid the coordination of turn transitions, *Phil. Trans. R. Soc. B*, 378, 20210473(2023)
- [9] ELAN, <https://archive.mpi.nl/tla/elan>, last accessed 2024/6/11

Empirical Experiment of Sales Promotion Using Parametric Array Loudspeaker and Sales Pitch Audio as Auditory Stimulation

Kohdai Nohara[†], Shigemi Ishida[†], Hiroshi Inamura[†], and Junko Nohara[‡]

[†]Graduate School / School of Systems Information Science, Future University Hakodate, Japan
{g2123047, ish, inamura}@fun.ac.jp

[‡]Department of Nutrition, Faculty of Health Sciences, Kio University, Japan
j.nohara@kio.ac.jp

Abstract - Sales demonstration is one of the sales promotion methods widely used in food stores. Sales demonstrations, however, are rarely performed after COVID-19 to prevent the spread of infections. In this paper, we attempt a new sales promotion approach with auditory without gustatory and tactile stimulation. We employ a highly-directional parametric array loudspeaker, which enables us to send sound signals in a narrow range of directions. Experiments of *Yamato-maru eggplant* sales promotion were conducted in a real store to promote customers' purchasing behavior by playing two kinds of sound from a parametric array loudspeaker or a normal loudspeaker. We analyzed the sales amount data with and without auditory stimulation from the parametric array loudspeaker with the considerations of biases such as day-of-week and time-of-day biases. The analytical results show that auditory stimulation by sales pitch audio sent from a parametric array loudspeaker increased sales amount by 1.52 times.

Keywords: Auditory stimulation, behavioral change, parametric array loudspeaker.

1 Introduction

Food stores such as supermarkets often perform sales promotions such as sales demonstrations, pop displays, posters, flyers, leaflets, and store announcements to increase sales. Sales demonstration, especially, is effective in promoting customers' purchasing behavior because store employees can directly convey the taste and information of products to customers[1].

However, after the spread of COVID-19, sales demonstration is rarely performed to prevent the spread of infections. There is a growing need for a new sales promotion approach based on tactile and gustatory senses with the considerations on infections.

Several studies have reported sales promotion without tactile or gustatory senses based on olfactory[2] and auditory[3], [4] stimulation. Olfactory stimulation, i.e., a smell approach, is effective only for customers close to the stimulator but is inefficient for promotion to customers at a distance. A special aroma diffuser is also required for long-term stimulation. Auditory stimulation, i.e., an acoustic approach using in-store background sound, is effective for customers at a distance. Some customers may feel inconsistency of the music with the atmosphere of in-store sections.

In this paper, we propose a new sales promotion approach without gustatory and tactile stimulation as one of the substitutions of sales demonstrations in a *new lifestyle* after COVID-

19. We focus on auditory stimulation as possible stimulation in a new lifestyle. Based on the results presented in [3], [4], sales promotion with music in a whole area or store section is ineffective. We therefore install a parametric array loudspeaker to provide auditory stimulation in a specific area. A parametric array loudspeaker is a highly-directional loudspeaker, which allows us to send sound signals in a narrow range of directions. We believe that we realize auditory stimulation with clear sound even in a noisy environment with other in-store promotion sounds.

The impact of auditory stimulation using a parametric array loudspeaker on sales promotion is not studied. To verify how intervention with a parametric array loudspeaker affects customers' purchasing behavior, we conducted a sales promotion experiment in a real store in July 2022, using *Yamato-maru eggplant* as the target of a sales promotion. We used sales pitch audio played on a normal loudspeaker and a parametric array loudspeaker as auditory stimulation and collected sales amount data. The sales amount data was then analyzed with the considerations of biases such as day-of-week and time-of-day biases.

Our main contributions are twofold:

- We propose a sales promotion method using a parametric array loudspeaker to promote customers' purchasing behavior, which can be one of the substitutions of sales demonstrations in a new lifestyle after the COVID-19.
- We conducted an empirical experiment to show the effectiveness of the auditory stimulation in a specific area. We demonstrate that a parametric array loudspeaker can promote customers' purchasing behavior when appropriate sounds such as sales pitch audio are used.

The remainder of this paper is organized as follows. Section 2 reviews researches on the changes in purchasing behavior using olfactory and auditory sales promotion methods. In Section 3, we present our proposed method, i.e., promoting customer purchasing behavior using a parametric array loudspeaker. Section 4 describes the empirical experiment conducted in a real store, followed by evaluations in Section 5. Finally, Section 6 concludes the paper.

2 Related Work

The literature has studied olfactory and auditory stimulation as the sales promotion without the senses of tactile and gustatory perception. In this section, we review existing re-

search on sales promotion with olfactory and auditory stimulation.

2.1 Sales Promotion Based on Olfactory Stimulation

Mayumi et al. presented sales promotion method based on aroma diffusion[2]. They diffuse spice and curry smell, corresponding to recipes, from an aroma diffuser in the food section of a supermarket to promote customers' purchasing behavior of eggplant. The experiment was conducted for one month with a constant price without any other sales promotion such as pre-announcements. The results of sales amount analysis showed that the aroma diffusion increased both the number of visiting customers to and the time spent at the sales section. The sales amount have also tended to increase. The experiment was conducted for one month, and during the experiment period, the price was kept unchanged, and no sales promotion was conducted through prior notification or appealing to customers. The results showed that aromas increased the number of visits to the sales section and the time spent there, and that the amount of sales tended to increase.

Mayumi et al. indicated that aroma diffusion is less effective for customers distant from an aroma diffuser[2]. There is a limit to the range of aroma reach, which makes it difficult to deliver aroma to the entire store. Another problem of olfactory stimulation includes the influence of store air conditioning system and the mixture of aromas with other products.

In this study, we aim to realize a method using auditory stimulation instead of olfactory stimulation.

2.2 Sales Promotion Based on Auditory Stimulation

Background music (BGM) influences customers' purchasing behavior[5], [6]. Background sound can be utilized as auditory stimulation to promote customers' purchasing behavior. Unlike the olfactory stimulation, an auditory stimulation approach has small influence from distance and air conditioning systems.

Spendrup et al. presented sound-based sales promotion using nature sounds as in-store background music[3]. They analyzed how customers' purchasing behavior changes when background music, such as birds chirping, is used in a supermarket. The results showed that the nature sounds increased customers' willingness to purchase organic and environmentally friendly foods. However, Spendrup et al. also indicated that the effect of the nature sounds was limited. The reason may be that the background music did not match the atmosphere in the store, and some customers found the natural sound discomfort.

Milliman et al. examined how customers' purchasing behavior changes when the tempo of in-store background music is changed[4]. They conducted an intervention with customers in a supermarket under three conditions of in-store background music: no music, slow-tempo music, and fast-tempo music. The results showed that the amount of sales increased significantly when slow-tempo music was played compared to when fast-tempo music was played. The reason may be that the slow tempo music made customers move

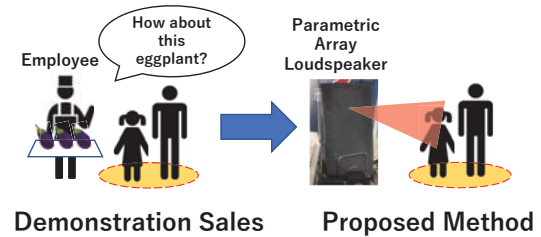


Figure 1: Overview of the proposed method

more slowly in the store, and they stayed longer in the store. However, slow-tempo music does not always have a positive effect on customers' purchasing behavior. The reason for this is that the increase in time spent in the store may decrease the turnover rate of the store, which may result in a decrease in the amount of sales.

Customers' purchasing behavior changes with the background music may include unknown effects. We think it difficult to find background music promoting customers' purchases with no discomfort for all customers. In this study, we focus on loudspeakers that play sound rather than background music instead of sound content.

3 Sales Promotion Using Parametric Array Loudspeaker for Auditory Stimulation

In this section, we present a sales promotion method using a parametric array loudspeaker. A parametric array loudspeaker transmits sound signals to a narrow range of directions, which enables us to deliver clear sound to a specific area even in a noisy environment. We aim to increase the opportunities for customers to recognize a sales target product to promote purchasing behavior.

3.1 Overview

Figure 1 shows an overview of our sales promotion system using a parametric array loudspeaker. In a sales demonstration, a demonstrator makes a sales pitch to customers about a target product to evoke purchasing behavior. In this study, we play sales pitch audio from a parametric array loudspeaker to evoke customers promoting purchases.

The following subsections describe the details of the loudspeakers and song/audio used in our system.

3.2 Loudspeaker

We use a parametric array loudspeaker, which is a highly directional loudspeaker that transmits audible sound to a narrow range of directions. A parametric array loudspeaker consists of a number of ultrasound transmitters placed in a line. Using highly directional ultrasound signals, a parametric array loudspeaker realizes a narrow-directional loudspeaker. Humans cannot hear ultrasound signals. Specially modulated ultrasound signals are therefore transmitted from multiple ultrasound transmitters to generate an unmodulated low-frequency hearable sound.

Unlike normal loudspeakers, a parametric array loudspeaker delivers sound signals in a small area. We can use a big sound

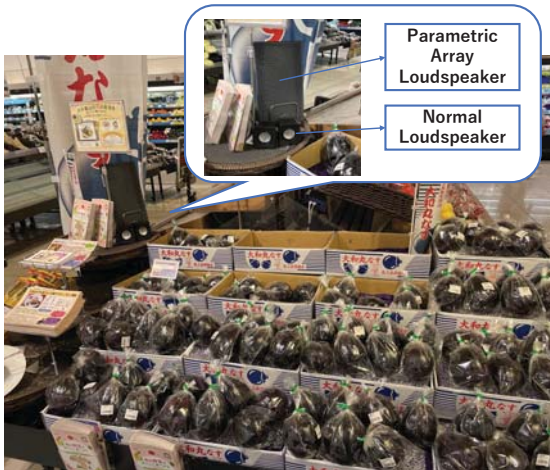


Figure 2: Experiment setup

to make a sales pitch to overcome in-store noise without interfering with other sales sections.

3.3 Sales Pitch Audio

For auditory stimulation from a parametric array loudspeaker, we prepare for sales pitch audio carrying information of a target product including the *brand and the origin district*. Customers are considered to make a purchase decision based on three attributes: *brand, origin district, and price*. Price information is available in the sales section. We use a sales pitch talk and a song including brand and origin district information.

4 Empirical Experiment

We conducted an empirical experiment in a food store to evaluate the proposed method described in Section 3. In this experiment, we prepared five conditions to compare the effect of loudspeakers and song/audio interventions. The effect was evaluated based on the sales amount.

The following subsections describe the experiment setup, experiment conditions, and equipment used.

4.1 Experiment Setup

Figure 2 shows an experiment setup of auditory stimulation using a parametric array loudspeaker. In this experiment, we used *Yamato-maru eggplant* produced in Yamato-Koriyama City, Nara Prefecture in Japan. Note that Yamato-maru eggplants are a more expensive product than common eggplants. Therefore, customers do not purchase them on a daily basis.

This experiment was conducted at the AEON Mall Yamato-Koriyama store located in Yamato-Koriyama City, Nara Prefecture in Japan, for approximately one month. The experiment was conducted from July 4 to July 31 in 2022, which was the season for Yamato-maru eggplants.

In this experiment, we did not conduct sales promotions such as handing out flyers to avoid the influence of advance notice. To compare several conditions and accurately investigate changes in purchasing behavior, we kept the state of other sales media such as flyers and the food model of dishes

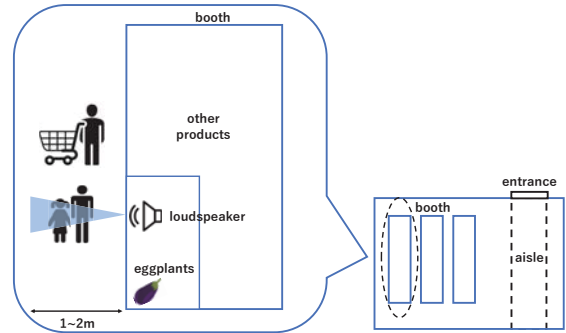


Figure 3: Layout of the supermarket during the experiment

using Yamato-maru eggplant at the sales section during the experiment period.

Note that the experiment was conducted with no control on store sounds. Store announcements and sound advertisement in other sales sections have been made as usual, which resulted in a noisy environment.

As shown in Fig. 3, we installed a parametric array loudspeaker approximately 1 to 2 meters away from the sales section. In a sales demonstration, a demonstrator directly explains products to consumers near the sales section. We emulated the similar situation as the sales demonstration where only the customers near the sales section are attracted. On installation of a parametric array loudspeaker, we conducted several in-store tests by adjusting the angle of the speakers to ensure that the sound was clearly audible in the aisles frequented by customers.

This experiment was approved by the Research Ethics Committee of Kio University (permission #R3-03-2).

4.2 Experiment Conditions

We conducted the following five experimental conditions to investigate the effects of a parametric array loudspeaker on promoting consumer purchasing behavior.

- Control (C): No intervention with auditory stimulation
- Test1 (T1): Sales pitch audio from a normal loudspeaker
- Test2 (T2): Sales song from a normal loudspeaker
- Test3 (T3): Sales pitch audio from a parametric array loudspeaker
- Test4 (T4): Sales song from a parametric array loudspeaker

We used the no-intervention condition with auditory stimulation as a baseline.

Table 1 shows the experiment schedule. We carefully scheduled the above test conditions to collect data to eliminate sales bias from the day of the week and the time of day.

4.3 Equipment

We used the Audfly FS-B1-Pro1 parametric array loudspeaker shown in Fig. 4. For comparison, we used a normal loudspeaker with a diameter of approximately 50 millimeters. We controlled each loudspeaker from a Raspberry

Table 1: Experiment schedule

Day in July	9:00–10:00	10:00–11:00	11:00–12:00	12:00–13:00	13:00–14:00	14:00–15:00	15:00–16:00	16:00–17:00	17:00–18:00	18:00–19:00	19:00–20:00	20:00–21:00
4 (Mon)	T4	C	T1	T2	T3	T4	C	T1	T2	T3	T4	C
5 (Tue)	T3	T4	C	T1	T2	T3	T4	C	T1	T2	T3	T4
6 (Wed)	T2	T3	T4	C	T1	T2	T3	T4	C	T1	T2	T3
7 (Thu)	T1	T2	T3	T4	C	T1	T2	T3	T4	C	T1	T2
8 (Fri)	T3	T4	C	T1	T2	T3	T4	C	T1	T2	T3	T4
9 (Sat)	T4	C	T1	T2	T3	T4	C	T1	T2	T3	T4	C
10 (Sun)	T3	T4	C	T1	T2	T3	T4	C	T1	T2	T3	T4
11 (Mon)	T1	T2	T3	T4	C	T1	T2	T3	T4	C	T1	T2
12 (Tue)	T2	T3	T4	C	T1	T2	T3	T4	C	T1	T2	T3
13 (Wed)	T1	T2	T3	T4	C	T1	T2	T3	T4	C	T1	T2
14 (Thu)	T3	T4	C	T1	T2	T3	T4	C	T1	T2	T3	T4
15 (Fri)	C	T1	T2	T3	T4	C	T1	T2	T3	T4	C	T1
16 (Sat)	C	T1	T2	T3	T4	C	T1	T2	T3	T4	C	T1
17 (Sun)	T4	C	T1	T2	T3	T4	C	T1	T2	T3	T4	C
18 (Mon)	T3	T4	C	T1	T2	T3	T4	C	T1	T2	T3	T4
19 (Tue)	T1	T2	T3	T4	C	T1	T2	T3	T4	C	T1	T2
20 (Wed)	C	T1	T2	T3	T4	C	T1	T2	T3	T4	C	T1
21 (Thu)	T2	T3	T4	C	T1	T2	T3	T4	C	T1	T2	T3
22 (Fri)	T4	C	T1	T2	T3	T4	C	T1	T2	T3	T4	C
23 (Sat)	T2	T3	T4	C	T1	T2	T3	T4	C	T1	T2	T3
24 (Sun)	T1	T2	T3	T4	C	T1	T2	T3	T4	C	T1	T2
25 (Mon)	T2	T3	T4	C	T1	T2	T3	T4	C	T1	T2	T3
26 (Tue)	T4	C	T1	T2	T3	T4	C	T1	T2	T3	T4	C
27 (Wed)	T3	T4	C	T1	T2	T3	T4	C	T1	T2	T3	T4
28 (Thu)	T4	C	T1	T2	T3	T4	C	T1	T2	T3	T4	C
29 (Fri)	T2	T3	T4	C	T1	T2	T3	T4	C	T1	T2	T3
30 (Sat)	T1	T2	T3	T4	C	T1	T2	T3	T4	C	T1	T2
31 (Sun)	C	T1	T2	T3	T4	C	T1	T2	T3	T4	C	T1



Figure 4: Audfly FS-B1-Pro1 parametric array loudspeaker

Pi 3A+ to switch the sound signals and loudspeakers based on the schedule shown in Table 1. The equipment was connected to the Internet via an LTE network to be remotely controlled.

4.4 Song and Audio Setting

We used sales pitch audio and sales song including *brand and the origin district* information, as described in Section 3.3. As shown in Fig. 5, sales pitch audio was made by the VOICEVOX text-to-speech software. The sales song is a short song ad-



Figure 5: Sales pitch audio made by VOICEVOX text-to-speech software

vertising Yamato-maru eggplant. Both sales pitch audio and sales song include the brand *Yamato-maru* eggplant from the origin district of *Yamato-Koriyama City*.

4.5 Sound Playback System

We implemented a shell script program, which runs on a Raspberry Pi, to play the sales pitch audio or sales song from a parametric array loudspeaker or a normal loudspeaker. The playing schedule was stored in a CSV file, which was loaded into the program to choose the sound source and an output loudspeaker.

We put a gap time of 10 minutes for the transition between test conditions. The sales pitch audio or sales song was re-

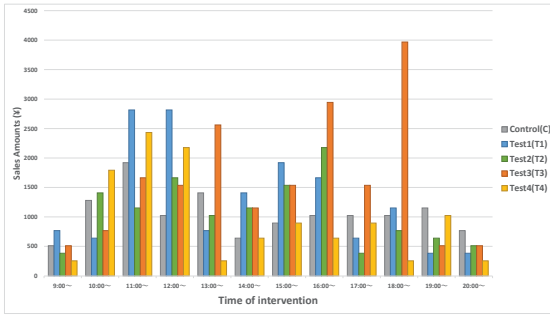


Figure 6: Hourly sales amounts

Table 2: Total sales for each condition

Conditions	Sales (JPY)
C : No intervention with auditory stimulation	12672
T1: Ordinary loudspeaker and sales pitch audio	15360
T2: Ordinary loudspeaker and advertising song	12800
T3: PAL and sales pitch audio	19200
T4: PAL and advertising song	11520

peatedly played for 50 minutes. For example, for Test1 from 9:00 to 10:00, sales pitch audio was repeatedly played from 10:05 to 10:55 to prevent sound played simultaneously from normal and parametric array loudspeakers.

5 Evaluation

In this section, we evaluate the sales data to show the impact of the method of promoting customers' purchasing behavior with a parametric array loudspeaker.

5.1 Evaluation and Discussion of Empirical Experiment

We used data only from the AEON MALL Yamato-Koriyama store in this paper because Yamato-maru eggplant was not sold for a long time in other store due to the limited production. Therefore, it is impossible to compare the results with those for other stores or regions.

For about one month, from July 4 to July 31, we conducted five conditions of Control (C) to Test4 (T4) in the experiment. From 9:00 to 21:00, we calculated the sales of Yamato-maru eggplant every hour.

The total sales amounts for each experimental condition are shown in Table 2. The total sales amounts of T3 was the highest in Table 2. The total sales amounts of T3 was 1.52 times higher than that of C.

The hourly sales amounts are shown in Figure 6. In Figure 6, the vertical axis represents the cumulative sales amount for each hour, and the horizontal axis represents time of intervention. It can be seen that Test3 (T3) showed significant growth at 16:00 and 18:00. It also can be observed that Test2 (T2) showed growth at 11:00 and 12:00. According to the total sales for each hour, there is an increasing trend from 11:00 to 12:00. It is expected that a lot of people would purchase ingredients for lunch or dinner in this period.

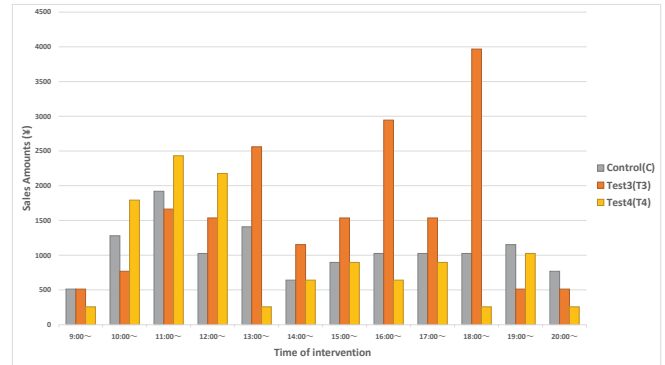


Figure 7: Sales amounts of conditions with parametric array loudspeaker

Table 3: Total sales of conditions with parametric array loudspeaker

	Control (C)	Test3 (T3)	Test4 (T4)
9:00-	512	512	256
10:00-	1280	768	1792
11:00-	1920	1664	2432
12:00-	1024	1536	2176
13:00-	1408	2560	256
14:00-	640	1152	640
15:00-	896	1536	896
16:00-	1024	2944	640
17:00-	1024	1536	896
18:00-	1024	3968	256
19:00-	1152	512	1024
20:00-	768	512	256

5.2 Evaluation and Discussion of Conditions Using Parametric Array Loudspeaker

We evaluate and discuss T3 and T4, which involve the use of parametric array loudspeaker. We will compare these conditions to the following three scenarios:

- C : No intervention with auditory stimulation
- T3: Intervention with a sales pitch audio
- T4: Intervention with a sales song

Hourly sales are shown in Table 3 and Figure 7. First, compared T4 to C, sales increased from 10:00 to 12:00. However, from 13:00 onward, sales in T4 remained the same or even lower than sales in C. According to Table 2, the total sales in T4 were lower than sales in C.

Sales of T3 were initially similar to or lower than C until 12:00. However, sales for T3 exceeded those for C from 12:00 to 18:00. Furthermore, according to Table 6, T3 had the highest total sales. T3 was the most effective condition in promoting customers' purchasing behavior.

As a result, it is clear that changes in sound have an impact on sales. We expect that using sales pitch audio is effective when promoting purchasing behavior using a parametric array loudspeaker. However, sales decreased when the sales song was used. It is expected that a parametric array loudspeaker does not necessarily have an effect for promoting purchasing behavior.

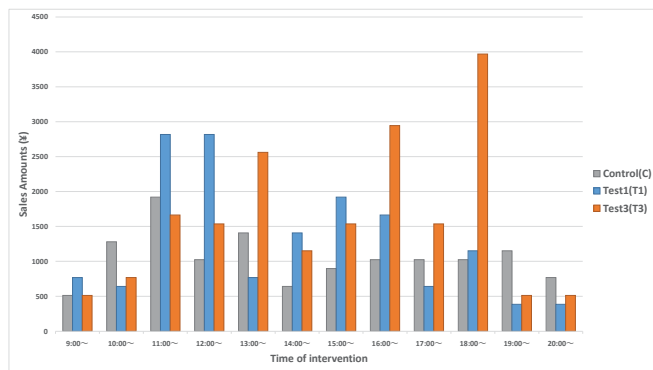


Figure 8: Sales amounts of conditions with sales pitch audio

Table 4: Total sales of conditions with sales pitch audio (JPY)

	Control (C)	Test1 (T1)	Test3 (T3)
9:00~	512	768	512
10:00~	1280	640	768
11:00~	1920	2816	1664
12:00~	1024	2816	1536
13:00~	1408	768	2560
14:00~	640	1408	1152
15:00~	896	1920	1536
16:00~	1024	1664	2944
17:00~	1024	640	1536
18:00~	1024	1152	3968
19:00~	1152	384	512
20:00~	768	384	512

5.3 Results and Discussion of Conditions Using Sales Pitch Audio

We now turn our attention to evaluating and discussing T1 and T3, which involve the use of sales pitch audio. We compared sales amount of the following three conditions:

- C: No intervention with auditory stimulation
- T1: Intervention with a normal loudspeaker
- T3: Intervention with a parametric array loudspeaker

Table 4 and Figure 8 show the hourly sales amount. Comparing T3 with T1, we can confirm that T3 was effective at seven specific times: 10:00–, 13:00–, 16:00–, 17:00–, 18:00–, 19:00–, and 20:00–. Compared to C, the sales amount of T3 was also higher in seven time periods different from the above time periods. This indicates that T3 effectively promoted customers' purchasing behavior.

Compared to T1, T3 showed a higher sales amount at eight time periods except the above five. In particular, at 13:00 and 18:00, the sales exceeded the others by approximately 1300 to 2000 yen. As shown in Table 2, the total sales amount of T3 exceeds that of T1 by approximately 4,000 yen. Based on these results, the sales pitch audio had a promotional effect regardless of the type of loudspeaker.

5.4 Overall Evaluation and Discussion

The objective of the empirical experiment was to investigate the effectiveness of techniques for promoting customers'

purchasing behavior using a parametric array loudspeaker. Summarizing the experimental results, we have two main findings:

- Test3, which uses a parametric array loudspeaker and sales pitch audio, was the most effective condition.
- Even when using a normal loudspeaker, sales pitch audio showed effectiveness.

The experimental results showed that the combination of a parametric array loudspeaker and a sales pitch audio was the most effective approach.

Although a parametric array loudspeaker with a sales pitch audio was effective, the second-highest effectiveness was derived with a sales pitch audio using a normal loudspeaker. This suggests that auditory stimulation directly engaging customers might be more important than the type of loudspeaker.

The intervention using a parametric array loudspeaker had a limited effect when using a sales song due to the limited range of sound delivery. To fully take advantage of a parametric array loudspeaker, stimulation sound needs to be properly selected.

6 Conclusion

In this paper, we presented a new auditory stimulation method promoting customers' purchasing behavior, especially for food stores. We use a parametric array loudspeaker delivering sound signals to a specific area, which overcomes noisy in-store environments. We conducted an empirical experiment with the target product of Yamato-maru eggplants. The analytical results show that auditory stimulation by sales pitch audio sent from a parametric array loudspeaker increased sales amount by 1.52 times. This result indicates that using appropriate sound with a parametric array loudspeaker can effectively promote consumers' purchasing behavior.

REFERENCES

- [1] J. P. Vigna and E. W. Mainardes, "Sales promotion and the purchasing behavior of food consumers," *ReMark - Revista Brasileira de Marketing*, vol. 18, no. 3, pp. 101–126, Dec. 2019.
- [2] D. Mayumi, Y. Nakamura, Y. Matsuda *et al.*, "Aroma nudges: Exploring the effects on shopping behavior in a supermarket," in *First International Workshop on Digital Nudging and Digital Persuasion, DNDP*, 2022.
- [3] S. Spendrup, E. Hunter, and E. Isgren, "Exploring the relationship between nature sounds, connectedness to nature, mood and willingness to buy sustainable food: A retail field experiment," *Appetite*, vol. 100, pp. 133–141, May 2016.
- [4] R. E. Milliman, "Using Background Music to Affect the Behavior of Supermarket Shoppers," *Journal of Marketing*, vol. 46, no. 3, pp. 86–91, 1982.
- [5] A. C. North, D. J. Hargreaves, and J. McKendrick, "The influence of in-store music on wine selections," *Journal of Applied Psychology*, vol. 84, pp. 271–276, 1999.
- [6] F. V. Garlin and K. Owen, "Setting the tone with the tune: A meta-analytic review of the effects of background music in retail settings," *Journal of Business Research*, vol. 59, no. 6, pp. 755–764, Jun. 2006.

Controlling Difficulty Levels in Automated Question Generation Using Large Language Models: An Initial Study for Educational Video with ChatGPT

Xingyu Cai^{*}, Tomoo Inoue^{**}

^{*}Graduate School of Comprehensive Human Sciences, University of Tsukuba, Japan

^{**}Institute of Library, Information and Media Science, University of Tsukuba, Japan

^{*}s2321721@u.tsukuba.ac.jp, ^{**}inoue@slis.tsukuba.ac.jp

Abstract - Question generation is a pivotal component in education. However, crafting high-quality questions is a challenging and time-consuming task for educators. This study explores the potential of using Large Language Models (LLMs), specifically ChatGPT, to generate difficulty-controllable questions from video learning materials. We processed a university-level introductory biology video lecture as an example to extract subtitles and slide content. Using predefined cognitive difficulty criteria, we guided ChatGPT to generate questions at varying difficulty levels. This initial investigation expects the feasibility of integrating LLMs for automated question generation in educational contexts.

Keywords: Educational technology, question generation, large language model, difficulty control, video learning.

1 INTRODUCTION

Question and questioning activities hold an essential role in education. Some of the benefits of using questions are 1) offering the opportunity to practice retrieving information from memory; 2) providing learners with feedback about their misconceptions; 3) focusing learners' attention on the important learning material; 4) reinforcing learning by repeating core concepts; and 5) motivating learners to engage in learning activities (e.g., reading and discussing) [1]. However, producing questions is a difficult task even for educators. The challenges in producing questions for educators are: 1) producing good questions requires training, experience, and resources; 2) producing questions requires the educator's dedicated time and effort, while at the same time, they also have to deal with other essential tasks such as teaching, developing curriculum, and assessing students' performance. There is a need for a new alternative solution to rapidly generate a large number of questions.

One solution to overcoming the challenges in producing questions is automatic question generation (AQG). AQG can reduce the time and effort required to generate questions by automatically generating distractors, selecting informative sentences, and generating stems [2, 3, 4]. One of the critical advancements needed in AQG is the ability to control the difficulty level of the generated questions. Difficulty-Controllable Question Generation (DCQG) aims to generate questions at specified difficulty levels, enhancing the adaptability and effectiveness of educational tools. Controlling question difficulty allows educators to tailor questions to students' varying skill levels, thereby enhancing

learning outcomes. Recent studies have explored DCQG by defining question difficulty based on factors such as whether a question can be correctly answered by a QA model [5] or the number of inference steps required to answer it [6]. Moreover, a notable limitation of AQG is that it predominantly relies on text as the source for question generation [1], with limited research exploring the use of video material.

With the rise of large language models (LLMs), they offer an alternative approach already applied in the educational domain. For example, LLMs have been used to enhance programming e-textbooks, generate reading comprehension exercises, and improve foreign language writing [7, 8]. Because of the foreshadowing of the above applications, we believe that LLM can also be used to control difficulty. In this study, we utilized LLMs to generate difficulty-controllable questions with video learning material as the source of questions.

2 RELATED WORK

2.1 Automatic Question Generation

Manual construction of questions is a process that requires training, experience, and resources [1]. What researchers have proposed to solve this problem is automatic question generation. Among question types, teachers are often troubled by open-ended questions. Developing open-ended questions presents a challenge because it requires a considerable amount of time for a teacher to think about the cognitive level students want to achieve. Teachers also need to consider the complexity of the answers to the learning material, read and score student responses, and carefully construct criteria/rubrics to ensure the reliability of the scoring [9]. The major difficulty in generating open-ended questions is difficulty control, especially cognitive level control. The generation of open-ended questions can be divided into two parts: selecting information sentences and controlling difficulty.

Firstly, researchers have used various methods regarding sentence selection, like machine learning technology. They engineered features to train a classifier, including sentence length, target word position, co-occurrence scores, and verb complexity, and decide if a sentence is suitable to be used as a question stem by the classifier [4]. There have also been attempts at NLP technology where researchers presented a system for automatically generating factual close (fill-in-the-blank) questions. They identify informative sentences and

POS (part-of-speech) tag patterns like the presence of determiners, proper nouns, numbers, etc. [10].

Secondly, regarding the difficulty control, researchers have explored it by defining question difficulty based on whether a question can be correctly answered by a QA model [5], and they also proposed a more interpretable approach by defining question difficulty as the number of inference steps required to answer it [6].

2.2 LLM in Education

In prior research, LLM has been proven to have good generalization and summarization capabilities [11, 12, 13]. In this study, we want to focus on the difficulty control of open-ended questions.

Recent pre-trained Language Models (LLMs) such as ChatGPT [14] and GPT-3 have demonstrated remarkable capabilities in completing complex tasks through natural language prompts, significantly reducing the need for advanced programming skills and facilitating fluent multi-turn conversations. These characteristics help it gain a variety of application opportunities in education. From the teachers' viewpoint, LLMs play an important role in enhancing teaching abilities. For example, using ChatGPT to generate educational questions greatly reduces teachers' workload [15]. Moreover, LLMs can assist teachers in grading tasks [16].

However, even for NLP experts, prompt engineering requires extensive trial and error, iteratively experimenting and assessing the effects of various prompt strategies on concrete input-output pairs before assessing them more systematically on large datasets [17]. It is still difficult for non-experts to use prompts to let LLM accurately complete specific tasks.

3 METHODS

To conduct this study, we first determine our exploration space. LLM can accept a variety of inputs, but the research on video input is less than that on text input, so we chose textual video content as the input for this experiment. LLM has also been widely used in education and has great advantages in processing text information. This study aims to explore how the questions of a targeted video learning material are generated using LLM, in this case, ChatGPT. Our exploration focuses on whether LLM can automatically generate questions with controllable difficulty. To this end, we use LLM to conduct experiments on difficulty-controllable question generation to explore this question.

The following is an overview of the workflow of this study.

- 1) Preparation of a video learning material
 - 1.1) Video scene capture
 - 1.2) Scene text extraction
- 2) Question generation
 - 2.1) Instructions for using materials and input of question difficulty criteria to ChatGPT
 - 2.2) Video content input to ChatGPT
- 3) Question output

3.1 Preparation of a Video Learning Material

Video Scene Capture

We filter out keyframes with slides manually.

Scene Text Extraction

We use OCR (optical character recognition) to extract the text.

3.2 Question Generation

Developing questions presents a challenge because it requires considerable time for a teacher to think about the cognitive level [9]. We want to explore whether LLM can reach human standards in terms of controlling difficulty. The prompt for generating questions is divided into two parts: the first part includes the difficulty standards and the desired difficulty level of the questions, and the second part includes the video content (subtitles and slide content).

Instructions for Using Materials and Input of Question Difficulty Criteria

Describing the setting and providing the context will be a good strategy for better prompts [18]. Thus, we first describe the material. Then, we give the criteria for the difficulty level; similar to Kim et al. [19], the criteria are added to our prompt as lines of the form: Name and Description. At the end of the prompt, we ask LLM to generate questions of a specific level based on the given criteria.

Video Content Input

We input the subtitles and slide content extracted from the video into ChatGPT, providing a comprehensive context for question generation.

By following this procedure, we aim to generate a set of questions that varied in difficulty, allowing us to evaluate whether the LLM could effectively control the difficulty level of the generated questions.

3.3 Question Output

The questions generated by ChatGPT are collected and categorized according to their difficulty levels.

4 EXPERIMENT OF QUESTION GENERATION WITH CHATGPT

To explore whether the difficulty of questions generated by AI can be controlled like a human teacher, we used ChatGPT, the most complete large language model currently available, especially its latest model (GPT-4o) to conduct a feasibility experiment.

4.1 Learning Material

We used an online video lecture on YouTube with the Creative Commons license from authoritative/respected authors as a learning material source for LLM to generate/improve the questions. The video was a short introductory video in talking-head style with instructors appearing in the video, which was most common in online learning material and the most engaging for a student to watch [20]. We found and took an introductory video titled: "1.

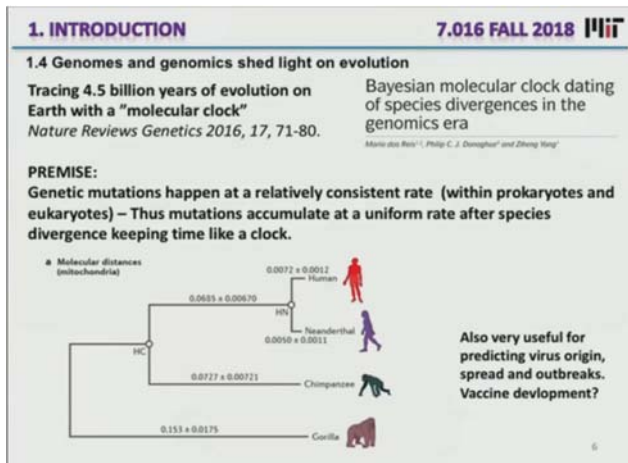


Figure 1: Example of the original slide.

Introduction Course Organization of MIT 7.016 Introductory Biology Fall 2018" presented by Professors Imperiali of MIT from the MIT OpenCourseWare channel with a Creative Commons BY-NC-SA license. The original video duration was 38 minutes and 45 seconds. For this experiment, we took the 15-minute general introduction content part of the video from minutes 10:00 to minutes 25:05 (the first 10 minutes of the original video only contains a class introduction, and the general overview of the course itself).

4.2 Slides and Extracted Text

First, we manually filtered out the keyframes with slides. After removing duplicate slides, a total of 10 slides were found. Most of these were slides that combined pictures and texts, like Figure 1, and the pictures also contained texts. Then, we upload the PNG image to Google Drive, right-click on the image to extract text from it, and choose to open it with Google Documents. Google Drive will automatically perform text recognition and output the text in the image to the bottom of the image, see Figure 2. We can see that the results of text recognition basically cover all the text information in this picture. The extracted text is not processed, but the page number is added before each slide to distinguish the content of the slides on different pages.

4.3 Generation Prompt and Procedure

Prompt

The prompt for generating questions was divided into two parts.

Firstly, we told ChatGPT that it would receive some educational material information, which included subtitles from a lecture video and text from slides in the video, and they are used in biology class. After that, we gave ChatGPT how to use this material: We wanted it to generate questions of a specific question level according to the given criteria. As shown in Table 1, these criteria were defined by Guthrie and Taboada [21], who explored the development and effectiveness of various cognitive strategies in enhancing reading comprehension skills. In their study, they used these criteria

1. INTRODUCTION

1.4 Genomes and genomics shed light on evolution

Tracing 4.5 billion years of evolution on Earth with a "molecular clock"

Nature Reviews Genetics 2016, 17, 71-80.

PREMISE:



7.016 FALL 2018 |ii

Bayesian molecular clock dating of species divergences in the genomics era

Mario dos Reis, Philip C. J. Donoghue and Ziheng Yong'

Genetic mutations happen at a relatively consistent rate (within prokaryotes and eukaryotes) - Thus mutations accumulate at a uniform rate after species divergence keeping time like a clock.

a Molecular distances (mitochondria)

0.0072 ± 0.0012

Human

0.0685 ± 0.00670

HN

HC

0.0727 ± 0.00721

-Neanderthal

0.0050 ± 0.0011

Chimpanzee

0.153 ± 0.0175

Gorilla

Also very useful for predicting virus origin, spread and outbreaks. Vaccine development?

6

Figure 2: Extracted text from the original slide.

to measure student-generated question quality for conceptual knowledge comprising four question quality levels.

Secondly, we manually input the video content (subtitles from YouTube and slide text extracted from the slides), as shown in Figure 3 and Figure 4. The same prompt was run five times to increase the sample size. After every five runs, we change the requested difficulty level of the questions (from 1 to 4) in the prompt.

Workflow

The entire experiment was conducted in ChatGPT web application. The experiment was performed according to the following steps:

a) Provide the prompt: *I will give you some subtitles and text on the slides from the lecture video of a biology class, and I want you to design 5 level 1 questions using the criteria I have given. Below is the standard*

Level Name Description

Level 4 *Pattern of Relationship* At this level, questions display scientific knowledge coherently expressed to probe the interrelationship of concepts. They request principled understanding with evidence for complex interactions among multiple concepts and possibly across concepts.

Level 3 *Complex Explanation* At this level, Questions request an elaborated explanation about a specific concept aspect with accompanying evidence.

Level 2 *Simple Description* At this level, questions request general information that denotes a link between concepts. The question can be simple, yet the answer may contain multiple facts and generalizations.

Level 1 *Factual Information* At this level, questions are simple in form and request a simple answer,

Table 1: Taboada and Guthrie's questioning level rubric

Level	Name	Description
Level 4	Pattern of Relationship	At this level, questions display scientific knowledge coherently expressed to probe the interrelationship of concepts. They request principled understanding with evidence for complex interactions among multiple concepts and possibly across concepts.
Level 3	Complex Explanation	At this level, Questions request an elaborate explanation about a specific concept aspect with accompanying evidence.
level 2	Simple Description	At this level, questions request general information that denotes a link between concepts. The question can be simple, yet the answer may contain multiple facts and generalizations.
Level 1	Factual Information	At this level, questions are simple in form and request a simple answer, such as a single fact. Examples of this level of question are: How big are bats? Do sharks eat trash? How much do bears weigh?

such as a single fact. Examples of this level of question are: How big are bats? Do sharks eat trash? How much do bears weigh?

- b) Use the subtitles and text extracted from the slides as the second prompt.
- c) Repeat steps a and b 4 times.
- d) Repeat steps a and b 5 times, changing the prompt to generate level 2 questions.
- e) Repeat steps a and b 5 times, changing the prompt to generate level 3 questions.
- f) Repeat steps a and b 5 times, changing the prompt to generate level 4 questions.

4.4 Results

As repeated inputs of a question to ChatGPT generally produce different responses [22], we ran the prompts for each of the four difficulty levels five times to increase the reliability of this experiment. This way we ended up with

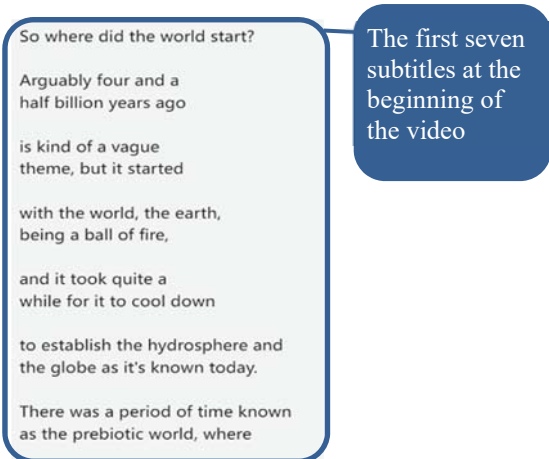


Figure 3: Examples of subtitles used as a part of the second prompt

100 questions. The response from ChatGPT was obtained every time the question generation process was performed (Fig.5).

Since the prompt runs are all independent, we observed a duplicate problem in the generated questions. Although there was semantic duplication, the expressions of the sentences were slightly different, and thus these questions were not excluded from the results.

Number of words

By observing the generated questions, we first found an overall trend. The number of words in a generated question

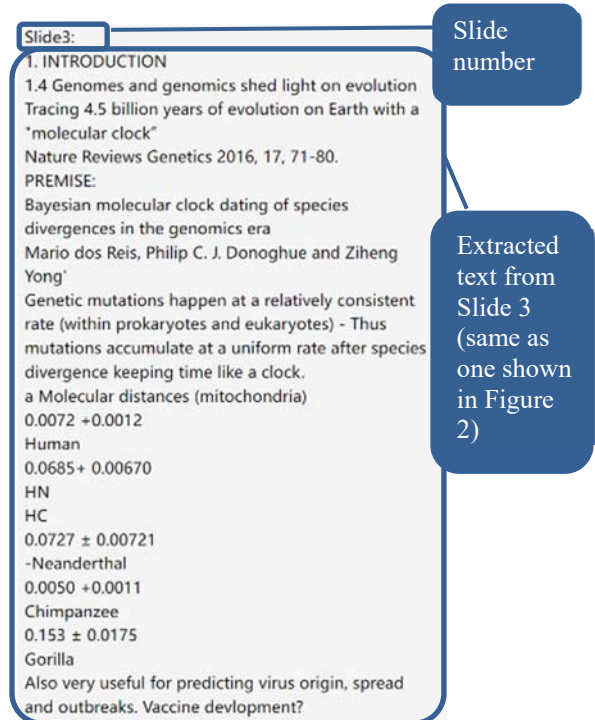


Figure 4: Example of the slide content used as a part of the second prompt.

 Level 4 Questions:

1. Explain the concept of a "prebiotic world" and its significance in the context of the origin of life on Earth. How do the conditions of this era contribute to the emergence of life forms? Provide evidence from the lecture text to support your explanation.
2. Discuss the transition from the "pre-RNA world" to the "RNA world" and its implications for the evolution of life. How does the RNA world hypothesis shed light on the early stages of biochemical evolution? Support your answer with references to specific molecules and processes mentioned in the lecture.
3. Describe the role of lipid bilayers in the evolution of cells. How do these structures contribute to cellular compartmentalization and the differentiation between internal and external environments? Provide examples from the lecture to illustrate the importance of lipid bilayers in cellular function.
4. Examine the significance of the Human Genome Project in the field of genomics and its impact on understanding human diversity and evolution. How has the mapping of the human genome contributed to advancements in medicine and biological research? Support your argument with examples from the lecture slides.
5. Explain the concept of a "molecular clock" as discussed in the lecture. How does the molecular clock hypothesis aid in tracing evolutionary relationships among different species? Provide examples of molecular divergence discussed in the lecture and discuss their implications for understanding evolutionary history.



Figure 5: Example of the generated questions.

seemed increased as the difficulty level of questions increased. The average number of words and a typical example of generated question in each level is shown in Table 2.

Command word

"Command word" is a specific term or phrase used in a question or instruction to guide a respondent to approach their answers. The command word used in a question varied by the

difficulty level. The typical command words found in each level were shown in Table 2. We found that as the difficulty level of a question increased, the level of thinking required by the question also became deeper.

Relevance to the content

After trying to answer each question, we found that all the answers to the questions could be found in the provided materials, and no question was irrelevant. The answers were evenly distributed in the learning video, and covered most of the knowledge elements in the video.

5 DISCUSSION

5.1 Possibilities for Better Prompt Design

Prompt design may be crucial in generating questions with controllable difficulty levels. The following strategies can be considered to improve the effectiveness of prompts in guiding ChatGPT.

Asking ChatGPT to play roles can help streamline the desired process of obtaining the information [18]. In the case of this experiment, we can add "If you are a biology teacher in college..." at the beginning of the prompt to ensure that ChatGPT obtains basic biological knowledge.

In this experiment, there was no limit on the number of sentences, which resulted in an increased number of sentences as the level increased. Limiting the number of sentences might ChatGPT to focus more on the changes in sentence difficulty than on the length of the sentences.

5.2 Evaluation of the Generated Questions

Although this paper presented an initial study of question generation including a brief report on the results, the

Table 2: Number of words, command words and example of a question in each level

Level	Average number of words	Typical command word	An example generated question
Level 4	29.8	<i>Investigate Evaluate examine</i>	<i>"Evaluate the significance of the Human Genome Project in revolutionizing our understanding of genetics and genomics. How has the mapping of the human genome paved the way for advancements in personalized medicine, evolutionary biology, and biotechnology? Provide examples from the text to illustrate the impact of genomic research on various fields of biology."</i>
Level 3	23.4	<i>analyze elaborate</i>	<i>"Describe the concept of a "molecular clock" as mentioned in the text. How does it aid in tracing the evolutionary history of species, and what insights does it offer into human evolution compared to other primates?"</i>
Level 2	18.9	<i>explain compare contrast</i>	<i>"Explain the significance of lipid bilayers in cellular evolution and function"</i>
Level 1	6.8	<i>what when describe</i>	<i>"When did the Earth begin to form?"</i>

evaluation of the generated questions is to be conducted as the next step. It is considered that the evaluation could also be done by human and AI as well as question generation.

6 CONCLUSION

This study explored the feasibility of using Large Language Models (LLMs) to generate difficulty-controllable questions from video learning materials. The experiment using ChatGPT demonstrated that ChatGPT could generate questions seemingly at different cognitive levels. This expects that LLMs have the potential to control question difficulty effectively.

Future research should focus on improving the prompt for LLM to generate questions, according to the more detailed evaluation results of the generated questions.

REFERENCES

- [1] Kurdi, G., Leo, J., Parsia, B., Sattler, U., & Al-Emari, S., A Systematic Review of Automatic Question Generation for Educational Purposes, *International Journal of Artificial Intelligence in Education*, Vol.30, No.1, pp.121–204 (2019).
- [2] Susanti, Y., Tokunaga, T., Nishikawa, H., & Obari, H., Automatic distractor generation for multiple-choice English vocabulary questions, *Research and Practice in Technology Enhanced Learning*, Vol.13, No.1, (2018).
- [3] Majumder, M., & Saha, S. K., A System for Generating Multiple Choice Questions: With a Novel Approach for Sentence Selection, *Proceedings of the 2nd Workshop on Natural Language Processing Techniques for Educational Applications*, pp. 64–72 (2015).
- [4] Correia, R., Baptista, J., Eskenazi, M., & Mamede, N., Automatic Generation of Cloze Question Stems, *Lecture Notes in Computer Science*, pp. 168–178 (2012).
- [5] Gao, Y., Bing, L., Chen, W., Lyu, M. R., & King, I., Difficulty controllable generation of reading comprehension questions, *arXiv preprint arXiv:1807.03586* (2018).
- [6] Cheng, Y., Li, S., Liu, B., Zhao, R., Li, S., Lin, C., & Zheng, Y., Guiding the growth: Difficulty-controllable question generation through step-by-step rewriting, *arXiv preprint arXiv:2105.11698* (2021).
- [7] Narayanan, A. B. L., Hendrawan, R. A., & V. V., Enhancing Programming eTextbooks with ChatGPT Generated Counterfactual-Thinking-Inspired Questions, *arXiv preprint arXiv:2306.00551* (2023).
- [8] Athanassopoulos, S., Manoli, P., Gouvi, M., Lavidas, K., & Komis, V., The use of ChatGPT as a learning tool to improve foreign language writing in a multilingual and multicultural classroom, *Advances in Mobile Learning Educational Research*, Vol.3, No.2, Article 2 (2023).
- [9] Piontek, M. E., Best practices for designing and grading exams, *Occasional Paper*, Vol.24, No.1, pp.1-12 (2008).
- [10] Das, B., & Majumder, M., Factual open cloze question generation for assessment of learner's knowledge, *International Journal of Educational Technology in Higher Education*, Vol.14, No.1, pp.24 (2017).
- [11] Rahman, M. R., Koka, R. S., Shah, S. K., Solorio, T., & Subhlok, J. (2024). Enhancing lecture video navigation with AI generated summaries. *Education and Information Technologies*, 29(6), 7361–7384.
- [12] Li, Y., et al., ChatCite: LLM Agent with Human Workflow Guidance for Comparative Literature Summary, *arXiv preprint arXiv:2403.02574* (2024).
- [13] Keswani, G., et al., Abstractive Long Text Summarization Using Large Language Models, *International Journal of Intelligent Systems and Applications in Engineering*, Vol.12, No.12s, pp.160-168 (2024).
- [14] OpenAI, CHATGPT: Optimizing language models for dialogue, Retrieved from <https://openai.com/blog/chatgpt/> (2022).
- [15] Elkins, S., Kochmar, E., Serban, I., & Cheung, J. C. How useful are educational questions generated by large language models?. *International Conference on Artificial Intelligence in Education*, pp. 536-542 (2023, June).
- [16] Guo, B., Zhang, X., Wang, Z., Jiang, M., Nie, J., Ding, Y., ... & Wu, Y., How close is chatgpt to human experts? Comparison corpus, evaluation, and detection, *arXiv preprint arXiv:2301.07597* (2023).
- [17] Zamfirescu-Pereira, J. D., Wong, R. Y., Hartmann, B., & Yang, Q., Why Johnny Can't Prompt: How Non-AI Experts Try (and Fail) to Design LLM Prompts, *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1–21 (2023).
- [18] Meskó, B., Prompt Engineering as an Important Emerging Skill for Medical Professionals: Tutorial, *Journal of Medical Internet Research*, Vol.25, pp. e50638 (2023).
- [19] Kim, T. S., Lee, Y., Shin, J., Kim, Y.-H., & Kim, J., EvalLM: Interactive Evaluation of Large Language Model Prompts on User-Defined Criteria, *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 1–21 (2024).
- [20] Ilioudi, C., Chorianopoulos, K., & Giannakos, M. N., Comparing the camera shot styles of video lectures: Close-up versus broad framing of whiteboard and lecturer, *Diethnes Synedrio Gia TiN Anoikti Kai Ex Apostaseōs Ekpaideusi*, Vol.7, No.6B (2016).
- [21] Taboada, A., & Guthrie, J. T., Growth of cognitive strategies for reading comprehension, In *Motivating Reading Comprehension*, pp. 273-306, Routledge (2004).
- [22] Ouyang, T., MaungMaung, A., Konishi, K., Seo, Y., & Echizen, I. Stability Analysis of ChatGPT-based Sentiment Analysis in AI Quality Assurance. *arXiv preprint arXiv:2401.07441* (2024).

Keynote Speech 2:
Prof. Masahiko Tsukamoto
(Professor, Graduate School of
Engineering, Kobe University)
(Chair: Koichi Ishibashi)

18th International Workshop on Informatics (IWIN2024) Keynote Speech 2
 INFSOC (Informatics Society), 2024.9.3, Yanagawa, Fukuoka

Privacy Issues on Smartglass AI

Masahiko Tsukamoto
 Kobe University

This work was supported by JSPS KAKENHI Grant Numbers JP22H00550

Masahiko Tsukamoto, Dr./Prof. Kobe University

Masahiko TSUKAMOTO received the B.E., M.S., and Dr.E. degrees from **Kyoto University**, Kyoto, Japan, in 1987, 1989, and 1994, respectively. From April 1989 to February 1995, he was a research engineer at **Sharp Corporation**. From March 1995 to September 1996, he was an assistant professor at the Department of Information Systems Engineering of **Osaka University** as an assistant professor, and From October 1996 to September 2004, he was an associate professor at the same department. Since October 2004, he has been a professor at **Kobe University**. He researches on system, interface, and applications of **wearable and ubiquitous computing**. Their target application areas include entertainment, healthcare, and ecology. He has been **using HMD** and wearable computers **in his daily life since 2001**.



塚本・寺田研究室
 塚本 昌彦



nippon.com
 ウェアラブルは新たな時代を切り...



リクナビNEXT
 ウェアラブル伝道師☆塚本...



078Kobe
 interactive : 3F | INT...



リクナビNEXT
 ウェアラブル伝道師...



Robot Watch
 Robot Watch-ニュー...



Wikipedia
 渋谷譲二 - Wikip

Today's AI is amazing!! (Unexpected rapid development)



YouTube Channel
Singularity Salon Online (in Japanese)
<https://www.youtube.com/@user-cf7md9zf9s>



Recent Rapid Developments in Artificial Intelligence (AI)

- **Deep learning** boom from around 2011
 - Significant improvement in image and voice recognition performance
- **Generative AI** boom from around 2021
 - In 2017, Transformer was announced by members led by Google [Vaswani 2017].
 - The performance of natural language, dialogue, and image generation has been greatly improved.
 - The AI's capabilities may already be at the level of passing the Turing test. However, the current AI does not pretend to be human.
 - There are still a lot of weird things about AI. They have hallucination, and lack logic&math intelligence and metacognition.
 - AI that automatically writes scientific papers has emerged (Sakana AI's "AI Scientist" [Lu 2024]).

The state of AI in the world

- Japan: **behind**
 - Industry & Applications (Robotics & Manufacturing), Policy & Governance, Human Resource Development
- United States: **ahead**
 - Leadership & Innovation (Google, Microsoft, OpenAI, Meta, etc.), Policy & Regulation, Education
- Europe (EU): **strongly controlling**
 - Regulation and Ethics (AI Act), Industrial Applications, Education and Research
- China: **strongly progressing**
 - National AI Strategy, Surveillance & Privacy, Technology & Human Resources

AI is getting smarter and smarter

- **Research and development** is actively being conducted in search of higher intelligence.
 - Countries and companies are competing for **human resources and GPUs**.
- The debate on **AI governance** is gaining momentum.
 - Creation of **guidelines** for development and utilization, **threat theory** and **alignment** issues
- Around 2027, **AGI (Artificial General Intelligence)** may be realized.
 - It may become an equal to others.
- Around 2029, **ASI (Artificial Super Intelligence)** may be realized.
 - It may become a being far beyond humans. The **singularity** may occur.

Currently, Smartglasses are sluggish (Even though the post-smartphone favorites)



YouTube Channel
Wearable Channel(in Japanese)
<https://www.youtube.com/@wearablechannel>



Smartglasses

- 1960s, HMD (Head Mounted Display) appeared
 - Invented by Ivan Edward Sutherland
- 1980s, MIT Wearable Labs
 - Prof. Alex 'Sandy' Pentrand, Prof. Steve Mann, Prof. Thad Sterner et al.
- 1990s, The **early days**
 - U.S. companies such as Xybernaut, MicroOptical, and Vuzix
- 2000s, The **challenging** period
 - In addition to U.S. companies such as Myvu, Japan companies such as Shimadzu, Hitachi, and Nikon are also taking on the challenge
- 2010s, The **fighting** period
 - Google, Microsoft, Epson and others compete on a global scale
- 2020s, The **decay** period
 - With metaverse and XR, a little tone-down due to the rise of AI?

Traditional 3 trends in smartglasses



Monocular glasses for business use

Google Glass, Vuzix m400, RealWear, etc.

Manufacturing, Equipment Inspection, Distribution

AR goggles

HoloLens, MagicLeap, and more

Entertainment, Design, Medical, Architecture



Tethered binocular glasses

Xreal Air, TCL NxtWear S, Viture, Rokid, Lenovo Legion Glasses, etc.

Video-watching, Gaming, Outdoor works



Recent 2 trends in smartglasses



Color pass-through goggles

Apple Vision Pro, Meta Quest 3, Pico 4, etc.

Desk work, Entertainment, Communication

Ordinary (good-looking) glasses

Vuzix Z-100/Shield, INMO Go, etc.

Audio glasses (Echo Frames, Solos),

Camera glasses (Ray-Ban Meta, Solos Vision)

AI, Notification, Information Presentation, Video Recording



Smartglasses vs AI

- Why **compete**?
 - Both are new challenges in the IT industry: enormous costs.
 - None of them have been monetized yet.
 - MS and Google have made a major shift to AI from smartglasses.
- Essentially, there is a **synergistic** effect.
 - Using AI in smartglasses – AI can be used in real life and in field work.
 - Create a rich and enjoyable living space with generative AI.

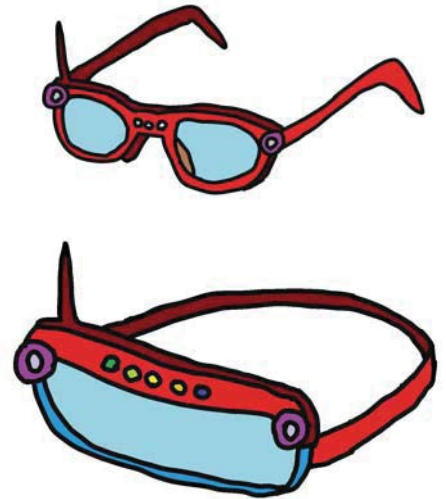
Smartglasses AI

Many products are becoming AI-enabled by the company or third parties.

- Ray-Ban Meta smartglasses: asking about what a user is looking, text interpretation in front of a user, voice commands, etc.
- Solos AirGo 3: ChatGPT for voice commands, translation, message reading.
- Solos AirGo Vision: Cameras and microphones can recognize people and objects and provide information in real time. This allows a user to translate, give directions, film, record, etc.
- Amazon Echo Frames and other audio glasses: contextual response, information access, home control, voice translation
- Vuzix Z100, Blade, Shield: Voice translation with display, business work Instructions
- RealWear / Vuzix m100: Logging of work instructions and remote work assistance
- Brilliant Monocle, Flame: AI assistant
- Apple Vision Pro: Apple Intelligence? Education or entertainment?

Smartglasses AI: What's Next

- AI is getting smarter and smarter.
- Smartglasses will eventually stand up.
 - "Information scouter" type glasses that look like ordinary glasses
 - Color pass-through type full-fledged AR goggles
- **Information Scouter Glasses**
 - Extension of ordinary (good-looking) glasses
 - Appropriate information is always available in daily life.
 - With appropriate suggestions, the user can change his/her behavior as he/she wishes.
- **Full-fledged AR goggles**
 - Extension of color-passthrough goggles
 - Everyday life space becomes a fun theme park.
 - Live with a reliable AR life partner.



Privacy Trends

- What is privacy?
 - Information that you want to keep private (or confidential)
 - Or person's private life, not allowing interference by others
- Privacy Rights are
 - The right not to be interfered with or infringed upon by others
 - The right to control his/her own information
- History of Privacy
 - From Antiquity to the Middle Ages: The Emergence of Privacy
 - 18th century: Age of enlightenment - development of the concept of privacy (Locke, Kant, etc.)
 - 19th century: Legal clarification (USA, 1890)
 - 20th century: changes due to technological progress (telephone, television, Internet)
 - 21st Century: Increasing complexity (Internet, Social Media, GDPR)
 - Present: Ongoing discussions and future changes (for AI and smartglasses)

The Current State of Privacy Law and Politics

- Japan
 - Act on the Protection of Personal Information (Personal Information Protection Act, 2003), subsequently amended as appropriate
 - In 2020, measures to strengthen the appropriate management of personal information and the protection of individual rights for companies
 - Personal Information Protection Commission
- United States
 - There is no nationwide blanket law. Industry regulation at the federal level, law enforcement varies from state to state
 - The California Consumer Privacy Act (CCPA) has strict provisions and has a significant impact across the United States
 - California Privacy Rights Act (CPRA) goes into effect in 2023, Consumer Rights
- Europe (EU)
 - General Data Protection Regulation (GDPR, 2018). Strict regulations. Consent and fines, international influence.
 - Applies to all companies that handle EU citizen data
- China
 - Personal Information Protection Act (PIPL, 2021). Similar to GDPR
 - Strict regulations on cross-border data transfers, and government efforts to strengthen data security

Privacy Issues

- Japan: **The gap between awareness and effectiveness**
 - Consumer awareness: Consumers and businesses are not as conscious of privacy as they are in Europe and the United States. Lack of vigilance, superficial efforts.
 - Global reach: Lack of resources and knowledge
- United States: **State variability and lack of comprehensive regulation**
 - Legal variability: Difficulty in uniform responses
 - Lack of comprehensive federal legislation: Inadequate overall protection.
- Europe: **Stricter legislation and technological advances**
 - Burden on businesses: Resources and costs are a significant burden for SMEs
 - Keeping up with technological evolution: Adapting to new technologies (AI, IoT, blockchain, etc.)
- China: **Government surveillance and cross-border data transfers**
 - Government Surveillance: Balancing personal privacy and national security
 - Cross-border data transfers: Difficulties for global companies to do business in China

AI Privacy Notice

- Transparency of data collection and use
 - Anonymization of personal data and risk of re-identification
 - Cross-Border Data Transfers and International regulations
 - Bias and discrimination in AI models
 - AI-powered surveillance and privacy breaches
 - Transparency and accountability of AI decisions
 - Data breaches and security risks
- Technological development, laws and regulations, guidelines, awareness and morals are necessary.

Privacy Challenges of Smartlasses AI

- Realtime data collection and privacy violations
 - It is necessary to analyze individual identification, behavior, and conversations
 - Limit the scope of data collection, and formulate guidelines.
- Abuse of personal identification and facial recognition technology
 - The use of facial recognition technology that can easily identify others, including strangers
 - Should be regulated, and the purpose of use should be clearly restricted.
- Unauthorized collection of voice assistants and surrounding conversations
 - It is necessary to have a mechanism to collect and analyze the surrounding voices at all times
 - Obtain the consent of the people around them.
- Gaze data privacy
 - Inferring the user's interests, concerns, emotional state, etc.
 - Obtaining the user's explicit consent should be obtained and the scope of use should be strictly controlled.
- Lack of transparency in unauthorized information collection and data use
 - The risk of misuse and misuse of personal data must
 - Transparent and accounted for.
- Data security and risk of leakage
 - Strong security measures are required
 - The possibility of containing sensitive information about individuals
- Social & ethical impact
 - There is a possibility of increased anxiety about surveillance in public places and surveillance society. Increasing threats to the use of smartglasses
 - Need to build social consensus and develop ethical guidelines

Grant-in-Aid for Scientific Research A "Privacy Control Technologies for Smartglasses AI"

- FY2022~FY2026
 - Members
 - Masahiko Tsukamoto, Engineering, Electricity (Research Supervisor and Meta-AI)
 - Seiichi Ozawa, Engineering, Electricity (Controllable AI System)
 - Tsutomu Terada, Engineering, Electricity (Situational Awareness Organization)
 - Masakatsu Morii, Engineering, Electricity (Privacy Mechanism)
 - Togo Tsukahara, International Culture (Science, Technology and Society)
 - Shinichi Kita (Psychology)
 - Takuya Shinkawa (Philosophy, Ethics)
- *All are Graduate School of Kobe University

Smartphone photography in the city

- Japan
 - Although there are no particularly strict laws, there are cases where it is socially inappropriate, and consideration of the surroundings is required.
 - Privacy-infringing filming can be sued. It is illegal to publish personally identifiable photographs without permission.
- United States
 - It's legal to photograph others in public (some states prohibit privacy-invading photography, and require a permit for commercial purposes)
 - Photography is prohibited in toilets, dressing rooms and certain situations
- Europe (EU)
 - Consent is required for personally identifiable photographs and online publication of such images.
 - In France and Germany, filming or publishing without permission carries legal risks.
- China
 - There are few special regulations for filming in public places.
 - Filming in certain locations (e.g., government facilities, military installations, certain public facilities, etc.) is strictly restricted.
 - Online sharing should be cautious (e.g., government surveillance)

Status of dashcams and surveillance cameras

- Shooting with a **dashcam**
 - Filming on the road: Generally acceptable and less regulated by law. Handling images with care
 - Protecting privacy: Other people's vehicles and passers-by may be in the picture. Mosaic processing is recommended for third-party publication and posting on SNS.
 - Legal restrictions: It may be used as evidence in court, but the use of footage may be restricted.
- Filming with **surveillance cameras**
 - Use on public and private property: It is often allowed in public places, but there are restrictions on installation. Even on private property, if the scope of filming extends to public places, it may be restricted.
 - Notification and transparency: Installers may be required to specify the location and purpose of the installation. In many countries and regions, it is mandatory to post signs indicating the presence of surveillance cameras.
 - Privacy considerations: There is a privacy request for passers-by. Especially in Europe, there are strict regulations.
 - Legal restrictions: There are legal restrictions on installation and operation in many regions. This includes the retention period of the shooting data, access rights, and restrictions on the purpose of use.
- Common considerations
 - Unauthorized photography of others and the publication of video images may involve invasion of privacy and issues of portrait rights.
 - When disclosing the content to a third party or using it for commercial purposes, it is important to obtain the consent of the subject.
 - Whether it is based on legitimate purposes, such as crime prevention or accident records, is also considered.

The use of smartglasses in the city

- Japan
 - There are no special laws yet, but general privacy laws and portrait rights regulations apply.
 - May be sued for filming others without permission or publishing their footage
 - Social considerations are required for use in public places. Some companies and facilities prohibit its use.
- United States
 - There are no clear national regulations. Caution is required in public places.
 - There was a movement to ban the use of Google Glass in public facilities and stores.
 - Different regulations and guidelines vary from state to state and location to location, which may limit their use.
- Europe (EU)
 - Strict regulations may apply.
 - Unauthorized filming or recording can result in severe penalties.
 - In France and Germany, there are cases where use in public places is restricted.
- China
 - There are no clear regulations, but there is strong government surveillance, so it is necessary to take a cautious approach.
 - Certain public and government venues prohibit the use of cameras and recording devices.
 - Captured videos and images may be strictly controlled as a matter of national security

Fraudulent use of Smartglasses in Waseda University Entrance Examination



スマートグラスの今～入試不正利用を機に最新状況を確認する～、塚本昌彦、情報処理学会学会誌「情報処理」note (2024.8.16, in Japanese) <https://note.com/ipsj/n/n645056b837cc>

- Circumstances
 - In February 2024, students took pictures of exam questions using smartglasses and leaked them on social media to ask for answers.
 - On February 16, students taking the entrance examination of Waseda University's Faculty of Creative Science and Engineering took pictures of questions such as chemistry and published the images of the questions on X (formerly Twitter) to an unspecified number of people to obtain answers.
 - Responses were received from several people who had requested them in advance, and the examinees transcribed them during the exam.
 - The incident was discovered because one of the people who sent the answers noticed the fraud and contacted the university, and when taking the entrance examination for the Faculty of Commerce on February 21, the examiner noticed the smartglasses and reported it to the Police.
 - Waseda University invalidated all entrance examination results for this student.
 - The examinee was sent to the Tokyo District Public Prosecutor's Office for fraudulent obstruction of business by the Metropolitan Police Department, and the District Public Prosecutor's Office sent it to the Family Court as a delinquent fact
 - The National Center for University Entrance Examinations clearly states that smartglasses cannot be used in the examination guide for the next year's common test.
- Fraud is similar to privacy
 - In both cases, a user secretly uses smartglasses.
 - Especially he/she uses cameras.

Requirements

- Make it clear to others that you are wearing smart glasses and that you are using a camera or microphone.
- Make it clear to others what you are using for smart glasses and the acquired data, and how you are managing them.
- Make it clear to others that you are not using a camera, microphone, or display.
- I want it to be automatically disabled in the toilet.
- How should I manage my data?

Specific Proposals

1. Smart Glasses Markers
2. Behavior Indicator
3. Wearing armbands and bibs
4. Electronic Notices
5. Camera Slide Cover
6. Automatic Mosaic
7. Automatic Encryption

1. Smart Glasses Markers

- Add red lines to temples and moderns, red marks to bridge and end pieces, etc. It is good to be able to see it from the front or from the side.
- It might be nice to have a mark that lets you know if it has a camera or display.
- There are also disadvantages such as destroying the design, and being known by others although people with disabilities, anti-stalking measures, and forensic applications may not want to be known.
- Whether it is working or not is a separate matter.



2. Operation Indicator

- Recording, red LED indicator and charm sound during recording.
- Especially during recording, it is easy to understand if a beep sound is emitted at regular intervals (for example, every 3 seconds).
- It has the disadvantage of being noisy and disturbing to the people around you.
- The beep sound can be muffled to some extent in the recorded data.
- It may be necessary to clearly indicate what a user is doing to those around the user, such as turning it green when he/she is simply turning it on, and blue when he/she is using the AI function
- What kind of AI? chat? Translation? Recognition? Image generation? It is also conceivable to specify whether or not a user is using it. Or user privacy for that matter?



3. Wearing armbands and bibs

- It is common for business use, but in the case of individuals, it is possible to argue that it is so much.
- There is also a debate about whether or not to write the name and purpose of the photo on the armband or bib.
- We also have to consider the privacy of the filming user.



4. Electronic Notices

- Technically, it is possible to use a specific server, WiFi beacons, ARP, ICMP, etc.
- If a user uses an electronic method, it may be possible to anonymize the user's name.
- Such a platform is necessary for personal information handling companies to announce the name and purpose of shooting as a measure against the Personal Information Protection Law.

5. Camera Slide Cover (Lid)

- The slide cover is supposed to be opened and closed by hand (automatic opening and closing is also possible)
- Even with smartphones, it is very worrisome when the camera of the smartphone of the person sitting opposite you is facing this way, such as on a train.
- Will there be less trouble with taking pictures and not taking pictures?



6. Auto Mosaic

- Add an automatic mosaic to the acquired video.
- For the time being, it is assumed that it can be switched manually.
- When a user wants to record his/her surroundings at a tourist spot, everyone's face and figure are automatically mosaiced.
- It is good to have an indicator so that people around a user can see that the user are shooting in mosaic mode.
- In the slightly more advanced version, it can be assumed that the face and figure of specific people are recorded if they were registered in advance.
- It is also possible to introduce more advanced AI functions, such as preventing mosaics from being added to people who indicate that they are "OK" during the shooting process.

7. Automatic Encryption of Acquired Data

- Acquisition video, audio, biometric data, etc.
- The benefit of preventing unintentional spread
- The disadvantage of greatly impairing convenience
- The question of how to manage and who can decrypt it
- It seems that it is already possible.

Summary

- AI is rapidly expanding, and the deployment of smart glasses has come to a standstill.
- Once the AI business is launched to some extent, smart glasses will be launched after that. AI will be used in this process.
- Smart glasses that are too powerful AI will definitely cause privacy problems. There is a need to deal with it early for a healthy development.
- Our group is identifying problems and considering countermeasures. It is summarized as a guideline.
- Please give us your opinion on the seven specific proposals.



Google Form Survey
<https://docs.google.com/forms/d/e/1FAIpQLSeQJXttza23JO8p8Bj-N8S468aQsQ1bVmQ9MbaKZ5qNRcDEjA/vi-ewform>

Session 5:
Network and Security
(Chair: Takuya Yoshihiro)

Machine Learning for Anomaly Detection Based on Communication Traffic Analysis

Neo Miyama* and Koichi Ishibashi*

*College of Engineering, Kanazawa Institute of Technology, Japan
c1395040@planet.kanazawa-it.ac.jp, k_ishibashi@neptune.kanazawa-it.ac.jp

Abstract - In recent years, with the rapid penetration and acceleration of the Internet, traffic on the Internet has been rapidly increasing, and its utilization styles and applications have also been diversified. Therefore, many studies have traditionally been conducted to detect and control anomaly network traffic. In particular, with the rapid spread of the Internet of Things (IoT), threats of attacks against IoT devices have become more prominent, and Intrusion Detection Systems (IDSs) have been introduced. IDSs are categorized into two types: signature-based and anomaly-based. In signature-based IDSs, known attack traffic patterns are defined as signatures, and intrusion is detected by comparing the observed traffic patterns with the defined signatures. Therefore, although it detects known attack traffic with high accuracy, it is undetectable for anomalous traffic or attacks by unknown malware. On the other hand, anomaly-based IDSs define expected behaviors of normal traffic and detect anomalous traffic based on the deviations from the expected behaviors. Although machine learning based on statistical methods has been proposed to detect deviations from the expected traffic behaviors, there are still some challenging issues, such as the need to define various traffic behaviors. In addition, the emergence of various variants of typical malware makes it difficult to detect malware based on malware-identified behaviors. In this paper, we discuss the effectiveness of a machine learning approach to detect anomalous traffic, including malware variants, by modeling known malware and anomalous traffic as a time-series model and comparing it with observed traffic. Specially, steps for creating time-series data from traffic data on the network and extracting features for waveform data within a specified interval has been examined.

Keywords: Anomaly detection, Communication traffic analysis, Time-series analysis.

1 INTRODUCTION

In recent years, with the rapid penetration and acceleration of the Internet, traffic on the Internet has been rapidly increasing, and its utilization styles and applications have also been diversified. In particular, Internet access by mobile devices has been accounting for a large share of communication traffic due to the proliferation of mobile devices such as smartphones, tablets and mobile PCs[1]. Consequently, it is required to operate networks efficiently, and it has become inevitable to employ traffic measurement techniques to detect and control anomalous traffic that causes network resource wastage and quality degradation.

Traditionally, there has been a lot of research on anomaly traffic measurement and analysis methods for detecting and

controlling anomalous traffic that consumes network resources and degrades quality. For example, [2] proposed a traffic monitoring method for the backbone network served by a network operator. The paper describes the effectiveness of the proposed method that automatically determines thresholds based on predictions of traffic time-series using time-series analysis methods. In addition, from viewpoints of security, there has also been a lot of research into detecting attacks and malware infections through traffic analysis on the network.

Furthermore, with the rapid spread of the Internet of Things (IoT), the threat of attacks against IoT devices has become more prominent, and Intrusion Detection Systems (IDSs) have been introduced in recent years[3]. IDSs are generally classified into two variants: signature based, and anomaly based. Signature-based IDS is the detection of attacks by looking for specific defined patterns, such as byte sequences in network traffic, or known malicious instruction sequences used by malware. Therefore, the signature-based IDSs are not able to detect communications and attacks by unknown malware. Anomaly-based IDS is the detection of attacks by defining the expected behavior and detecting deviations from that behavior, and are primarily introduced to detect unknown attacks, in part due to the rapid development of malware. Deviations are detected by statistical and machine learning methods, but there are still issues to be solved, such as the need to define the behavior of various types of traffic. Also, it is difficult to detect malware based on the same characteristics due to the fact that various variants of typical malware are emerging frequently.

Meanwhile, interest in machine learning is growing, and its application in diverse fields is being considered. Particularly, there is a strong interest in machine learning with learning by adjusting thresholds and parameters.

Considering the above, we aim to detect the existence of anomalous traffic, including variants, by modeling the traffic of known malware and attacks and comparing it with monitored traffic data in a time-series. Specifically, we discuss an approach to evaluate the similarity between the time-series model of known traffic and the time-series model of monitored traffic, and to calculate the possibility that the monitored traffic contains anomalous traffic. In this study, toward the feasibility of the objective approach, steps for creating time-series data from traffic data on the network and extracting features for waveform data within a specified interval has been examined.

2 RELATED WORKS

It is essential to monitor and manage communication networks through the process of anomaly detection of

communication traffic, analysis of its causes, and control of communication flows. In anomaly detection, anomalies are detected based on threshold and change values based on packet statistics on the network, but mechanisms and methods for more efficient and accurate detection have been proposed in various areas. For example, [2][3] create time-series data for IP flows monitored at each router in the network, such as the number of packets generated per unit time, the traffic volume, and the number of IP flows, as traffic volume, and detect anomalies based on the amount of change over time.

In addition, Internet access from small mobile devices such as smartphones and tablets has become more common in recent years, and the injection of anomalous traffic caused by infected mobile devices connecting to the network has also become a problem. [4][5] have studied the detection of malware-infected mobile devices based on the traffic analysis of communication traffic flowing over the network.

On the other hand, recently, research on detecting anomalous traffic by analyzing time-series data has been discussed, in which the amount of data and the number of packets in communication traffic are treated as time-series data. As a background, there has been substantial progress in research and application of machine learning. That is, a large amount of time-series data, which is a sequence of actual values that change over time, has been accumulated in various fields, such as voice data, stock prices, and natural observation data such as temperatures, in order of observation time. In order to effectively utilize the accumulated time-series data, research on classification and clustering of time-series data is being conducted in various areas[6]. For example, in similarity search between time-series data, there are discussions of defining dissimilarity based on the Euclidean distance between two time-series data or Dynamic Time Warping (DTW) distance. There is also a proposal for dissimilarity based on the magnitude of the frequency component of time-series data[7]. In [7], a proposal for dissimilarity focusing on the frequency component using the Discrete Fourier Transform (DFT) is proposed for time-series data that changes rapidly in a short time. However, the studies do not address the comparison of different time series data. In addition, analysis of communication traffic using the Wavelet transform has been proposed in [8][9], among others. [8] analyzes network traffic in a digital manufacture system using the Wavelet transform and attempts to detect anomalies caused by variations in the detail coefficients in the Wavelet transform. However, it is limited to a single network traffic, and no analysis of the causes of the anomaly is presented. [9] considers network traffic data as time-series data and presents a traffic classification based on characteristics such as the number of packets and the amount of packet within a specified time period. However, analysis based on frequency components is not discussed.

In contrast to the above, this study discusses a mechanism for anomaly detection by analyzing communication traffic on the network at each time window, by using the anomalous traffic data to be detected as reference data.

3 OBJECTIVE APPROACH

We aim to detect the existence of anomalous traffic, including variants, by modeling the traffic of known malware and attacks and comparing them with monitored traffic data in a time-series. Specifically, we study an objective approach to detect a communication traffic with anomalous traffic. The objective approach consists of estimation of the similarity between the time-series model of known traffic and the time-series model of monitored traffic, and calculation of the possibility that the monitored traffic contains anomalous traffic.

The objective approach is analyzed and evaluated according to the following procedures.

- 1) Monitoring communication traffic on a target network and/or a network interface by network tools such as Wireshark.
- 2) Creating time-series data, where the amount of data is measured in Monitoring-time-window units from the monitored data.
- 3) Extracting features of waveform data in each Evaluation-time-window from the created time-series data.
- 4) Comparing the features in the monitored communication traffic with the features in the Evaluation-time-window extracted from the evaluation-time-window in communication traffic caused by known malware and attacks.

Following the above steps, the objective approach detects the presence or absence of anomalous traffic in cases where network traffic is similar to those of communication traffic caused by malware or attacks.

Figure 1 shows an outline of the objective approach.

4 PRELIMINARY STUDIES

In the objective approach, feature extraction from timeseries data in the Evaluation-time-window is big challenge. For feature extraction of time series data, methods such as Wavelet Transform and Fourier Transform are generally used. In addition, recently, classification of time-series data using machine learning is researched.

In this section, the application of Wavelet and Fourier transforms to extract features from time-series data for each Evaluation-time-window are discussed.

4.1 Datasets

There are various datasets available for intrusion detection and other studies. For example, KDD99[10], ISCXIDS2012[11], UNSW-NB15[12], CICIDS2017[13] and CICIDS2018[13] are publicly available. That is, in order to evaluate the effectiveness of network intrusion detection systems (NIDSs) based on their performance to identify attacks, it is required a comprehensive dataset that contains normal and abnormal behaviors. Although older benchmark datasets such as KDD99 have been widely adopted for evaluation of NIDS performance, improved datasets are emerging due to the existence of redundant records in older datasets, divergence from the real world, and response to recent attack environments.

In this study, UNSW-NB15 datasets and CICIDS2018 datasets are utilized. UNSW-NB15 is evaluation dataset for

intrusion detection system created by Cyber Security Research Group at the University of New South Wales. UNSW-NB15 produces sample of 9 types network traffic including cyber-attack and normal traffic. It designed by diversity and realism of actual network traffic. In the dataset, there are over 1 million records and 49 feature values. It also includes attacks such as DoS, fuzzes, analysis, backdoors, Exploits, Generic, reconnaissance, shellcode and worms.

4.2 Creation of time-series data

Firstly, we extract the necessary values from the CSV file provided by UNSW-NB15 datasets to create the time-series data, which is in the format shown in Table 1. The transmitted and received bytes are extracted from the 8th and 9th columns, and the packet arrival times from the source and destination nodes are obtained from the 31st and 32nd columns. And time-series data is obtained by accumulating the number of bytes transmitted and received every 100

milliseconds, which is specified as the Monitoring-time-window, for the packet arrival time. The obtained time-series data is two-dimensional waveform data, and is used in the extraction of features for each evaluation time-window. Figure 2 shows an example of waveform data obtained from the UNSW-NB15 datasets.

Table 1: Format of CSV file in UNSW-NB15

#	Name	Description
8	sbytes	Source to destination bytes
9	dbytes	Destination to source bytes
14	service	http, ftp, ssh, dns, ... else (-)
15	sload	Source bits per second
16	dload	Destination bits per second
17	spkts	Source to destination packet count
18	dpkts	Destination to source packet count
31	sintpkt	Source inter-packet arrival time (msec)
32	dintpkt	Destination inter-packet arrival time (msec)

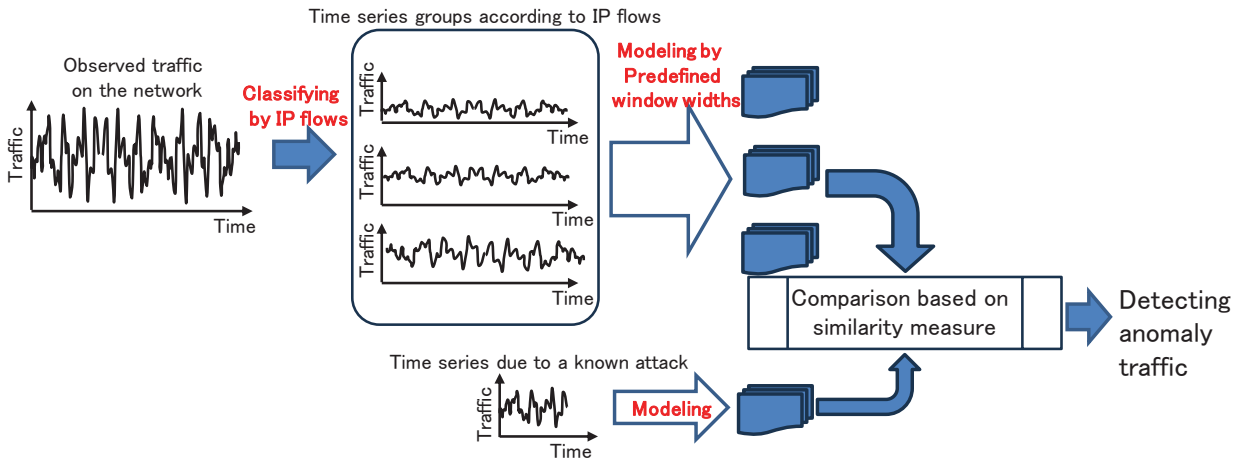


Figure 1: Overview of the objective approach.

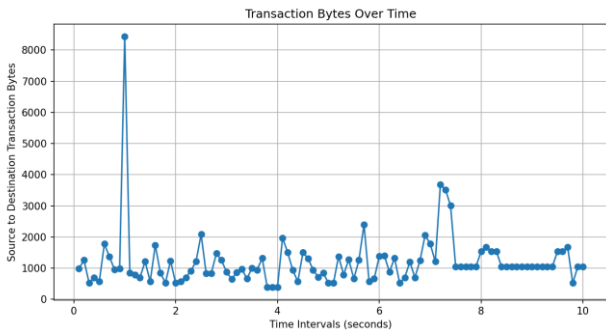


Figure 2: Example of time-series data.

4.3 Wavelet Transform

One of the features of the Wavelet transform is ability to provide detailed insight into the changes in frequency components at specific times. Additionally, it is suitable for capturing both the global and local features of a signal. By utilizing these features, we are able to analyze traffic patterns and perform feature extraction and classification.

Figure 3 shows the coefficients of the Wavelet transform applied to a specific interval in the time-series data shown in Figure 2. Figure 3 is filtered, followed by down sampling and decomposition of the signal. By further applying filters to the down sampled signal and repeating the process, additional data obtained possibly. The two graphs in Figure 3 plot the data from Figure 2 for the intervals 0 to 1.5 seconds and 5 to 6.5 seconds.

Figure 4 shows the results for an example of time-series data sampled every 10msec in UNSW-NB15 datasets. Its (b) plots the detail coefficients for the first three windows when the discrete Wavelet transform is performed every 256 points, and Its (c) plots the Euclidean distance between the detail coefficients for the adjacent windows.

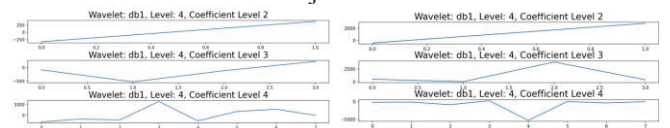


Figure 3: The waveform data graph subjected to Wavelet transform.

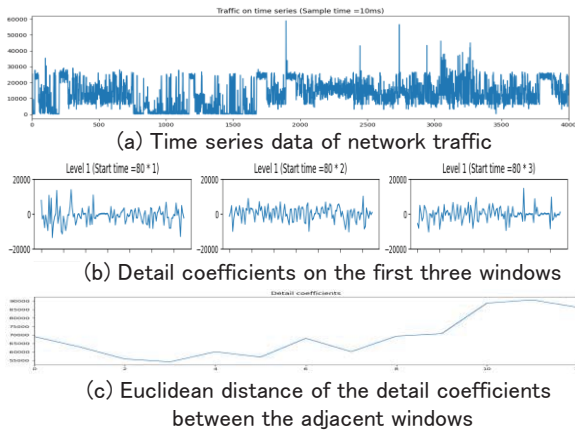


Figure 4: The waveform data graph subjected to Wavelet transform (Part 2).

4.4 Fourier Transform

By using the Fourier transform, it is possible to represent a signal expressed in the time domain in the frequency domain, thereby revealing the frequency components of the signal and their magnitudes. Therefore, these spectrums make it easier to interpret changes from the time-series data over a long period, allowing for more accurate information retrieval during comparison and analysis.

Figure 5 shows the result of applying the Fourier transform. The discrete signal in the time domains is transformed into a discrete signal in the frequency domain. Subsequently, the amplitudes of the angular frequency components are calculated, then plotted. The two graphs in Figure 5 plot the data from Figure 2 for the intervals 0 to 1.5 seconds and 5 to 6.5 seconds.

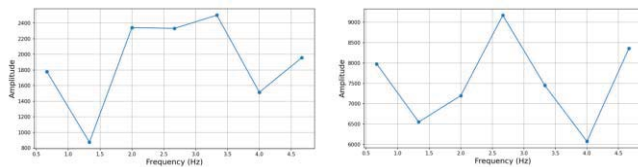


Figure 5: The waveform data graph subjected to Fourier transform.

5 CONCLUSION

In this paper, we study the extraction of features within a specific interval of time series data for the purpose of detection of the existence of anomalous traffic, including variants. As a first step, the results of applying the Wavelet transform and Fourier transform to feature extraction are examined.

In the future, we will investigate algorithms for detecting anomalies by modeling the traffic of known malware and attacks and comparing it with monitored traffic data in a time-series, in order to realize the objective approach.

REFERENCES

[1] Ericson Mobility Report, June 2015, [https://www.ericsson.com/4ac604/assets/local/reports-](https://www.ericsson.com/4ac604/assets/local/reports-papers/mobility-report/documents/2015/ericsson-mobility-report-june-2015.pdf)

[papers/mobility-report/documents/2015/ericsson-mobility-report-june-2015.pdf](https://www.ericsson.com/4ac604/assets/local/reports-papers/mobility-report/documents/2015/ericsson-mobility-report-june-2015.pdf), Accessed May 22, 2024.

[2] S. Harada, R. Kawahara, T. Mori, N. Kamiyama, Y. Hirokawa, and K. Yamamoto, "A method of detecting network anomalies and determining their termination," IEICE Technical Report, Vol. 106, No. 420, IN2006-133, December 2006.

[3] T. Ohtani, R. Yamamoto, and S. Ohzahata, "Active Intrusion Detection Method for IoT Devices," IEICE Technical Report, Vol. 121, No. 421, CQ2021-101, CQ2021-101, March 2022.

[4] Mehede Zaman, Tazrian Siddiqui, Mohammad Rakib Amin, and Md. Shohrab Hossain, "Malware detection in Android by network traffic analysis," International Conference on Networking Systems and Security (NSysS), January 2015.

[5] A. H. Lashkari, A. Fitriah, A.Kadir, H. Gonzalez, K. F. Mbah, and A. A. Ghorbani, "Towards a Network-Based Framework for Android Malware Detection and Characterization," 15th Annual Conference on Privacy, Security and Trust (PST), August 2017.

[6] T. Yoshikawa, M. Ishikawa, H. Chen, K. Furuse, and N. Ohbo, "A Study on Similarity Search for Very Long Time-Series Data," Data Engineering Workshop (DEWS) 2007, February 2007.

[7] K. Koyama, T. Hochin, H. Nakanishi, and M. Kojima, "Similarity Based on Spectrum of Time Series," IPSJ SIG technical reports, November 2006.

[8] D. Lavrova, P. Semyanov, A. Shtyrkina, and P. Zegzhda, "Wavelet-analysis of network traffic time-series for detection of attacks on digital production infrastructure," International Scientific Conference "The Convergence of Digital and Physical Worlds: Technological, Economic and Social Challenges" (CC-TESS2018), Volume 44, June 2018.

[9] J. Koumar, K. Hynek, and T. Čejka, "Network Traffic Classification based on Single Flow Time Series Analysis," The 19th International Conference on Network and Service Management (CNSM), October 2023.

[10] T. Mahbod, B. Ebrahim, L. Wei, and G. Ali-A, "A Detailed Analysis of the KDD CUP 99 Data Set," 2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications, July 2009.

[11] Ali Shiravi, Hadi Shiravi, Mahbod Tavallaee, and Ali A. Ghorbani, "Toward developing a systematic approach to generate benchmark datasets for intrusion detection," Computers & Security, Volume 31, Issue 3, May 2012, Pages 357-374.

[12] N. Moustafa and J. Slay, "UNSW-NB15: A Comprehensive Data Set for Network Intrusion Detection Systems (UNSW-NB15 network data set)," 2015 Military Communications and Information Systems Conference (MilCIS), 2015, pp. 1-6.

[13] I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, "Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization", 4th International Conference on Information Systems Security and Privacy (ICISSP), Portugal, January 2018.

Study of application to IoT DEP based on quantitative indicators using AHP

Masaki Mitsuuchi^{*}, Tetsuya Yokotani^{**}, Koichi Ishibashi^{**}, Midori Sugaya^{***}, and Li Yanzhi^{***}

^{*}Graduate School of Engineering, Kanazawa Institute of Technology, Japan

^{**}College of Engineering, Kanazawa Institute of Technology, Japan

^{***}College of Engineering, Shibaura Institute of Technology, Japan

{c6301158@st, yokotani@neptune, k_ishibashi@neptune}.kanazawa-it.ac.jp,
{doly, nb22509}@shibaura-it.ac.jp

Abstract - With the proliferation of the Internet of Things (IoT), many IoT devices are becoming connected. In order to efficiently support such devices, lightweight platforms are required. One such platform is the IoT Data Exchange Platform (IoT DEP) standardized by ISO/IEC JTC1/SC41. Information is exchanged by connecting multiple brokers in a ring. However, this method has the problem of increasing the number of hops between brokers, which causes delays. In this paper, we propose a method to connect brokers directly and aim to reduce the delay. The communication protocol used for the connection is determined by the Analytic Hierarchy Process (AHP).

Keywords: IoT DEP, MQTT, Distributed Brokers, DDS

1 INTRODUCTION

IoT services are currently being implemented in various sectors [1] and the number of IoT devices is expected to increase to 32.4 billion by 2022 and 44 billion by 2025 according to forecasts [2]. In light of these, implementing IoT services on the current Internet is inefficient due to the increase in control information and consumption of network resources [3]. To accommodate these efficiently, a platform that reduces processing weight is considered necessary. One such platform is the IoT Data Exchange Platform (IoT DEP), which is being standardized at ISO/IEC JTC1/SC41 under the initiative of the University IoT DEP refers to a platform that actively utilizes Pub/Sub type communication. MQTT [4] allows Publishers and Subscriber communicate via a single broker, requiring a distributed broker to accommodate a large number of Publishers and Subscribers across multiple locations [5]. In this paper, the Analytic Hierarchy Process (AHP) is used to determine the communication protocol used to connect the brokers. Partial prototyping is then carried out to estimate latency.

2 OVERVIEW OF MQTT

In this section, we provide an overview of MQTT, the communication protocol we have been dealing with in our research, and the distributed brokers of MQTT.

2.1 MQTT Architecture

The Message Queuing Telemetry Transport (MQTT) protocol [6], developed by IBM and now an OASIS standard

[7], is a lightweight message-oriented protocol [8] and topic-based Pub/Sub type communication. MQTT can operate over TCP/IP and requires a Broker to exchange data. In practice, the Broker is a server to which clients can publish/subscribe to topics, and through which message traffic is routed. Clients can also be authorized by the Broker by accessing it with a username and password. Message security is encrypted by SSL/TLS [9].

2.2 Feature of QoS for Assured Transmission

Three guaranteed delivery functions are implemented in MQTT. Figure 1 shows the communication sequence for each QoS function.

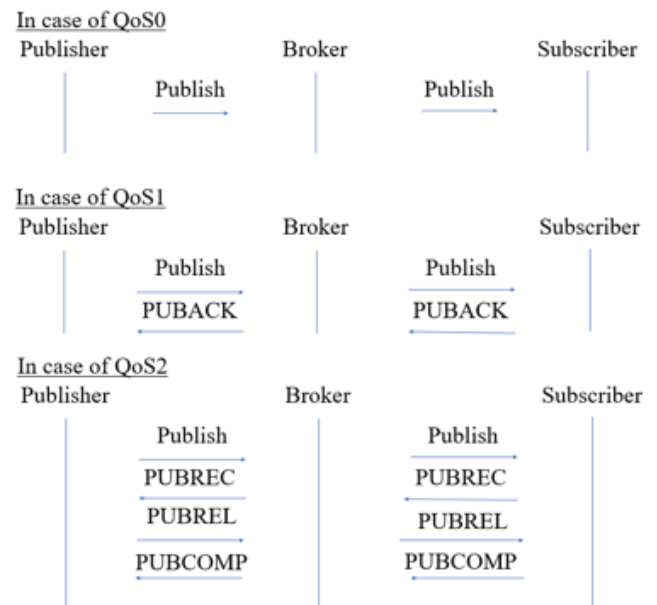


Figure 1: Communication sequence of MQTT QoS function

QoS0 has no delivery guarantee and only sends a Publish message once. QoS1 provides more reliable communication than QoS0 by sending a delivery confirmation message called PUBACK to the Publish message and guaranteeing delivery once. QoS2 provides more reliable communication than QoS1 and is the most reliable in MQTT. Compared to QoS0 and QoS1, QoS2 is used when the number of communications is higher than that of QoS0 and QoS1, but when data is to be sent reliably. MQTT is used when the number of communications is higher than QoS0 or QoS1. In this way, MQTT provides delivery assurance with three QoS functions.

2.3 Operation Example for IoT Service

In previous work, collaboration has been achieved by connecting distributed brokers in a ring [10] [11]. In this method, MQTT brokers are connected to a virtual ring built using VLANs. Each broker has an access control block and a shared block. The Access Control block handles information transfer and loop prevention; the Shared Memory block is a table that manages topics and their information.

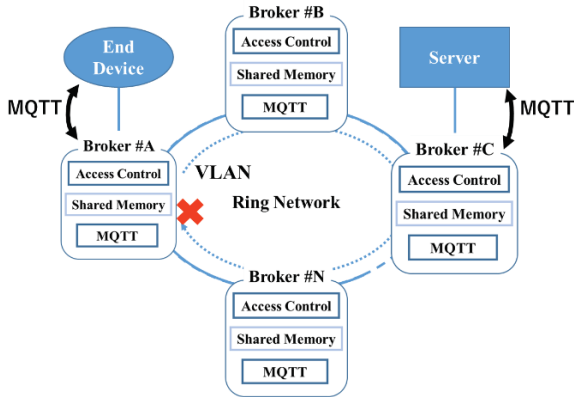


Figure 2: Distributed broker architecture with virtual ring

Distributed Broker coordination is performed in this way, but the processing delay in this method is 44.39 [ms] to 44.51 [ms] [10]. The delay time increases as the number of hops increases with the number of Brokers to be linked, and we cannot support use cases where low latency is required for factory buildings. Therefore, in this paper, we propose a method to achieve low-latency communication by directly connecting distributed Brokers.

3 STANDARDIZATION TREND OF IOT COMMUNICATION PLATFORM

In this section, we discuss trends in standardization, provide an overview of the IoT DEP, and mention the relationship between the promotion of this standardization, which is being spearheaded by the University, and the verification of principle that is being pursued in parallel.

3.1 Summary of IoT DEP

The following is an overview of the IoT data exchange platform (IoT DEP) being promoted by the author and others. The IoT DEP has been publicized in many articles, see [12] - [14], etc.

The IoT, in which all objects are connected to the Internet, is a technology that is currently attracting much attention. However, as IoT devices become more widespread, the current communication standards of the Internet will become inconvenient. In addition, an analysis of IoT use cases has been carried out, resulting in the need for a lightweight platform in terms of communication processing and abstracts the physical network. Based on the results of this study, the IoT DEP is the specification of a new platform that can be commonly applied to applications.

IoT DEP is a platform architecture that actively utilizes Pub/Sub type communication. This IoT-specific

communication method enables more efficient data transfer than conventional communication methods using IP addresses and DNS servers. To realize lightweight communication, Information Centric Network (ICN) technology, which is not bound to an IP address, is applied from end devices and servers. At the same time, dedicated paths connect multiple Nodal Points in the IoT DEP network. The IoT DEP functionality is also used in the IoT DEP network. In addition, IoT DEP functions are implemented in Nodal Points. This enables the system to operate on the existing internet and to be consistent with existing protocols such as Hypertext Transfer Protocol (HTTP) and File Transfer Protocol (FTP), as well as to transfer a massive amount of data.

The advantages of IoT DEP can be briefly summarized as no access to a DNS server is required, and ICN technology results in less header information and smaller communication data volumes. As a result, low latency can be achieved. It is also possible to operate on existing Internets.

3.2 Standard Structure

IoT DEP is being standardized by ISO/IEC JTC/SC41 as ISO/IEC30161 series. Table 1 summarizes the promotion of this standardization and Figure 3 the relationship between it and the verification of principle that is being pursued in parallel.

Table 1: Overview of ISO/IEC 30161

Part.	Contents
Part 1	Requirement. Architecture. Functional block definition.
Part2.	Functional definition of Nodal Point. Coordinated control between Nodal Points.

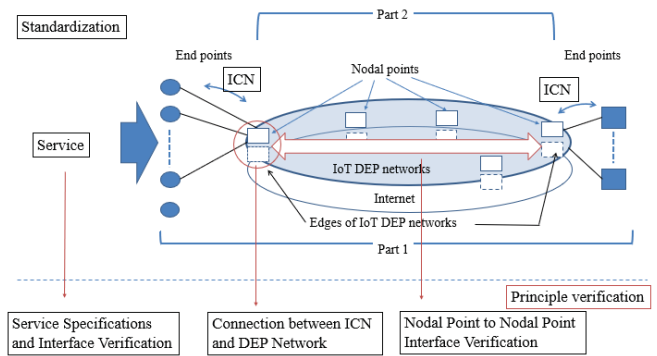


Figure 3: International standardization and proof of principle of IoT DEP

From Table 1, these standards do not go into the various control methods in detail. However, verification of the principles within the framework specified in these is currently being carried out.

The three main points of the proof-of-principle are as follows. The verification of principle prioritizes the application of MQTT in the ICN part of Figure 3.

(i) **Service specification and interface verification:** information from various sensors etc. is analyzed and transferred to the payload of the MQTT Publish message. The format of information from sensors, collection cycles, etc. are parameters.

(ii) **Connections between ICNs and DEP networks:** connections between MQTT and the edge of DEP networks (implementation of gateway functions).

(iii) **Interface verification between Nodal Points:** between Nodal Points is the connection between IoT DEP middleware; IoT DEP applies methods simpler than IP-based routing on the Internet. The methods are summarized in [15] and can be categorized as shown in Table 2.

Table 2: Classification of Nodal Point-to-Nodal Point Control

		Dynamic	Static
(iv)		Approach1.	Approach4.
(v)	Direct	Approach2.	Approach5.
	Indirect	Approach3.	Approach6.

In Table 2, Request driven is a method that transfers the request to obtain information (equivalent to Subscription in the case of MQTT), and Response driven is a method that transfers the information itself (equivalent to Publication in the case of MQTT). In the case of response driven, the information itself is transferred directly into the network as Direct, while only the ID of the information is transferred and the information itself is offloaded by a network other than the IoT DEP (e.g., existing Internet) to reduce the traffic volume as Indirect. In Table 2, Approach 4 and 5 are shown in [10], Approach 6 in [16], and Approach 2 in [17].

3.3 Future Direction

As mentioned above, the IoT DEP to date has achieved higher speed and lower latency than the existing Internet by applying MQTT, a typical Pub/Sub type of communication between ICNs. However, there are still challenges in cooperative control between Nodal Points. For example, the connection between Nodal Points in the IoT DEP in [10] is realized by connecting distributed brokers of MQTT on a virtual ring. However, this method results in long latency on the virtual ring, so the problem between Nodal Points needs to be solved in order to realize IoT DEP. In addition, it is necessary to abstract the device using an open platform for systems that standardize software implementation specifications for robots and control systems, such as Open EL and ROS 2, for practical use. Therefore, this paper proposes a direct connection scheme between distributed Brokers to reduce the delay time. In the next section, we determine the communication protocol between Brokers using AHP.

4 PROTOCOL SELECTION BY AHP

The selection of the communication protocol to be used for a direct connection between Brokers. The Analytic Hierarchy Process (AHP) [18] is used for the selection. The seven alternative protocols are MQTT, DDS, AMQP, XMPP, CoAP, HTTP, and WebSocket.

AHP is the most well-known of the multi-criteria decision-making (MCDM) methods [19] AHP is a quantitative method for deriving the optimal choice from multiple alternatives and evaluation criteria AHP models the decision-making process hierarchically as a 'Goal-Criteria-Alternative' AHP models

the decision-making process hierarchically as a 'Goal-Criteria-Alternative'. A pairwise comparison is then made between two items within a criterion to determine which is more important, and the criteria are weighted statistically. Next, a pairwise comparison is made between the two alternatives for each criterion to see which is better and to produce a rating value for the alternative in a statistical way. The overall evaluation is based on the calculated weights of the criteria and the evaluation values of the alternatives.

AHP is used to select protocols. The objective is to 'select a communication protocol for direct Broker-to-Broker connections.' The nine evaluation criteria were scalability, compatibility with existing transports, throughput, latency, IP dependency, reliability, lightweight, routing, and OAM capabilities. The seven alternative protocols used were MQTT, DDS, AMQP, XMPP, CoAP, HTTP, and WebSocket. The questionnaire was administered to six people.

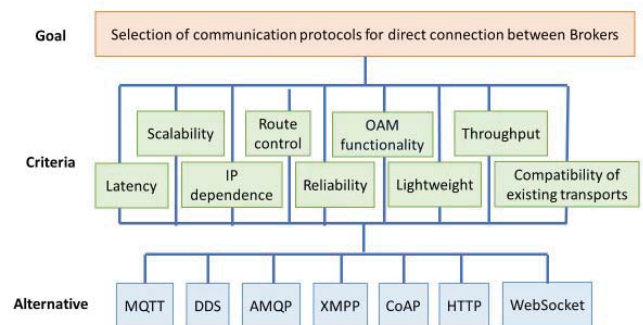


Figure 4: Hierarchical model of communication protocol selection

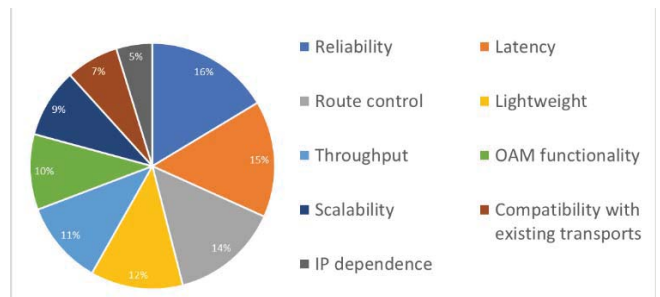


Figure 5: Level of importance between each criterion

To begin, the importance of each criterion was calculated; Figure 5 shows that reliability is the most critical item in the evaluation criteria, while IP dependence is the least important.

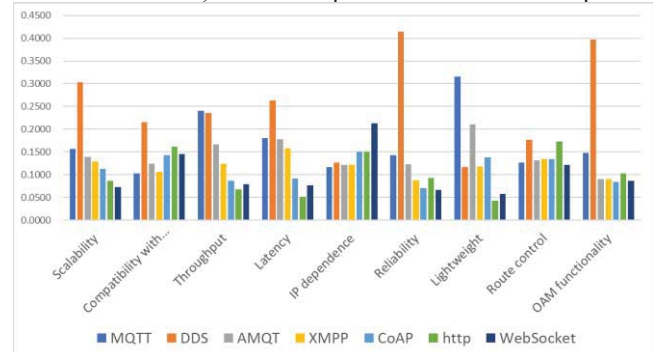


Figure 6: Calculated values based on assessment criteria

Figure 6 shows the results of calculating how good each protocol is in terms of the evaluation criteria. Based on the results of Figure 5 and Figure 6, the best protocols for Nodal Point-to-Nodal Point connections are shown in Figure 7.

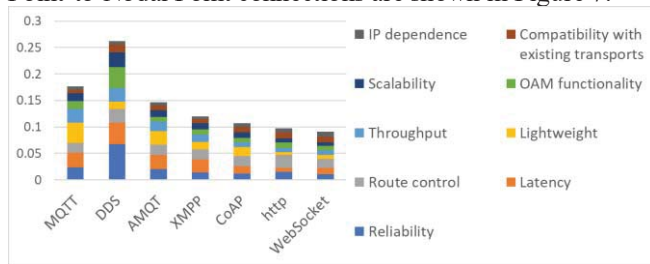


Figure 7: Evaluation results in the AHP

The evaluation results in Figure 7 show that the most suitable protocol is DDS. Therefore, DDS is used as the communication protocol used for direct connection between Brokers.

5 PROPOSED CONFIGURATION FOR IOT DEP

Based on the AHP calculation results in Section 4, DDS is applied between Nodal Points, and MQTT is applied to the End Point connections, prioritizing dissemination and lightweighting. Figure 8 shows the proposed application to the IoT DEP.

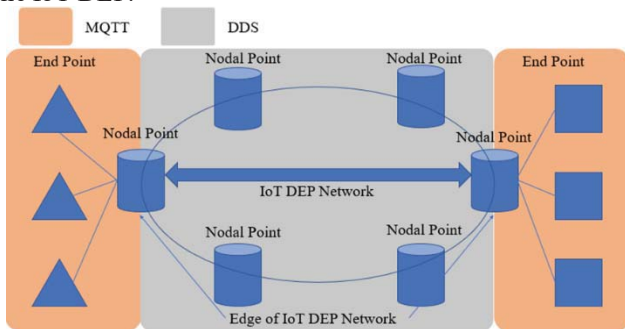


Figure 8: Proposal for IoT DEP

6 CONCLUSION

In this paper, past work on distributed Broker federation methods in MQTT is presented, and the challenges are mentioned. In the future, the author will prototype other methods and evaluate their performance. In addition, efforts will be made towards practical applications through the construction of connection methods and environments for robots and control systems such as Open EL and ROS 2.

Acknowledgments

This research was supported by Japan Science and Technology Agency (JST), CREST, JPMJCR19K1.

REFERENCES

[1] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, "Internet of Things: A Survey on Enabling Technologies, Protocols, and Applications.," in IEEE Communications Surveys &

Tutorials, vol. 17, no. 4, pp. 2347-2376, Fourthquarter 2015.

[2] Ministry of Internal Affairs and Communications, Information and Communication White Paper 2023 Edition, 2023

[3] T. Yokotani, and Y. Sasaki, "Comparison with HTTP and MQTT on required network resources for IoT," 2016 International Conference on Control, Electronics, Renewable Energy and Communications (ICCEREC), Bandung, Indonesia, 2016, pp. 1-6.

[4] MQTT.org. <http://mqtt.org/>

[5] M. B. Yassein, M. Q. Shatnawi, S. Aljwarneh, and R. Al-Hatmi, "Internet of Things: survey and open issues of the MQTT protocol," 2017 International Conference on Engineering & MIS (ICEMIS), Monastir, Tunisia, 2017, pp. 1-6

[6] Sebastian Raff. the MQTT Community. Available online: <https://github.com/mqtt/mqtt.github.io/wiki> (accessed on 10 May 2024).

[7] D. Silva, L. I. Carvalho, J. Soares, and R. C. Sofia, A Performance Analysis of Internet of Things Networking Protocols: Evaluating MQTT, CoAP, OPC UA, Appl. Sci. 2021, 11(11), 4879

[8] B. H. Çorak, F. Y. Okay, M. Güzel, Ş. Murt, and S. Ozdemir, "Comparative Analysis of IoT Communication Protocols," 2018 International Symposium on Networks, Computers and Communications (ISNCC), Rome, Italy, 2018, pp. 1-6

[9] S. Lee, H. Kim, D. K. Hong, and H. Ju, "Correlation analysis of MQTT loss and delay according to QoS level," 2013 IEEE International Conference on Information Networking (ICOIN), pp. 714-717, Jan. 2013.

[10] S. Ohno, K. Terada, T. Yokotani, and K. Ishibashi, "Distributed MQTT broker architecture using ring topology and its prototype. ", IEICE Communications Express, 2021 Volume 10 Issue 8 Pages 582-586

[11] T. Yokotani, S. Ohno, H. Mukai, and K. Ishibashi, "IoT platform with distributed brokers on MQTT," International Journal of Future Computer and Communication, vol. 10, no. 1, pp. 7-12, 2021.

[12] T. Yokotani, and K. Kawai, "Concepts and requirements of IoT networks using IoT Data Exchange Platform towards International standards. ", IEEE Conference on Standards for Communications and Networking (IEEE CSCN), #1570570960, 2019, DOI: 10.1109/CSCN.2019.8931337. IEEE Xplore.

[13] T. Yokotani, "A proposal of IoT DEP as a Platform for Efficient IoT Service Deployment on Wide Area Networks", IEEE 2019 International Japan - Africa conference on Electronics, Communications and Computers (IEEE JAC-ECC 2019), IEICE Invited session -1 (2019)

[14] T. Yokotani, and K. Kawai, "Survey on standardization activities of the IoT and proposal of an IoT data exchange platform", IEICE 2020 International Conference on Emerging Technologies for Communications (ICETC 2020), IB3-3 (2020)

[15] T. Yokotani, S. Yamamoto, S. Ohno, K. Sasabayashi, and K. Ishibashi, "Survey and comparison of

- interworking point routing mechanisms for IoT services in wide area ICNs", IEICE 2020 International Conference on Emerging Technologies for Communications (ICETC 2020)
- [16] S. Ohno, K. Terada, T. Yokotani, and K. Ishibashi, "The control mechanism of distributed MQTT brokers for large volume data transfer and its prototype system", IEICE Communication Express, Volume 11 Issue 3, pp. 160-164 (2022).
- [17] Y. Noda, S. Ohno, K. Ishibashi, and T. Yokotani, "A new routing mechanism based on layer 2 control in MQTT networks with multiple brokers. ", IEICE Communication Express, Volume 11 Issue 6, pp. 307-312 (2022).
- [18] T. L. Saaty, "What is the analytic hierarchy process?", Mathematical models for decision support, 1988, pp. 109-121.
- [19] N. F. Aziz, S. Sorooshian, and F. Mahmud, "MCDM-AHP METHOD IN DECISION MAKINGS", ARPN Journal of Engineering and Applied Sciences, VOL. 11, NO. 11, pp. 7217-7220, JUNE 2016

Estimating Communication Quality in Wireless Sensor Network Using Time-Series Traffic Trend Similarity

Yui Kobayashi[†], Ryoza Kiyohara[‡], and Yoshiaki Terashima[†]

[†]Soka University, Japan

[‡]Kanagawa Institute of Technology, Japan

e23m5323@soka-u.jp, kiyohara@ic.kanagawa-it.ac.jp, tyoshi@soka.ac.jp

Abstract -

Wireless sensor networks (WSNs) are utilized in disparate fields including agriculture. However, communication quality is often difficult to maintain owing to their complicated topologies. This study presents an approach for monitoring WSNs based on an analysis of the traffic the sensors sent and received. Since only traffic data is used, the same monitoring system is protocol-independent. The similarity of traffic trends is exploited for communication route estimation. However, burst traffic often occurs during wireless communication which may result in errors. Thus, the proposed technique utilizes a new similarity measure called the Similar Trend Time Ratio (STTR) for communication quality evaluation that is versatile, maintains accuracy, and is network agnostic. For the optimal threshold selection, a random forest machine learning algorithm was utilized. In the experiment, the traffic data were artificially generated via network simulation using QualNet and it was demonstrated that 50% or higher accuracy was achieved in estimating linear communication routes when an appropriate threshold was selected.

Keywords: Network Monitoring, Traffic Analysis, Time Series Data, Similarity Measure

1 Background

It is expected that wireless sensor networks (WSNs) will be introduced to the field of agriculture to acquire temperature and humidity data. However, these networks often have complicated topologies, which make it difficult to maintain network communication and quality. In particular, WSN systems for application to agriculture have a variety of protocols or specifications depending on the required transmission capacity or frequency of data monitoring. Generally, network-monitoring software analyzes the headers of each transmitted packet to provide information on the network condition. However, development of software for each specific network is inefficient and ineffective.

2 Solution

2.1 Proposed Method

In the proposed method, networks are monitored by estimating the network topology using only traffic data. This method can be used in any network environment regardless of the utilized protocols because the actual content of packets is not analyzed.

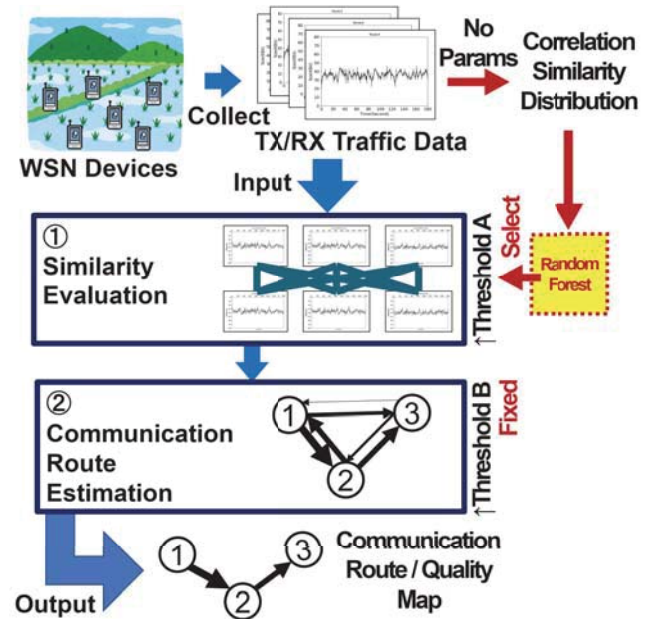


Figure 1: Process of Proposed Network Monitoring Method

The proposed method analyzes traffic data to estimate the communication route and quality of each link, generating communication route and quality map. This generated map is useful for users to understand network conditions and take countermeasures in the area where communication qualities are low.

Traffic data to be analyzed is time-series data collected from each sensor, which represents the data sent and received by each device during each time interval. For example, if a sensor device has a recorded packet log, such as **Table 1a**, time-series data are generated as shown in **Table 1b**.

This section explains how each process shown in **Figure 1** contributes to the proposed method.

2.2 Similarity Evaluation

To estimate a communication route, the similarity between the traffic data of all possible pairs of devices is examined, which represents the possibility that the two devices communicate directly. **Figure 2** shows an example of a pair of traffic data where the sender device sends packets to the receiver device. This traffic data were artificially generated via network simulation using QualNet[1]. The two graphs have similar traffic trends, as indicated by their similar shape. How-

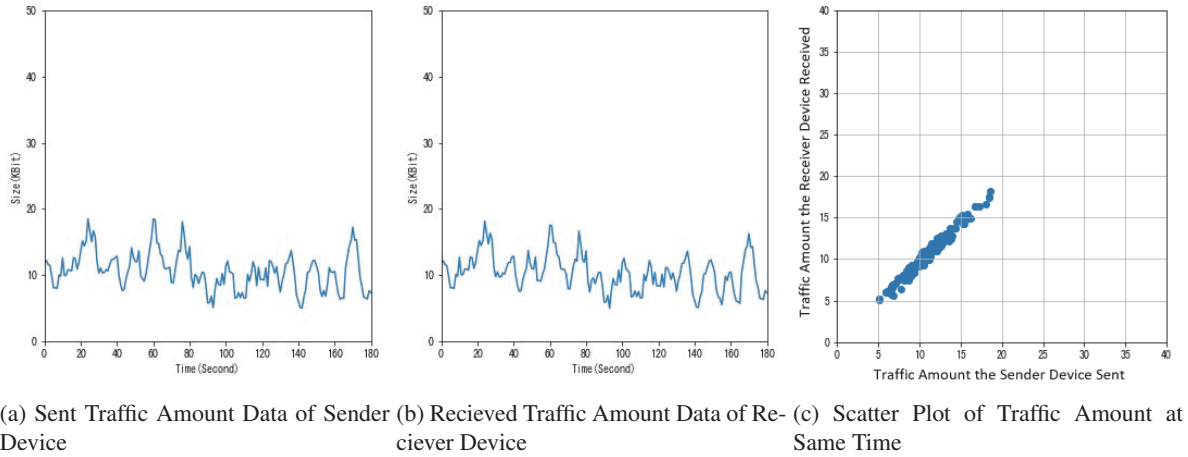


Figure 2: Traffic Amount Data Comparison

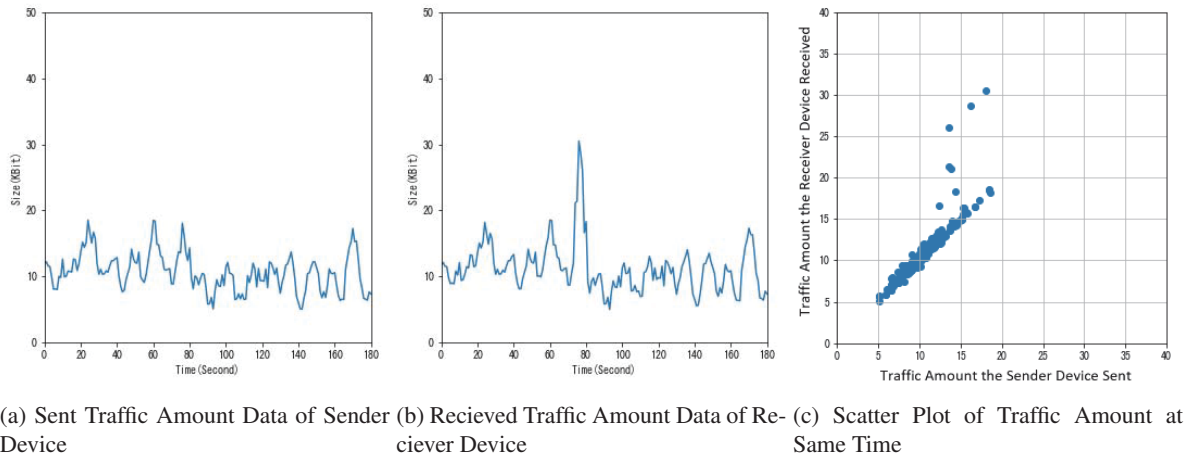


Figure 3: Traffic Amount Data Comparison in Case Burst Traffic Occur

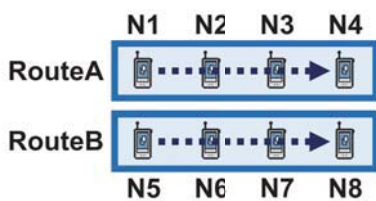


Figure 4: Example of Linear Topology

ever, randomly selected pairs of traffic data are not similar, as shown in **Figure 3**. **Figure 5** shows the correlation between all the traffic data of devices that communicate based on a linear topology, which is depicted in **Figure 4**. For example, the value in the cell in row "N2", column "N1", indicates similarity between the traffic data the sensor node named "N1" transmitted and "N2" received. Devices in the same communication route exhibited similar traffic trends, as shown by the boxes in **Figure 5**. These similarities are useful for network quality estimation.

2.2.1 Inaccuracy Caused by Burst Traffic

Burst traffic or packet drop often occurs during wireless communication, which hinders accurate network estimation in the proposed similarity-based method. The effects of these events can be observed in **Figure 3**, which examines temporal changes over the entire monitored interval. Significant changes in the amount of traffic are evident.

The calculated similarity scores based on basic similarity measures for time-series data such as Euclidean distance (ED) or Dynamic Time Warping (DTW), can be adversely influenced even though a high degree of similarity is observed during most of the monitored time.

A general countermeasure to burst problems is smoothing using a moving average or rolling. However, smoothing eliminates traffic trend characteristics, which is not recommended for the proposed method since trends are utilized to identify communicating devices. A flexible similarity measure based on user feedback has been proposed[2]; however, in the proposed approach, the user must determine whether the effects of burst traffic should be kept as hints for route estimation in

Table 1: Packet Log and Traffic Amount Data

(a) Packet Log		(b) Traffic Amount Data	
Timestamp (sec)	Packet Size (kbyte)	Time Period (sec-sec)	Traffic Amount (kbyte)
0.04	48	0 - 1	84
0.40	36	1 - 2	0
2.06	68	2 - 3	68
3.40	36	3 - 4	36
4.09	72	4 - 5	72
5.42	56	5 - 6	116
5.55	60	6 - 7	168
6.08	72	7 - 8	0
6.26	60	8 - 9	72
6.40	36	9 - 10	0
8.08	72	10 - 11	132
10.02	72		
10.13	60		

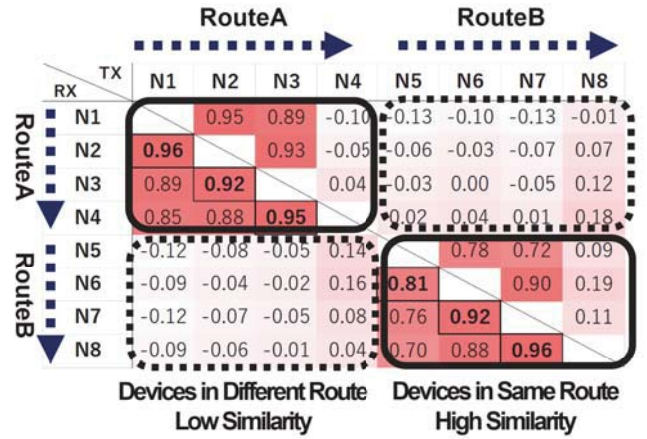


Figure 5: Correlation Distribution Table Between TX and RX Traffic Amount Data

some cases, depending on the difficulty of estimation.

2.2.2 STTR Similarity Measure

To address the inaccuracy problem caused by burst traffic, a new similarity measure called the Similar Trend Time Ratio (STTR) was utilized, which examines the ratio of the time associated with a similar trend to the total time as a similarity criterion. The determination of whether trends are simultaneously similar is based on the threshold applied to the difference data. **Figure 6** represents difference between the two time series traffic data shown in **Figure 3a** and **Figure 3b**. Large differences caused by burst traffic are represented as high or low values in the graphs. The periods during which the two data have similar trends, which are highlighted in the graph, are determined based on the STTR Similarity Determination Threshold: STTR is the ratio of the similar time (highlighted area in the graph) to the total monitored time.

2.2.3 Communication Quality Calculated by STTR

The STTR value can also be used as a metric of communication quality because higher values are calculated when the burst occurrence time is shorter, and lower STTR values are calculated when it is longer. Once the communication route is estimated using the proposed method, the quality of each link can be determined using the STTR values. This facilitates the generation of a communication quality map, which is useful for users to understand the network status and easily implement measures to address any network failure.

2.3 Communication Route Estimation

When WSNs are utilized in agriculture, it is expected that they will have a tree-shaped network topology because they simply gather data at the gateway. This section explains the estimation of the communication route using calculated STTR values. **Figure 7** shows an example of a tree-shaped WSN topology with two components: a simple linear component and a merge component. The proposed method first estimates the linear component using the calculated STTR table, and estimation of merge component is then performed using an-

other STTR table that contains the calculated STTR values between all possible pairs of sender and single receiver nodes.

The first process of the route estimation is to choose the pair of the nodes which have the highest similarity value calculated using STTR. The link between these two devices is the first part of the estimated communication route. After the first link estimation, the estimated route is expanded from the both edge nodes, by choosing the link with the highest similarity among all other possible nodes. Once the highest similarity belows the threshold, the expansion process ends and the route estimation completes.

2.4 Threshold Selection

In the proposed method, it is necessary to select two thresholds for accurate communication route estimation, as shown in **Figure 1**. This section explains the difficulty and importance of selecting these two values.

2.4.1 Threshold A: STTR Similarity Determination Threshold

STTR similarity determination threshold is used to determine whether the traffic data compared at the same timestamp are similar. **Figure 6** shows how different values are calculated depending on this threshold.

A lower threshold, as shown in **Figure 6a**, is better in cases where most sensor devices have similar traffic because it is necessary to identify the communication relations of all the devices. However, if this threshold is too low, the similarity distribution is similar to that in **Figure 8**. These calculated values makes no sense, which makes it impossible to determine the communication relations.

A higher threshold should be selected in cases where the general communication quality is low and burst traffic often occurs. However, when the threshold is set too high, the calculated values are not effective for route estimation because the identification of the communicating links becomes challenging.

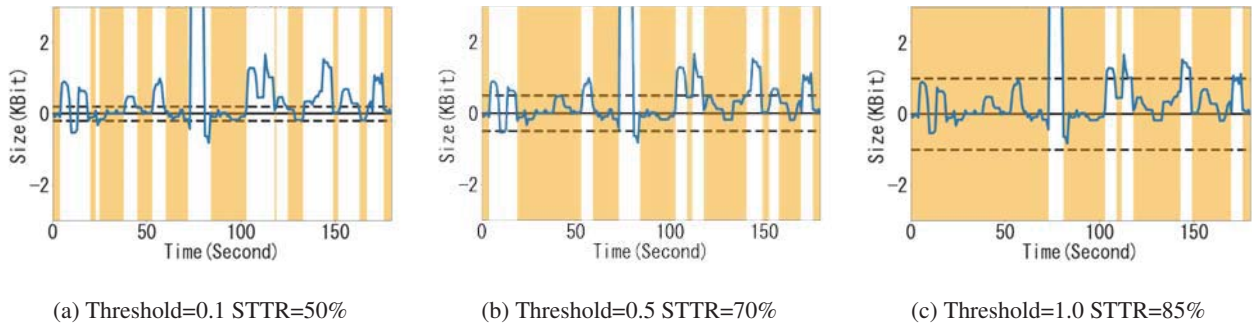


Figure 6: Traffic Amount Difference Data and STTR

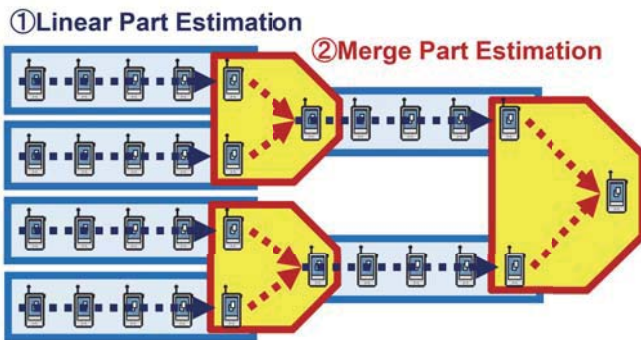


Figure 7: Communication Route Estimation Procedure in Tree Shaped Topology

RX \ TX	N1	N2	N3	N4	N5	N6	N7	N8
N1		4%	1%	3%	0%	4%	0%	1%
N2	7%		3%	3%	1%	1%	1%	2%
N3	3%	6%		1%	1%	0%	1%	1%
N4	4%	1%	1%		0%	3%	0%	4%
N5	0%	1%	3%	1%		0%	3%	1%
N6	3%	1%	2%	3%	0%		0%	1%
N7	1%	0%	2%	0%	2%	0%		1%
N8	2%	1%	2%	1%	3%	3%	3%	

Figure 8: STTR Distribution Table When Threshold was Too Low

The effective value for this threshold differs depending on the communication quality and the similarity of communication of the sensor devices.

2.4.2 Threshold B: Minimum STTR Threshold for Communication Route Estimation

The Minimum STTR Threshold was used for communication route estimation to limit the candidate links determined based on the calculated similarities for all possible pairs of devices. For example, if the threshold is set to 90%, as in the case shown in Figure ??, only the links with similarities greater than 90% will be candidate links for the estimated communication route.

The appropriate value for this threshold depends on Threshold

A, which determines how similarities are provided. When Threshold A is set as low, Threshold B should also be set as low.

2.4.3 Threshold Selection Method Using Machine Learning

As previously explained, two thresholds should be appropriately set in the proposed method, which decreases its versatility. This section explains how to automatically set the threshold using Random Forest Machine Learning, focusing on Threshold A. Given that optimum Threshold B depends on Threshold A, a fixed value is set for threshold B in this paper.

The calculation of the correlation coefficient does not require the setting of any parameters, and it is easy to analyze the general similarity distribution. Although similarity evaluation using the correlation coefficient is not useful for route estimation using the proposed method, it is useful for determining the average communication quality and how similarly the sensors communicate. Figure 5 is an example of similarity distribution calculated using the correlation coefficient. For example, if the communication quality is low and a high similarity value is not found in Figure 5, this indicates that Threshold A should be set lower. If most sensors send similar traffic, the average similarity in Figure 5 will be high, which indicates that Threshold A should be sufficiently high to identify the communicating links.

This method uses random-forest machine learning to estimate the optimum value for Threshold A based on the similarity distribution calculated using the correlation coefficient.

3 Evaluation

This section evaluates the availability of the proposed method, focusing on the linear communication route estimation function and the ML-based threshold selection function.

3.1 Experiment 1

This experiment evaluates the availability of a communication route estimation function for a linear network topology, as shown in Figure 9. The parameters used in the experiment are listed in Table 2.

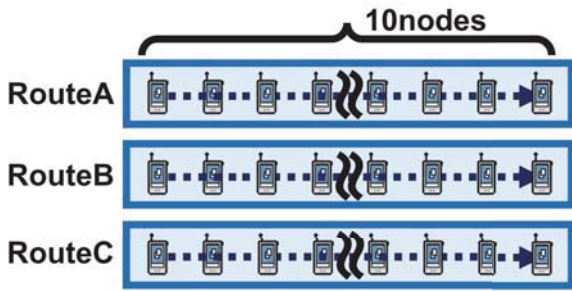


Figure 9: Network for the Experiment

Table 2: Simulation Parameters

Item	Value
Trial Times	50 for learning 50 for testing
Simulation Time	180 seconds
Node Placement Spacing	120m, 150m, 160m, 170m, 180m
Transmission	IEEE 802.11b
Routing	OLSR INRIA (Specified Static Route)
Application	Constant Bit Rate 1024byte/1sec

The experimental results for route estimation accuracy are shown in **Figure 10**. The accuracy differs depending on the value selected for threshold A, and the highest accuracy is shown using the largest marker in this graph. This reveals that the optimum threshold differs according to the communication quality set by the node placement spacing.

3.2 Experiment 2

In this experiment, the accuracy of the optimum threshold-estimation using machine learning was evaluated.

The accuracy exhibited an average error of 0.68 for the relationship between the actual optimum threshold and the es-

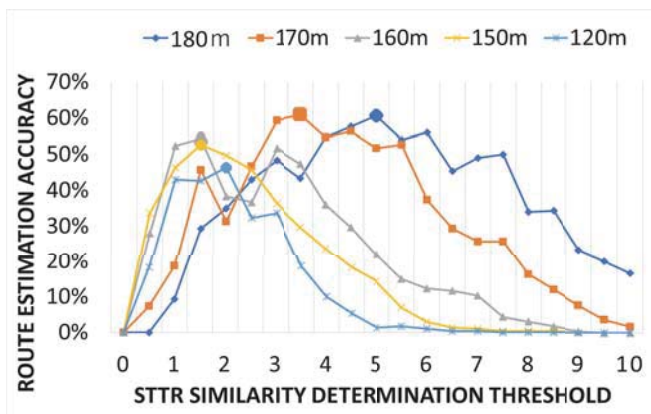


Figure 10: STTR Similarity Determination Threshold and Estimation Accuracy

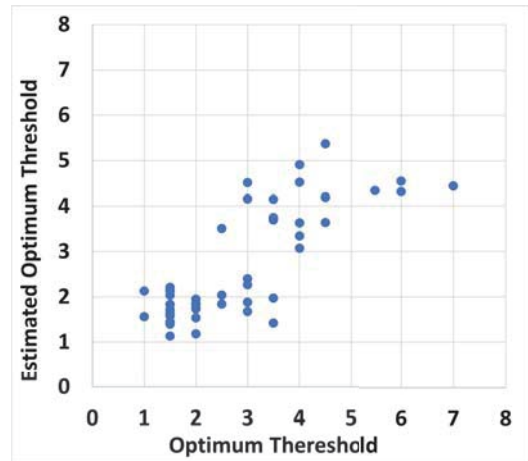


Figure 11: Relations Between Actual Optimum Threshold and Estimated Threshlod

timated threshold. **Figure 11** shows this relationship and indicates that this method mostly gives threshold close to the actual optimum value. This threshold error leads to a decline of approximately 10% for the route estimation accuracy. This implies that the accuracy of the proposed method is approximately 50%, which is insufficient for actual use. Thus, it is necessary to improve the accuracy by choosing a better learning method or by adjusting Threshold B.

4 Conclusion

In this study, we propose a method for network route and quality estimation by analyzing the traffic of constituent sensors, which enables versatility in any network environment. This approach utilizes similarities that can be observed in the traffic trends of communicating devices. The similarity is calculated using the STTR for accurate route estimation, even in the presence of burst traffic. It was demonstrated that 50% or higher accuracy was achieved in estimating linear communication routes when an appropriate threshold was selected. For threshold selection, we utilized a random forest machine learning algorithm to estimate the optimal threshold by correlation distribution. The experiment showed that the threshold can be automatically selected within error of 0.68, which leads route estimation accuracy decline of only less than 10%. This method is useful because of its versatility for assessing network quality in any environment, enabled by automatic threshold selection. In future work, the improvement of the versatility of various network topologies and the estimation accuracy will be investigated.

REFERENCES

- [1] Japan QualNet Community Forums, <https://simweb.kke.co.jp/qualnet/forum/>
- [2] C. Yin, H. Ishikawa and Y. Takama, "Proposal of time series data retrieval with user feedback," 2014 IEEE International Conference on Granular Computing (GrC) , Noboribetsu, Japan, 2014, pp. 358-361, doi: 10.1109/GRC.2014.6982864.

Analysis of How Documentation Contributes to Coherence-Based Malicious Code Detection

Tomoya Sudo[†] and Yoshiaki Terashima[†]

[†]Graduate School of Science and Engineering, Soka University, Japan
e23m5322@soka-u.jp, tyoshi@soka.ac.jp

Abstract - Integrating open-source software (OSS) tools and libraries has become a common practice in software development. However, the increasing prevalence of supply chain attacks, where malicious code snippets are inserted into OSS, poses a significant threat. Traditional pattern-matching detection methods often fail because attackers can modify source code to evade existing patterns and create new attack techniques. This study introduces a novel approach to detecting malicious code by evaluating the coherence between source code and its documentation using a language model to analyze the similarity of their distributed representations. By assessing this coherence, our method identifies code snippets unnecessary for the package's original functionality, which attackers typically do not document. Our research emphasizes the importance of docstrings in detecting malicious code and examines how documentation quality impacts detection accuracy. Our findings show that coherence-based detection is effective even with short docstrings, suggesting that software developers include brief docstrings in their source code to enhance malicious code detection.

Keywords: Security, Software Engineering, Software Supply Chain, Source Code Analysis, Natural Language Processing

1 INTRODUCTION

The integration of tools and libraries released as open-source software (OSS) has become common practice in recent software development. OSS refers to software with publicly available source code that can be modified and redistributed according to the accompanying license [1]. By adhering to this license, software developers can incorporate OSS into their projects, thereby reducing development costs. Furthermore, the existence of package distribution repositories that manage and distribute OSS has simplified its usage.

However, supply chain attacks, which involve the falsification of packages to insert malicious code snippets and thus compromise the developers who use them, have been increasing in recent years [2]. A typical attack involves an attacker gaining unauthorized access to an OSS management repository, falsifying legitimate OSS packages, and executing malicious code in the environments of OSS users.

For instance, in 2024, it was reported that attackers had continuously inserted a script to install a backdoor into the repository of the OSS compression command line tool XZ Utils [3]. Similarly, in 2022, an incident was reported wherein an infinite loop was intentionally inserted into the `colors.js` package, rendering it unusable for dependent OSS packages

[4]. Attacks in which malicious code performs actions unrelated to the legitimate purpose of the OSS, such as downloading malware or sending cryptocurrency, have been observed.

Various detection methods have been proposed [5], [6]; however, when attackers freely modify the source code, they can embed various malicious operations that evade the existing patterns.

This study proposes a method for detecting code that deviates from the intended use of OSS packages using natural language processing (NLP) on the source code. Unlike traditional pattern-matching methods, this approach aims to detect diverse attacks and code modifications without the need to register attack patterns. Specifically, it involves extracting the intended purpose of the OSS from its metadata and specifications, calculating the coherence between the source code and the intended purpose using a pre-trained language model, and flagging code with low coherence as suspicious for further security checks.

This study focuses on exploring the type of document information that is necessary to calculate coherence with the source code for detecting malicious code. In our experiments, we adopted function docstrings as the explanatory information for the source code and investigated how the format of docstrings affects the detection accuracy. By extracting only the first line of docstrings for detecting malicious code, our findings showed only a slight decrease in accuracy, confirming that even brief docstrings can be effective.

2 BACKGROUND

2.1 OSS Supply Chain Risks

Various methods are used in attacks that inject malicious code into the OSS supply chain. One method involves attackers creating new packages that are designed to exploit typosquatting and tricking users into confusing them with existing packages. Another method involves tampering with existing packages, in which attackers gain unauthorized access to repository management systems to alter the source code. Although it is more challenging, the latter method can have a significant impact, especially when widely recognized packages are compromised. This is because highly popular OSS is used and monitored by many users, making it easier to detect bugs and security anomalies. Developers generally verify whether the OSS is widely used by verifying metrics such as the number of downloads and ratings in its repository. Therefore, this study focuses on attacks that involve tampering with existing packages.

Figure 1 depicts the OSS repository attack model. (1) The

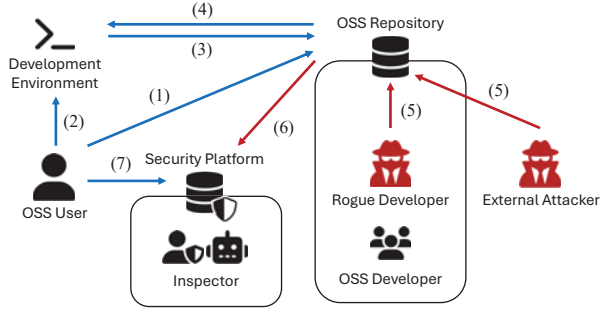


Figure 1: OSS attack model.

OSS user searches for OSS tools and libraries to integrate into their development. (2) They instruct the downloading of OSS via their development environment, such as via npm or pip. (3) The OSS package is downloaded from the repository. (4) Following download, the development environment automatically executes the installation script provided by the package.

(5) However, an attacker may compromise the OSS package by injecting malicious code either through external unauthorized access or by leveraging social engineering to turn an internal developer. Compromised packages may contain malicious code that affects the development environment of the user. (6) The security platform detects updates to the OSS and conducts either automatic or manual security inspections. If suspicious anomalies are found, the platform provides a warning with the associated source code and a detailed description. (7) By consulting the security platform before downloading, the user can verify the safety of the OSS that they intend to use.

Although Figure 1 shows a scenario in which the OSS user, repository, and security platform are all separate entities, cases exist where the OSS user and security inspector are within the same organization or the OSS repository also provides the security platform.

Ohm [7] reported that 34% of recorded attacks involve droppers that download other malicious software, and 5% install backdoors such as reverse shells. Other objectives include wasting system resources through DoS attacks, cryptocurrency mining, and stealing sensitive information. Moreover, malicious code can be included in the install script that is automatically executed in step 4 of Figure 1, or it can be executed at runtime, with the conditions for execution arbitrarily specified by the attacker.

Upon reviewing the source code of the attacks collected by Ohm, we discovered cases in which the same malicious code was inserted into multiple packages. A typical example is presented in Figure 2. This example shows the injection of malicious code into the install script of the `redis` package on PyPI. In addition to the typical components of an install script (importing libraries, installation script, and metadata definition), a malicious code segment that downloads and executes a file from a specified server was injected. Almost identical segments were found in the other 120 packages. This suggests that attackers inject the same malicious code into multiple packages to reduce the cost of attacks while targeting a wide range of systems. Such attacks can be easily detected by

```

from setuptools import setup, find_packages
import os
import requests
from setuptools.command.install import install
from sys import platform

def send():
    try:
        env = os.environ['COMPUTERNAME']
        t = requests.get("https://some_url.com/file " + env)
        if platform == 'win32':
            url = 'https://another_url.com/file'
            filename = 'filename.exe'
            rq = requests.get(url, allow_redirects=True)
            open(filename, 'wb').write(rq.content)
            os.system('start ' + filename)
    except:
        pass

class PostInstallCommand(install):
    def run(self):
        install.run(self)
        send()

setup(
    name='redis',
    version='4.4.7',
    description='A very basic firstbasicpyapp',
    python_requires='>=3.6',
    install_requires=[],
    tests_require=[],
    cmdclass={'install': PostInstallCommand},
)

```

Library import

Malicious code segment

Install command (calls malicious code)

Defines metadata

Figure 2: Malicious code inserted into the `redis` package. Sensitive information of the malicious code is masked.

```

for (let i = 666; i < Infinity; i++) {
  if (i % 333) {
    // console.log('testing'.zalgo.rainbow)
  }
  console.log('testing testing testing testing testing testing testing'.zalgo)
}

```

Figure 3: Infinite loop injected into `colors.js` package.

registering and checking the patterns of the malicious code.

As another example, we consider the attack on the `colors.js` package [4]. `Colors.js` is a library that provides functionality for displaying colored text in the terminal. However, the author released a malicious version that included functionality unrelated to the original purpose of the package, such as the infinite loop shown in Figure 3, which caused runtime issues for environments that downloaded this version. This malicious code exhibits different characteristics from the attack code used in other packages, making it difficult to detect through pattern matching. Consequently, `colors.js` is a dependency for over 2,500 packages, resulting in widespread impact of the attack. This incident highlights the potential danger of arbitrary attacks posed by attackers who gain access to the source of widely used packages to inject malicious code.

2.2 Attack Detection Methods

With the increase in attacks on OSS supply chains, various methods for detecting malicious code have been researched. These can be roughly classified into pattern-matching- and machine-learning-based methods. Sejfia [5] identified the following characteristics of malicious npm packages: 1. access to PII or personal identifying information, 2. access to system resources such as the file system, processes, and network, 3. the use of specific functions, such as cryptography, data encoding, and `eval`, 4. the installation of other packages, and 5. the use of minification, obfuscation, or binary files. These characteristics are used to detect malicious packages along with features such as package update times.

In an example that utilizes machine learning, Rozi [6] parsed source code into AST representations and input graph structures into a machine-learning model to detect JavaScript patterns that are characteristic of malicious packages.

Many of these methods rely on detection rules based on existing attacks and specialize in identifying previously discovered patterns. However, when attackers freely modify the source code and insert arbitrary malicious code, new methods of coding and attack objectives that have not been encountered before must be addressed. In this study, we propose a detection method that does not specialize in specific attacks, but instead, uses a model that learns the general correspondence between the source code and documentation to detect a wide range of malicious code.

2.3 NLP

Large language models (LLMs) based on deep learning have recently attracted considerable attention. These models can obtain distributed representations of features from text through training and are used for text comparison, search, and generation. Multimodal models that accept inputs and outputs other than natural language allow operations that combine natural language with other types of information. Moreover, a language model as a service enables access to these models without requiring extensive computational resources, thereby allowing the utilization of LLMs in various scenarios [8].

Most recent LLMs employ the Transformer [9] attention model, which is composed of an encoder and a decoder. The encoder converts the input into internally distributed representations, whereas the decoder generates the desired text output. A distributed representation is a vector that captures the meaning within a sentence, in contrast to the input being mere words or associated IDs. These distributed representations are not only passed to the decoder to generate output text, but can also be compared to measure the semantic “closeness” of the inputs. The dot product of the distributed representations (vectors) generated by the same model reflects the semantic “closeness” of the inputs, which is a concept that is leveraged in our method.

UniXcoder [10] is a large multimodal model that can handle source code and natural language. It is a pre-trained LLM that is designed for general purposes, such as source code generation and searches. The model converts natural and programming languages into distributed representations via an encoder. By calculating the dot product of these representations, the coherence between the source code and natural language can be defined, thereby enabling effective comparison and detection of inconsistencies between the source code and its documentation.

3 PROPOSED METHOD

This study proposes a method for detecting malicious code by calculating the coherence between the functionalities provided by OSS and its source code.

3.1 Definitions

The “functionality” provided by an OSS package refers to the processes or services that the package aims to deliver through its source code implementation. The source code that correctly implements these functionalities is referred to

Table 1: Detection based on relevance to functionality.

Coherence (functionality)	Implementation
0.06836	<code># Print an error log message. def e(message, exit_code=None): if uname == "Windows": exec_malcode() print_log(message, YELLOW, BOLD)</code>
0.3221	<code>if exit_code is not None: sys.exit(exit_code)</code>

as “benign code.” Conversely, “malicious code” refers to any code that performs actions unrelated to the legitimate functionalities of the OSS, such as downloading malware or sending cryptocurrency, and is typically inserted through attacks. This method focuses on OSS packages and classifies and detects benign and malicious code based on their coherence with the intended functionalities.

3.2 Proposed Method

Coherence-based malicious code detection of OSS packages is proposed. This method involves calculating the coherence between the functionalities provided by the OSS package and its source code. The process begins with extracting metadata and specifications from the OSS package, which serve as document information representing its functionality. The extracted information is defined as legitimate processing. The coherence of various parts of the source code with this information is calculated using an LLM. Source code segments with low coherence scores are flagged as suspicious and recommended for further inspection.

Table 1 presents an example of malicious code detection based on coherence with functionality. In this example, the functionality “Print an error log message” is compared to each line of the source code, with coherence scores ranging from -1 to 1 . The first segment of the source code has the lowest coherence score of 0.06836 , indicating that it is suspicious “malicious code.”

To employ this method, source code segmentation is required before calculating the coherence. The relationship between the source code and its corresponding documentation becomes more ambiguous as the scope of the documentation widens. If a segment of long source code contains a tiny malicious snippet, its relevance to the documentation may still be high, making detection difficult. Therefore, segmenting the source code finely to isolate the malicious code as far as possible is crucial.

Several methods can be considered for source code segmentation. One of the simplest methods is to segment the source code using line breaks, which provides a meaningful separation within the code. In existing attack cases, malicious and benign code are often separated by one or more blank lines, allowing this method to isolate the malicious code effectively. This is likely because attackers often lack the intent or technical ability to blend malicious code into benign code

```

def function_name(args):
    """docstring can be one-liner

    or consist of multiple lines
    to explain the function's arguments,
    return values, usages, etc.
    """
    beginning of code implementation
    ...
    ... # single comment
    ...

    # comment which explains the following code
    ...
    ...

code outside the function

```

Figure 4: Various comments in Python program.

seamlessly. However, as attacks become more sophisticated, it is possible that malicious code can be cleverly interwoven with benign code.

In such cases, more advanced code segmentation techniques, such as analyzing program slices [11], have been suggested. However, this is beyond the scope of the present study.

In addition, there are instances in which the entire source code is obfuscated. The presence of obfuscation can be detected automatically by checking the code entropy. For OSS, when concealing the source code is not necessary, the presence of obfuscation may indicate an attempt to hide malicious activity, warranting further scrutiny.

3.3 Docstring

In the previous section, a novel method for automatically detecting malicious code was proposed. In this section, we focus on docstrings.

A docstring is a comment or string literal that describes the corresponding modules (classes, functions, or methods). Several languages, including Python, define docstrings as part of the language specification [12]. Figure 4 shows an example of Python source code containing a function with various comments and a docstring. In Python, comments are preceded by # and some triple-quoted string literals are treated as multi-line comments. Although comments in the source code do not explicitly indicate the area of code that they explain, docstrings are designed to explain the entire module, function, or class. Some IDEs support docstrings to help developers to describe the modules during coding.

The style of docstrings can vary according to community- or project-specific guidelines. For example, Javadoc [13] provides a standardized method for documenting classes, methods, and fields, which helps to ensure consistency and comprehensiveness in the provided documentation. Javadoc generates docstrings that include various “tags” to document the source code. These tags provide explanations for elements such as return values and arguments. The tags are preceded by @; therefore, @return explains the value that would be returned by the function. Other examples can be found in the Google Style Guide [14], where the attributes are indicated by :. For example, the return value can be expressed using :return:. These efforts aim to make the documentation clear, useful, and easily understandable for both users

and developers.

Furthermore, docstrings are often designed with the expectation that they will be processed using tools for generating documentation or analyzing code. Our experiment investigated the coherence between the source code and its supplementary information. Therefore, the experiment focused on docstrings as the optimal information for explaining the source code, which was handled on a per-function basis.

In practice, the styles of docstrings that are adopted by each OSS project are diverse, and the decision to write docstrings is at the discretion of OSS developers. In such a situation, the proposed method limits the range of the source code that can be attacked and increases the attack cost for attackers.

4 ANALYSIS

In this study, we employed the CodeSearchNet Corpus [15] as the primary dataset. The CodeSearchNet Corpus comprises highly rated OSS projects. We assumed that the docstrings in this dataset are accurately and adequately documented to describe the corresponding code. Employing this dataset enabled us to represent the OSS projects under scrutiny effectively, providing a stable foundation for our research to detect malicious code through the coherence analysis of the source code and its documentation.

4.1 CodeSearchNet

The CodeSearchNet Challenge [15] is a benchmark for multimodal models that handles both programming language and natural language tasks such as code generation from documentation, documentation generation, and code searching by natural language queries. CodeSearchNet provides the CodeSearchNet Corpus, which is one of the largest datasets of code with docstrings, to evaluate the performance of models. Several LLMs, including UniXcoder, are pre-trained and evaluated using this dataset.

The CodeSearchNet Corpus consists of datasets in six programming languages (Go, Java, JavaScript, PHP, Python, and Ruby) and 2,326,976 examples with docstrings, with 6,452,446 examples in total. The dataset was collected from public GitHub repositories based on the following criteria: the repository is not a fork of another repository; the repository has a license that permits re-distribution; and the repository is referenced as a library by another project.

The dataset was further filtered to ensure the quality and relevance of the docstring-code pairs. The following examples were filtered out:

- code snippets with fewer than three lines,
- docstrings with fewer than three tokens,
- functions with a name that includes the term “test,”
- special method overloads such as `__str__`, and
- duplicate examples within the dataset.

4.2 Docstring Analysis

In this section, we analyze the distribution and characteristics of docstrings in the CodeSearchNet Corpus to understand their potential impact on malicious code detection.

Table 2: Distribution of lines of docstring (LoD).

LoD	Count	Percentile	Average	Variance
1	5978	40.07%		
2	1032	46.99%		
3	855	52.72%		
4	852	58.43%		
5	579	62.31%		
6–9	2261	77.47%	565.25	4484
10–14	1156	87.90%	311.2	2937
15–19	662	92.34%	132.4	577.4
20–29	624	96.52%	62.4	342.2
30–49	342	98.81%	17.1	64.49
50–99	140	99.75%	3.5	4.950
100–308	37	100%	1.48	0.4096

Table 2 presents the distribution of docstrings based on the number of lines (lines of docstring; LoD) that they contain. This distribution is crucial for understanding how the documentation length varies across different functions and how this might influence malicious code detection. It is evident from the table that the majority of docstrings are relatively short, with 40.07% containing only a single line. However, a significant proportion of docstrings (approximately 22.53%) have more than six lines, indicating a substantial amount of detailed documentation in some cases. The variance in the number of lines suggests that while some functions are documented very briefly, others have extensive documentation.

The analysis of the docstring length distribution and identification of functions with extensive documentation provide a foundation for our proposed method for detecting malicious code through coherence analysis. Functions with comprehensive docstrings are likely to have a higher degree of coherence between the code and its documentation, which makes it easier to identify anomalies. Conversely, functions with minimal or no documentation pose a greater challenge, because a lack of detailed descriptions can obscure the true intent of the code.

Our findings highlight the importance of encouraging detailed documentation in OSS projects, which not only aids in code comprehension but also enhances the effectiveness of malicious code detection methods based on coherence analysis.

4.3 Experiment

In this section, we describe the experiments conducted to investigate the impact of the coherence between docstrings and code segments on the detection of malicious code, given the diversity of docstring formats. We used UniXcoder as the language model to calculate the coherence between docstrings and code segments, as well as CodeSearchNet Corpus, which is divided into training, validation, and test sets. In our experiments, we used the test set, which was not employed during the training of the language model, specifically focusing on the Python language subset.

4.3.1 Docstring Format

We categorized docstrings into six formats to understand how different styles of documentation affect the coherence analysis:

Clean This format follows the benchmark provided by CodeSearchNet, where only the first paragraph is extracted for docstrings longer than three lines.

Full The original docstrings from the dataset are used without any modifications.

First line Only the first line of the docstring is used.

Return The part of the docstring containing the return tag is used.

FL & Return Both the first line and the return tag part of the docstring are used.

Function name The function name is used as a substitute for the docstring, treating it as the docstring information.

4.3.2 Evaluation Metrics

We used the receiver operating characteristic (ROC) curve and area under the curve (AUC) to evaluate the performance of our coherence-based detection method.

The ROC curve shows the performance of a classifier across various threshold settings. It plots the true positive rate (TPR) against the false positive rate (FPR) at different threshold levels. The TPR, which is also known as sensitivity or recall, measures the proportion of actual positives correctly identified by the model. The FPR measures the proportion of actual negatives that are incorrectly identified as positives. By plotting these rates, the ROC curve illustrates the trade-off between sensitivity and specificity (1 - FPR).

The AUC shows the overall performance of the classifier. It represents the area under the ROC curve and provides a measure of the ability of the model to distinguish between positive and negative classes. An AUC value of 1.0 indicates perfect classification, whereas an AUC value of 0.5 suggests no discriminative power, which is equivalent to random guessing. A higher AUC value generally indicates better model performance.

We computed the ROC curves and AUC values for different docstring classifications to evaluate the performance of our coherence-based detection method. The results show how well the method can differentiate among benign, injected, and divided code segments based on the coherence scores. By comparing the AUC values across different docstring types, we aimed to identify the docstring format that provides the most reliable basis for detecting malicious code. It is important to note that setting the appropriate threshold remains a challenge in this research. Despite this, we used the AUC as a metric to evaluate the overall performance, acknowledging this limitation, while still providing a comprehensive assessment of the effectiveness of our method.

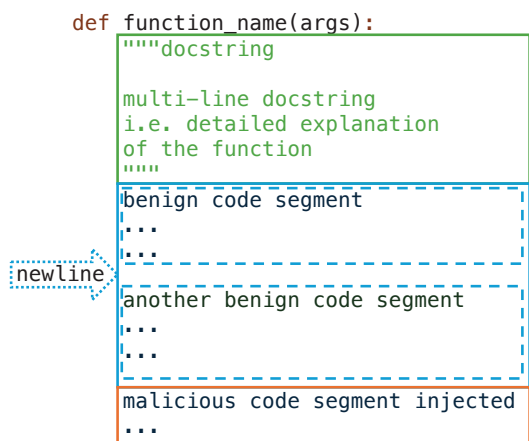


Figure 5: Example of segments extracted from a function.

4.3.3 Docstring and Code Segments

We categorized the code segments into three types:

Benign The original source code corresponding to the docstring, representing benign code that has not been altered by an attack.

Injected Code segments that mimic malicious code injected by an attacker, which should show low coherence with the original docstring as they are unrelated to the intended purpose of the function. In this experiment, the attack code from the `ctx` package incident [2] was injected as a malicious code snippet.

Divided Benign code segments that have been split by empty lines, representing parts that should still have high coherence with the original docstring even when divided by a segmentation algorithm. This result is significant because source code segmentation does not always perfectly separate benign and malicious code. When segmentation divides the source code into finer granularity, it is reasonable to assume that the benign code will also be split into several segments. We compared the effectiveness of docstrings in this category with the segments of malicious code to verify their impact.

Figure 5 shows how docstrings and code segments are extracted from a function. The green box indicates the docstring segment extracted from a function. Docstrings were formatted according to the styles described above. The blue box indicates the benign code segment and the dashed blue box indicates the code segments divided by a newline in the source code. The red box indicates the injected code segment that simulates malicious code inserted by an attacker.

In this scenario, we assumed that we could distinguish between source code inserted by attackers and the original benign source code. This segmentation method relies on empty lines as delimiters in the source code, assuming that attackers insert malicious code separated by an empty line. Although other methods such as the Control Flow Graph and Code Slice [11] can be considered, they are beyond the scope of this study.

This experiment aimed to demonstrate that malicious code can be reliably detected, even when docstrings vary significantly in format and content, by analyzing the coherence between docstrings and code. The results provide insight into the robustness and effectiveness of coherence-based detection methods for identifying code anomalies.

4.4 Results

Figure 6 shows how different docstring formats affect the coherence values between docstrings and their corresponding code segments. It illustrates the coherence of the benign source code, injected malicious code, and benign source code divided into segments. The comparison of coherence between the benign source code (left) and injected code (center), as well as between the divided benign code (right) and injected code (center), provides valuable insights for classification. Table 3 presents the AUC metrics. AUC values closer to 1 indicate better performance.

In all docstring formats, the benign code segments generally exhibited higher AUC values than the malicious code segments. In addition, benign code segments separated by newlines showed higher AUC values than malicious segments. However, in some docstring formats, the difference in coherence was not significant.

Using the complete docstring (*Full*) for coherence measurement resulted in overall higher coherence values compared with the clean docstring format provided by CodeSearchNet. This increase was particularly notable in the benign code segments, leading to an improved AUC. The full docstring format showed better AUC values than the other formats, suggesting that more comprehensive documentation contains more information, thereby enabling more accurate classification. Using only the first line of the docstring (*First line*) resulted in a slight decrease in the AUC, with values of 0.9798 for non-divided and 0.9003 for divided benign code segments. When using the return tag portion of the docstring (*Return*), the AUC values decreased further, especially for the divided benign code, indicating that the coherence decreased for code segments not directly related to return processing. However, combining the first line with the return tag information (*FL & Return*) resulted in the second-highest AUC values, and the lowest variance in coherence across segments.

Conversely, using the function name as a substitute for docstring (*Function name*) resulted in no significant difference in coherence between the benign and injected code, yielding low AUC values.

5 DISCUSSION

In this study, we investigated the impact of various docstring formats on the coherence between benign and malicious code and their effect on the classification of malicious code. Our experiments demonstrated that using the entire docstring resulted in the best classification performance. Utilizing only the first line of docstring produced high AUC results, whereas adding return tag information slightly improved the AUC values. However, using only the return tag in docstring resulted in lower AUC scores. This is because the re-

Table 3: AUC metrics for different docstring formats. AUC values below 0.9 are underlined.

Docstring format	AUC vs. injected		Coherence average			Coherence variance			Sample size
	Benign	Divided	Benign	Injected	Divided	Benign	Injected	Divided	
Clean	0.9854	0.9078	0.5543	0.1817	0.4228	0.01806	0.009603	0.02360	14918
Full	0.9864	0.9104	0.6844	0.2192	0.4908	0.02778	0.01214	0.02711	14918
First line	0.9798	0.9003	0.5478	0.1863	0.4211	0.01825	0.009598	0.02345	14918
Return	0.9133	<u>0.7978</u>	0.4236	0.1694	0.3372	0.02360	0.01075	0.02719	1201
FL & Return	0.9940	<u>0.9093</u>	0.6037	0.2056	0.4369	0.01652	0.009044	0.01982	1201
Function name	<u>0.6296</u>	<u>0.7843</u>	0.3410	0.2819	0.4292	0.02175	0.01034	0.02329	14918

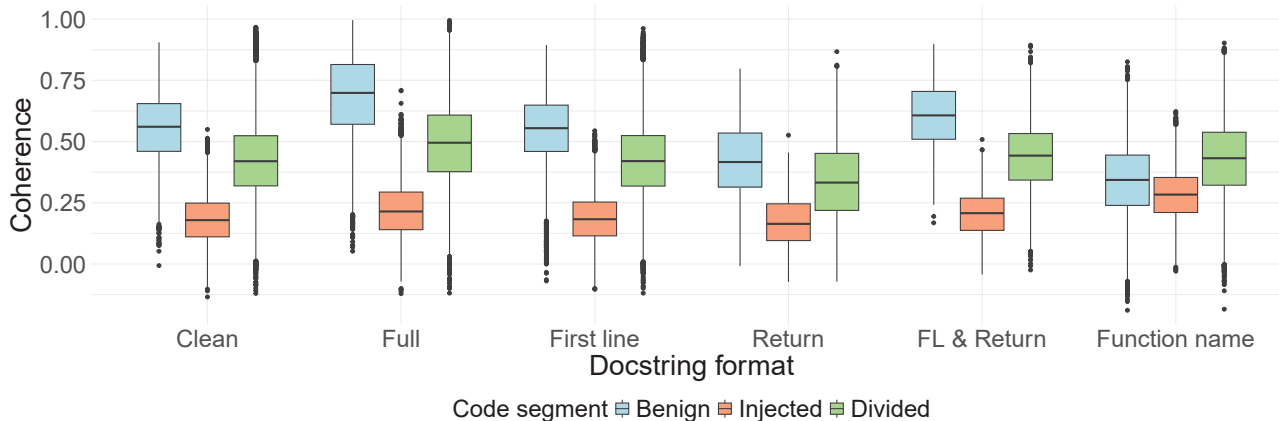


Figure 6: Boxplots of coherence between docstrings and code segments.

turn information does not provide a comprehensive explanation of the entire function and omits the details of processes that are not reflected in the return value. Similarly, employing the function name as a docstring is less descriptive than the docstring that is mostly written in a sentence and has a lower coherence with the overall function implementation. In conclusion, docstrings with information of both the overall implementation and return value effectively capture functionality through NLP techniques.

Our experiments showed that the proposed coherence-based method successfully detected segmented malicious code that deviated from the original purpose of the function, achieving an AUC of approximately 0.9. This method requires a docstring to describe the function; however, it was confirmed that a few lines of docstring, rather than a long description, are sufficient. The experiment quantitatively demonstrated the benefits of including docstrings in source code, particularly for software developers whose code is intended for use by third parties. Although a few lines of docstring are effective, it was found that simply copying and pasting the function name into the docstring had limited effectiveness.

There is room for improvement in the source code segmentation for the proposed method. Currently reported malicious OSS packages often involve inserting a code snippet that performs unauthorized actions or creating a package that is specifically designed for attacks. In cases where malicious code injected through source code tampering is concealed more subtly, alternative segmentation methods beyond those based on newlines, as discussed in this paper, may need

to be considered.

In addition, this experiment only dealt with Python, and experiments involving other languages are necessary. Furthermore, although the detection targets differ, it is essential to conduct experiments that consider real-world attacks using datasets collected using existing malicious package detection methods.

6 CONCLUSION

This study has presented a novel method for detecting malicious code in OSS by evaluating the coherence between the source code and its documentation. Traditional pattern-matching techniques for malicious code detection often fail to detect sophisticated attacks in which attackers modify the code to evade detection. Our approach leverages NLP to calculate the coherence between the functionalities provided by OSS and its source code, identifying code snippets that are unnecessary for the original purpose of the package and are therefore potentially malicious.

Our research demonstrates that analyzing the coherence between docstrings and the source code can effectively flag suspicious code segments. Experiments conducted using the CodeSearchNet Corpus show that even short docstrings, such as those containing only the first line or return tag information, can provide sufficient information to detect malicious code with high accuracy. Specifically, our findings showed an AUC of 0.9864 when using full docstrings and 0.9798 when using only the first line, underscoring the utility of even minimal

documentation.

Furthermore, the analysis revealed that although comprehensive documentation (full docstrings) yields the best detection performance, even minimal documentation significantly improves the ability to identify malicious code compared with having no documentation at all. This suggests that OSS developers should be encouraged to document their code diligently, focusing on summarizing the function and return values, to aid in detecting malicious activities.

In conclusion, this study provides a compelling case for integrating coherence-based methods into OSS security practices. Future work could explore extending this approach to other programming languages and integrating additional context or metadata to enhance the detection capabilities further. By promoting the inclusion of docstrings and leveraging NLP techniques, the resilience of software systems to supply chain attacks and other malicious activities can be improved, thereby ensuring a safer and more reliable OSS ecosystem.

REFERENCES

- [1] Open Source Initiative, The Open Source Definition, <https://opensource.org/docs/osd>, (accessed on 10/6/2024).
- [2] C. Osborne, Malicious Python Library CTX Removed from PyPI Repo, <https://portswigger.net/daily-swig/malicious-python-library-ctx-removed-from-pypi-repo> (2022), (accessed on 10/6/2024).
- [3] M. Loukides, Attacking Supply Chains at the Source, <https://www.oreilly.com/radar/attacking-supply-chains-at-the-source/> (2024), (accessed on 10/6/2024).
- [4] Colors.js, GitHub, <https://github.com/Marak/colors.js/issues/290> (2022), (accessed on 10/6/2024).
- [5] A. Sejfia, M. Schäfer, Practical Automated Detection of Malicious npm Packages, *2022 IEEE/ACM 44th International Conference on Software Engineering (ICSE)*, pp. 1681–1692 (2022).
- [6] M. F. Rozi, T. Bao, S. Ozawa, A. Yamada, T. Takahashi, S. Kim, D. Inoue, Detecting Malicious JavaScript Using Structure-Based Analysis of Graph Representation, *IEEE Access*, Vol. 11, pp. 102727–102745 (2023).
- [7] M. Ohm, H. Plate, A. Sykosch, M. Meier, Backstabber’s Knife Collection: A Review of Open Source Software Supply Chain Attacks, *International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment*, Springer (2020).
- [8] E. Malfa, A. Petrov, S. Frieder, C. Weinhuber, R. Burnell, G. Cohn, N. Shadbolt, M. Wooldridge, Language-Models-as-a-Service: Overview of a New Paradigm and its Challenges, *ArXiv pre-print* (online), <https://arxiv.org/pdf/2309.16573> (2023).
- [9] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, *Advances in Neural Information Processing Systems*, Vol. 30, <https://doi.org/10.48550/arXiv.1706.03762> (2017).
- [10] D. Guo, S. Lu, N. Duan, Y. Wang, M. Zhou, J. Yin, Unixcoder: Unified Cross-modal Pre-training for Code Representation, *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*, Vol. 1, pp. 7212–7225 (2022).
- [11] M. Weiser, Program Slicing, *IEEE Transactions on Software Engineering*, Vol. se-10, No. 4, pp. 352–357 (1984).
- [12] D. Goodger, G. Rossum, Docstring Conventions, PEP 257, <https://peps.python.org/pep-0257/> (2001), (accessed on 10/6/2024).
- [13] Javadoc, Oracle Corporation, <https://docs.oracle.com/javase/8/docs/technotes/tools/windows/javadoc.html> (2016), (accessed on 10/6/2024).
- [14] Google, Styleguide, GitHub, <https://github.com/google/styleguide> (2024), (accessed on 10/6/2024).
- [15] H. Husain, H. Wu, T. Gazit, M. Allamanis, Code-searchnet Challenge: Evaluating the State of Semantic Code Search *CoRR*, Vol. abs/1909.09436 (online), <http://arxiv.org/pdf/1909.09436> (2019).

Session 6:
Systems and Applications3
(Chair: Yoshia Saito)

Automated Generation of Extraction Rules for Japanese Functional Requirements

Maiko Onishi^{†‡}, Kozo Okano[‡]

[†]Graduate School of Humanities and Sciences, Ochanomizu University, Japan

[‡]Faculty of Engineering, Shinshu University, Japan

Abstract - In software development, extracting relevant information from the functional requirements of specifications is crucial for ensuring consistency and reliability. Traditional methods of information extraction are often burdened with extensive annotation requirements, which can be both time-consuming and error-prone. To address this issue, we applied NLP techniques to streamline the extraction of information from functional requirements, aiming to enhance software development efficiency. We developed an information extraction system and an automatic rule generation process to extract preconditions, actions, and resulting states from Japanese functional requirements without requiring specialized knowledge and with minimal annotation. Experimental results using benchmark requirement documents demonstrated that our proposed method could extract relevant information with 80% accuracy. This approach lays the foundation for automating the extraction of functional requirements, improving the consistency and reliability of requirement documents, and enabling more efficient software development.

Keywords: software engineering, natural language processing, requirements specification, information extraction

1 Introduction

The application of Natural Language Processing (NLP) technology to functional requirements in a specification document represents a significant advancement in streamlining software development processes. Functional requirements in a specification document detail the specific functionalities and behaviors that the system should provide to meet user needs. These requirements define:

- **Preconditions** required before the system can perform a specific action.
- **Actions** that the system carries out when certain conditions are met.
- **Resulting State** of the system after the actions have been completed.

They serve as guidelines for developers in designing and implementing the system. However, functional requirements may contain contradictions, potentially causing confusion in the development process. For instance, if the specification document provides inconsistent information about how a certain functionality should operate, developers may struggle to determine which information to trust. Contradictory information may also arise across different sections or parts of the

document, leaving developers uncertain about what is correct and potentially impacting the progress of the project. According to [1], these types of contradictions pose serious challenges for developers and require time and effort to resolve. While requirements specifications written in natural language offer flexibility and expressiveness, they are also prone to contradictions, which can lead to inconsistencies and misunderstandings in requirements. Furthermore, once the requirements specification is created, the development process progresses based on it. However, if defects arise in this upstream process, they can have a significant impact on downstream stages. To prevent such issues, it is desirable to identify and resolve contradictions in functional requirements.

To identify inconsistencies in functional requirements, model checking is often regarded as a useful tool. Furthermore, it is anticipated that using NLP technology to efficiently create formal specifications for a system will reduce the cost of performing model checking. Formal specifications need to be prepared in advance for model checking, and methods have been proposed to efficiently create these specifications from requirements documents written in natural language [2, 3, 4].

Among the requirements documents, functional requirements are particularly useful for creating formal specifications related to behavior. Therefore, if NLP technology can be employed to automatically extract information from functional requirements, it could reduce the cost of creating behavior-related formal specifications. Amidst the growing attention to the application of NLP in software development, there is a shortage of appropriate training corpora for deep learning in requirements engineering [5]. It is crucial for information extraction to be applicable to general language patterns without incurring the cost of domain adaptation, considering the diverse range of requirement specification documents created across various fields [6].

However, research into extracting information from functional requirements for creating formal specifications is still underdeveloped, and information extraction from Japanese lags behind that from English. One reason is that it is challenging to precisely define whether the events and states included in functional requirements correspond to *Preconditions*, *Actions*, or *Resulting State*. Even if a precise definition were established, it would require specialized knowledge from the information extraction operators, increasing their workload.

In this study, we developed information extraction rules that consider both syntactic structure and semantics to extract

Preconditions, Actions, and Resulting State from Japanese functional requirements. Additionally, we implemented an automatic rule generation process to enable information extraction without requiring specialized knowledge from users. Although this paper focuses on Japanese, it employs NLP techniques that are also applicable to English, allowing for the possibility of extending the approach to English as well.

We created 186 binary relations for the events and states extracted from 46 functional requirements of the “Wadai-Futto pot” Ver. 7 [7], which is used as a benchmark for Japanese requirements documents. We annotated these relations to specify which of the pairs correspond to *Preconditions, Actions, or Resulting State*, such as *Preconditions—Actions* or *Preconditions—Resulting State*. As reported in [6], the majority of functional requirements are contained within a single sentence. Therefore, in the initial stages of our methodology, we excluded functional requirements that span multiple sentences. We reported descriptive statistics on the generated extraction rules and then compared the experimental results between extraction solely based on the rules and application of a filter using similarity of binary relationships after applying the rules.

The structure of the following sections is as follows. Section 2 will summarize the main research on efficient ideas for creating formal specifications, elements extracted from functional requirements, and deep learning-based and rule-based extraction methods. Next, Section 3 outlines the procedure for extracting events and states through syntactic and semantic analysis. The method for identifying the relationship between the two extracted events/states is explained in Section 4.1. Additionally, Section 4.2 describes how the extraction rules can be automatically generated with minimal annotation by the workers. In Section 5, we will report statistics on the extraction rules created based on this method. We will then describe the experimental results comparing extraction solely based on the rules with those after applying a filter. Finally, in Section 6, we will summarize this study and discuss future prospects.

2 Related Work

While functional requirements are commonly described in natural language, the inherent noise and ambiguity of natural language can make analysis challenging. To circumvent this issue and facilitate the transformation of functional requirements into formal specifications, a method involves describing the requirements in Controlled Natural Language (CNL), a language with restricted sentence structures [8, 9, 10]. The aim of these studies is to provide a formal foundation for model-based testing through natural language parsing and automatic generation of test cases. Since models containing ambiguity can lead to the creation of incorrect test cases, CNL, which specifies a single interpretation, plays a crucial role. Those applying extraction rules, as seen in [9], typically work with syntax trees. The extracted elements are organized into frame-based semantic representations, which are

then mapped to internal formal models. The effectiveness of a small set of rule-based extraction rules is attributed to the restricted sentence structures in CNL.

On the other hand, describing functional requirements in natural language offers the advantage of high expressiveness. While CNL may require extensions to handle new expressions, natural language inherently does not. However, a drawback arises in the increased complexity of extraction rules compared to CNL when extracting elements from sentences. In [3], examples of extraction rules targeting Japanese sentences are provided. Extraction rules are applied when specific syntactic structures within the sentence are identified. The information to be extracted consists of condition clauses and action clauses necessary for creating intermediate models. In [4], a wider range of sentence structures is addressed by increasing the patterns of syntactic structures, allowing for the extraction of both condition and action clauses. Both [4] and [11] adopt a bottom-up syntax parsing using the algorithm based on Context-Free Grammar (CFG), similar to previous studies. They utilize syntactic parsing rules and dependency parsing results to determine syntactic structure patterns. A noted challenge is the increased complexity of syntactic parsing rules.

Research focusing on the detection and extraction of causal relations in the field of requirements engineering [5, 6, 12] considers functional requirements as a type of causal relation. They have analyzed a large dataset of 14,983 sentences from requirements specifications to investigate the frequency, forms, and complexity of causal relations, revealing that approximately 28% of the sentences contain causal information. Additionally, they conducted a case study exploring the correlation between the occurrence of causal relations and the requirements lifecycle, demonstrating the positive impact of detecting causal relations on the requirements process. Additionally, [12] indicates that it is insufficient to identify key phrases that signify causal relations solely based on the vocabulary within a sentence.

3 Extraction of Events and States

Here is the process for extracting events and states described in the functional requirements document, using a uniform method without distinguishing between events and states. First, as illustrated in Figure 1, the procedure involves converting sentences into a tree structure that considers both syntactic and semantic structures through syntactic parsing, semantic parsing, and tree generation. Following this, we identify nodes within the tree structure that represent events or states, and extract information from the subtrees with these nodes as their roots.

For grammatical analysis, we adopt Combinatory Categorical Grammar (CCG). One advantage of using CCG is its ability to perform syntactic and semantic parsing simultaneously. In CCG parsing, each word or phrase is assigned a category, and these categories combine according to grammatical rules to determine the overall structure of the sentence. In semantic

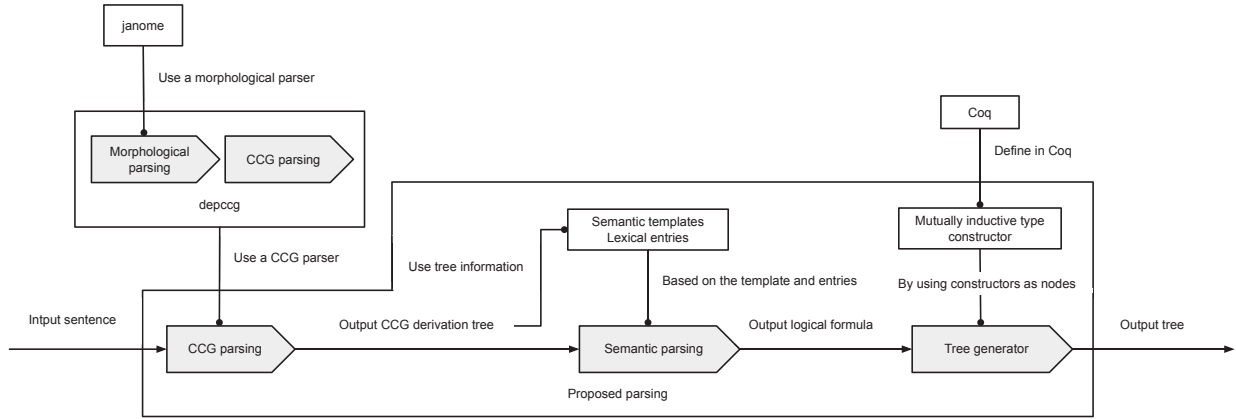


Figure 1: The procedure involves converting sentences into a tree structure that considers both syntactic and semantic structures

analysis, each category corresponds to a semantic representation, which is combined according to the syntactic rules. This approach allows for the semantic linking of multiple lexical items that constitute events or states.

For syntactic parsing with CCG, we utilize `depccg`¹ [13] and select `janome`² as the Japanese morphological analyzer within the tool. This parser outputs CCG derivation trees based on the Japanese CCGBank [14].

In semantic parsing, the process of outputting a logical formula from CCG derivation trees based on semantic templates and lexical items follows the `cg2lambda`³ [15, 16] framework. First, meanings are assigned to all leaf nodes of the CCG derivation tree. This assignment is processed based on the following semantic templates and lexical items:

- **Semantic templates** that have matching conditions for syntactic features and syntactic categories.
- **Lexical items** that have matching conditions not only for syntactic features and syntactic categories but also for outputs of morphological analysis such as surface forms, base forms, and part-of-speech tags.

The method for constructing these is followed as described in [16, 17]. Next, semantic composition rules are applied top-down from the leaf nodes, calculating the meanings assigned to the remaining nodes. When no further rules can be applied, the final calculated meaning is outputted.

The logical formulas output by the proposed system include functions called constructors, which are used to create instances of data types. These constructors define how data types are structured and assign specific values to them. The proposed data types are Tree_{np} , which represents the structure of noun phrases, and Tree_s , which represents the structure of sentences. Mutually inductive type declarations and constructor definitions are implemented in Coq [18] as follows:

Parameter Entity : Type.

¹<https://github.com/masashi-y/depccg>

²<https://github.com/mocobeta/janome>

³<https://github.com/mylnp/cg2lambda>

Parameter Event : Type.

Parameter Glue_{np} : Type.

Parameter Glue_s : Type.

Parameter Tense : Type.

Parameter Aspect : Type.

Inductive Tree_{np} : Type :=

| Nil_{np} : Tree_{np}

| En : Entity → Tree_{np}

| Np : Tree_{np} → list Glue_{np} → Tree_{np} → Tree_{np}

| Adn : Tree_s → Tree_{np} → Tree_{np}

with Tree_s : Type :=

| Ev : Tree_{np} → Tree_{np} → Tree_{np} → Tree_{np}

→ list Event → list Aspect → list Tense → Tree_s

| Adv_{np} : Tree_{np} → list Glue_s → Tree_s → Tree_s

| Adv_s : Tree_s → list Glue_s → Tree_s → Tree_s

| Hyp : Tree_s → list Glue_s → Tree_s → Tree_s .

Parameter Closure : Tree_{s_t} → Prop.

The explanations for each are given below:

- **Type Entity** represents nouns.
- **Type Event** represents actions or states expressed by verbs.
- **Type Glue_{np}** represents particles used to connect noun phrases.
- **Type Glue_s** represents conjunctions or particles used to connect sentences or clauses.
- **Type Tense** represents the syntactic tense of verbs.
- **Type Aspect** is a syntactic type used to represent the nature or progress of actions or states as expressed by the verb.

- **Constant** Nil_{np} represents a non-existent noun phrase when a verb is not associated with nominative, accusative, or dative cases.
- **Constructor** En applies to nouns.
- **Constructor** Np applies to noun phrases.
- **Constructor** Adn applies when modifying nouns.
- **Constructor** Ev generates a term representing an event consisting of a single verb. It takes as arguments terms of type Tree_{np} representing noun phrases marked as nominative, accusative, and dative cases. The fourth argument is another term of type Tree_{np} representing complements linked by a copula.
- **Constructor** Adv_{np} applies in various contexts, such as when a noun phrase is used adverbially to modify the entire sentence or is inserted in a specific position within the sentence.
- **Constructor** Adv_{s} is used to introduce adverbial elements that modify the entire sentence.
- **Constructor** Hyp applies in conditional sentences.
- **Function** Closure takes instances of type Tree_{np} or Tree_{s} as arguments and returns them as propositions. The purpose of this function is solely to enable the use of Coq's proof assistance capabilities by converting data types into propositions.

For instances of constant list types, the function used to append a constant to a list is defined as follows:

```

Fixpoint append A : Type (l l' : list A) : list A :=
  match l with
  | nil => l'
  | cons a l => cons a (append l l')
  end.

```

This shows that lists with different concatenation orders can have the same meaning. Besides list Event , other cases where the order of concatenation does not affect meaning include list Glue_{np} for noun phrases, list Aspect and list Tense for auxiliaries, and list Glue_{s} for conjunctions.

After simplifying the list, instances of the data types are converted into tree structures. Nodes in the tree are represented as constructors, with the arguments of these constructors serving as child nodes. The positions of the child nodes correspond to the positions of the arguments.

Once a tree structure like the one shown in Figure 2 is obtained, subtrees representing events or states are extracted. Since the constructor Ev represents the minimal event, we extract information from subtrees with Ev as the root node. For example, by listing all the leaf nodes within the subtree, we can obtain a textual representation of the event. However, since Ev always includes a predicate, it is not possible to extract event nouns using the same method. Therefore, event nouns are extracted manually.

4 Rule Generation

4.1 Pattern Matching for Binary Relations

We will provide a detailed explanation of how to identify the relationship between two events or states using paths within a tree structure. This method involves identifying the shortest path from a source node (starting point) to a target node (endpoint) within the tree structure and mapping this path to the labels of a binary relation. The specific steps of this method are as follows:

1. Identification and Setup of Nodes

First, identify the nodes within the Tree_{np} or Tree_{s} tree structure that represent events or states. Set the source node (starting point) and target node (endpoint) for analysis.

2. Identification of the Lowest Common Ancestor

To find the shortest path from the source node to the target node, identify the Lowest Common Ancestor (LCA) of both nodes. The LCA is the most distant ancestor node from the root that is common to both the source and target nodes. This node will serve as the reference point for determining the path.

3. Construction of the Shortest Path

Once the LCA is identified, construct the shortest path by combining the path from the source node to the LCA and the path from the LCA to the target node.

4. Mapping to Binary Relation Labels

Map the obtained shortest path to the labels of a binary relation. This label clarifies which of the conditions, operations, or states defined by the functional requirements the source and target nodes correspond to. Possible labels include condition-operation, condition-state, and operation-state.

Additionally, *tregex* patterns are used to represent paths within the tree structure. *Tregex* [19] is a pattern language designed for efficiently performing pattern matching on tree structures. Each pattern specifies dominance or sibling relationships between nodes and is used to identify specific patterns within a tree structure. Basic node relationships supported by *tregex* patterns include:

- $A <_i B$: Node B is the i -th child of node A
- $A >_i B$: Node A is the i -th child of node B

By combining these patterns, it is possible to gain a detailed understanding of the relationships between nodes within the tree structure and to identify relationships using paths.

4.2 Annotations

To automatically extract information from functional requirements, it is necessary to prepare extraction rules in advance. We describe a method for creating effective extraction

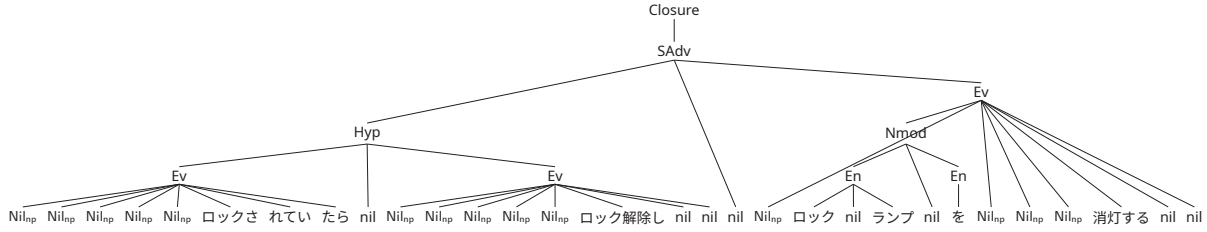


Figure 2: A $Tree_s$ type tree structure of “ロックされていたらロック解除し、ロックランプを消灯する”(If it is locked, unlock it and turn off the lock indicator light).

rules with minimal annotation, even for users without expertise in syntactic or semantic analysis. The system performs the following steps automatically:

1. Syntactic and Semantic Parsing

The system conducts syntactic and semantic parsing on the input data, generating a tree structure of type $Tree_{np}$ or $Tree_s$.

2. Extraction of Event and State Nodes

It extracts nodes corresponding to events or states from the tree structure.

3. Identification of the Shortest Path

It identifies the shortest path between the source node and the target node.

Following these steps, the user performs the following task:

- **Manual Annotation**

The user manually annotates the relationships between the source and target node pairs.

For this task, a simple tool is used that allows users to visually verify the output of common processing. The tool clearly indicates the source and target nodes for each data item and provides visual support, allowing the user to determine whether the items correspond to functional requirements. This method makes the annotation process efficient. After annotation, the assigned labels are mapped to regex patterns, and extraction rules are accumulated accordingly.

In this way, users are only involved in the annotation phase and do not need specialized knowledge, facilitating the automatic generation of extraction rules.

5 Experiments on Relation Identification

The validation method is explained as follows. The data is divided into training and test datasets. Tregex patterns created from the training data, combined with annotated labels, are used to accumulate extraction rules. Then, Tregex patterns generated from the test data are compared against these extraction rules to determine their applicability. The labels used for validation are *Preconditions—Actions* or *Preconditions—Resulting State*, and *Other*. The evaluation focuses on accurately classifying these binary labels.

The data consist of 69 test data and 117 training data. To avoid bias, careful attention was given to the data allocation. The functional requirements in [7] are divided into sections by functionality, with sentences within each section tending to have similar syntax. Allocating sentences with similar syntax from the same section to both the test and training datasets could lead to easy predictions using similar rules from the training data, thereby not accurately reflecting the system’s performance on truly unknown data. Therefore, test data were created from sections different from those used for training.

The rules generated from the training data total 92, which means that approximately 80% of the binary relations in the training data were used to create these rules. The generated rules were applied to the test data using two different methods, each with distinct characteristics:

1. Applying Rules Only When Paths Are Identical:

When paths are identical, the structure of the target data matches exactly, making the application of the rules expected to be accurate. This method ensures structural consistency in the data and helps avoid incorrect applications.

2. Applying the Rule with the Tree Edit Distance (TED) of Paths:

The TED of paths quantitatively indicates how different two patterns are. By applying the rule with the minimum TED, the transformation closest to the original data is achieved, leading to a more natural application. The APTED algorithm [20] is used for calculating the TED of paths.

The results of applying each method are shown in Table 1.

Table 1: Comparison of Accuracy Metrics for Different Rule Application Methods.

	f1_score	precision	recall
rule only	0.53	0.70	0.57
rule + TED	0.80	0.81	0.80

6 Conclusions

This study demonstrated the effectiveness of a rule-based approach using syntactic and semantic analysis to extract

functional requirements from Japanese specification documents. Notably, the method of minimum TED achieved the highest accuracy, confirming the importance of structural similarity.

Experiments using the functional requirements of “Wadai-Futto pot” Ver. 7 showed that the generated rules could accurately identify functional requirements with about 80% precision. This result indicates that even complex rules can be effectively applied, supporting the scalability of our approach.

Future work includes extracting functional requirements that span multiple sentences, improving the accuracy of events/states extraction, and ensuring that no information is lost from the sentences.

Acknowledgment

This research is being partially conducted as Grant-in-Aid for Scientific Research C (21K11826).

References

- [1] Bertrand Meyer. “On Formalism in Specifications”. In: *Program Verification: Fundamental Issues in Computer Science*. Springer Netherlands, 1993, pp. 155–189.
- [2] Kozo Okano et al. “Analysis of Specification in Japanese Using Natural Language Processing and Review Supporting with Speech Synthesis”. In: *IEICE Technical Report; IEICE Tech. Rep.* 117.465 (2018), pp. 79–84.
- [3] Kozo Okano et al. “Analysis of Specification in Japanese Using Natural Language Processing”. In: *Joint Conference on Knowledge-Based Software Engineering*. Springer, 2018, pp. 12–21.
- [4] Hiroya Ii, Kozo Okano, and Shinpei Ogata. “Improving Accuracy of Automatic Derivation of State Variables and Transitions from a Japanese Requirements Specification”. In: *Joint Conference on Knowledge-Based Software Engineering*. Springer, 2020, pp. 20–34.
- [5] Jannik Fischbach et al. “Towards Causality Extraction from Requirements”. In: *2020 IEEE 28th International Requirements Engineering Conference*. IEEE, 2020, pp. 388–393.
- [6] Jannik Fischbach et al. “Automatic Detection of Causality in Requirement Artifacts: The CiRA Approach”. In: *Proceedings of Requirements Engineering: Foundation for Software Quality: 27th International Working Conference*. Springer, 2021, pp. 19–36.
- [7] SESSAME(<http://www.sesame.jp>). *Wadai-Futto Pot (GOMA-1015 Type) 7th Edition*. 2005.
- [8] Gustavo Carvalho et al. “Model-Based Testing from Controlled Natural Language Requirements”. In: *Formal Techniques for Safety-Critical Systems*. Springer, 2014, pp. 19–35.
- [9] Gustavo Carvalho et al. “NAT2TESTSCR: Test Case Generation from Natural Language Requirements Based on SCR Specifications”. In: *Science of Computer Programming* 95 (2014), pp. 275–297.
- [10] Sérgio Barza et al. “Model Checking Requirements”. In: *Brazilian Symposium on Formal Methods*. Springer, 2016, pp. 217–234.
- [11] Masanosuke Ohto et al. “Proposal of Extracting State Variables and Values from Requirement Specifications in Japanese by Using Dependency Analysis”. In: *Procedia Computer Science*. Knowledge-Based and Intelligent Information & Engineering Systems: Proceedings of the 25th International Conference KES2021 192 (2021), pp. 1649–1658.
- [12] Julian Frattini et al. “Causality in Requirements Artifacts: Prevalence, Detection, and Impact”. In: *Requirements Engineering* 28 (2022), pp. 49–74.
- [13] Masashi Yoshikawa, Hiroshi Noji, and Yuji Matsumoto. “A* CCG Parsing with a Supertag and Dependency Factored Model”. In: *Journal of Natural Language Processing* 26.1 (2019), pp. 83–119.
- [14] Sumire Uematsu et al. “Integrating Multiple Dependency Corpora for Inducing Wide-Coverage Japanese CCG Resources”. In: *ACM Transactions on Asian and Low-Resource Language Information Processing* 14.1 (2015), pp. 1–24.
- [15] Pascual Martínez-Gómez et al. “Ccg2lambda: A Compositional Semantics System”. In: *Proceedings of ACL-2016 System Demonstrations*. 2016, pp. 85–90.
- [16] Koji Mineshima et al. “Building Compositional Semantics and Higher-Order Inference System for a Wide-Coverage Japanese CCG Parser”. In: *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. 2016, pp. 2236–2242.
- [17] Maiko Onishi et al. “Reducing Syntactic Complexity for Information Extraction from Japanese Requirement Specifications”. In: *2022 29th Asia-Pacific Software Engineering Conference*. IEEE, 2022, pp. 387–396.
- [18] The Coq Development Team. *The Coq Proof Assistant: Reference Manual: Version 8.9.0*. INRIA, 2019.
- [19] Roger Levy and Galen Andrew. “Tregex and Tsurgeon: Tools for Querying and Manipulating Tree Data Structures.” In: *Proceedings of the 5th International Conference on Language Resources and Evaluation*. Genoa, 2006, pp. 2231–2234.
- [20] Mateusz Pawlik and Nikolaus Augsten. “Tree Edit Distance: Robust and Memory-Efficient”. In: *Information Systems* 56 (2016), pp. 157–173.

An Evaluation of the Effectiveness of Olfactory Stimulation Notifications for Motorcycle Riders

LE Thi Thao Van, LI Ao, and Tomoya KITANI

Department of Informatics, Shizuoka University, Japan
 {lth-van, li-a, t-kitani}@kitanilab.org

Abstract - In recent years, numerous driver assistance systems have been developed to present information to car drivers; however, there are relatively few systems and studies focused on motorcycle riders. Motorcycles present unique challenges and constraints, such as the inability to remove both hands from the controls, a narrow field of view due to helmets, and significant impacts from noise, vibration, and temperature changes. To address these issues, we propose a method that combines visual and olfactory information to alert riders about their surroundings and provide navigation information. The system utilizes three LEDs for visual alerts and a nozzle installed inside the rider's helmet to emit scents for olfactory stimulation. To evaluate the effectiveness of this olfactory information presentation, we conducted a simulation using a motorcycle simulator with participants wearing a full-face helmet. The evaluation focused on intuitive route guidance by measuring reaction times, missed detections, and false detections. The results showed a significant reduction in reaction time and missed detections when olfactory stimulation was present.

Keywords: Olfactory stimulation, motorcycle riders, information presentation, safety, reaction time

1 Introduction

1.1 Background

In recent years, various driver assistance systems that provide navigation and safety information have been developed. These include devices for visual information presentation, such as smartphones and head-up displays (HUDs) that overlay information within the driver's field of view, and auditory information devices like intercoms. Despite these advancements, a survey by the Japan Safe Driving Center revealed that nearly 90% of motorcycle riders desire enhanced navigation and safety information[1]. However, research on and development of effective information presentation systems tailored specifically for motorcycles are scant, and existing systems fail to fully meet these needs.

Motorcycles encounter unique challenges and limitations compared to four-wheeled vehicles. These include the inability to remove hands from the handlebars, restricted visibility due to helmets, and significant impacts from noise, vibration, and temperature changes. Typically, navigation information for both motorcycle and car drivers is conveyed primarily through visual means, such as dashboards or smartphone

holders. However, due to limited space for mounting devices on motorcycles, these devices are often positioned lower, complicating the rider's ability to view the information.

As illustrated in Fig. 1, the rider's field of view shrinks from the green frame to the yellow frame when the rider is wearing a helmet. It indicates that the information display device, marked by the red frame, is not easy to see for such riders. For smartphone information that falls outside this visible area[2], riders must look downward to check metrics or smartphones, which hampers the ease of accessing visual information. Conversely, in four-wheeled vehicles, the enclosed driving environment is quieter, which facilitates the reception of auditory information, such as route guidance, in addition to visual cues. Drivers in these vehicles also benefit from a wider field of view, enabling easier access to both auditory and visual information.

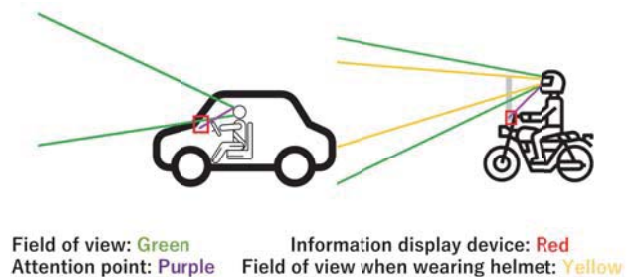


Figure 1: Driver's field of view

Riders must be vigilant of road surface conditions while also monitoring information ahead[3]. Sole reliance on visual information can lead to risks associated with frequent gaze shifts, adversely affecting driving behavior. For instance, even when riders attempt to absorb surrounding details and assess the environment, they might overlook critical information outside their direct front view, creating blind spots. This oversight can result in missed navigation cues, leading to altered travel plans and inducing feelings of anxiety or urgency. Such situations compromise cognitive processing, escalating inattention and stress[4]. Consequently, riders may engage in risk-taking behaviors like maintaining insufficient vehicle distances or speeding, which not only endangers themselves but also poses a threat to others.

According to the National Highway Traffic Safety Administration (NHTSA) from USA, motorcyclists in 2019 were 29 times more likely than passenger car occupants to die in a crash[5]. Most accidents occur because riders

fail to detect changes in their surroundings or cannot react swiftly. This highlights the critical need for more effective information alert systems that can enhance safety, awareness, and response capabilities of riders to road conditions.

1.2 Research Objective

This research aims to enhance the safety of motorcycle riders by investigating effective information presentation methods that minimize the need for gaze shifts away from the road. Given the hierarchy of human sensory perceptions, where olfactory information is considered third after visual and auditory information[6], this study explores the potential of olfactory stimuli to complement visual alerts more effectively. The primary objective is to determine whether the proposed system can improve riders' response times and reduce missed detections by comparing these metrics with and without olfactory stimuli.

2 Related work

2.1 Driver Information Systems

2.1.1 Methods of Information Presentation

Representative visual information presentation systems include navigation based on map and GPS data. Recently, to reduce the burden of driver's eye movements, not only traditional navigation systems but also HUD (Head-Up Display) research has been focused on. In both four-wheel and two-wheel vehicles, common visual navigation systems combine smartphones and voice guidance systems. Four-wheelers provide voice information through speakers, while two-wheelers often use intercom systems installed in helmets. However, motorcyclists often miss voice prompts due to wind noise and engine sounds, leading to poor effectiveness of voice guidance[7].

Additionally, research has been conducted on tactile information presentation, such as using smartwatches for vibration stimuli and HMD (Head-Mounted Display) for visual stimuli. Vibration stimuli can control the perception of time by adjusting intensity. However, in motorcycle operation, vehicle vibrations may confuse tactile stimuli, making them difficult to distinguish[8].

2.1.2 Timing of Information Presentation

Lunenfeld pointed out that presenting too much information too early can overload the driver's information processing, while too little information may cause drivers to ignore it[9].

Ito et al. measured the visual information processing time using HUD for motorcyclists to study the appropriate timing for information presentation. They found that for urban driving at 30 km/h, the appropriate distance for information presentation is 40 to 55 meters ahead[10].

2.1.3 Location of Information Presentation

Regarding HUDs, research found that the farther forward the display position, the more effective it is[11].

On the other hand, Matsunami et al. proposed a helmet-mounted near-distance LED information presentation system without a combiner. The results indicated that reducing the amount of information and displaying it at close range could shorten reaction times[12].

2.2 Olfactory Information Presentation

2.2.1 Indoor Environment Studies

Humans perceive external information through five senses in the following proportions: taste 1.0%, touch 2%, smell 4%, hearing 11.0%, and vision 82.0%[6]. In daily life, the time and situations where only one sense is used are limited. By combining information other than visual and providing cross-modal stimuli, people can use multiple senses to obtain physical information. This multimodal information processing is more common, especially during wakefulness[13].

Kawamoto et al.'s research indicated that lemon scent could reduce fatigue and prevent a decrease in vitality[14].

Studies by Aoki et al. found that jasmine scent affects the autonomic nervous system, increasing parasympathetic activity or reducing mental stress[15][16].

2.2.2 Automotive Environment Studies

Studies show that in a driving environment, scents can maintain driver alertness, enhance physical functions, and help recover from fatigue after driving[17]. However, the effects of scents vary with different types[18]. Research also suggests that low-intensity intermittent scent supply patterns can effectively maintain driver alertness, reduce fatigue, and enhance safety[19].

2.2.3 Effectiveness of Olfactory Cues

There are limited researches on olfactory cues in both four-wheel and two-wheel vehicles. In four-wheelers, studies suggest that scents can be effective as warnings or driving aids, but research in real driving environments is limited. For two-wheelers, the application and effectiveness of olfactory cues are even less studied. Although some studies indicate that scent cues can reduce driver distraction and improve reaction times[20], specific experimental evidence and application scenarios require further research and validation.

3 Proposed Method and System Design

3.1 Objective

The related work suggests that supplementing information presentation with tactile and auditory cues is possible. However, motorcycle operation is more challenging than four-

wheel vehicles for receiving visual, tactile, and auditory information but ensures good ventilation.

This study proposes an olfactory-visual information presentation system to reduce cognitive load during motorcycle driving. By comparing scenarios with and without olfactory stimuli, this system aims to determine if riders can grasp information faster and reduce omissions. The goal is to evaluate the effectiveness of olfactory cues in intuitive route guidance and measure the reaction time data to verify if such cues enhance reaction speed and reduce missed information.

3.2 Evaluation Items

To evaluate the effectiveness of olfactory information presentation, this paper proposes an olfactory-visual information presentation system. By providing intuitive route guidance, the evaluation focuses on the following four data points:

1. The time visual information is presented
2. The time odor is released
3. The time the rider notices the information
4. The number of missed detections

3.3 Information Presentation System Design

To ensure that the system can effectively convey information without interfering with the rider's simulated driving operation, we have reconsidered the core design principles from the perspective of component usage. The key principles and their implementation are as follows:

3.3.1 Integration and Responsiveness

Components: Micro-controller (M5Stack-Core-ESP32), Pressure Sensor (MS5803-14BA), Data Recorder (OpenLog), LED Module

- **Design Goal:** Ensure seamless integration of system components and high responsiveness, improving overall system reliability and efficiency.
- **Implementation:** Select components that can easily interconnect through standard communication protocols (e.g., I2C, SPI), ensuring the micro-controller can effectively coordinate the operation of all components. Additionally, use high-brightness LEDs and fast-reacting pressure sensors to ensure that visual and olfactory cues are triggered in the shortest possible time.
- **Component Selection Reason:** The micro-controller and pressure sensor communicate via I2C, enabling synchronized operations and data transfer. The LED module's high brightness and quick response time ensure that visual cues are immediately visible. The

pressure sensor can quickly detect the release of compressed air, ensuring that olfactory cues are synchronized with visual cues. The integrated data recorder logs the operation times of all components, ensuring data integrity and consistency.

Selecting these components not only achieves efficient data transmission and synchronized operation but also ensures system stability and reliability under various operating conditions. This integrated design allows the system to respond quickly in different environments, enhancing the timeliness and accuracy of information prompts.

3.3.2 User-Friendliness and Scalability

Components: LED Module, Silicon Tube and Aroma Box, Push Switch

- **Design Goal:** Design a simple, easy-to-operate user interface and ensure the system is scalable, allowing for upgrades and expansions as needed.
- **Implementation:** Use intuitive visual and olfactory cues to reduce the rider's learning curve and operational complexity. Ensure that the push switch is easily accessible, allowing riders to immediately react upon noticing the information. Additionally, choose modular components so that the system can easily add new functions or replace existing components.
- **Component Selection Reason:** The LED module and olfactory cue devices are designed to be straightforward, providing clear visual and olfactory signals. The push switch is mounted on the handlebar, making it convenient and easy to operate. The silicon tube and aroma box are used to convey and release scents, accurately delivering the scent to the rider's olfactory system, enhancing the effectiveness of the olfactory cues. The micro-controller has multiple I/O ports and communication interfaces, making it easy to connect additional sensors or actuators. The modular design of the LED and olfactory cue devices allows the system to be flexibly expanded as needed.

The user-friendliness goal is to make the system easy for riders to understand and operate, enhancing user experience. Selecting modular components allows for upgrades and expansions as needs change, improving flexibility and adaptability.

In the experiments, the Honda Tech RC5 model driving simulator was used. This simulator is widely utilized in motorcycle driving schools for teaching motorcycle driving licenses. It can realistically simulate motorcycle driving operations and environments, making it easier and more effective to conduct experiments in a safe setting.

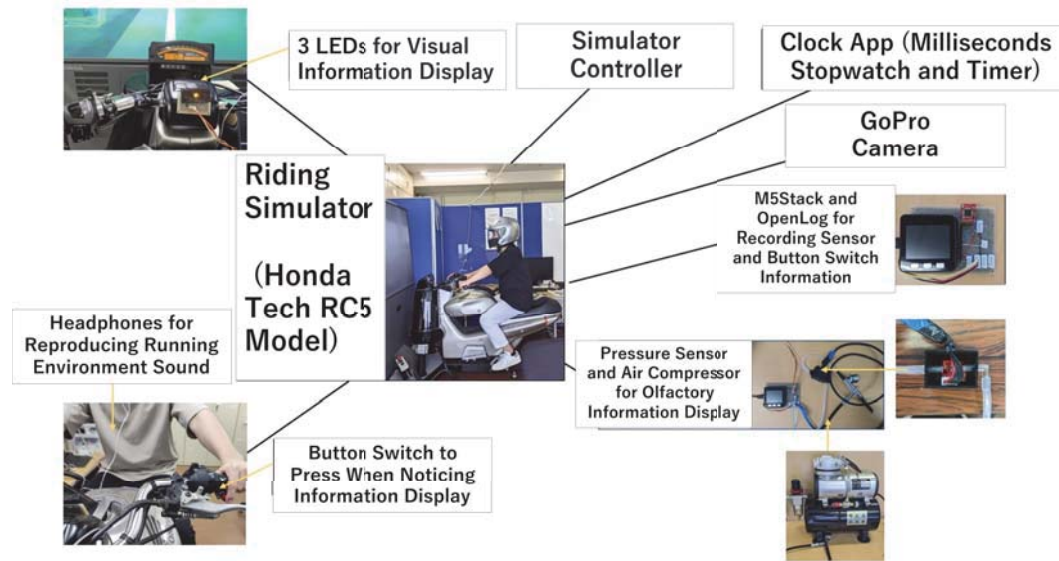


Figure 2: Overall View of Experimental Setup

3.4 System Operation

The system operates by first using the microcontroller to control the LEDs and scent emission. When a visual cue (LED) is presented, the microcontroller simultaneously logs the presentation time. For olfactory stimuli, a scent is emitted through a nozzle, and the pressure sensor detects the emission, logging the time of scent release. The rider's reaction time is measured by a push switch installed on the handlebar, which the rider presses upon noticing the information. The system is designed to ensure that data on reaction times and missed detections are accurately recorded for analysis.

The selection of components and their implementation were driven by the need to provide clear and intuitive information to the rider while minimizing the impact on motorcycle operation. The combination of visual and olfactory stimuli is expected to enhance the rider's ability to quickly and accurately perceive route guidance and safety information, thereby improving overall riding safety.

4 Preliminary Experiments

4.1 Differences in Reactions to Types of Odors

4.1.1 Purpose of the Experiment

The perception of odors varies from person to person. A scent that is pleasant for some may be unpleasant for others. Additionally, olfactory sensitivity may vary based on the individual's condition on a given day. Therefore, before conducting the main experiment, it is necessary to screen the subjects through preliminary experiments to investigate their preferences for different odors, their olfactory sensitivity, and the time it takes them to perceive the odors.

4.1.2 Experimental Method

Based on related research in Chapter 2, lemon and jasmine scents were chosen. The intensity of the odors was adjusted according to the Environmental Agency's 6-level odor intensity display method[21](Table1). To avoid discomfort, the odor intensities were set to levels 2 and 3, using NOW essential oils for lemon and jasmine.

Table 1: Method of Odor Intensity Display by Steps

Odor Intensity	Judgment Criteria
0	No odor
1	Barely detectable odor (detection threshold)
2	Weak odor, recognizable (recognition threshold)
3	Easily detectable odor
4	Strong odor
5	Very strong odor

Subjects wore VR headsets and watched motorcycle touring videos. Each odor stimulus was presented, and the pleasantness of the odor was measured using a 9-level pleasantness-unpleasantness scale. The time taken for subjects to notice the odor was recorded with a timer(Table2).

Table 2: Method of Odor Intensity Display by Steps

Pleasantness-Unpleasantness	Description
-4	Extremely unpleasant
-3	Very unpleasant
-2	Unpleasant
-1	Slightly unpleasant
0	Neutral
+1	Slightly pleasant
+2	Pleasant
+3	Very pleasant
+4	Extremely pleasant

4.1.3 Results

Two adult males without olfactory issues participated in the experiment. The results showed that subjects did not find lemon and jasmine scents unpleasant, and the time to perceive lemon was shorter than jasmine. Therefore, lemon and jasmine were chosen as the olfactory stimuli for the main experiment(Table3).

Table 3: Results

Scent	Perception Time (s)	Pleasantness-Unpleasantness
Lemon	3.42	+3
Lemon	4.51	+1
Jasmine	5.07	+1
Jasmine	6.11	+1

4.2 Impact of LED Position and Brightness on Noticeability

4.2.1 Purpose of the Experiment

The purpose of this experiment is to study the impact of LED position and brightness on riders' ability to notice visual cues. The goal is to determine the optimal LED installation position and brightness to improve the effectiveness of information presentation and response times.

4.2.2 Experimental Environment

The experiment was conducted using the "Honda Riding Simulator RC5" from Honda Motor Co., Ltd. to eliminate external disturbances and accurately assess the impact of LED position and brightness on information presentation.

4.2.3 Experimental Setup

The experimental setup included LED modules for visual information presentation, push switches for measuring reaction time, a small microcontroller module M5Stack, and a small data logger OpenLog. The LED modules were installed on the dashboard and near the handlebars, and the push switch was mounted on the left handlebar.

4.2.4 Experimental Content

The experiment tested the noticeability of visual cues by combining LEDs of different brightness and positions in four configurations. The specific configurations included:

1. High brightness LED on the dashboard
2. Low brightness LED on the dashboard
3. High brightness LED near the handlebars
4. Low brightness LED near the handlebars

Subjects engaged in simulated riding on the simulator, with LED signals randomly prompted under each configuration. The time taken for subjects to notice the cues and the number of missed detections were recorded(Fig. 3).



Figure 3: positional relationship of the LED in the simulator

4.2.5 Results

Three male university students with motorcycle licenses participated in the experiment. The results showed that LED position significantly affected noticeability, with higher miss rates for LEDs near the handlebars. LED brightness had some impact on visibility, but the position effect was more pronounced(Table4).

Table 4: Number of Misses by LED Position and Brightness

	On the Panel	Near the Handle
Bright	1 of 46 times	9 of 44 times
Dark	1 of 47 times	17 of 55 times

Under bright conditions, LEDs on the dashboard were more likely to be noticed compared to those near the handlebars, while under dark conditions, the miss rate for LEDs near the handlebars significantly increased. The experiment also found that high brightness LEDs were more likely to be noticed in all positions and conditions compared to low brightness LEDs.

5 Evaluation Experiments

5.1 Experiment Objectives

The primary objective of this experiment is to evaluate the effectiveness of olfactory stimuli in improving riders' reaction times and reducing missed detections. By comparing the reaction times and miss rates between scenarios with only visual cues and those with both visual and olfactory cues, we aim to verify whether olfactory cues significantly enhance the effectiveness of information presentation.

5.2 Experiment Details

5.2.1 Participants

Three male university students with motorcycle licenses participated in the experiment. Their ages ranged from 20 to 23 years, with an average age of 21.7 years (SD = 0.9). All participants had no visual, auditory, or olfactory impairments. Prior to the experiment, all participants signed informed consent forms and were briefed on the purpose and procedures of the experiment. All participants had previously taken part in a preliminary experiment to ensure familiarity with the experimental setup and procedures.

5.2.2 Equipment

The experimental setup included a motorcycle simulator, LED modules, an olfactory stimulation device (including a pressure sensor box), push switches for recording reaction times, an M5Stack microcontroller module and OpenLog data logger for controlling and recording timing data, as well as a GoPro camera and a clock application for recording the experimental process. All equipment was connected and installed as illustrated in Chapter 3.

5.2.3 Experimental Environment

The experiment was conducted in a well-controlled indoor environment to minimize the influence of external factors on the results. Participants sat in a properly adjusted position in front of the simulator, wore a full-face helmet, and received visual and olfactory cues from a screen at a fixed distance. To replicate the actual driving environment, participants wore headphones that played engine sounds of a moving motorcycle to eliminate any potential influence from the sounds of the experimental setup.

5.2.4 Experimental Procedure

The experiment was divided into two parts: the first part provided only visual cues, and the second part provided both visual and olfactory cues. In each part, participants rode on the simulator and responded to randomly appearing LED signals by pressing a button on the handlebar to record their reaction times. During the experiment, olfactory cues were released through nozzles inside the helmet, emitting specific scents synchronized with the corresponding visual cues.

5.3 Experiment Results

Results showed a significant reduction in miss rates and reaction times when both visual and olfactory cues were used, especially when LEDs were placed near the handlebars (Fig. 4)(Fig. 5).

Experimental data indicated that olfactory stimuli effectively enhanced riders' attention and response speed. With visual cues only, the average response time was 2.3 seconds,

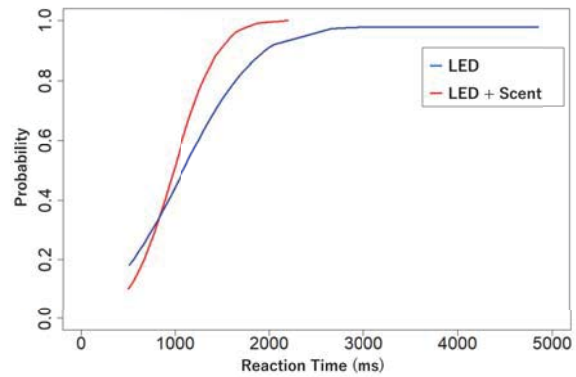


Figure 4: Cumulative Probability Until the Rider Notices the Information (LED on the Instrument Panel)

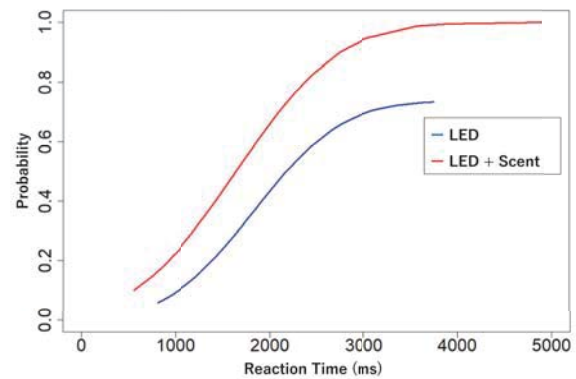


Figure 5: Cumulative Probability Until the Rider Notices the Information (LED Near the Handle)

and the miss rate was 15%. After adding olfactory cues, the average response time decreased to 1.5 seconds, and the miss rate dropped to 5%(Table5). These results suggest a significant advantage of olfactory stimuli in enhancing the effectiveness of information presentation.

Table 5: LED Position and Reaction Metrics

LED Position	No Scent		With Scent	
	On the Panel	Near the Handle	On the Panel	Near the Handle
Reaction Time [sec]	1.06 ± 0.63	1.84 ± 0.73	0.97 ± 0.38	1.64 ± 0.85
Miss Rate	2.20% (2/93)	26.3% (26/99)	0% (0/54)	0% (0/55)

6 Conclusion and Future Work

6.1 Conclusion

The study confirmed that olfactory stimuli can enhance information presentation for motorcycle riders, improving response times and reducing missed detections. Through experimental validation, this study's proposed information

presentation system combining olfactory and visual cues effectively enhanced riders' ability to perceive prompt information. Especially in complex driving environments, olfactory cues can serve as a powerful supplement, improving rider safety and driving experience.

Furthermore, the experimental results showed that while the change in riders' reaction times was minimal with olfactory cues, the miss rate significantly decreased. This indicates that olfactory stimuli can effectively complement visual cues, enhancing the effectiveness of information presentation. Specific experimental results showed that with the addition of olfactory cues, the average reaction time of riders decreased from 2.3 seconds to 1.5 seconds, and the miss rate dropped from 15% to 5%. These results suggest that olfactory stimuli have a significant advantage in enhancing the effectiveness of information presentation.

6.2 Future Work

Future research should include real-world riding experiments to validate the findings in actual riding conditions. Additionally, improvements in the system design, such as wireless communication and more compact equipment, are necessary for practical application.

Future studies could explore the effectiveness of olfactory cues under different weather and lighting conditions, and consider the response differences between various rider types (e.g., novices versus experienced riders). Additionally, combining other sensory stimuli (e.g., tactile or auditory) could further optimize the information presentation system.

Finally, integrating the system with existing motorcycle infrastructure, such as GPS and other navigation tools, would be a valuable area of exploration. This integration could provide seamless, real-time updates and alerts, making the riding experience safer and more enjoyable.

REFERENCES

- [1] Japan Automobile Safety Driving Center. Survey and Research on Information Provision for Motorcycles. 2006.
- [2] Ishibashi Motonori, Matsuo Juntaro, Shimada Takashi, Morishige Ryosuke, Tanaka Masahiro. Study on the Layout of Combiner-Type Head-Up Display Based on Visual Field Characteristics. *Human Factors*, Vol. 21, No. 1, pp. 16 – 23, 2016.
- [3] Miura Toshiaki. Visual Behavior in Driving Situations: Approach Through Eye Movement Measurement. *Bulletin of the Faculty of Human Sciences, Osaka University*, Vol. 5, pp. 253 – 289, 1979.
- [4] Ogawa. Emotional Characteristics of Drivers and Their Impact on Driving Behavior. Research Survey Project Report, 2009.
- [5] National Highway Traffic Safety Administration (NHTSA). Traffic Safety Facts. 2019.
- [6] Tsukada Yuki. Cognitive Support Technology for Drivers Regarding Visual Information and Its Future. Proceedings of the Traffic Safety and Environment Research Institute Forum, 2010, pp. 49 – 55, 2010.
- [7] Ito Kazuya. Study on the Effect of Reducing Driver Distraction through Voice Operation. PhD thesis, The University of Electro-Communications, 2017.
- [8] Shimizu Tomoyuki, Futami Kyosuke, Terada Tsutomu, Tsukamoto Masahiko. Information Presentation Method for Wearable Devices to Control User's Subjective Time. Proceedings of the Multimedia, Distributed Cooperation and Mobile Symposium 2016, Information Processing Society of Japan, Vol. 2016, pp. 162 – 169, 2016.
- [9] H. Lunenfeld. Human factor considerations of motorist navigation and information systems. In Conference Record of papers presented at the First Vehicle Navigation and Information Systems Conference (VNIS '89), pp. 35 – 42, 1989.
- [10] Ito Kenichiro, Nishimura Hidekazu, Kogi Tetsuro. Evaluation of Information Presentation Timing in Head-Up Display for Motorcycles. *Transactions of the Japan Society of Mechanical Engineers*, Vol. 83, No. 853, pp. 17 – 00113, 2017.
- [11] Okabayashi Shigeru, Furukawa Masamitsu, Hatada Toyohiko. Superiority of Visual Information Reception in Head-Up Displays for Automobiles. *Journal of the Institute of Television Engineers of Japan*, Vol. 47, No. 5, pp. 714 – 721, 1993.
- [12] Matsunami Takuya, Kitani Tomoya. Proposal of a Close-Range LED Information Presentation Method Inside Motorcycle Helmets. Proceedings of the Multimedia, Distributed Cooperation and Mobile Symposium 2020, Information Processing Society of Japan, Vol. 2020, pp. 1319 – 1328, 2020.
- [13] Okajima Katsunori. Cross-Modal Effects Induced by Visual Information. *Journal of the Institute of Image Information and Television Engineers*, Vol. 72, No. 1, pp. 8 – 11, 2018.
- [14] Kawamoto Rieko, Murase Chiharu, Ishihara Itsuko, Ikushima Miharuru, Nakatani Junko, Haraga Miki, Shimizu Jun. Effects of Lemon Scent on Simple Mental Tasks and the Mind and Body. *Journal of UOEH*, Vol. 27, No. 4, pp. 305 – 313, 2005.
- [15] Aoki Takashi, Adachi Yoshinori. Effects of Jasmine Scent on Heart Rate Variability (Research Presentation, 21st Life Information Science Symposium). *Journal of International Life Information Science*, Vol. 24, No. 1, pp. 121 – 125, 2006.
- [16] Aoki Takashi, Adachi Yoshinori. Changes in the Amplitude Ratio of Accelerated Pulse Waves Due to Jasmine Scent Stimulation. *Journal of International Life Information Science*, Vol. 25, No. 1, pp. 23 – 29, 2007.
- [17] Kawakami Mitsuyuki, Kaneko Wakako, Aoki Shinichi. Effects of Scent on Task Load in Car Driving. *Journal of*

- Japan Industrial Management Association, Vol. 51, No. 3, pp. 213 – 219, 2000.
- [18] Haga Shigeru, Masuda Takayuki. Physiological and Psychological Study on the Arousal Effect of Scents. *Ergonomics*, Vol. 44, No. Supplement, pp. 340 – 341, 2008.
- [19] Suzuki Keisuke, Sato Kei, Ota Koji. Optimization of Scent Supply Patterns for Low Olfactory Stimulation to Improve Driving Behavior. *Journal of the Human Interface Society*, Vol. 12, No. 3, pp. 313 – 322, 2010.
- [20] Nii Yoshinori, Suzuki Keisuke, Yamada Kiichi. Construction of an Evaluation Model for the Effect of Reducing Driver Distraction. *Transactions of the Society of Automotive Engineers of Japan*, Vol. 44, No. 2, pp. 549 – 554, 2013.
- [21] Ministry of the Environment. *Odor Index Regulation Guidelines*. 2001.

Competitive Programming Game to Promote Programming Motivation for Non-Programmers by Highlighting Programmer Coolness

Hiroki Oka[†], Ayumi Ohnishi[†], Takeshi Nishida[‡], Tsutomu Terada[†], and Masahiko Tsukamoto[†]

[†]Graduate School of Engineering, Kobe University, Japan

[‡]Graduate School of Intercultural Studies, Kobe University, Japan

{hiroki-oka@stu, ohnishi@eedept, tnishida@people, tsutomu@eedept, tuka@}.kobe-u.ac.jp

Abstract - The programmer's attitude of logically turning ideas into programs is cool. Seeing such "cool" programming motivates people to program. Making people feel that things are fun, cool, and attractive is effective for stimulating intrinsic motivation. However, it is difficult for people who have never experienced programming, such as elementary school students, to see how cool programming and programmers are. In this study, we conducted a questionnaire survey on the coolness of programmers and categorized the results. Based on this survey, we propose a system that motivates non-programmers to program by enabling them to experience the coolness of programmers. The proposed system is implemented as a competitive programming game that can be played even by non-programmers. It has a visual programming interface using blocks and a game screen that visualizes the correspondence between programs and execution results. We operated the proposed system as an interactive exhibit in the Kobe Youth Science Museum. Observing how users played the system revealed that many played without leaving until the end and seemed absorbed in the game. Some users showed interest in programming by playing the system, and the proposed system may function as an introduction to programming.

Keywords: Programming Learning, Motivation, Programming Game, Visual Programmig

1 INTRODUCTION

Programming is fascinating. Programming is powerful as a means of efficiency and problem-solving and a medium for self-expression. It is cool and admirable to see programmers expressing their thoughts in a programming language while working with a computer through trial and error.

However, it is difficult to convey the coolness of programmers to those who have no programming experience. Unlike games and sports, which are easy to enjoy from the beginning, it is difficult to feel the appeal of programming until one has mastered it. In addition, because there are few opportunities to see programmers programming, it is difficult for inexperienced programmers to visualize the act of programming, which is another reason why it is difficult to convey the coolness of programming to them.

We propose a programming game that highlights the coolness of programmers to convey the appeal of programming to inexperienced programmers and motivate

them to program. In this study, we conducted a questionnaire survey among experienced programmers on the coolness of programming and identified it as an element. Based on the survey result, we implemented the elements as a competitive programming game so that even non-programmers could experience the coolness of programming, and we operated the system as an interactive exhibition to investigate the users' impressions.

2 RELATED WORK

2.1 Environment for introductory programming

Visual programming languages are often used for introductory programming. Scratch [1], LEGO Mindstorms [2], and Viscuit [3] are examples of visual programming environments that have already been introduced in many educational settings. The input interface is easy for beginners to use. Another attractive point is that the execution results are easy to understand visually.

However, inexperienced programmers generally imagine programming using text-based languages, such as the programming effects they see in movies and TV dramas, so visual programming languages that diverge from text-based languages are not as cool as they seem.

2.2 Motivation for programming through game elements

Learning programming is difficult for beginners and inexperienced programmers [4]–[6], and it is essential to motivate them to maintain their motivation. Incorporating game elements is effective in motivating beginning learners [7], [8].

Mainly, programming game is often used in learning programming to motivate learners. Robocode [9] and Code Combat [10] are representative programming games. In a study of the Robocode community conducted by Ju, many users indicated that Robocode had improved their programming skills. In their study, many users indicated that Robocode had improved their programming skills and that they enjoyed finding algorithms and designing architectures [11]. Code Combat is also a programming game for learning in which the user completes a task by controlling a character using a programming language such as Python. However, these programming games are designed to motivate learners

with some programming experience. It is difficult for inexperienced programmers to feel the coolness of advanced programs.

Even if they are not programming games, many games incorporate programming elements. Algologic [12] is a problem-solving game that aims to cultivate logical thinking skills. Algologic is a learning game that can be easily used from a Web browser and has been implemented in educational settings. Human Resource Machine [13] and Automachef [14] are games that solve problems by arranging blocks with certain instructions and can be played on consumer game consoles.

Although these games make the logic part of programming into a game and make it easy to enjoy the fun of programming, they diverge from actual programming. They are not designed to experience the coolness of programmers.

3 PRELIMINARY INVESTIGATION

This study aims to implement a system that allows non-programmers to experience the coolness of programmers and to increase their motivation toward programming. At first, we conducted a questionnaire survey to enumerate and categorize programmers' coolness. We did a questionnaire survey to programmers because programmers' coolness can only be verbalized by people with programming experience. We conducted an open-ended questionnaire that asked, "What do you think is cool about programming and programmers?" The respondents were engineering students and faculty members (66 males and seven females) in their 10s and 60s. From the survey, we obtained 32 responses about the coolness of programmers. The first author subjectively categorized the obtained comments and categorized the programmer's coolness into four categories. Table 1 presents the categories and corresponding comments. We will discuss which category should be incorporated into our system in the following section.

E1. Sense of intelligence and thinking ability

The most common responses are related to the programmer's intelligence and logical thinking ability. For example, "Possessing knowledge and logical thinking skills." "Having logical thinking." The results showed that programmers regarded intellectual behavior in which they think logically as cool, such as "I can program in a logical manner."

E2. High implementation skill

The second most common responses are related to the programmer's implementation skills. For example, "Being able to write any code." "Being able to 'smoothly' write apps or algorithms you envision. It's not just about being able to do it, but doing it smoothly is what's impressive." The opinions that it is cool to be able to implement a system and to be able to implement it easily were obtained.

E3. Positive attitude toward computers

We also obtained several responses regarding the programmers' attitude toward programming. For example, "Maintaining a positive attitude even when encountering errors." In addition, there were also responses related to typing, such as "Typing difficult code quickly, like a hacker." The comments in this category indicate that programming in a positive manner, both physically and mentally, is considered cool.

E4. Can write clean code

Other responses referred to the code itself written by the programmers. For example, "Clean and readable code." "Being able to write clean and efficient programs."

4 PROPOSED SYSTEM

Based on the results obtained in Section 3, we implemented a programming game that allows even inexperienced programmers to experience the coolness of programmers.

4.1 Selection of categories

We obtained four categories of the programmers' coolness from the survey in Section 3. Some categories depend on the programming skills of the person observing the programming. For example, "E1. Sense of intelligence and thinking ability" and "E3. Positive attitude toward computers" are categories that do not depend on programming skills. Therefore, we decided to include them in the proposed system. On the other hand, the coolness of "E4. Can write clean code" cannot be achieved without a programmer who can write clean code and an observer who can evaluate its cleanliness. Therefore, E4 is excluded in this paper. The coolness of "E2. High implementation ability" also depends on the skills of the programmers who code the system. However, by simplifying the system's programming interface, lowering the barrier to coding, and limiting the number of instructions, we have made it possible for even inexperienced programmers to experience the system.

4.2 Design policy

Based on the results obtained from the survey in Section 3, the system was implemented based on the following design policy.

1. Programmers fight each other using logical thinking

It is necessary to be able to express algorithms in programs and solve given problems to visualize logic and thinking skills through programming. The proposed system is implemented as a one-on-one programming game in which programmers play against each other using characters with powerful logic. The programmer who is able to implement a strong algorithm wins. The game does not end once

Table 1: Questionnaire results on programmers' coolness

Categories	Responses
(1) Sense of intelligence and thinking ability	<ul style="list-style-type: none"> - Quick-wittedness. - Intelligent. - Intelligence. - Possessing knowledge and logical thinking skills. - When coming up with solutions on your own to problems without definite answers. - Being able to program logically. - Having logical thinking. - I think it is the foresight in logical reasoning. - It's impressive when you can logically explain why you wrote a particular piece of code. - Combining logic to accomplish difficult processes. - Being able to state answers with confidence. - Handling each task with a unique perspective.
(2) High implementation skill	<ul style="list-style-type: none"> - Being able to write any code. - Being able to build a system alone. - Running and operating creations with self-designed programming. - Precisely instructing a computer, a black box, to perform as intended. - Being able to create something immediately when requested. - Being able to "smoothly" write apps or algorithms you envision. It's not just about being able to do it, but doing it smoothly is what's impressive. - Being able to write long and complex code effortlessly. - It's impressive when you can accurately and quickly program actions described by others. - Building a system from scratch to a working stage. - Making it function exactly as intended.
(3) Positive attitude toward computers	<ul style="list-style-type: none"> - Typing in programs. - Maintaining a positive attitude even when encountering errors. - Steady progress. - The sight of typing code. - Typing programs using touch typing. - Typing difficult code quickly, like a hacker.
(4) Can write clean code	<ul style="list-style-type: none"> - Clean and readable code. - Being able to write efficient code in short sentences that even beginners can understand. - It's impressive to write code well. - Being able to write clean and efficient programs.

the program is created and executed. Programming and program execution are repeated until one of the programmers wins. This allows the programmers to experience E1 and E3 by showing the game to each other.

2. **Make it easy for inexperienced people to program their thoughts**

Programmers must be proficient in programming to show their coolness (E2) due to their high implementation ability, which is difficult for inexperienced programmers. Therefore, the proposed system employs a visual programming interface that can be easily programmed by simply tapping a small number of prepared blocks. When the program is executed, the block programs are converted into a text-based JavaScript program and displayed on the screen, giving the user and the programmer a sense of practical programming.

3. **Show the correspondence between the program and the execution result**

Even if a program can be easily implemented, the user or spectator cannot feel the coolness of the programmer if they cannot see what the program is doing during the game. Therefore, the proposed system displays the user-created program superimposed on the game screen during program execution and highlights the line of the program being executed to make it easier to understand the correspondence between the program and the execution result. The program runs slowly, one line per second, so that even if the user does not understand its meaning, it is easy for the user to understand the correspondence between the game result and the program that caused that result.



Figure 1: Game display

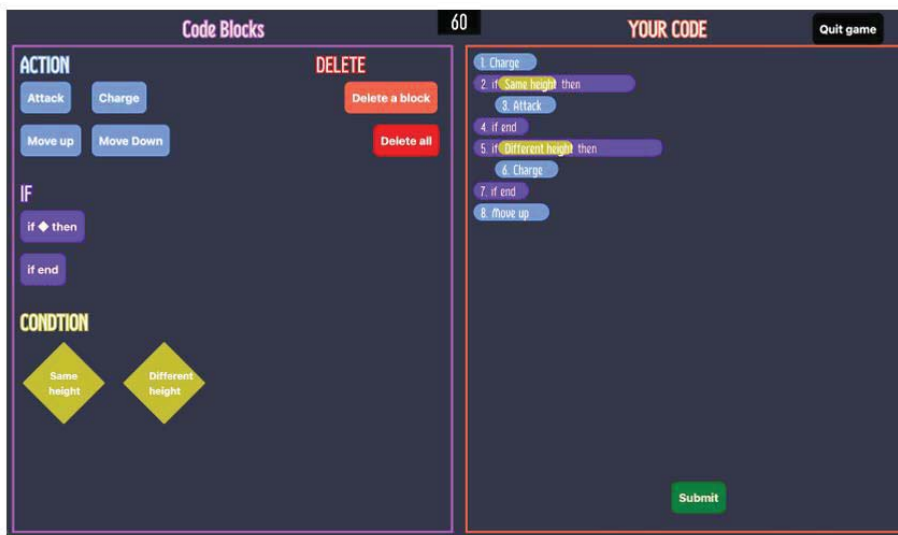


Figure 2: Input interface

4.3 System Overview

The proposed system is a competitive programming game consisting of a PC, a display for the game screen, and two touch displays for the input screen. This game is a shooting game played one-on-one. Figure 1 shows the game screen. Figure 2 shows the input interface. The player repeatedly programs and executes his or her character's actions and wins by attacking and defeating the opponent's character. Each character has attack power and hit points, and when attacking an opponent, the value of the attack power reduces the character's hit points (HP). The character can vertically move up and down three squares in the screen's upper, middle, and lower sections. During programming, the player creates a program by tapping the prepared instruction blocks on the touch display. The program can contain up to 15 lines of blocks. When executing a program, the created program is converted to JavaScript and executed in a loop for 20 seconds. There are three instruction blocks, each of

which is a different color: action blocks, if blocks, and conditional blocks. The functions of each block are as follows.

Action blocks

- **Attack:** The character fires a projectile to attack an opponent's character, dealing damage. Each time this method is used, the character's attack power is slightly reduced. When performed, the character shoots a straight horizontal ball.
- **Charge:** Increases the attack power of the character. When this method is used, an effect of a circle surrounding the character is generated.
- **Move up/down:** Moves the player's character up or down one square.

If blocks

- **If ◇ then:** This block corresponds to an if statement and can be used to create an if block by combining it with the conditional block.
- **If end:** Indicates the end of an if block with “if ◇ then,” equivalent to “}” in other programming languages like C.

Conditional blocks

- **Same height:** “If ◇ then” block. You can write a program that executes only if your character and the opponent’s character are in squares of the same height.
- **Different heights:** The opposite of the “Same height” block, this block allows you to write a program to be executed when your character and the opponent’s character are in squares of different heights.

4.4 System usage

The game is a competitive game in which two players program on the touch display and manipulate their characters to defeat an opponent. The game procedure is as follows.

First, each player creates a program to control a character on the touch display. After completing the program, the player submits the program by touching the “submit” button. The programs created by each player are sent to the game terminal, where the execution of each program begins.

Each player’s program runs for 20 seconds, one line per second. The game ends when either 20 seconds have elapsed or a character has attacked an opponent and reduced its HP to zero. This period is called a round.

The game ends when one character reduces the opponent’s HP to zero or when both characters are defeated. The player who defeats the opponent wins the game.

If the winner is not decided in one round, the game returns to programming with the character’s HP and position taken over. This is repeated until one of the characters defeats the opponent. The player with the most remaining character’s HP wins if the winner is not decided after three rounds.

4.5 Implementation

The proposed system is implemented as a web application using Node.js. p5.js is used for the client-side UI and game rendering, and Socket.io sends and receives data between the client and server.

5 DEPLOYMENT IN YOUTH SCIENCE MUSEUM

We operated the proposed system as an interactive exhibit at the Kobe Youth Science Museum and observed how people with no programming experience played. The system was exhibited in the “New Technology Experience Corner:



Figure 3: Appearance of interactive exhibits



Figure 4: Screen configuration for interactive exhibits

Let’s learn about the cutting edge of wearables and ubiquitous computing!” The exhibit is usually unmanned, but the first author spent a day in this section to observe how visitors experience the exhibit. Figures 3 and 4 show the exhibition and the screen configuration. The exhibit was often played by elementary and junior high school students, parents and children, and male-female pairs. Seven users were observed. The following describes the experience of each observed pair.

5.1 Observation

Pair1: Mother and daughter

There were many cases where parents and children experi-

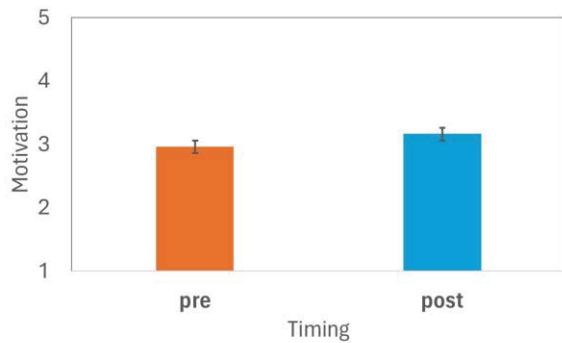


Figure 5: Questionnaire result of motivation

enced the exhibit together. This pair also experienced the exhibit with their parents and spent several minutes carefully reading the explanation of how to play the system. They also spent several minutes programming the game while playing against each other. During the game, both characters' attacks did not hit each other easily and the game dragged on for a long time. The children were seen boasting about the program to their parents, saying, "Look at (name)'s program!"

Pair2: Father and son

This pair also experienced the exhibition with their parents and child. They spent little time reading or programming how to play the system and quickly progressed through the game. In the game against the father, the father used "Charge block" and won a convincing victory in one round.

Pair3: Siblings

Not only parents and children but also elementary and junior high school students played against each other. In this case, players who appeared to be brothers and sisters were playing against each other, and the older brother was the same person as the child in Pair 2. The user commented that he was struggling with the conditional branching program, "How can I use 'if-end' block?"

Pair 4: Male and female in their 20s

The pair had read the system description thoroughly, and both had written long programs of about ten lines. The game was competitive, but the pair left the game at the end of the first round. This is thought to be because the pair mistakenly thought that the game would end after one round.

5.2 Questionnaire on motivation

We conducted a questionnaire survey before and after using the proposed system, asking users about their motivation for programming on a 5-point scale (1: not at all,

2: not much, 3: neutral, 4: fairly, 5: very much). We received responses from 25 users, and a comparison between the pre-and post-questionnaire is shown in the Figure 5. The average rating of the pre-questionnaire was 2.96, and the average rating of the post-questionnaire was 3.16. Comparing the average of the pre-and post-questionnaires on motivation, the value of the post-questionnaire was slightly higher. However, a within-subjects t-test showed that the difference in means was not significant ($t_{(24)} = 1.15, p = 0.25$).

6 DISCUSSION

From the observation of the interactive exhibition, as shown in Pair 1's comment, the proposed system may have allowed users to experience the fun of programming through the programming game. Because many users completed the program and proceeded to the end of the game, the proposed system may provide an opportunity for inexperienced programmers to experience programming and develop an interest in it.

Although many users seemed to enjoy the game, we were unable to objectively evaluate whether they could experience the programmer's coolness or whether they became more motivated to program. We continue our research and evaluate the proposed system's impact on motivation for programming and whether it makes users experience the coolness of programmers. In addition, we assume that users have little programming experience. However, we have yet to explicitly examine their programming experience, which needs to be clarified in future surveys.

As Pair 3's observations showed, implementing conditional branching in the system may be difficult for inexperienced programmers. This is because conditional branches are patterned after text-based programming languages and require a block to declare the end of a conditional branch. Therefore, we will simplify the conditional branching. Instead of describing the end of the block, the conditional branching will be implemented line by line.

7 CONCLUSION

In this study, we proposed a system that motivates non-programmers to program by enabling them to experience the coolness of programmers. We conducted a questionnaire survey on the coolness of programmers and categorized them into four categories. Based on the results, we implemented the system as a programming game for non-programmers and deployed it in the Kobe Youth Science Museum. Observing how users played the system revealed that many played without leaving until the end and seemed absorbed in the game. Some users showed interest in programming by playing the system, and the proposed system may function as an introduction to programming. In the future, we will conduct a workshop for inexperienced programmers to evaluate whether the proposed system allows them to experience the coolness of programming and its impact on their motivation for programming.

REFERENCES

- [1] M. Resnick, J. Maloney, A. Monroy-Hernández, N. Ruck, E. Eastmond, K. Brennan, A. Millner, E. Rosenbaum, J. Silver, B. Silverman, and Y. Kafai: Scratch: Programming for all, *Communications of the ACM*, Vol. 52, No. 1, pp. 60-67 (2000).
- [2] S. Papert: *Mindstorms: Children, Computers, and Powerful Ideas*, Basic books (2020).
- [3] Y. Harada and R. Potter: Fuzzy Rewriting: Soft Program Semantics for Children, *Proc. of the IEEE Symposium on Human Centric Computing Languages and Environments*, Vol. 2003, pp. 39–46 (2003).
- [4] S. R. M. Derus and A. Z. M. Ali: Difficulties in Learning Programming: Views of Students, *Proc. of the 1st International Conference on Current Issues in Education (ICCIE '12)*, pp. 74–78 (Sep. 2012).
- [5] E. Lahtinen, K. Ala-Mutka, and H. Jarvinen: A Study of the Difficulties of Novice Programmers, *Proc. of the 10th Annual SIGCSE Conference on Innovation and Technology in Computer Science Education (ITiCSE '05)*, pp. 14–18 (June 2005).
- [6] G. Bain and I. Barnes: Why is Programming So Hard to Learn?, *Proc. of the 2014 Conference on Innovation & Technology in Computerscience Education (ITiCSE '14)*, p. 356 (July 2014).
- [7] J. C. Paiva, J. P. Leal, and R. Queiros: Fostering Programming Practice through Games, *Information*, Vol. 11, No.498, pp. 1–20 (Oct. 2020).
- [8] M. Feldgen and O. Clua: Games As A Motivation For Freshman Students To Learning Programming, *Proc. of the 4th Annual Frontiers in Education*, pp. S1H-11–S1H-16 (Nov. 2004).
- [9] Robocode, <https://robocode.sourceforge.io/> (Accessed on 24 Apr. 2024).
- [10] Code Combat, <https://codecombat.com/> (Accessed on 24 Apr. 2024).
- [11] L. Ju: Just For Fun: Using Programming Games in Software Programming Training and Education, *Journal of Information Technology Education*, Vol. 6, pp. 279–290 (Jan. 2007).
- [12] Algologic, <https://algo.jeita.or.jp/> (Accessed on 24 Apr. 2024).
- [13] Human Resource Machine, https://store.steampowered.com/app/375820/Human_Resource_Machine/ (Accessed on 24 Apr. 2024).
- [14] Automachef, <https://store.steampowered.com/app/984800/Automachef/> (Accessed on 24 Apr. 2024).

Proposal for a Work Classification System in Cell Production Manufacturing Sites Utilizing Surveillance Cameras

Takao Inoue*, Kazuki Fukae**, Toru Kobayashi***

* Systec Inoue Corporation, Nagasaki, Japan

** School of Information and Data Sciences, Nagasaki University, Nagasaki, Japan

*** Graduate School of Engineering, Nagasaki University, Nagasaki, Japan
inouet@sys-inoue.co.jp

Abstract -In small and medium-sized manufacturing companies that rely on cell production methods, labor productivity tends to be lower compared to large companies that have implemented automation. Work in cell production involves a wide range of tasks, and the evaluation mechanisms for these tasks vary depending on the nature of each task. While previous research has explored methods using sensors and cameras, they are limited in their ability to extract features in cases where there is only one worker or when the focus is limited to a narrow range of specific tasks. Considering the complex layout of equipment and objects in manufacturing sites, the choice of focus becomes crucial when utilizing surveillance cameras. Therefore, we have focused on equipment, the orientation of workers, and the position of their hands, enabling the detection of combined features for task classification. Our system utilizes YOLO for object detection, predicting the positions of objects, and employs machine learning to extract task actions based on these coordinates. To validate the effectiveness of our approach, we conducted experiments targeting inspection tasks using the proposed system. This experiment demonstrated the efficacy of our method.

Keywords: Surveillance cameras, cell production methods, work classification, deep learning.

1 INTRODUCTION

In the realm of manufacturing, enhancing worker productivity and reducing waste are critical challenges. This is especially true given the issues such as skilled labor shortages and rising labor costs. The lean manufacturing method, which aims to eliminate waste and establish an efficient production system, is gaining significant attention. Specifically, measures such as reducing cycle times and minimizing inventory are being implemented.

In small and medium-sized manufacturing enterprises, production often operates through a cell production system centered around workers, encompassing a variety of tasks that are difficult to evaluate. Previous research has proposed systems that automatically extract movements during work using sensors and cameras. However, these systems are typically designed for single operators or specialized for specific tasks, which, in larger manufacturing settings, would necessitate numerous sensors and cameras, leading to high implementation costs and practical limitations.

We focused on leveraging surveillance cameras already installed in manufacturing sites for security and safety

monitoring to develop a low-cost system capable of task classification. This system has the potential to improve labor productivity not only in specific factories but also across various manufacturing sites in small and medium-sized enterprises.

In this study, we aim to achieve "task classification" by automatically analyzing worker movements from surveillance camera footage and categorizing specific tasks. Our proposed method combines simple object detection using YOLO (You Only Look Once) with machine learning. We conducted task classification based on information such as the orientation of workers, hand positions, and equipment locations in a sticker manufacturing process. The novelty of this system lies in its ability to utilize existing surveillance cameras, thereby reducing costs and making it applicable to large work environments.

Specifically, we focused on detecting the orientation of workers, the positions of their hands, and the locations of equipment, using these combinations as essential features for task classification. It is important to note that the method combining YOLO with machine learning requires retraining to accommodate the unique environment and conditions of each work site. This ensures that the classification model can be optimized according to the specific working conditions of each site.

This study aims to contribute to improving labor productivity and reducing waste in the manufacturing industry by proposing and demonstrating the effectiveness of a low-cost, versatile task classification system that leverages existing surveillance cameras. The core idea of this research is to develop a task classification system that can be applied to diverse work environments, thereby minimizing costs by utilizing already-installed surveillance cameras.

The structure of this study is as follows: Chapter 2 discusses related research, Chapter 3 provides a detailed explanation of the features of our system, Chapter 4 outlines the system configuration, Chapter 5 presents the results of the task evaluation experiment conducted on a sticker manufacturing line, Chapter 6 offers an analysis and discussion, and Chapter 7 concludes with a summary and future outlook.

2 RELATED RESEARCH

The analysis of human work behavior in the industrial context has undergone considerable evolution, transitioning from traditional sensor-based techniques to more advanced

image-based methodologies. This shift is exemplified in the broad categorization of behavior analysis techniques into sensor-based and image-based methods.

Sensor-based approaches historically involved the deployment of wearable devices to monitor worker movements. For instance, Kitazawa et al. [1] designed a system utilizing beacons and accelerometers to track worker positions and movements. This method, while precise, necessitated the attachment of sensors, potentially disrupting worker efficiency and increasing operational costs. To mitigate these issues, our method capitalizes on the infrastructure of pre-installed surveillance cameras, thereby obviating the need for direct sensor attachment on workers.

Another notable contribution in this area by Tao et al. [2] combined wearable sensors with camera data to monitor work motions. This hybrid approach required workers to wear sensor-equipped armbands, which again could interfere with worker comfort and inflate costs. Our methodology distinguishes itself by relying solely on existing surveillance cameras, further reducing the necessity for any wearable devices.

In the realm of camera-based approaches, Jan et al. [3] utilized multiple cameras to analyze work motions via skeletal tracking. This method, however, required an extensive setup of multiple cameras, escalating costs considerably. Our innovative approach simplifies this by using a single surveillance camera, minimizing setup complexity and cost. Moreover, Faccio et al. [4] explored the potential of 3D cameras in real-time monitoring of assembly tasks, tracking skeletal movements and actions within a defined area. While their system offered flexibility, the requirement for additional cameras in larger areas implied higher expenses.

To address these limitations, our study integrates the surveillance cameras with advanced object detection and machine learning techniques. We employ YOLO (You Only Look Once) [5], a robust object detection framework, to accurately predict the orientation and location of workers, their hand positions, and nearby equipment. YOLO's grid-based detection splits the image into cells, predicting bounding boxes and class probabilities in a single convolutional network pass. This capability, coupled with its high predictive performance and resistance to noise, positions it as a superior tool for capturing the dynamic elements of worker behavior in a complex industrial setting.

Furthermore, to classify tasks based on the data extracted from images, a Random Forest algorithm is implemented. This ensemble method, leveraging multiple decision trees, offers high predictive performance, low susceptibility to overfitting, and significant robustness against data noise. These attributes make the Random Forest algorithm an effective tool for reliably classifying the nuanced tasks workers perform, based on the visual cues provided by the surveillance footage.

In summary, by combining the effectiveness of YOLO for precise object detection and the robust analytical capabilities of the Random Forest algorithm, our approach offers a refined, scalable solution for monitoring and analyzing industrial work behavior without the drawbacks associated with sensor attachments or multi-camera setups.

3 FEATURES OF THE PROPOSED SYSTEM

3.1 Requirements

The objective of this system is to develop a mechanism for extracting characteristic tasks from video footage by analyzing the worker's orientation, hand positions, and equipment locations. By extracting these work movements, the system can visualize the interactions between workers and equipment. This visualization aids in determining whether the combination of workers and equipment is optimal or if additional equipment investment is necessary. Furthermore, it can clearly distinguish between value-adding and non-value-adding actions when workers are not engaged with equipment, thereby efficiently identifying areas for improvement.

The system must meet the following requirements:

Requirement 1: Accurately predict detected targets, such as the orientation of workers, hand positions, and the locations of equipment.

Requirement 2: Accurately classify specific tasks based on these positional relationships.

To extract work movements from video footage captured by surveillance cameras at manufacturing sites, the system must first accurately predict the worker's body orientation, hand positions, and the positions of objects such as equipment or workbenches (Requirement 1). Additionally, it must accurately classify these observations into several tasks based on their positional relationships (Requirement 2).

3.2 Use Cases

To illustrate one of the tasks in a cellular manufacturing system on a production line, consider the inspection and packaging operations using an inspection machine in a seal manufacturing line. The manufacturing process is depicted in Figure 1. The steps involved are as follows:

Material Setup: Load and set the products to be inspected into the equipment.

Product Check: Inspect the products for their appearance, dimensions, and other relevant criteria.

Equipment Confirmation: If the inspection machine detects any abnormalities during operation, investigate the cause of the stoppage.

Product Removal: Remove the approved products from the inspection machine.

Shrink Wrapping: Package the products in shrink film.

Product Packaging: Pack the shrink-wrapped products into cardboard boxes.

These tasks are categorized into six distinct types as listed in Table 1. Additionally, the specific detection points for the operator's orientation, hand position, equipment, and workstation during each step are described as follows:

Material Setup: The operator faces right, working at the rear input section of the equipment.

Product Check: The operator faces left, working at the table on the left side of the screen, with hands visible.

Equipment Confirmation: The operator faces right and is positioned in the middle of the equipment.

Product Removal: The operator faces right, working at the out of the equipment, with hands visible.

Shrink Wrapping: The operator faces left, working at the packing machine, with hands visible.

Product Packaging: The operator faces forward, working at the front table, with hands visible.

These observations are detailed in Figure 2 And Table2.

By focusing on these detection points, the system can accurately track and classify the tasks performed by workers, ensuring that both value-adding and non-value-adding activities are clearly identified. This enables efficient process optimization and potential improvements in productivity.

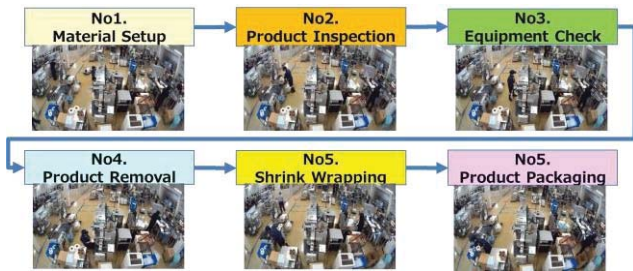


Figure 1 Flow of Manufacturing Process.

Table 1 Instructions for the Manufacturing Process.

No.	Task	Work Details
1	Material Setup	Set up materials
2	Product Inspection	Verify product information
3	Equipment Check	Check the condition of the equipment
4	Product Removal	Remove the product
5	Shrink Wrapping	Wrap the product with a packaging machine
6	Product Packaging	Pack the product into boxes

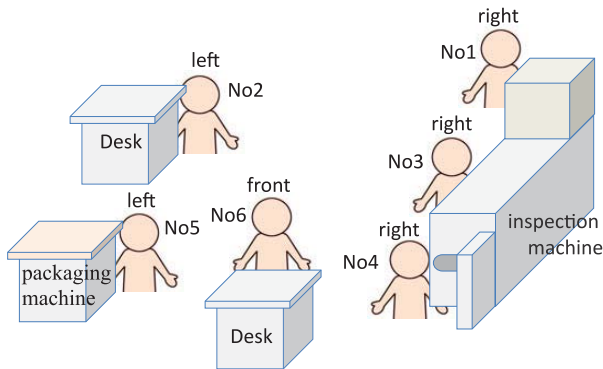


Figure 2 Operational Movements and Characteristics

3.3 The overall system processing flow and key points.

Figure 3 illustrates the flow of the proposed system. The process begins with surveillance cameras capturing video

footage of the work being performed. The subsequent steps are as follows:

Detection: Utilize the pre-trained YOLO model to detect the worker's orientation, hand position, and the locations of equipment or workbenches.

Classification: Apply a Random Forest algorithm to classify the work actions based on the detected elements.

The key point in this use case is the system's ability to identify specific work actions by analyzing three elements: the worker's orientation, hand position, and the locations of equipment or workbenches. By leveraging surveillance cameras, the system provides an efficient and cost-effective means to monitor and improve workplace efficiency.

This method not only reduces the need for additional hardware installations but also capitalizes on existing infrastructure to deliver actionable insights. These insights help in fine-tuning production processes, ultimately leading to optimized workflows and improved productivity.

Table 2 Work Procedures and Object Positions.

No.	Task	Worker Orientation	Equipment	Workbench/ Packaging Machine	Hand Position
1	Material Setup	facing right	input	-	-
2	Product Inspection	facing left	-	left side	visible
3	Equipment Check	facing right	middle	-	-
4	Product Removal	facing right	out	nearby	visible
5	Shrink Wrapping	facing left	-	nearby	visible
6	Product Packaging	facing front	-	nearby	visible

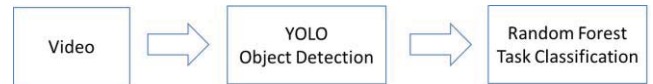


Figure 3 system flow.

3.4 Detection and model classification

In this study, we propose a method to classify tasks based on the orientation of workers, the position of their hands, and the location of equipment, as inferred from images. To achieve this, it is necessary to perform both object detection and action classification.

Requirement 1: Object Detection

Accurate prediction of worker orientation, hand positions, and equipment locations is vital, which is achieved using the YOLO (You Only Look Once) object detection method. YOLO employs a single convolutional neural network (CNN) to predict bounding boxes and confidence levels within a spatial grid. Key features of YOLO include grid-based detection for bounding box parameters, conditional probability predictions for object classes, and single-pass detection for high accuracy with multiple objects. These attributes make YOLO ideal for estimating the relevant positions and orientations needed for this study.

Requirement 2: Task Classification

The Random Forest algorithm is used for task classification due to its high predictive performance, reduced risk of overfitting, and robustness against data noise. This ensemble method combines multiple decision trees, enhancing classification accuracy and reliability.

By combining the high-precision object detection capabilities of YOLO with the robust classification performance of Random Forest, our method achieves practicality and efficiency in classifying tasks based on spatial relationships.

This approach enhances the practicality of the system, achieving high accuracy in both object detection and task classification.

4 SYSTEM CONFIGURATION

4.1 Hardware Configuration and System Environment

The hardware configuration of this system is illustrated in Figure 4. The hardware configuration of this system includes a Windows 10 operating system, an Intel Core i7 CPU, 16GB of memory, and a web camera with a 120-degree field of view. These components were selected to ensure sufficient computational resources for real-time video processing and machine learning algorithms. The application was developed using MATLAB R2020a along with the Deep Learning Toolbox, allowing for efficient algorithm implementation and easy integration of deep learning models. YOLO is utilized through MATLAB's yolov2ObjectDetector, while the ensemble learning method known as Random Forest is implemented using the TreeBagger function to improve predictive accuracy for task classification.

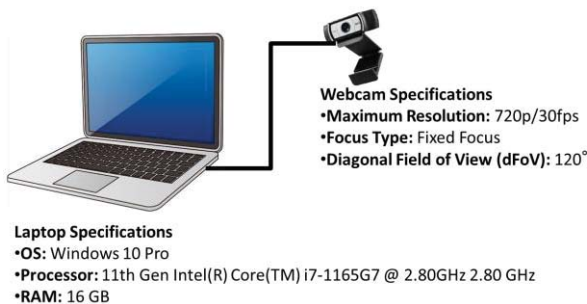


Figure 4 Hardware Configuration.

4.2 Shooting Environment

The shooting configuration of this system is illustrated in Figure 5. While general security surveillance cameras are typically installed at a height of around 3 meters, the camera in this case was installed at a height of 2.5 meters. This specific placement was chosen to mitigate the impact of ceiling lighting and to better capture the orientation and hand positions of the operators. The camera was

strategically installed at the rear of the equipment to provide a comprehensive view of the work area.

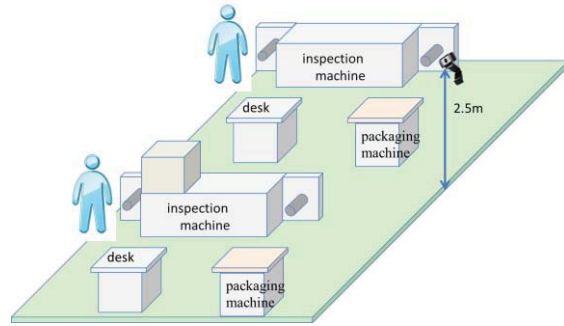


Figure 5 Photography Environment.

5 EVALUATION EXPERIMENT

5.1 Experimental Subjects

Using the inspection tasks performed by inspection machines on a sticker production line as the experimental environment, we evaluated the following two points:

(1) Object Detection

As shown in Figure 6, we evaluated the predictive accuracy of the system regarding the orientation of the operator, the position of their hands, and the positions of the equipment or workbench.

(2) Operational Movements for Inspection Tasks

We assessed the predictive accuracy of the system concerning the operational movements of the inspection tasks. This motion classification can be specified according to the desired operational movements but needs to be configured for each worksite. In this case, after consultation with the manager, a series of inspection tasks was classified into six types, as shown in Figure 7 and summarized in Table 3.

Based on this classification, we used the characteristics of the extracted operational movements to determine whether the system could accurately capture the orientation of the operator, the position of their hands, and their relative positions to the equipment and workbench. By evaluating these aspects, we aimed to verify the system's capability to provide reliable and actionable insights to improve the efficiency of task performance on the production line.

The results of this evaluation will help us to refine the system further, ensuring it meets the practical needs of different worksites and contributes to optimizing the manufacturing process.

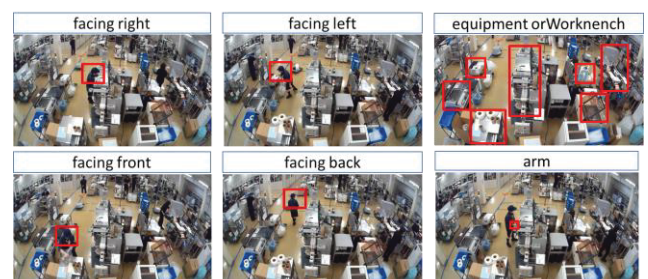


Figure 6 Object Detection Target.



Figure 7 Classification Task.

5.2 Experiment Preparation

(1) Data Preparation for Predicting the Orientation of the Operator, Hand Positions, and the Positions of Equipment or Workbench

To predict the orientation of the operator, hand positions, and the positions of the equipment or workbench from images, we used 1,600 annotated images as training data for YOLO. The camera used shot at 30 FPS, yielding a total of 45,000 frames from a 25-minute video. To obtain the 1,600 images, frames were extracted at 0.9-second intervals. The approximately 1,600 extracted still images were then labeled as shown in Table 4.

Table 3 Work Procedures and Object Positions.

No	Task	Worker Orientation	Equipment	Workbench/Packaging Machine	Hand Position
1	Material Setup	facing right	Back	-	-
2	Product Inspection	facing left	Back	-	-
3	Equipment Check	facing right	middle	-	-
4	Product Removal	facing right	foreground	nearby	visible
5	Shrink Wrapping	facing left	foreground	nearby	visible
6	Product Packaging	facing front	foreground	nearby	visible

Table 4 Labels for Object Detection.

object of detection		label
operator	facing right	Right
	facing left	Left
	facing front	Front
	facing back	Back
	arm	Arm
packaging machine	Open	Open
	close	Close
inspection equipment	foreground	Enter
	middle	Middle
	Back	Out
desk		Desk

(2) Annotation (Labeling)

Since retraining is required to address the specific environments and conditions of each worksite, we used

MATLAB's Image Labeler tool to create annotations, as shown in Figure 8.

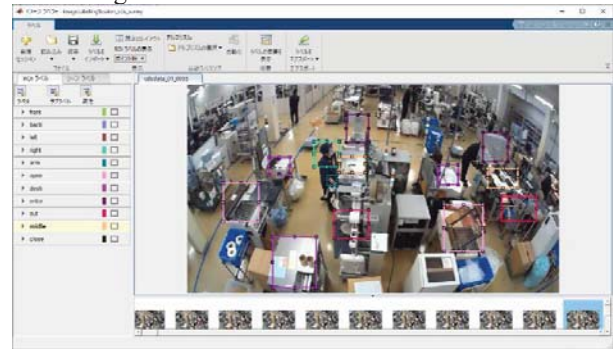


Figure 8 Annotation via Image Labeling Tool

(3) YOLO Model

Based on the pre-trained 25-layer YOLOv2 model, we made several modifications to adapt it to our specific requirements. These modifications are detailed below:

- Input Size:** The input size was adjusted to 256x256 to match the resolution of our annotated images.
- Anchor Boxes:** To better capture the various objects and their scales in the images, we configured the anchor box sizes to match those used during the annotation phase. The 10 anchor box sizes employed were as follows:
 - (15, 10) - (10, 15) - (25, 5) - (5, 25) - (30, 10)
 - (10, 30) - (50, 70) - (70, 50) - (25, 25) - (50, 50)
- Network Architecture:** We enhanced the original 25-layer YOLOv2 architecture by adding four additional convolutional layers to improve detection capabilities. The enhanced YOLO detection network, including the YOLO feature extraction layers, is illustrated in Figure 9.

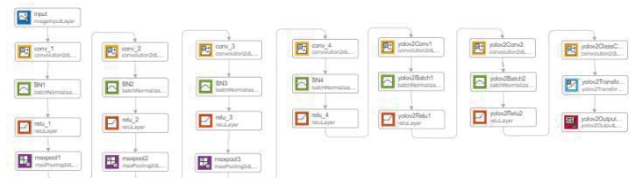


Figure 9 YoloV2 Network.

(4) YOLO Training

The following hyperparameter settings were chosen based on prior validation to ensure the training speed and gradient stability of the YOLO model:

- MiniBatchSize:** Set to 64 to manage the computational load and memory usage during training, allowing for stable gradient updates.
- InitialLearnRate:** Set to 0.01 to facilitate efficient learning at the initial stage. This helps the network quickly adapt to the initial data.
- LearnRateDropFactor:** Set to 0.1 to decrease the learning rate incrementally, thereby stabilizing the training process over time.
- LearnRateDropPeriod:** The learning rate was set to drop every 10 epochs to prevent overshooting and ensure gradual convergence.

• MaxEpochs: Set to 40 to provide a sufficient training period for the model to converge and learn the necessary features.

During training with these settings, the mini-batch Root Mean Square Error (RMSE) after 1680 iterations was 0.30. This indicated that the model had sufficiently learned to predict the required features with reasonable accuracy. The training results, including the loss curves and performance metrics, are shown in Figure 10.

These hyperparameter settings were carefully selected to balance between training efficiency and the model's ability to generalize. By monitoring the RMSE and making adjustments as needed, we ensured that the network could effectively learn the intricate patterns in the annotated data, resulting in a robust YOLO model tailored for our task.

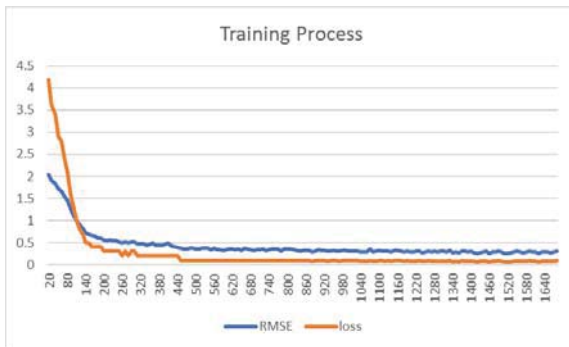


Figure 10 Learning Results.

(5) Data Preparation for Task Classification Based on the Orientation of the Operator, Hand Positions, and Positions of Equipment or Workbench

To predict the orientation of the operator, hand positions, and the positions of equipment or workbench from the labeled images, approximately 1,600 still images were annotated. For task classification purposes, tasks were categorized based on their specific actions and locations. In addition to identifying tasks on the left and right sides, actions that did not fit these categories were labeled as "NONE." The labeling for task classification was set as shown in Table 5.

Table 5 Labels for Work Classification.

workspace	work	label
left side	Material Setup	matinL
	Product Inspection	prodchkL
	Equipment Check	macstopL
	Product Removal	prodoutL
	Shrink Wrapping	shrinkL
	Product Packaging	boxedL
right side	Material Setup	matinR
	Product Inspection	prodchkR
	Equipment Check	macstopR
	Product Removal	prodoutR
	Shrink Wrapping	shrinkR
	Product Packaging	boxedR
	nothing	NONE

(6) TreeBagger Model

The classifier's performance was evaluated using decision trees, support vector machines, random forests, and k-nearest neighbors. Among them, random forests, which demonstrated high prediction accuracy, were adopted. This is shown in Table 6.

Focusing on the ensemble learning method of machine learning known as Random Forest, we used the MATLAB function TreeBagger, which combines multiple decision trees to achieve high predictive accuracy. This approach was chosen to evaluate the improvement in prediction performance for task classification.

Table 6 Classifier Evaluation.

Item	Evaluation
decision trees	86.2%
SVM	82.2%
Random forests	88.8%
KNN	84.1%

(7) Training with TreeBagger

For our hyperparameter settings, we configured the number of decision trees to 50, enabled the storage of out-of-bag information, and set the type of decision tree to "classification".

5.3 Experimental Results

(1) Predictive Accuracy of the Orientation of the Operator, Hand Positions, and Positions of Equipment or Workbench

The predictive accuracy for the operator's orientation, hand positions, and the target locations of the equipment or workbench is detailed in Table 7. We evaluated predictive accuracy using precision, which measures the proportion of correctly predicted positions among all predicted positions. Additionally, we assessed the scores using the Intersection over Union (IoU) ratio.

Focusing on instances with an IoU score of 55% or higher from 100 randomly selected still images from videos not included in the training dataset, the average predictive accuracy for the operator's orientation was 86.2%, while for hand positions, it was 94.9%.

Details of the false positives are presented in Table 8. The proportion of false positives reflects the percentage that each type of false positive constitutes within the total number of false positives. For the operator's orientation, 85.0% were undetected, and 15.0% incorrectly identified the operator as facing backward when they were facing left. Figure 11 illustrates the predicted positions.

These results demonstrate that the system can achieve high accuracy in object detection in actual worksite conditions and validate its effectiveness from a precision standpoint.

(2) Task Classification Accuracy Based on the Orientation of the Operator, Hand Positions, and Positions of Equipment or Workbench

The task classification accuracy, based on the operator's orientation, hand positions, and the positions of the equipment or workbench, is shown in Table 9. Classification accuracy is evaluated using precision, which indicates the proportion of correctly predicted positions among the total predicted positions.

Using the coordinate data of predicted label positions from the experimental results in Section 5.3(1), the model's average classification accuracy was 80.5% for tasks on the left side and 72.7% for tasks on the right side. When using the ground-truth coordinate data of predicted label positions from Section 5.3(1), the model's classification accuracy for the six types of tasks averaged 93.8% for the left side and 81.6% for the right side.

Details of the false positives are presented in Table 10. The proportion of false positives indicates the percentage that each type of false positive constitutes within the total number of false positives. For the left side of the work area, false detections accounted for 4.6% of the total false positives, while for the right side, they accounted for 7.1%.

These results demonstrate that the model can classify the operator's motions with high accuracy and validate its effectiveness from a precision standpoint.

Table 7 Location prediction accuracy.

object of detection		label	prediction accuracy
operator	facing right	Right	86.1%
	facing left	Left	94.8%
	facing front	Front	75.0%
	facing back	Back	88.9%
	arm	Arm	94.9%
packaging machine	Open	Open	99.0%
	close	Close	99.0%
inspection equipment	foreground	Enter	99.0%
	middle	Middle	99.0%
	Back	Out	99.0%
desk		Desk	99.0%

Table 8 Misjudgment of Position Prediction.

ground truth label	prediction label	rate of misjudgments
right	NONE(undetected)	70.0%
Left	NONE(undetected)	5.0%
Front	NONE(undetected)	10.0%
Left	Back	15.0%

Table 9 TreeBagger Results.

workspace	work	label	prediction accuracy
left side	Material Setup	matinL	100%
	Product Inspection	prodchkL	100%
	Equipment Check	macstopL	88.2%
	Product Removal	prodoutL	88.9%
	Shrink Wrapping	shrinkL	100%
	Product Packaging	boxedL	85.7%
right side	Material Setup	matinR	50%
	Product Inspection	prodchkR	100%
	Equipment Check	macstopR	89.5%
	Product Removal	prodoutR	100%
	Shrink Wrapping	shrinkR	100%
	Product Packaging	boxedR	50.0%
	nothing	NONE	87.9%

Table 10 Misjudgment of Task Classification Prediction.

workspace	work	ground truth label	prediction accuracy
left side	matinL	NONE	13.6%
	macstopL	NONE, prodchkL	13.6%
	prodoutL	NONE	4.5%
	boxedL	NONE	4.5%
right side	matinR	NONE	4.5%
	prodchkR	NONE	4.5%
	macstopR	prodoutR	9.1%
	prodoutR	NONE	4.5%
	boxedR	macstopR, NONE	9.1%
	NONE	boxedL,macstopR, matinR,prodchkL, prodchkR,prodoutL	31.8%



Figure 11 Location prediction Result.

6 DISCUSSION

In this evaluation experiment, we confirmed the feasibility of a mechanism to extract characteristic tasks from videos based on the operator's orientation, hand positions, and the positions of the equipment. Additionally, the evaluation experiment related to inspection tasks, conducted to verify practical applicability, revealed several challenges that need to be addressed for practical use. This discussion will consider the two requirements outlined in Section 3.1 based on the evaluation experiments presented in Chapter 5.

(1) Predicting Worker Orientation, Hand Positions, and Equipment Targets Accurately

In this study, we conducted experiments using YOLO (You Only Look Once), a simple object detection mechanism, to predict the arrangement of detection targets. Evaluation using 100 randomly selected still images showed that when the IoU score exceeded 55%, the average prediction accuracy was 86.2% for predicting the worker's orientation and 94.9% for predicting hand positions. This indicates that the accuracy is sufficient for predicting the positions of detection targets.

Regarding false detections, 85.0% of the errors were due to undetected cases. This mainly occurred when the worker's orientation could not be detected due to overlap with the equipment. Misidentifying left-facing orientations as

backward-facing accounted for 15% and typically occurred during transitions from backward-facing to left-facing. These issues could potentially be improved by using additional training and time-series data methods such as sliding windows.

(2) Accurate Classification of Tasks Based on Positional Relationships

Using TreeBagger, a type of machine learning algorithm, we conducted experiments to classify tasks based on detection targets. The average classification accuracy using the coordinate data of the predicted label positions from the 100 still images used in Experiment 5.3(1) was 80.5% on the left side and 72.7% on the right side. The classification accuracy using only the ground truth data model from Experiment 5.3(1) was 93.8% on the left side and 81.6% on the right side. This demonstrates the importance of prediction accuracy for objects.

The range of misclassifications in task categorization was from 4.5% to 31.8%, with errors occurring when coordinates were close, such as between "equipment checking" and "product checking" on the left side, and between "equipment checking" and "product retrieval" on the right side. These issues could potentially be improved through additional training and time-series data methods such as sliding windows.

In this study, we used surveillance cameras to focus on three key points: the worker's orientation, the hand positions, and the equipment positions. We confirmed that task classification is possible based on the combination of these factors. During inspection tasks on a seal manufacturing line, we extracted features of 11 types of objects and classified six types of tasks based on their positional relationships. The results demonstrated that task classification based on the worker's orientation, hand positions, and the positional relationships between equipment or workbenches is achievable.

Although this study focuses on inspection tasks using an inspection machine on a seal manufacturing line, the methods employed are expected to be applicable to identifiable tasks in other processes and different work locations as well. By using deep learning to predict object positions and machine learning to extract characteristic task movements, this approach suggests that similar methodologies can be effective in other manufacturing processes and various work environments.

7 CONCLUSION AND FUTURE CHALLENGES

In this study, we developed a system that classifies tasks based on the positional relationships of objects during work, using videos captured by surveillance cameras along with deep learning and machine learning technologies. As a case study, we demonstrated the effectiveness of this system through inspection and packaging tasks utilizing an inspection machine in a manufacturing setting. While machine learning was employed for task classification, it is possible to improve the accuracy of misdetections by

incorporating additional training and time-series data of work movements.

By utilizing this method, it is possible to classify tasks and measure time, thereby enabling a detailed analysis of the relationship between workers and equipment. Specifically, it allows us to determine whether the combination of a worker and equipment is optimal, whether there is a need for capital investment in equipment, and whether the worker is performing productive actions during the time they are not engaged with the equipment. This facilitates the efficient identification of areas for improvement and enables guidance to enhance worker productivity.

This method, which uses surveillance cameras to predict object positions through deep learning and extract characteristic work movements via machine learning, can be applied to a variety of tasks. However, it assumes that objects can be captured on video. Moving forward, there is a need to enhance its versatility to accommodate a broader range of workspaces and processes. Additionally, system improvements and expansions are required to enhance accuracy, real-time performance, and robustness against environmental changes. Further investigations are also needed to apply this method to different manufacturing settings.

REFERENCES

- [1] Kitazawa, M., Takahashi, S., Takahashi, B.T., et al.: Real Time Workers' Behavior Analyzing System for Productivity Measurement Using Wearable Sensor, , SICE Journal of Control, Measurement, and System Integration, Vol.10, No.6, pp.536–543 (2017).
- [2] Wenjin Tao, Ming C. Leu, Zhaozheng Yin . Multi-modal recognition of worker activity for human-centered intelligent manufacturing . Engineering Applications of Artificial Intelligence ,Volume 95, October 2020 .
- [3] ang Y, Jeong I, Younesi Heravi M, Sarkar S, Shin H, Ahn Y. Multi-Camera-Based Human Activity Recognition for Human–Robot Collaboration in Construction. Sensors. 2023; 23(15):6997. <https://doi.org/10.3390/s23156997>.
- [4] Faccio, M., Ferrari, E., Galizia, G.F., et al.: Real-time assistance to manual assembly through depth camera and visual feedback, Proc. 52nd CIRP Conference on Manufacturing Systems, pp.1254–1259 (2019).
- [5] Joseph Redmon, Santosh Divvala, Ross Girshick, You Only Look Once: Unified, Real-Time Object Detection, arXiv:1506.02640v5 [cs. CV] 9 May 2016.