IWIN2025

International Workshop on Informatics

Proceedings of
International Workshop on Informatics

August 31 - September 3, 2025
Akiu-onsen, Japan

Informatics Society

Sponsored by Informatics Society

IWIN2025

International Workshop on Informatics

Proceedings of
International Workshop on Informatics

August 31 - September 3, 2025
Akiu-onsen, Japan

Informatics Society

Sponsored by Informatics Society

Publication office:
Informatics Laboratory
3-41, Tsujimachi, Kitaku, Nagoya 462-0032, Japan
Publisher:
Tadanori Mizuno, President of Informatics Laboratory
ISBN:
978-4-902523-52-2

Printed in Japan

Table of Contents

Session 1: Health and Wellbeing Technologies ···· 1

(Chair: Katsuhiko Kaji) (8:45 - 10:25, Sep. 1)

(1) Development and Evaluation of a Personalized Health Support System Using Real-Time Data ···· 3

Masami Shinoda, Yuuto Takahashi, Ryosuke Takahashi, Haruto Iwase, and Kanae Matsui

(2) Design and Implementation of a Behavior Change Promotion System to Improve Skin Quality ···· 13

Ryosuke Takahashi, Hibiki Kaneko, Masami Shinoda, Haruto Iwase, and Kanae Matsui

(3) A Study on Sleep Behavior Improvement Using Wearable Device Data ···· 29

Haruto Iwase, Ryosuke Takahashi, Masami Shinoda, and Kanae Matsui

(4) FocusSense: Construction and Evaluation of a Multimodal Concentration Estimation Application ···· 41

Noriyuki Tanaka, Ko Watanabe, Shoya Ishimaru, Andreas Dengel, Shingo Ata, and Manato Fujimoto

Session 2: AI and Intelligent Systems ······················ 49

(Chair: Hayato Tomisu) (10:40 - 11:55, Sep. 1)

(5) A Method for Classifying Kumite Techniques and Detecting Preliminary Actions during Kumite Matches ······ 51
Kwangyun Kim, Shuhei Tsuchida, Tsutomu Terada, and Masahiko Tsukamoto

(6) Direction-of-Arrival Estimation of Impulsive Sounds Using the Spatial Audio Technology of Ambisonics ······ 59
Shingo Sawazaki, Jinhui Chen, and Takuya Yoshihiro

(7) Exception Oriented Test Case Generation and Evaluation using GitHub Copilot ······ 63
Sogo Yano, Kozo Okano, and Shinpei Ogata

Keynote Speech 1 ······ 71

(Chair: Tetsuya Yokotani) (13:00 - 13:50, Sep. 1)

(I) How to deal with evolving AI - The role of NTT's R&D of generative AI, IOWN and other advanced technologies - ········ No distribution of materials
Mr. Shingo Kinoshita

Session 3: User Behavior and Interaction Design ··· 73

(Chair: Kei Hiroi) (14:05 - 15:20, Sep. 1)

(8) A Cycling Route Recommendation System based on AI-Predicted Narrative Scores ··· 75
Hayato Tomisu, Shota Morita, Naoto Kai, and Tomoki Yoshihisa

(9) Designing Traces as Social Cues Support Visitor Exploration ··· 81
Ayaka Negishi, Hiroki Echigo, and Minoru Kobayashi

(10) Analysis of Gaze Direction and Turn-Taking in the Multitasking Videoconferences ··· 89
Taketo Imagawa, Atsuto Kurokochi, Koki Yanagii, Kazuyuki Iso, Masayuki Ihara, and Minoru Kobayashi

Session 4: Multimedia and Sensor Systems ··· 95

(Chair: Manato Fujimoto) (15:35 - 16:25, Sep. 1)

(11) Design and Control Olfactory Displays based on Diffusion and Flow of Aroma Gas ··· 97
Koichi Onuki, Motofumi Hattori, Yohei Seta, and Yuichi Bannai

(12) An Empirical Study of Broadcaster-following POV Heatmap based on MR technologies in 360-degree Internet Live Broadcasting ··· 103
Yoshia Saito and Junichiro Suto

Session 5: Smart Cities and Transportation ······· 111

(Chair: Masashi Saito) (8:30 - 10:10, Sep. 2)

(13) Real-Time Position Estimation of Non-Line-of-Sight Obstacles Using Reflected Images on Reflective Surfaces ······ 113
Baili Sheng and Yusuke Takatori

(14) Predicting Unexpected Traffic Congestion in Urban Scenarios Using Planned Route Information ······ 119
Shota Okazaki and Takuya Yoshihiro

(15) Road Congestion Prediction Method using Multimodal Machine Learning 123
Mikiko Sode

(16) Propose of Visit Promotion System Based on Visit Prediction and Common Topics ······ 127
Kohei Hayashi and Katsuhiko Kaji

Session 6: Human-Computer Interaction and Education ………………………………………… 137

(Chair: Kanae Matsui) (10:25 - 12:05, Sep. 2)

(17) Rhythm Rally: Entertainment Sport Combining Table Tennis and Rhythm Game ……………………………………………… 139
Koki Fuseya and Katsuhiko Kaji

(18) Empowering End-Users with Generative AI: A New Paradigm for Tool Development in Higher Education Portals ……………………… 149
Akihiro Hayashi, Michi Komura, Midori Ishihara, and Hajime Kaneko

(19) Challenges of a Self-Learning Support System Using Augmented Reality 157
Yuga Ito, Yuta Takenaka, and Shoji Sano

(20) Educational Question Insertion into Lecture Videos Using Large Language Models: An Initial Exploration of Chat GPT Prompt Design for Question Content and Insertion Points ……………………………… 165
Shuo Liu and Tomoo Inoue

Keynote Speech 2 ……………………………………… 171

(Chair: Shoji Sano) (13:00 - 13:50, Sep. 2)

(II) Advancing Smart Home Interoperability: Our Current Efforts with ECHONET Consortium……………………………………………… 173
Dr. Van Cu Pham

Session 7: Network and Security Technologies ·193

(Chair: Yoshitaka Nakamura) (14:05 - 15:45, Sep. 2)

(21) An Evaluation and Analysis of a Zero Trust-Based IoT Security Framework ··· 195
Hibiki Oizumi, Nobuhiro Kobayashi, and Ryozo Kiyohara

(22) A Proposal for LDoS Attack Method Using MPTCP Signal Manipulation ··· 197
Hiromichi Hagiwara, Hiroshi Inamura, and Shigemi Ishida

(23) Survey and Challenge on ICN for Industrial IoT Services ···················205
Atsuko Yokotani, Yuuki Hatanaka, Koichi Ishibashi, Tetsuya Yokotani, and Satoshi Ohzahata

(24) Anomaly Detection of Network Traffic Based on Frequency Analysis with Consideration of Time Scales ·· 211
Rintaro Okamoto and Koichi Ishibashi

Session 8: Computing in Society and Culture 215

(Chair: Takuya Yoshihiro) (16:00 - 16:50, Sep. 2)

(25) Automatic Translation from PlantUML description to NuSMV using LLM ··· 217
Kansei Inoue, Shinpei Ogata, Toshifusa Sekizawa, and Kozo Okano

(26) Emotional Parameters for Evacuation Agents via Live Earthquake Stream Analysis ··· 225
Kei Hiroi, Masami Shinoda, Kanae Matsui, Akihito Kohiga, Sho Fukaya, and Yoichi Shinoda

Message from the General Chairs

We are delighted to welcome all of you to Akiu, Japan, for the 19th International Workshop on Informatics (IWIN 2025). 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, Hokkaido in Japan, and Yanagawa in Japan respectively. This workshop aims to foster collaboration and education among students, alongside discussions on the latest research topics. In this workshop, we hope that students collaborate beyond their communities. Moreover, they will be motivated to support globalization by attending this workshop.

In IWIN 2025, 26 papers have been accepted after peer review by the program committee. Based on the papers, eight technical sessions have been organized in a single-track format, which highlighted the latest research results in areas such as Health and Wellbeing Technologies, AI and Intelligent Systems, User Behavior and Interaction Design, Network and Security Technologies, Computing in Society and Culture.IWIN2025 will also welcome two keynote speakers: Mr. Shingo Kinoshita, Senior Vice President of Research and Development Planning Department, NTT. Dr. Van Cu Pham, Research Assistant Professor, Center for Digitalization Endeavors, Japan Advanced Institute of Science and Technology. 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 collaborate with a large group of professionals worldwide, contributing to the workshop's great success. We are looking forward to seeing you all in the workshop. We hope you enjoy IWIN 2025.

September 2025

Tetsuya Yokotani
Shoji Sano

Organizing Committee

General Co-Chairs

Tetsuya Yokotani (Kanazawa Institute of Technology, Japan)
Shoji Sano (Kanazawa Institute of Technology, Japan)

Steering Committee

Hitoshi Aida (The University of Tokyo, Japan)
Toru Hasegawa (Shimane 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

Hironobu Abe (Tokyo Denki University, Japan)

Keiichi Abe
(Kanagawa Institute of Technology, Japan)

Masaki Endo (Polytechnic University, Japan)

Manato Fujimoto
(Osaka Metropolitan University, Japan)

Yusuke Gotoh (Okayama University, Japan)

Akihiro Hayashi
(Shizuoka Institute of Science and Technology, Japan)

Kei Hiroi (Kyoto University, Japan)

Hiroshi Inamura (Future University Hakodate, Japan)

Tomoo Inoue (University of Tsukuba, Japan)

Kazuyuki Iso
(Tokyo Information Design Professional University, Japan)

Katsuhiko Kaji (Aichi Institute of Technology, Japan)

Masaji Katagiri
(Professional University of Information and Management for Innovation, Japan)

Yoshinobu Kawabe
(Aichi Institute of Technology, Japan)

Tomoya Kitani (Shizuoka University, Japan)

Minoru Kobayashi (Meiji University, Japan)

Yuki Koizumi (The University of Osaka, Japan)

Kanae Matsui (Tokyo Denki University, Japan)

Masaki Nagata (Shizuoka University, Japan)

Ayano Naito (Aichi Institute of Technology, Japan)

Ayumi Ohnishi (Kobe University, Japan)

Kozo Okano (Shinshu University, Japan)

Yoshia Saito (Iwate Prefectural University, Japan)

Fumiaki Sato (Toho University, Japan)

Tetsuya Shigeyasu (Prefectural University of Hiroshima, Japan)

Masaaki Shirase (Future University Hakodate, Japan)

Mikiko Sode
(National Institute of Technology, Niihama College, Japan)

Hideyuki Takahashi
(Tohoku Gakuin University, Japan)

Yuichi Tokunaga
(Kanazawa Institute of Technology, Japan)

Akira Uchiyama (The University of Osaka, Japan)

Takaaki Umedu (Shiga University, Japan)

Takayasu Yamaguchi
(Akita Prefectural University, Japan)

Yasuyuki Yanagida (Meijo University, Japan)

Tomoyuki Yashiro
(Chiba Institute of Technology, Japan)

Takuya Yoshihiro (Wakayama University, Japan)

Tomoki Yoshihisa (Shiga University, Japan)

Shinichiro Mori (Chiba Institute of Technology, Japan)

Session 1:
Health and Wellbeing Technologies

(Chair: Katsuhiko Kaji)

Development and Evaluation of a Personalized Health Support System Using Real-Time Data

Masami Shinoda[†], Yuuto Takahashi[†], Ryosuke Takahashi[†], Haruto Iwase[†], and Kanae Matsui[†‡]

[†]Graduate School of System Design and Technology , Tokyo Denki University, Japan
[‡]Expolis Co., Ltd
{24amj18, 23amj16, 24amj24, 25amj05}@ms.dendai.ac.jp matsui@mail.dendai.ac.jp

Abstract - According to the 2023 statistical report by the Ministry of Health, Labour and Welfare, the total number of patients with lifestyle-related diseases remains at a high level. To address this issue, personalized lifestyle improvement initiatives utilizing Information and Communication Technology (ICT) have been proposed. This study aims to develop a system that leverages wearable devices and mobile terminals to tackle this challenge. The system is designed to provide personalized advice based on data collected from users to promote behavior change and address individual health concerns. To evaluate the effectiveness of the proposed system, an experiment was conducted with male and female participants in their early 20s. As a result, an increase in step count was observed in 50% of the participants, and the post-experiment questionnaire confirmed that the system effectively promoted behavioral changes in their daily lives.

Keywords: Wearable devices, real-time data, health behavior change, personalized support, mobile health, user experience.

1 Introduction

Japan's healthcare system is currently facing a multitude of challenges, including the increasing burden on physicians, regional disparities in medical resources, and the rising number of patients with chronic diseases due to population aging. According to the Ministry of Health, Labour and Welfare, the maldistribution of physicians remains a longstanding issue, with significant shortages in rural and underserved areas. Furthermore, the national healthcare expenditure reached 47.3 trillion yen in FY2023, and projections estimate a further increase to over 76 trillion yen by 2040, threatening the sustainability of the healthcare infrastructure.

In response to these issues, the Japanese government has been promoting the use of digital technologies, particularly the Personal Health Record (PHR) system. PHRs are expected to improve the quality of healthcare by enabling individuals to manage their health and medical data and share it across medical institutions. However, current systems often emphasize data collection and visualization, while insufficiently supporting users in actively changing their health behaviors.

To address this gap, a system that not only collects data but also provides personalized, timely advice based on a user's real-time health condition is essential. This requires accurate sensing, analysis of individual behavior and health status, and adaptive guidance tailored to each user's stage of behavioral change.

In this study, we propose a personalized health support system using wearable devices and mobile terminals to promote self-management and the prevention of lifestyle-related diseases. The system provides users with individualized feedback based on real-time data such as heart rate, step count, calorie consumption, goal setting, and mental state. Through dynamic, context-aware advice and intuitive visual feedback, the system aims to foster user engagement, raise health awareness, and promote sustainable behavioral change.

The objectives of this research are threefold:

1. To promote health awareness and behavioral change through visual feedback and personalized advice;
2. To support user-driven goal setting and continuous motivation for health behavior;
3. To provide optimized feedback based on individual physical and mental data, as well as behavioral change models.

To evaluate the effectiveness of the proposed system, a field experiment was conducted involving participants in their early twenties. This paper presents the design, implementation, and experimental evaluation of the system and discusses its potential for enhancing health promotion and reducing the burden on healthcare systems.

2 Related Work

2.1 Theoretical Frameworks of Behavioral Change

To support sustained health behavior change, several theoretical frameworks have been proposed and applied in both academic and practical health contexts. Among them, the Transtheoretical Model (TTM) and Social Cognitive Theory (SCT) have been particularly influential in the design of digital health systems.

2.1.1 Transtheoretical Model

The Transtheoretical Model, proposed by Prochaska et al. [9], conceptualizes behavioral change as a progression through five distinct stages: precontemplation, contemplation, preparation, action, and maintenance. Each stage corresponds to a different level of readiness for change, and appropriate interventions must be tailored to the user's current stage to be effective.

Cognitive and behavioral processes such as consciousness raising, self-reevaluation, reinforcement management, and stimulus control play vital roles in helping individuals advance through the stages. In Japan, the Ministry of Health, Labour and Welfare (MHLW) has adopted this model in public health strategies, such as exercise promotion guidelines [5].

2.1.2 Social Cognitive Theory

Social Cognitive Theory, proposed by Bandura [10], highlights the role of self-efficacy in health behavior change. Self-efficacy refers to the belief in one's ability to perform specific actions to achieve desired outcomes. The MHLW emphasizes two key approaches to enhancing self-efficacy:

- **Mastery experiences:** Encouraging individuals to set achievable goals and accumulate successful experiences, thereby reinforcing positive beliefs.
- **Vicarious experiences:** Allowing users to observe peers with similar backgrounds succeed, thus boosting their confidence.

These frameworks provide theoretical underpinnings for designing systems that offer stage-appropriate interventions and motivational support.

2.2 Wearable and Mobile Health Applications

2.2.1 Wearable Data Collection and Analysis

The growing availability of wearable devices has enabled continuous, non-invasive collection of real-time physiological data, including step count, heart rate, and energy expenditure [12]. These devices often integrate multiple sensors, such as accelerometers, gyroscopes, and GPS, allowing for accurate estimation of physical activity levels.

Studies have shown that combining heart rate and acceleration data improves the precision of exercise intensity classification. Compared with traditional methods such as the doubly labeled water method, wearable-based systems are more accessible and cost-effective while maintaining acceptable accuracy.

2.2.2 Mobile Health Apps and Coaching Support

Mobile applications for health behavior change have become widely used tools, offering users personalized feedback and goal tracking. Monteiro-Guerra et al. [13] reviewed 17 digital coaching systems and emphasized the importance of real-time personalization and the integration of behavioral change theories.

Moreover, usability and perceived relevance have been found to significantly impact the effectiveness of these systems [14]. Successful applications often provide educational content, reminders, and social reinforcement to guide users toward long-term behavior change.

2.3 Positioning of This Research

Based on the findings from prior studies, this research builds upon established behavioral theories (TTM and SCT) and incorporates both physical and mental health dimensions, as defined by WHO [15]. Our system is distinguished by:

- Providing personalized advice based on behavioral stages, mental status, and biosensor data;
- Supporting user-driven goal setting and progress monitoring;
- Delivering real-time feedback through commonly used platforms (e.g., LINE).

This study aims to develop a system that continuously supports health behavior change by combining personalized recommendations with visual analytics, validated through user-centered field experiments.

3 Proposed Method

3.1 Overview of the System

The proposed system is designed to support users in adopting and sustaining healthier behaviors through personalized feedback based on real-time data. It integrates wearable devices, mobile platforms, and a cloud-based processing pipeline to monitor physiological and behavioral data. The system aims to enhance physical and mental health by offering tailored advice based on individual conditions and behavioral stages.

Three core objectives guided the system's design:

1. **Enhancement of health awareness and behavioral support:** Providing intuitive graphs and motivational advice to improve health literacy and behavior.
2. **User-driven goal setting and progress tracking:** Enabling users to define personalized goals (e.g., daily steps) and visualize progress.
3. **Personalized feedback delivery:** Using behavioral models and real-time biosensor data to generate optimized, time-sensitive health guidance.

A conceptual diagram of the system architecture is shown in Fig. 1.

3.2 System Architecture

The system comprises three primary components:

- **Sensor Module:** A wearable device (Fitbit Sense 2) collects heart rate, step count, calorie consumption, and skin temperature data.
- **Processing Module:** Cloud infrastructure using Heroku, Node.js, Python, and MongoDB for data acquisition, analysis, and advice generation.
- **Presentation Module:** Health information, graphs, and advice are delivered to users via LINE messaging and notification APIs.

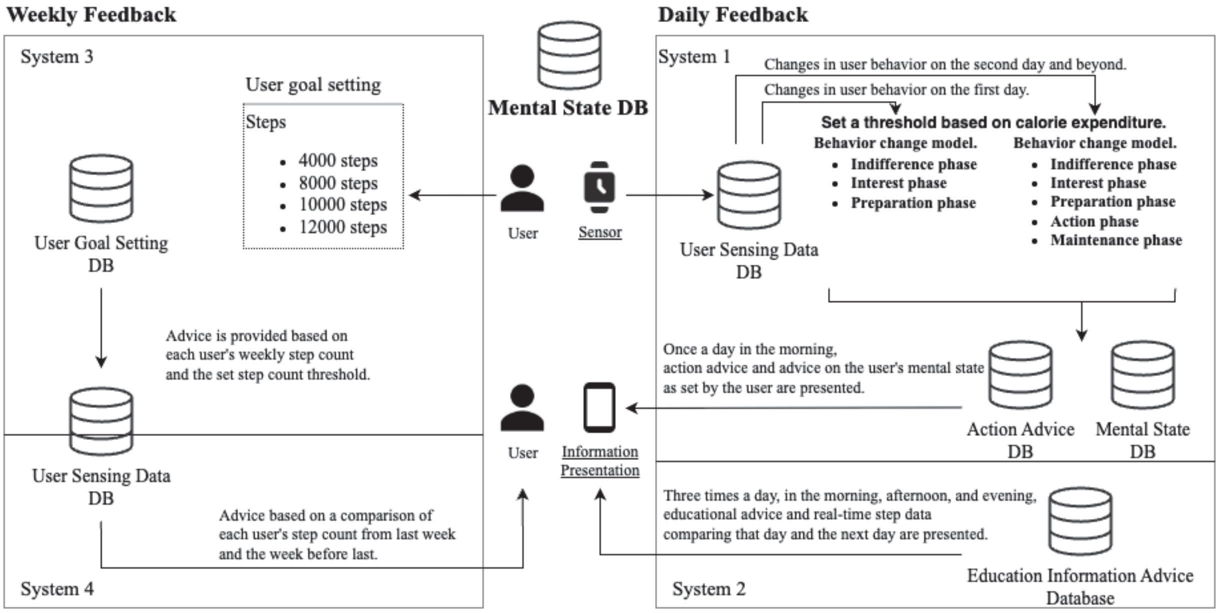

Weekly Feedback
System 3
User goal setting
Steps
4000 steps
8000 steps
10000 steps
12000 steps
User Goal Setting DB
Advice is provided based on each user's weekly step count and the set step count threshold.
User Sensing Data DB
Advice based on a comparison of each user's step count from last week and the week before last.
System 4
Mental State DB
User
Sensor
User
Information Presentation
Daily Feedback
System 1
Changes in user behavior on the second day and beyond.
Changes in user behavior on the first day.
Set a threshold based on calorie expenditure.
Behavior change model.
Indifference phase
Interest phase
Preparation phase
Behavior change model.
Indifference phase
Interest phase
Preparation phase
Action phase
Maintenance phase
User Sensing Data DB
Once a day in the morning, action advice and advice on the user's mental state as set by the user are presented.
Action Advice DB
Mental State DB
Three times a day, in the morning, afternoon, and evening, educational advice and real-time step data comparing that day and the next day are presented.
Education Information Advice Database
System 2

Figure 1: System Configuration Chart

3.2.1 Sensor Module

Fitbit Sense 2 was selected for its wide range of sensors and rich developer API. It supports continuous monitoring of physical activity using:

- 3-axis accelerometer and gyroscope;
- Heart rate monitor;
- Optical and electrical sensors (ECG, EDA);
- GPS and environmental sensors.

The data are synced to the Fitbit cloud and accessed using OAuth 2.0 via smartphone Bluetooth connections.

3.2.2 Processing Module

The processing module performs the following functions:

- **Data Acquisition:** Using Node.js and Python to fetch and store Fitbit data via APIs and web automation (Selenium).
- **Storage:** MongoDB stores real-time biometric data indexed by user ID. Google Sheets stores behavioral stage transitions and advice history.
- **Advice Generation:** Based on data trends, user-defined goals, and mental status, the system generates both *Behavioral Advice* and *Educational Advice*.
- **Stage Evaluation:** Behavioral stages are updated according to caloric thresholds and time series trends, following TTM logic.

3.2.3 Presentation Module

The output module consists of two communication modes:

- **Interactive Dialog:** Implemented via LINE Messaging API, allowing users to register information, set goals, and ask questions.
- **Scheduled Push Notifications:** Advice and visualizations (e.g., step graphs, goal progress) are pushed using LINE Notify API at scheduled times.

Users receive four types of graphs:

1. Weekly comparison of steps;
2. Daily goal achievement status;
3. Hourly step count and heart rate;
4. Hourly step count and total steps.

Each graph includes automatic interpretation and feedback.

3.3 Behavioral and Educational Advice Generation

The system delivers two types of advice daily:

- **Behavioral Advice (10:00 AM):** Based on current step count, BMI, caloric expenditure, and psychological status.
- **Educational Advice (8:00 AM, 12:00 PM, 7:00 PM):** Time-sensitive tips that reinforce self-efficacy and health literacy.

Advice is generated using rule-based templates populated with real-time data and personalized conditions, including behavioral stage and mental state (from 5-point surveys). The logic incorporates references from government health guidelines and WHO definitions.

3.4 Personalization Framework

Each user's advice is generated by matching conditions to a predefined advice database. For example:

- A user in the *Contemplation* stage with low step count receives motivational walking tips.
- A user with high stress (score 1–2) receives mental relaxation guidance, such as breathing exercises.

This personalization logic ensures that both physical and psychological dimensions are addressed based on evidence-backed models.

4 Implementation

This section details the technical implementation of the proposed health behavior change support system, focusing on data acquisition, processing logic, personalized feedback generation, and user interaction mechanisms. The system is built using cloud-based and mobile-friendly technologies to enable real-time feedback and continuous engagement.

4.1 System Infrastructure

The backend of the system is deployed on the Heroku platform, utilizing `Node.js` for web server operations and `Python` for data processing and advice generation. MongoDB Atlas serves as the primary database for storing biometric and behavioral data.

4.1.1 Data Acquisition

Sensor data are retrieved from the Fitbit Web API via OAuth 2.0 authentication. Heart rate, step count, and calorie consumption are fetched regularly and stored in MongoDB documents indexed by user ID and timestamp.

For example, Table 1 shows a sample of stored biometric data:

User-generated data such as goals and psychological assessments are stored in Google Sheets, linked by a common user identifier. The system also tracks behavioral stage transitions, determined by multi-day trends in calorie consumption, with thresholds set to 2400 kcal/day for males and 2200 kcal/day for females.

4.1.2 Data Flow

The system performs the following workflow:

1. Fetch biometric data via API (Fitbit + Selenium).
2. Evaluate data against thresholds (e.g., caloric baseline, daily goals).
3. Determine behavioral stage transition based on activity trends.
4. Generate advice using rule-based templates and context-aware data.
5. Push personalized messages and graphs via LINE APIs.

Google Apps Script is used to format incoming responses from users (e.g., Google Forms for advice feedback) and update spreadsheets accordingly.

4.2 User Interaction Design

4.2.1 Interactive Dialogue via LINE

Users communicate with the system through a LINE chat interface, which supports:

- Registration of user ID, height, and weight.
- Goal setting and updates (e.g., steps per day).
- Educational system explanations via rich menus and cards.

Information is displayed in a modular ”Box” format using the LINE Flex Message Simulator. This format helps convey system instructions, behavioral stage concepts, and health category breakdowns in an accessible manner.

4.2.2 Scheduled Notifications

The system uses the LINE Notify API to send personalized advice and graphs at fixed times:

- Behavioral Advice: 10:00 AM daily
- Educational Advice: 8:00 AM, 12:00 PM, 7:00 PM daily
- Graphs and Progress Reports: Weekly on Mondays, and daily summaries

Each message includes both text and images (graph snapshots), ensuring high engagement and readability.

4.3 Advice Generation Logic

Advice content is categorized as:

- **Behavioral Advice:** Based on step count, caloric trends, BMI, and psychological state (mental score: 1–5).
- **Educational Advice:** Based on time-of-day, behavioral stage, and health category (physical/mental). Includes both merits and risks (e.g., benefits of walking vs. risks of inactivity).

Templates are dynamically selected based on rule-based matching of user data to predefined conditions (e.g., “low steps + low motivation = recommend light activity with encouragement”).

4.4 Visualization and Feedback

The system generates four main types of graphs using matplotlib:

1. **Weekly Step Comparison:** Bar graphs comparing steps over two weeks.
2. **Goal Achievement Chart:** Daily step totals vs. target values.

Table 1: Sample Biometric Data Record

Date	Time	Heart Rate (bpm)	Step Count	Calories (kcal)
2024-11-12	13:00	82	218	16.51
2024-11-18	13:02	79	13	2.77

3. **Hourly Steps and Heart Rate:** Combined bar and line graphs.

4. **Hourly Steps and Total Steps:** Dual-axis display of activity level.

Each graph is accompanied by a tailored feedback message, e.g., encouragement if progress is evident, or a gentle reminder if step counts have dropped. These messages are based on predefined feedback rules and sent along with the graph images.

5 Experiment

To evaluate the effectiveness of the proposed health behavior change support system, a field experiment was conducted with university students over a four-week period. This section describes the experimental design, participant demographics, procedures, and data collection methods.

5.1 Objective

The goal of the experiment was to assess how personalized, real-time advice delivered via the system could influence users' health awareness, daily physical activity, and engagement with behavior change. Both quantitative and qualitative data were collected to evaluate system performance and user perception.

5.2 Participants

Ten students from Tokyo Denki University's Tokyo Senju campus participated in the experiment. All participants were in their early 20s. The group included 9 male and 1 female students. Participants were required to own a smartphone compatible with the Fitbit app and LINE messaging platform.

5.3 Procedure

The experiment was divided into two consecutive phases of two weeks each:

- **Period A (Baseline Phase):** Participants wore Fitbit Sense 2 devices while maintaining their usual lifestyle. No advice or feedback was provided during this phase.

- **Period B (Intervention Phase):** In addition to wearing the device, participants received personalized health advice, graphs, and motivational messages via LINE based on real-time data.

Table 2 summarizes the timeline of the experiment.

Table 2: Experiment Schedule

Phase	Duration
Period A (Baseline)	Nov 11, 2024 – Nov 24, 2024
Period B (Intervention)	Nov 25, 2024 – Dec 8, 2024

5.4 Data Collection

Multiple sources of data were collected during the experiment:

1. **Biometric Data:** Step count, heart rate, and calories were continuously recorded via Fitbit Sense 2 and stored in MongoDB.

2. **User Surveys:**
   - *Behavioral Advice Feedback:* Daily 3-question survey on perceived usefulness and motivation.
   - *Educational Advice Feedback:* Daily 2-question survey on perceived helpfulness.
   - *Pre-/Post-Program Surveys:* Conducted before and after the 4-week trial to assess changes in health awareness, behavior, and user experience.

Each advice message delivered during Period B included a Google Forms link for collecting feedback. Survey results were linked with behavioral data through anonymized user IDs.

5.5 Pre-/Post-Survey Design

The pre-program survey evaluated participants' baseline awareness and behavior across the following aspects:

- Perceived health status and interest in health;
- Daily habits: exercise, sleep, diet, stress level;
- Experience with health apps or programs.

The post-program survey focused on the following areas:

- Perceived behavior and awareness change;
- Evaluation of advice (behavioral and educational);
- Perceived usability, visibility, and satisfaction with system features;
- Overall willingness to continue using the system.

6 Results

This section presents the results of the four-week field experiment with ten participants. Both quantitative and qualitative evaluations were conducted to assess behavioral changes, the effectiveness of the advice system, and user experience.

6.1 Quantitative Evaluation

6.1.1 Step Count Analysis

Figure 2 shows a box plot comparison of average daily step counts during the baseline (Period A) and intervention (Period B) phases.

- In Period A, the median was 7,440 steps, and the mean was 7,835 steps.
- In Period B, the median increased to 8,268 steps, with a mean of 8,268 steps as well.

These results suggest a positive trend in physical activity during the intervention period.

6.1.2 Percentage Change in Step Count

Table 3 summarizes the percentage change in average daily step count between the two periods.

Half of the participants (5 out of 10) showed an increase in step count, with the largest improvement observed in Participant B (+69%). Others showed either a slight decrease or no significant change.

6.2 Qualitative Evaluation

6.2.1 Advice Usefulness (Likert Scale)

Behavioral and educational advice were rated on a 5-point Likert scale. Summary statistics are shown below:

- **Behavioral Advice Usefulness:** Mean = 3.89, SD = 0.54, $t(92) = 15.72$, $p < 0.001$
- **Educational Advice Usefulness:** Mean = 3.81, SD = 0.69, $t(371) = 22.63$, $p < 0.001$
- **Motivational Impact of Behavioral Advice:** Mean = 3.80, SD = 0.43, $t(92) = 18.13$, $p < 0.001$

All results were statistically significant and exceeded the hypothesized mean value of 3.0, indicating that the advice was perceived as both helpful and motivating.

6.2.2 Pre- and Post-Program Survey Results

Table 4 and 5 show changes in self-reported health behavior and awareness between pre- and post-intervention.

These findings indicate improvements in sleep, exercise, and stress tolerance. However, no improvement was observed in eating habit regularity.

6.2.3 User Feedback on System Features

Feature evaluation results showed the following:

- **Graph Features:** 100% positive response.
- **Behavioral Advice:** 90% positive response.
- **Educational Advice:** 70% positive, 20% negative.

Participants especially appreciated the visual nature of the graphs and their relevance to personal progress.

7 Discussion

This section discusses the implications of the experimental results, the effectiveness of the proposed system, and the potential for future enhancement in real-world health behavior change support.

7.1 Impact on Physical Activity

The experiment revealed an overall increase in physical activity during the intervention phase. The median and average step counts rose across participants, and 50% of users increased their step counts compared to the baseline period. These findings suggest that the personalized, real-time advice and visual feedback provided by the system encouraged participants to be more conscious of their physical behavior and take action.

Participants who set specific daily step goals (e.g., 8,000–9,000 steps) achieved greater improvements. This supports prior research indicating that goal setting can significantly enhance motivation and behavioral change.

7.2 Effectiveness of Feedback Mechanisms

The high usefulness scores for behavioral and educational advice indicate that daily feedback, when personalized, can reinforce motivation and improve self-efficacy. The positive results align with the behavioral change frameworks integrated into the system:

- **Transtheoretical Model (TTM):** The system guided users through behavior stages (e.g., contemplation → action) via step-specific advice and goal tracking.
- **Social Cognitive Theory (SCT):** Feedback encouraged self-monitoring, reflection, and incremental mastery, key factors in increasing self-efficacy.

Educational advice was most effective in the morning and evening, possibly due to users' higher receptiveness during these periods. However, some users found repetitive educational content less useful, suggesting a need for adaptive content delivery and variation.

7.3 Visualization and Motivation

Graph-based feedback was particularly well-received. The weekly comparison and goal achievement charts allowed users to visually confirm progress, which contributed to a sense of accomplishment. All participants gave positive feedback on the graphical features, with several noting that visualizations helped sustain motivation more than text alone.

This highlights the importance of intuitive and aesthetically effective data visualization in behavior change interventions.

7.4 Psychological Engagement and Mental State Support

By integrating a 5-point self-assessment of mental fatigue and matching advice accordingly, the system provided not only physical but also psychological support. Participants

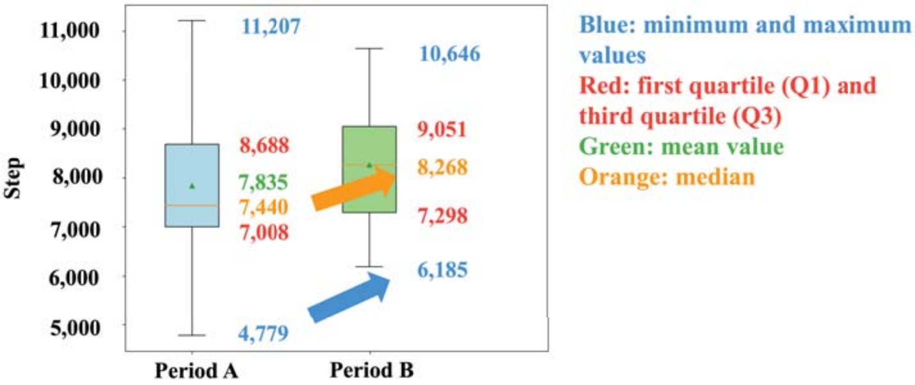


Figure 2: Comparison of Step Counts Between Period A and Period B

Table 3: Step Count Changes Between Period A and Period B

Participant	Period A (steps)	Period B (steps)	Change (%)
A	6,753	6,587	-2
B	4,779	8,038	+69
C	7,343	9,041	+23
D	8,515	9,061	+6
E	11,530	10,646	-8
F	7,004	6,185	-12
G	7,023	7,041	0
H	7,538	9,593	+27
I	8,746	7,556	-14
J	9,444	8,938	-5

with lower mental state scores received stress-reducing advice such as deep breathing, walking meditation, or sleep improvement tips. These were positively rated and may have contributed to improved stress tolerance (+13%) in post-survey results.

However, the static nature of the advice logic could be enhanced by introducing context-aware or AI-driven adjustments, such as tracking mood trends over time or responding to inactivity.

7.5 Limitations

While the system demonstrated effectiveness, several limitations remain:

- **Short duration:** A 4-week experiment may not capture long-term sustainability of behavior change.
- **Homogeneous population:** The participant group was limited to university students in their 20s. Generalizability to older adults or patients with chronic conditions requires further validation.
- **Advice personalization rules:** Current advice generation is rule-based. Advanced personalization using machine learning or adaptive dialogue systems may yield better outcomes.
- **Content fatigue:** Repetitive advice and rigid timing may reduce user engagement over time.

7.6 Implications for Real-World Applications

This study demonstrates the feasibility of combining real-time biometric sensing with personalized behavioral feedback via widely used platforms like LINE. The approach has the potential to:

- Reduce the burden on healthcare providers by promoting self-care;
- Serve as a preventive tool for lifestyle diseases;
- Be adapted to municipal-level health promotion programs.

Future deployment could incorporate seasonal trends, social support features, and integration with broader healthcare systems or PHR platforms.

8 Conclusion

In this study, we proposed and implemented a personalized health support system that utilizes real-time biometric data from wearable devices and mobile platforms to promote behavioral change and health awareness. The system provided customized feedback—including both behavioral and educational advice—tailored to each user's physical condition, psychological state, and stage of behavioral change.

Through a four-week experimental study involving ten university students, the system demonstrated measurable improve-

Table 4: Health Awareness Before and After the Program

Metric	Pre-Program	Post-Program	Change
Satisfaction with Health Status	3.45	4.20	+0.75
Interest in Health	3.60	4.40	+0.80

Table 5: Behavioral Indicators Before and After the Program

Item	Period A	Period B	Change (%)
Sleep Duration (hours)	6.4	6.8	+7.9
Exercise Frequency	2.9	3.5	+21
Regularity of Eating Habits	3.3	3.2	-3
Stress Sensitivity (inverse)	3.0	3.4	+13

ments in participants' step counts, health awareness, stress tolerance, and user satisfaction. Key findings include:

- A positive shift in physical activity levels for 50% of participants;
- High perceived usefulness and motivational effect of advice content;
- Improved awareness of health-related behaviors such as exercise, sleep, and stress management;
- Strong user engagement and satisfaction with visual feedback mechanisms.

The system's integration of theoretical frameworks such as the Transtheoretical Model (TTM) and Social Cognitive Theory (SCT) enabled it to deliver advice that was both evidence-based and contextually relevant. Real-time communication via the LINE messaging platform proved effective for timely and accessible health interventions.

While the system shows promise, limitations such as the short duration of the study, the homogeneous participant demographic, and rule-based personalization logic were identified. Future work will aim to address these limitations by:

- Expanding participant diversity across age groups and health conditions;
- Prolonging the intervention period to assess long-term behavior change;
- Introducing adaptive feedback generation using machine learning and conversational AI;
- Incorporating environmental and contextual data to further personalize health recommendations.

This research contributes to the growing field of digital health by demonstrating a practical, scalable method for delivering continuous, personalized health support. The proposed system has potential applications in public health promotion, workplace wellness programs, and preventive care initiatives led by local governments or healthcare providers.

REFERENCES

[1] Ministry of Health, Labour and Welfare, "Medical workforce distribution and recruitment capacity by region," 2023. (in Japanese)
[2] Ministry of Health, Labour and Welfare, "Overview of National Medical Care Expenditures, FY2023," 2023. (in Japanese)
[3] Ministry of Health, Labour and Welfare, "Projection of Future Medical Expenses Toward FY2040," 2023. (in Japanese)
[4] Ministry of Internal Affairs and Communications, "PHR infrastructure development and data utilization," 2023. (in Japanese)
[5] Ministry of Health, Labour and Welfare, "Exercise and Physical Activity Guide for Health Promotion 2023," 2023. (in Japanese)
[6] Ministry of Health, Labour and Welfare, "Health Japan 21 (Second Term)," 2021. (in Japanese)
[7] Ministry of Agriculture, Forestry and Fisheries, "Dietary Reference Intakes for Japanese (2020 Edition)," 2020. (in Japanese)
[8] J. Prochaska and C. DiClemente, "The Transtheoretical Model of Health Behavior Change," Am. J. Health Promot., vol. 12, no. 1, pp. 38–48, 1983.
[9] J. Prochaska and C. Velicer, "The transtheoretical model of health behavior change," Am. J. Health Promot., vol. 12, no. 1, pp. 38–48, 1997.
[10] A. Bandura, "Social Foundations of Thought and Action: A Social Cognitive Theory," Prentice Hall, 1986.
[11] Ministry of Health, Labour and Welfare, "Trends in wearable device-based health data collection," 2023. (in Japanese)
[12] S. Amagasa et al., "Evaluation of physical activity using smartphones and wearable devices in the medical and healthcare fields: current status and future prospects," Jpn. J. Public Health, vol. 68, no. 9, pp. 585–596, 2021. (in Japanese)
[13] F. Monteiro-Guerra et al., "Persuasive health behavior change in mobile health applications: A systematic review," J. Med. Internet Res., vol. 21, no. 6, pp. e117, 2019.
[14] T. Shimazaki et al., "Usability of mobile media in health behavior change," Japan J. Health Promotion, vol. 20, no. 1, pp. 33–41, 2021. (in Japanese)

[15] World Health Organization, “Constitution of the World Health Organization,” Basic Documents, Forty-fifth edition, Supplement, Oct. 2006.

[16] Fitbit, “Fitbit Sense 2 – Smartwatch Specifications,” [Online]. Available: https://www.fitbit.com/global/us/products/smartwatches/sense2. (accessed Jan. 21, 2025).

[17] Fitbit Developer Platform, “Fitbit API Reference,” [Online]. Available: https://dev.fitbit.com/build/reference/web-api/. (accessed June. 11, 2025).

[18] Heroku, “Heroku Platform,” [Online]. Available: https://www.heroku.com/. (accessed Jan. 21, 2025).

[19] The OAuth 2.0 Authorization Framework, RFC 6749, IETF, 2012. [Online]. Available: https://tools.ietf.org/html/rfc6749. (accessed June. 11, 2025).

[20] MongoDB Inc., “MongoDB: The Developer Data Platform,” [Online]. Available: https://www.mongodb.com/. (accessed June. 11, 2025).

[21] SeleniumHQ, “Selenium WebDriver,” [Online]. Available: https://www.selenium.dev/documentation/webdriver/. (accessed June. 11, 2025).

[22] Ministry of Agriculture, Forestry and Fisheries, “Physical Activity Standards for Promoting Behavioral Change,” 2020. (in Japanese)

[23] LINE Corporation, “Flex Message Simulator,” [Online]. Available: https://developers.line.biz/flex-simulator/. (accessed June. 11, 2025).

[24] K. Yamamoto and A. Sato, “The effects of pedometer-based walking programs on behavior change,” Jpn. J. Health Promotion, vol. 16, no. 2, pp. 75–82, 2013. (in Japanese)

[25] H. Suzuki et al., “A study on walking motivation through wearable devices,” Jpn. J. Sports Sci., vol. 21, no. 3, pp. 112–118, 2021. (in Japanese)

[26] Ministry of Health, Labour and Welfare, “Exercise and Physical Activity Guide for Health Promotion 2023,” 2023. (in Japanese)

[27] Ministry of Health, Labour and Welfare, “Dietary Reference Intakes for Japanese (2020 Edition),” 2020. (in Japanese)

[28] Ministry of Health, Labour and Welfare, “Sleep Guidelines for Health Promotion 2023,” 2023. (in Japanese)

Design and Implementation of a Behavior Change Promotion System to Improve Skin Quality

Ryosuke Takahashi†, Hibiki Kaneko†, Masami Shinoda†, Haruto Iwase†, and Kanae Matsui†‡

†Graduate School of System Design and Technology , Tokyo Denki University, Japan
‡Expolis Co., Ltd
{24amj24, 23amj06, 24amj18, 25amj05}@ms.dendai.ac.jp matsui@mail.dendai.ac.jp

Abstract - Recent advances in sensing technology have enabled the continuous collection of biometric data, fostering systems that provide personalized feedback for behavior change. While such systems have seen adoption in healthcare and lifestyle support, skincare management still relies heavily on dermatological consultations or commercial solutions. Considering that skin quality is sensitive to diurnal and environmental factors, timely and personalized care guidance is essential.

This study proposes a behavior change support system that delivers personalized skincare advice using sensor-derived skin moisture content and skin sebum content data. The system visualizes fluctuations in skin quality and provides three daily and one weekly message, offering tailored actions based on individual profiles to support habit formation and self-care.

We conducted a demonstration involving 27 participants (26 in their 20s and one in their 50s). Data collected from wearable sensors and self-reported questionnaires were used to evaluate the system's effectiveness in promoting skin improvement and behavior change.

Results showed that 55.56% of participants improved their skin quality, and all reported taking at least one recommended action. However, only 29.63% changed behaviors related to humidity and stress management, indicating the need for more context-specific feedback in future iterations.

Keywords: skin quality improvement, real-time data, behavior support, personalized care, sensor monitoring.

1 Introduction

This chapter describes the research background and objectives of this study.

1.1 Research Background

With advancements in sensing technology, systems that deliver personalized information and encourage behavior change based on real-time data from wearable devices and smartphones have gained attention. Such systems are increasingly adopted in healthcare and lifestyle improvement, where timely feedback and reminders facilitate effective and sustained behavior change[1].

In contrast, skin quality improvement still often depends on dermatological diagnosis and expert advice, despite the availability of daily skincare and professional treatment options. According to the Ministry of Health, Labour and Welfare, medical expenses related to dermatology reached approximately 333.2 billion yen in fiscal year 2022, indicating that many people are using dermatology services to improve their skin quality[2]. In addition, the "2024 Consumer White Paper" by the Consumer Affairs Agency reports an increase in complaints about cosmetics in cases involving damages over 10,000 yen, with 18,662 cases recorded from April to December 2023[3]. This suggests that self-directed approaches using commercially available cosmetics often fail to produce the desired effect, as it is difficult to find methods suited to individual skin conditions. Furthermore, since skin quality is easily affected by diurnal variations and environmental factors, daily monitoring is required[4].

Against this background, collecting skin quality data and providing individually optimized information based on it are essential for accurately capturing daily changes and implementing effective care.

1.2 Research Objective

This study aims to construct a behavior change support system that utilizes skin quality data obtained through sensor devices to provide improvement information tailored to individual skin types. Through this, the system seeks to provide an environment in which users can easily practice appropriate care for skin quality improvement in their daily lives, aiming for efficient management and continuous improvement of skin quality. Specifically, the research focuses on the following two objectives:

1. Visualization of skin quality and promotion of behavior change:

   Based on skin quality data collected by sensor devices, the system will present the information in an easily understandable visual format, enabling users to objectively recognize their skin quality. This is expected to raise user awareness and promote behavior change toward improving skin quality.

2. Provision of individually optimized improvement information:

   Based on the collected skin data, the system will propose care methods and improvement strategies suited to each user at the appropriate time, thereby achieving effective and efficient support for skin quality improvement.

To evaluate the achievement of these objectives, a demonstration experiment using the proposed system will be conducted to verify the user's skin quality improvement, system functionality, and the usefulness of the provided advice[5].

2 Related Work

2.1 Skin Quality Measurement Devices

A device used for measuring skin quality is the “Ipsalizer” by IPSA Co., Ltd[6]. This device quantifies skin moisture content and skin sebum content levels and is used to scientifically analyze skin quality. The measurement is conducted alongside interviews by skin specialists (receptionists), who propose optimal care methods based on each customer’s lifestyle and skin concerns.

During counseling, users remove makeup prior to measurement, and the results are used to provide detailed explanations of skin quality and recommended skincare methods. The session takes about 30 minutes, allowing users to gain a deep understanding of their skin state. Through this process, customers are able to discover the skincare items and methods best suited to their own skin.

Measurements with this device are conducted during in-store counseling, enabling customers to accurately grasp their skin quality and perform care accordingly. The use of such skin measurement devices plays an essential role in supporting individually optimized beauty recommendations.

2.2 Behavior Change Support Systems

This section reviews studies that promote behavior change by providing information based on sensing data.

2.2.1 Workstyle Improvement Support System

Tsuji et al.[7]developed and tested an application that collects workplace employee behavior data and provides personalized workstyle advice. In their study, behavior data was collected using wearable badge-type sensors. These sensors detect face-to-face interactions using infrared transmission and reception, measure body movement and rhythm via accelerometers, and record fine motor actions such as typing to evaluate desk work concentration.

The application delivers information in three stages. First, it visualizes the behavior logs to provide users an opportunity for self-reflection. Second, it conducts a type diagnosis based on behavior patterns to suggest workstyles that enhance organizational engagement. Third, it presents specific improvement suggestions based on both behavior logs and diagnosis results—such as “morning conversations contribute to increased engagement.”

This study demonstrated that the combination of badge-type sensors and the application is effective for improving workstyle, with potential contributions to voluntary behavior change and overall organizational performance. It validates the effectiveness of behavior change support using wearable devices and data analysis, offering a new approach to improving workplace environments.

2.2.2 Support System for Office Workers with Diabetes

Francesc et al.[8]studied the effectiveness of a mobile health (mHealth) intervention targeting office workers with type 2 diabetes. Data was collected using the Workforce Sitting Questionnaire (WSQ), the Spanish Brief Physical Activity Assessment Tool (SBPAAT), and the activPAL3TM device, gathering both subjective and objective data on sedentary behavior and physical activity. In addition, clinical data such as blood glucose and HbA1c levels were collected, along with mental well-being using the Warwick-Edinburgh Mental Well-being Scale (WEMWBS), and assessments of job performance and stress using the Work Limitations Questionnaire (WLQ) and Job Content Questionnaire (JCQ).

The information delivery system was implemented via a smartphone app and website, providing real-time feedback on sitting, standing, and walking times. It also supported behavior change through health information, goal setting, action planning, social support, and encouraging messages.

As a result, reductions in sedentary time and increases in physical activity were observed, along with improvements in HbA1c levels, mental health, work productivity, and stress management. The mHealth intervention proved effective in promoting behavior change and increasing physical activity in office workers with type 2 diabetes, with personalized feedback showing particularly strong effects.

This study highlights the importance of understanding users’ health conditions and behavior patterns in detail and providing real-time feedback. It demonstrates that multidimensional data collection and analysis are essential for developing effective intervention strategies.

2.3 Claims of This Study

Behavior change research utilizing sensing data has been conducted for a wide range of purposes, including workplace environment improvement and chronic disease management. These studies demonstrate that appropriately designed sensing systems and personalized feedback effectively promote user behavior change.

On the other hand, conventional skin measurement devices typically require expert counseling and in-store visits to perform measurements. In contrast, this study proposes a system that enables continuous skin monitoring within daily life. This approach allows individual users to understand their skin quality and take appropriate care without the need for regular store visits.

Furthermore, while traditional devices rely heavily on expert consultation and personalized recommendations, this study supplements such expert knowledge through literature-based information delivery. By providing scientifically grounded advice, the system offers reliable guidance without depending on skin specialists.

This study aims to provide personalized information based on sensing data in accordance with each user’s skin quality, thereby promoting behavior change. This approach serves as a demonstration of the effectiveness of sensing technology in the novel domain of skin quality maintenance and improvement.

This research explores new possibilities for behavior change support systems utilizing sensing data and is expected to contribute to the advancement of personalized healthcare. Furthermore, the insights gained from this study can be applied

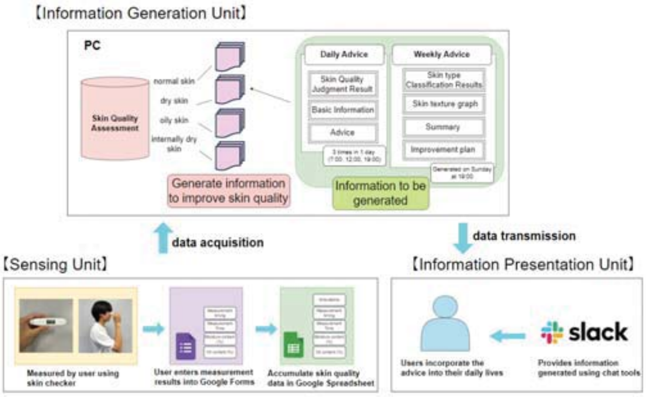

【Information Generation Unit】
PC
Skin Quality Assessment
normal skin
dry skin
oily skin
internally dry skin
Generate information to improve skin quality
Daily Advice
Skin Quality Judgment Result
Basic Information
Advice
3 times in 1 day (7:00, 12:00, 19:00)
Weekly Advice
Skin type Classification Results
Skin texture graph
Summary
Improvement plan
Generated on Sunday at 19:00
Information to be generated
data acquisition
data transmission
【Sensing Unit】
Measured by user using skin checker
User enters measurement results into Google Forms
Accumulate skin quality data in Google Spreadsheet
【Information Presentation Unit】
slack
Users incorporate the advice into their daily lives
Provides information generated using chat tools

Figure 1: System Configuration Chart

to the design and implementation of other behavior change support systems in the health and beauty domains.

3 Proposed Method

3.1 System Overview

The proposed system is based on data obtained from a skin quality measurement device and is designed to generate and deliver information aimed at improving or maintaining users' skin quality. The main features of the system are as follows:

- Continuous collection and analysis of skin quality data
- Generation of personalized information based on individual skin quality
- Promotion of behavior change through effective information presentation

Users are expected to refer to the provided information to select and implement appropriate actions for improving their skin quality. Through this process, the system is anticipated to enhance users' self-management capabilities and support continuous skin quality improvement.

3.2 System Configuration

The proposed system consists of the following three main components:

1. Sensing Module: A device that collects skin quality data
2. information Generation Module: A processing unit that interprets, stores, and generates information from the collected data
3. Information Presentation Module: An interface that effectively delivers the generated information to the user

The interaction among these components enables continuous and effective behavior change support. The overall system architecture is illustrated in Fig. 1.

3.2.1 Sensing Module

This system uses the "Skin Checker" provided by Ryohada Laboratory as the sensing device for measuring skin quality[9]. The Skin Checker is capable of measuring skin moisture content, skin sebum content, elasticity, and tone. Fig. 2 shows the Skin Checker used in this study.

The Skin Checker employs bioelectrical impedance analysis (BIA) to instantly calculate skin moisture content and skin sebum content levels by utilizing the electrical conductivity of moisture and skin sebum content. A weak electric current flows from the two skin sensors at the device's tip to measure skin moisture content in the stratum corneum and skin sebum content in the epidermis. The device is operated by pressing and holding the power button and applying the sensor to the measurement area for 3 seconds. Upon completion, four values are displayed on an LCD monitor. The device measures 12 cm in length, weighs 40 grams, and uses a CR2032 button battery instead of a rechargeable system, enabling approximately 2.5 years of use with two daily measurements.

In this study, users were instructed to use the Skin Checker three times per day—after waking up, before lunch, and before bedtime-and to manually enter their results into a Google Form, as the device is not equipped with network connectivity and does not support automatic data transmission. The recorded skin moisture content and skin sebum content data were then saved into a Google Spreadsheet.

As shown in Fig. 3, participants were instructed to measure the thin and sensitive area around the eyes. The contents of the questionnaire using Google Forms are shown in Table 1. Users are asked to input the measurement time, measurement timing, skin moisture content level, and skin sebum content. The data entered are stored in a Google Spreadsheet as illustrated in Table 2.

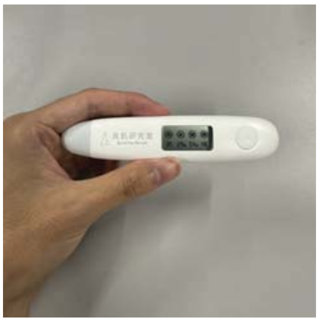


Figure 2: Skin Checker

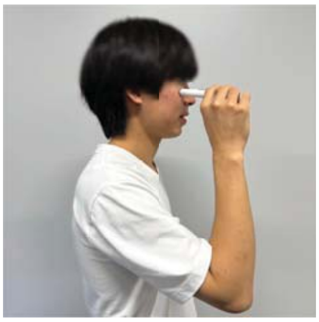


Figure 3: Measurement Site

3.2.2 Information Generation Module

In the information generation module, the collected and stored data from the sensing module are used to classify skin quality into four categories. Based on these classifications, the system generates both daily and weekly advice aimed at improving or maintaining optimal skin quality.

Daily advice is generated three times per day—at 7:00, 12:00, and 19:00—and consists of the following two types of information:

Table 1: Questionnaire Content

No.	Survey Contents
1	Please indicate the timing of the measurement.
2	Please specify the time of the measurement.
3	Please enter the skin moisture content level of your skin.
4	Please enter the skin sebum content of your skin.

Table 2: Data stored in spreadsheets

Timestamp	Measurement Timing	Measurement Time	skin moisture content(%)	skin sebum content(%)
2024/06/17 6:33:14	After waking up	5:55:00	48	47
2024/06/17 11:01:45	Before lunch	11:00:00	62	41
2024/06/17 23:06:51	Before bedtime	23:00:00	55	47
2024/06/18 6:58:30	After waking up	5:55:00	40	50
2024/06/18 11:34:43	Before lunch	11:00:00	46	57
2024/06/18 23:25:21	Before bedtime	23:00:00	55	48

1. skin quality classification results based on user data (hereafter referred to as Generated Information 1)
2. Basic skin care information and advice (hereafter referred to as Generated Information 2)

Weekly advice is generated at 19:00 on Sundays and includes the following four types of information:

1. Classification results based on the user's skin data from the past week (Generated Information 3)
2. A plotted graph of skin data from the past week (Generated Information 4)
3. Trends in skin moisture content and skin sebum content levels over the past week (Generated Information 5)
4. Improvement suggestions to enhance or maintain skin quality, based on the past week's data (Generated Information 6)

Skin classification is based on the scientific skin typing method proposed by Shiseido Co., Ltd. [10]. The details are provided in Table 3. The following section describes the details of the generated information outlined above.

In Generated Information 1, the user's skin type is determined based on the latest skin moisture content and skin sebum content stored in the spreadsheet, following a conditional branching method developed according to the classification described above, and the skin type judgment result is generated. Table 4 shows the details of the conditional branching and the resulting skin type assessment output. This information is presented in a concise format to enable users to quickly recognize their skin type when the information is provided.

In Generated Information 2, basic information and advice regarding skin type are generated based on the skin type judgment result generated in Generated Information 1. An example of the information generated for users with oily skin is shown in Table 5. Note that basic information and advice are stored in an Excel file and the appropriate information is extracted randomly for each skin type.

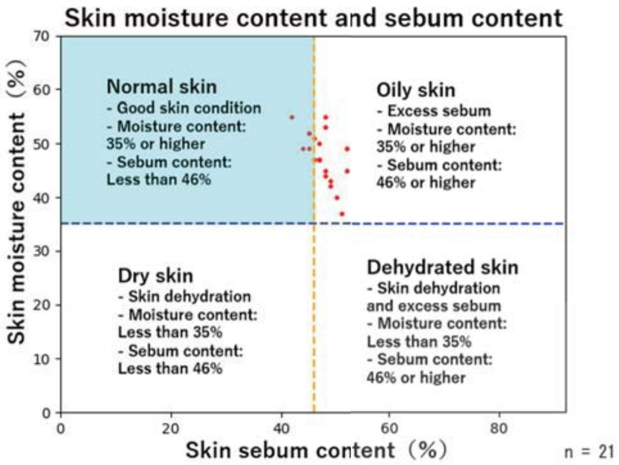


Figure 4: Generated information 4

In Generated Information 3, the proportion of each skin type is calculated from the past one week of skin quality data stored in the spreadsheet, the most frequent skin type is identified, and a skin type classification result is generated. The generated skin type classification result is shown in Table 6. This information is simplified so that the user can instantly recognize their classified skin type when receiving the provided information.

In Generated Information 4, a skin quality graph is created by plotting the skin data from the past week stored in a spreadsheet. The vertical axis of the graph represents the percentage of skin moisture content and the horizontal axis represents the percentage of skin sebum content. The user skin quality data are plotted as red dots. In addition, to help users recognize the ideal skin type, normal skin, the background of the normal skin area is colored light blue to distinguish it from other skin types. An example of the generated skin quality graph is shown in Fig. 4.

Table 3: Questionnaire Content

Skin Type	skin moisture content (%)	skin sebum content (%)	Description
Normal skin	≥ 35	<47	Balanced skin moisture content and skin sebum content levels; a healthy skin quality with few issues.
Dry skin	<35	<47	Low skin moisture content; prone to dryness and itching.
Oily skin	≥ 35	≥ 47	Excessive skin sebum content production; shiny appearance, enlarged pores, and prone to acne.
Dehydrated skin	<35	≥ 47	Lacks skin moisture content and is prone to dryness and itching, but also exhibits excessive skin sebum content production and is prone to acne

Table 4: Generated Information 1

skin moisture content (%)	skin sebum content (%)	Description
≥ 35	<47	As of 11:00 on July 1, 2024, your skin type was Normal skin.
<35	<47	As of 11:00 on July 1, 2024, your skin type was Dry skin.
≥ 35	≥ 47	As of 11:00 on July 1, 2024, your skin type was skin Oily skin.
<35	≥ 47	As of 11:00 on July 1, 2024, your skin type was Dehydrated skin.

Table 5: Generated Information 2

Basic Information	Advice
Oily skin is known to be caused by diets high in sugar and fat, which can increase skin sebum content production. A well-balanced diet is important for maintaining healthy skin.	For breakfast, include low-fat, high-protein foods. Egg whites, oatmeal, and fruits are good choices. Adding yogurt and berries also helps to intake antioxidants and support skin health.

Table 6: Generated Information 3

Most Frequent Skin Type	Skin Type Classification Result
Normal skin	Your skin type is classified as normal skin.
Dry skin	Your skin type is classified as dry skin.
Oily skin	Your skin type is classified as oily skin.
Dehydrated skin	Your skin type is classified as Dehydrated skin.

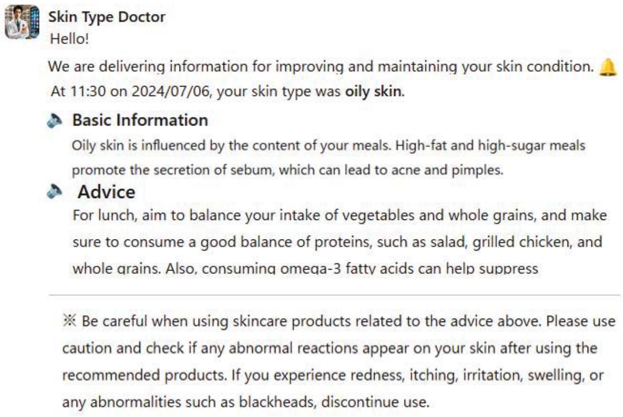
In Generated Information 5, the trends in skin moisture content and skin sebum content over the past week, stored in a spreadsheet, are analyzed. Five types of information are generated. The first is the range and average value of skin moisture content over the past week, along with an evaluation of the average. The second is the range and average value of skin sebum content over the past week, and its evaluation. The third is the average skin moisture content and skin sebum content levels after waking up, and an evaluation of these averages. The fourth is the average skin moisture content and skin sebum content content levels before lunch, and their evaluation. The fifth is the average skin moisture content and skin sebum content levels before bedtime, along with corresponding evaluations. Examples of the generated information are shown in Table 7.

In Generated Information 6, improvement suggestions are generated to maintain or enhance skin quality based on skin quality data accumulated over the past week. A total of nine types of information are generated. The first type comprises improvement suggestions related to skincare. Skincare refers to direct care from the outside of the skin, and appropriate methods contribute to the improvement of skin quality [11]. The second type involves suggestions related to skin moisture content. skin moisture content significantly affects the skin's skin moisture content retention condition, and insufficient skin moisture content can lead to dry skin or reduced barrier function [12]. The third type addresses dietary habits. Diet plays an internal role in supporting the skin's health condition by balancing nutritional intake [13]. The fourth type provides suggestions regarding smoking. Smoking impairs blood flow and accelerates skin aging, and it is necessary to mitigate these effects [14]. The fifth type involves suggestions regarding stress. Stress disrupts hormonal balance and may cause increased skin sebum content secretion or inflammation [15]. The sixth type deals with sleep. Sleep is a vital factor for ensuring skin repair time, and lack of sleep may lead to skin problems [16]. The seventh type addresses room temperature. Room temperature affects skin dryness and the balance of skin sebum content, making appropriate temperature control important [17]. The eighth type concerns humidity. Humidity is related to the retention of skin moisture content in the skin; insufficient humidity can lead to dryness, thus proper humidity control is required [17]. The ninth type offers suggestions regarding ultraviolet (UV) exposure. UV radiation is a major external factor that damages the skin, and countermeasures are necessary to prevent sunburn and aging [18].

An example of the information generated for individuals with oily skin is shown in Table 8.

3.2.3 Information Presentation Module

In this system, the information delivery functions of Daily Advice and Weekly Advice described in the previous section were implemented. For the method of information delivery in this system, the communication chat tool Slack [19] was utilized. On Slack, a dedicated channel was created for each participant, and the generated information was configured to be delivered by a bot named "skin quality Doctor." The following describes the details of the aforementioned information delivery functions.



Skin Type Doctor
Hello!
We are delivering information for improving and maintaining your skin condition. 🔔
At 11:30 on 2024/07/06, your skin type was **oily skin.**

Basic Information
Oily skin is influenced by the content of your meals. High-fat and high-sugar meals promote the secretion of sebum, which can lead to acne and pimples.

Advice
For lunch, aim to balance your intake of vegetables and whole grains, and make sure to consume a good balance of proteins, such as salad, grilled chicken, and whole grains. Also, consuming omega-3 fatty acids can help suppress

※ Be careful when using skincare products related to the advice above. Please use caution and check if any abnormal reactions appear on your skin after using the recommended products. If you experience redness, itching, irritation, swelling, or any abnormalities such as blackheads, discontinue use.

Figure 5: Daily Advice

Daily Advice is scheduled to be executed regularly at 7:00, 12:00, and 19:00 every day. Since this function is intended for daily use, it was designed with a simple interface. An example of the actual information delivery is shown in Fig. 5.

Weekly Advice is scheduled to be executed regularly at 19:00 on Sundays. Since this function involves a large volume of content, readability was improved through the use of paragraph and section divisions, emphasis, and icons. An example of the actual information delivery is shown in Fig. 6.

4 Demonstration Experiment

This chapter describes the demonstration experiment conducted using the proposed system.

4.1 Overview of the Demonstration Experiment

The purpose of this demonstration experiment was to collect users' skin quality data using the Skin Checker through the proposed system, utilize the collected data for generating skin improvement information, and promote behavior change. Additionally, the experiment evaluated the availability of the collected data and generated information, as well as the effectiveness and impact of the proposed system in improving skin quality through interviews conducted before and after the experiment.

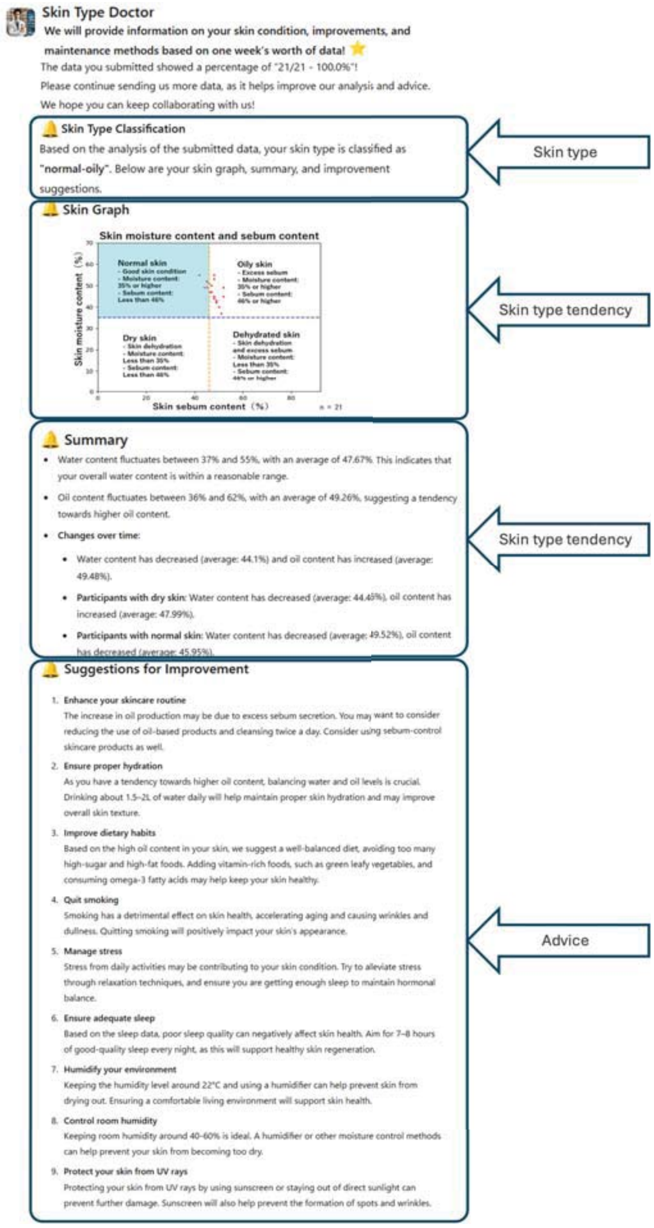
To examine the effect of the proposed system's information presentation on behavior change, the demonstration experiment was divided into two periods: Period A, during which the system was not used, and Period B, during which participants used the proposed system to improve their skin quality. In the first trial, Period A lasted from June 17 to June 23, 2024, and Period B from June 24 to July 7, 2024. Nine participants in their twenties took part in the experiment. In the second trial, Period A spanned from November 18 to November 24, 2024, and Period B from November 25 to December 8, 2024. A total of 18 participants joined: 17 in their twenties and one in their fifties.

Table 7: Generated Information 5

No.	Details
1	skin moisture content ranged from 37% to 55%, with an average of 47.6%. Overall, the skin moisture content level shows a trend toward the ideal range.
2	Skin sebum content ranged from 42% to 52%, with an average of 47.4%. Overall, the skin sebum content level tends to be on the higher side.
3	After waking, skin moisture content tends to decrease (average: 44.1%) and skin sebum content tends to increase (average: 48.9%).
4	Before lunch, skin moisture content tends to decrease (average: 46.4%) and skin sebum content tends to increase (average: 47.6%).
5	Before bedtime, skin moisture content tends to increase (average: 52.1%) and skin sebum content tends to decrease (average: 45.9%).

Table 8: Generated Information 6

No.	Details
1	Many measurement results show a high level of skin sebum content, which indicates a tendency toward skin sebum content skin. In the case of oily skin, it is important to remove excess skin sebum content. Use gentle cleansers and wash your face twice a day. Additionally, using wiping lotions and skin sebum content control cosmetics is also effective.
2	Although the measurements show a tendency for high skin moisture content, proper skin moisture content during the day is important to maintain skin health. Aim to drink more than 8 glasses of water a day to maintain body skin moisture content. This makes skin moisturizing more effective.
3	The measurement data show a high level of skin sebum content, so a balanced diet should be maintained. Avoid meals high in sugar and fat, and actively consume vegetables and fruits rich in vitamins and minerals to maintain skin health. Foods rich in omega-3 fatty acids such as fish and nuts are also recommended.
4	Smoking has a negative impact on skin health, so it is recommended to quit smoking. Smoking accelerates skin aging and causes wrinkles and dryness.
5	From the time of measurement data, stress during daytime activities may be affecting the skin. Incorporate yoga, meditation, and moderate exercise to reduce stress. Stress disrupts hormonal balance and causes skin problems.
6	Looking at the measurement data before bedtime and after waking up, skin moisture content and skin sebum content are moderately maintained, but securing high-quality sleep is even more important. Aim for 7-8 hours of quality sleep per day, keep bedding clean, and avoid stimulating the skin.
7	Maintaining the room temperature between 18–22°C prevents skin dryness. Use air conditioners and humidifiers appropriately to maintain a comfortable environment.
8	It is desirable to maintain indoor humidity between 40–60%. Use a humidifier to adjust the humidity and prevent skin dryness.
9	Based on the measurement time zone, UV protection during the day is important. When going out, use sunscreen and protect your skin with hats and sunglasses. Ultraviolet rays accelerate skin aging and cause spots and wrinkles.


Skin type
Skin type tendency
Skin type tendency
Advice
Skin moisture content and sebum content
Skin moisture content (%)
Skin sebum content (%)
n = 21
Normal skin
- Good skin condition
- Moisture content: 35% or higher
- Sebum content: Less than 46%
Oily skin
- Excess sebum
- Moisture content: 35% or higher
- Sebum content: 46% or higher
Dry skin
- Skin dehydration
- Moisture content: Less than 35%
- Sebum content: Less than 46%
Dehydrated skin
- Skin dehydration and excess sebum
- Moisture content: Less than 35%
- Sebum content: 46% or higher
0
10
20
30
40
50
60
70
0
20
40
60
80

Skin Type Doctor

We will provide information on your skin condition, improvements, and maintenance methods based on one week's worth of data! ⭐

The data you submitted showed a percentage of "21/21 - 100.0%"!

Please continue sending us more data, as it helps improve our analysis and advice.

We hope you can keep collaborating with us!

🔔 **Skin Type Classification**

Based on the analysis of the submitted data, your skin type is classified as "**normal-oily**". Below are your skin graph, summary, and improvement suggestions.

🔔 **Skin Graph**

🔔 **Summary**

- Water content fluctuates between 37% and 55%, with an average of 47.67%. This indicates that your overall water content is within a reasonable range.
- Oil content fluctuates between 36% and 62%, with an average of 49.26%, suggesting a tendency towards higher oil content.
- **Changes over time:**
  - Water content has decreased (average: 44.1%) and oil content has increased (average: 49.48%).
  - **Participants with dry skin:** Water content has decreased (average: 44.45%), oil content has increased (average: 47.99%).
  - **Participants with normal skin:** Water content has decreased (average: 49.52%), oil content has decreased (average: 45.95%).

🔔 **Suggestions for Improvement**

1. **Enhance your skincare routine**
   The increase in oil production may be due to excess sebum secretion. You may want to consider reducing the use of oil-based products and cleansing twice a day. Consider using sebum-control skincare products as well.
2. **Ensure proper hydration**
   As you have a tendency towards higher oil content, balancing water and oil levels is crucial. Drinking about 1.5–2L of water daily will help maintain proper skin hydration and may improve overall skin texture.
3. **Improve dietary habits**
   Based on the high oil content in your skin, we suggest a well-balanced diet, avoiding too many high-sugar and high-fat foods. Adding vitamin-rich foods, such as green leafy vegetables, and consuming omega-3 fatty acids may help keep your skin healthy.
4. **Quit smoking**
   Smoking has a detrimental effect on skin health, accelerating aging and causing wrinkles and dullness. Quitting smoking will positively impact your skin's appearance.
5. **Manage stress**
   Stress from daily activities may be contributing to your skin condition. Try to alleviate stress through relaxation techniques, and ensure you are getting enough sleep to maintain hormonal balance.
6. **Ensure adequate sleep**
   Based on the sleep data, poor sleep quality can negatively affect skin health. Aim for 7–8 hours of good-quality sleep every night, as this will support healthy skin regeneration.
7. **Humidify your environment**
   Keeping the humidity level around 22°C and using a humidifier can help prevent skin from drying out. Ensuring a comfortable living environment will support skin health.
8. **Control room humidity**
   Keeping room humidity around 40-60% is ideal. A humidifier or other moisture control methods can help prevent your skin from becoming too dry.
9. **Protect your skin from UV rays**
   Protecting your skin from UV rays by using sunscreen or staying out of direct sunlight can prevent further damage. Sunscreen will also help prevent the formation of spots and wrinkles.

Figure 6: Weekly Advice

4.2 Participant Attributes

This section describes the details of the experiment participants. In the first demonstration experiment, nine individuals in their twenties participated. The attributes of these participants, based on pre-experiment questionnaires, are shown in Table 9.

In the second demonstration experiment, 17 participants in their twenties and one participant in their fifties took part. Their attributes, as obtained from the pre-experiment questionnaires, are shown in Table 10.

4.3 Experimental Method

This section describes the experimental methodology of the demonstration experiment.

4.3.1 Intervention Method of the Experiment

In the proposed system, participants were instructed to use the Skin Checker three times a day—after waking up, before lunch, and before bedtime—and input their measurement results into a Google Form, thereby collecting skin quality data. To provide information to the participants, they were invited to join a dedicated Slack workspace for the experiment, where personalized information was delivered three times daily (at 7:00, 12:00, and 19:00) in their individual channels.

4.3.2 Pre-Experiment Questionnaire

To verify whether behavior change toward skin improvement occurred during the experiment, questionnaires were conducted before and after the experiment. Google Forms [20] was used for the survey.

The pre-experiment questionnaire investigated participants' skin quality, lifestyle habits, stress factors, skincare practices, and the influence of environmental factors on their skin. The questionnaire first collected basic information such as age, gender, height, and weight. Participants were then asked about their skin type and concerns, as well as their lifestyle habits, including sleep duration, beverage intake, number of meals per day, and frequency of consumption of fruits, vegetables, dairy products, seafood, soy products, nuts, processed foods, sweets, and alcoholic beverages.

Additionally, the survey asked about the frequency of stress over the past month, symptoms experienced during stress, its causes and coping methods, awareness of stress's impact on skin quality, and related skin symptoms. Furthermore, participants were asked whether they usually practiced skincare, the types of products used, skincare routines and frequency, and their use of sunscreen when spending time outdoors during the day.

Regarding environmental factors, questions were asked about the frequency of air conditioner and heater usage indoors, awareness of air dryness, and opinions on how these factors affect the skin. The questionnaire also investigated whether participants had allergic reactions to cosmetics or food, their level of interest in improving skin quality, and any actions they were taking for that purpose.

Table 9: Questionnaire Content

Participant ID	Age	Gender	Skin Type
P1	23	Male	Dry skin
P2	22	Male	Oily skin
P3	21	Male	Oily skin
P4	21	Male	Oily skin
P5	21	Male	Unknown
P6	22	Male	Dehydrated skin
P7	21	Female	Unknown
P8	21	Male	Oily skin
P9	21	Female	Nomal skin

Table 10: Attributes of the 2nd Demonstration Experiment Collaborators

Participant ID	Age	Gender	Skin Type
P10	23	Male	Normal skin
P11	23	Male	Oily skin
P12	22	Male	Oily skin
P13	21	Male	Oily skin
P14	21	Female	Normal skin
P15	22	Male	Oily skin
P16	22	Male	Oily skin
P17	21	Male	Dehydrated skin
P18	22	Male	Dehydrated skin
P19	21	Female	Dehydrated skin
P20	20	Male	Oily skin
P21	23	Female	Oily skin
P22	54	Female	Dry skin
P23	23	Female	Dehydrated skin
P24	22	Female	Normal skin
P25	20	Male	Oily skin
P26	20	Male	Oily skin
P27	21	Male	Dehydrated skin

4.3.3 Post-Experiment Questionnaire

The post-experiment questionnaire investigated the usefulness of the proposed system and changes in participants' awareness and behavior. Participants were asked about changes in their awareness and behavior before and after using the proposed system in relation to lifestyle and health management factors such as skincare, skin moisture content, diet, stress management, sleep, room temperature, humidity, and UV protection.

The questionnaire also assessed the usefulness of the daily advice provided each day and the weekly advice provided once a week, as well as the appropriateness of their frequency. In addition, participants were invited to provide free-text responses regarding the effectiveness of the system in improving skin quality, areas for improvement, and desired additional features.

5 Evaluation

This chapter evaluates whether the proposed system contributed to the improvement of skin quality among the participants through the demonstration experiment.

5.1 System Evaluation

This section evaluates whether the proportion of normal skin—considered the ideal skin type—increased as a result of using the system, by comparing the proportions during Period A and Period B of the demonstration experiment. Table 11 shows the proportion of normal skin and the comparison results for each period in the first demonstration experiment, and Table 12 shows the corresponding data for the second demonstration experiment. In the first demonstration experiment, the proportion of normal skin increased for 6 participants and decreased for 3 participants from Period A to Period B. In the second demonstration experiment, the proportion of normal skin increased for 9 participants, remained unchanged for 6 participants, and decreased for 3 participants from Period A to Period B. Overall, 55.56% of participants showed an increase in the proportion of normal skin, 22.22% showed no change, and 22.22% showed a decrease.

5.2 Questionnaire Evaluation

This section describes the evaluation results obtained from the questionnaire administered to participants after the demonstration experiment.

5.2.1 Behavior Change Induced by the System

This subsection evaluates whether the system induced behavior change in users regarding skin quality improvement. A post-experiment questionnaire was used to investigate whether the presented information led to behavior change. The responses were binary: "Yes" or "No", with "Yes" indicating that behavior change occurred.

Fig. 7 shows the survey results for the first demonstration experiment, Fig. 8for the second experiment, and Fig. 9 for the overall results.

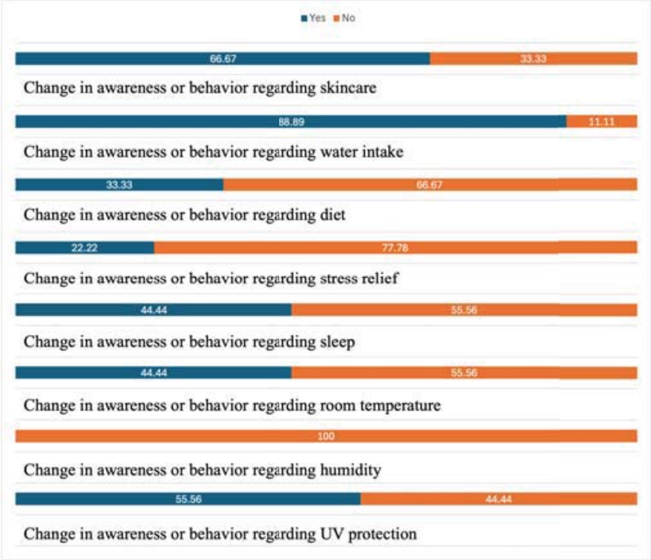


Figure 7: Responses Related to Changes in Attitudes and Behavior (Experiment 1)

In the first demonstration experiment, 66.67% of participants showed changes in awareness or behavior related to skincare, 88.89% in skin moisture content, 33.33% in diet, 22.22% in stress management, 44.44% in sleep, 44.44% in room temperature control, 0% in humidity control, and 55.56% in UV protection.

In the second demonstration experiment, 83.33% of participants reported changes in awareness or behavior related to skincare, 100% in skin moisture content, 61.11% in diet, 33.33% in stress management, 66.67% in sleep, 38.89% in room temperature control, 44.44% in humidity control, and 44.44% in UV protection.

Overall, 77.78% of participants reported changes in awareness or behavior related to skincare, 96.30% in skin moisture content, 51.85% in diet, 29.63% in stress management, 59.26% in sleep, 40.74% in room temperature control, 29.63% in humidity control, and 48.15% in UV protection.

Additionally, participants who reported changes in awareness or behavior regarding skincare adopted the following actions:

- Used cleansing products
- Used moisturizing cream
- Use toner
- Became more mindful of moisturizing often-neglected areas of skin
- Used sunscreen
- Started applying toner in the morning
- Became more conscious about drinking water
- Used beauty oils
- Used serums
- Frequently used facial devices

Table 11: Percentage of Normal Skin in Period A and Period B (Experiment 1)

Participant ID	Percentage of Normal Skin in Period A (%)	Percentage of Normal Skin in Period B (%)	Difference (B - A) (%)
P1	88.24	81.82	-6.42
P2	12.50	55.56	+43.06
P3	11.11	28.57	+17.46
P4	31.58	17.07	-14.51
P5	30.00	61.76	+31.76
P6	0.00	4.65	+4.65
P7	80.00	33.33	-46.67
P8	6.25	21.21	+14.96
P9	19.05	30.77	+11.72

Table 12: Percentage of Normal Skin in Period A and Period B (Experiment 2)

Participant ID	Percentage of Normal Skin in Period A (%)	Percentage of Normal Skin in Period B (%)	Difference (B - A) (%)
P10	5.56	37.50	+31.94
P11	0.00	0.00	0.00
P12	5.00	6.98	+1.98
P13	0.00	6.67	+6.67
P14	9.09	10.71	+1.62
P15	0.00	10.71	+10.71
P16	0.00	0.00	0.00
P17	41.18	48.65	+7.47
P18	0.00	0.00	0.00
P19	0.00	0.00	0.00
P20	9.09	0.00	-9.09
P21	14.29	47.37	+33.08
P22	70.00	82.50	+12.50
P23	0.00	0.00	0.00
P24	21.05	0.00	-21.05
P25	0.00	0.00	0.00
P26	0.00	5.00	+5.00
P27	9.09	0.00	-9.09

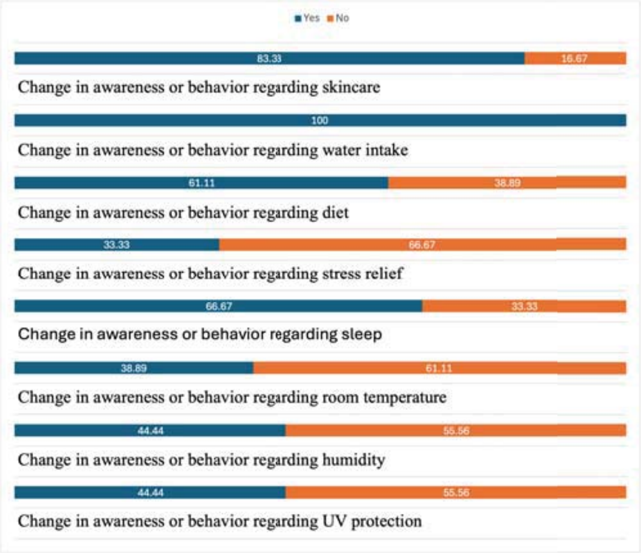


Figure 8: Responses Related to Changes in Attitudes and Behavior (Experiment 2)

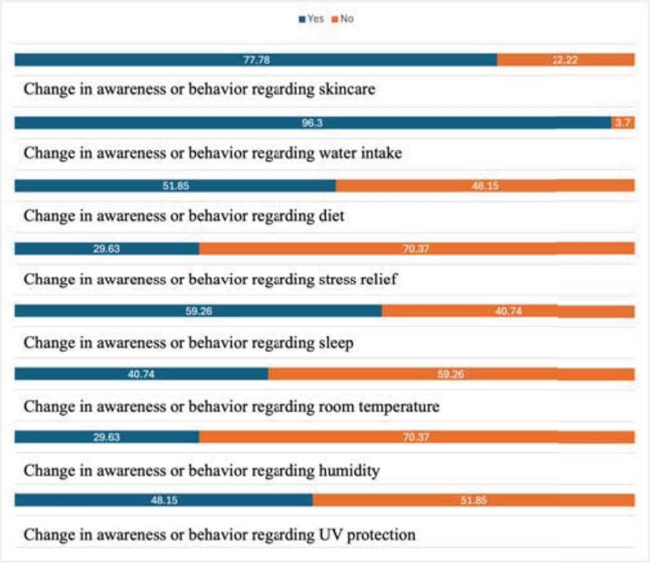


Figure 9: Responses Related to Changes in Attitudes and Behavior

Participants who reported changes in skin moisture content behavior adopted the following actions:

- Drank a glass of water upon waking up
- Drank a glass of water before going to bed
- Incorporated water-rich vegetables and fruits into meals
- Frequently drank water and herbal tea

Participants who reported dietary behavior change adopted the following actions:

- Consumed seafood rich in omega-3 fatty acids
- Had antioxidant-rich lunches including salad, vegetables, and fish
- Ate dinners rich in quality protein such as chicken and legumes
- Ate fruits
- Consumed yogurt
- Ate citrus fruits and kiwis rich in vitamin C
- Ate nuts rich in vitamin E
- Took vitamin C supplements

Participants who reported behavior change in stress management adopted the following actions:

- Started taking 20-minute walks at night to stay active
- Took short walks or did light stretches during lunch breaks
- Spent evenings reading or doing light stretching to relax

Participants who reported behavior change in sleep adopted the following actions:

- Ensured 7–9 hours of high-quality sleep each night
- Tried to go to bed earlier
- Created a relaxing sleep environment
- Refrained from using smartphones and computers before bed
- Tried to get enough sleep even when going to bed late

Participants who reported behavior change in room temperature control adopted the following actions:

- Checked that room temperature was maintained at 20–22°C in the morning
- Maintained room temperature at 22–24°C during the day

- Maintained room temperature at 20–22°C at night to ensure a comfortable sleep environment

Participants who reported behavior change in humidity control adopted the following actions:

- Checked room humidity in the morning and used a humidifier if necessary
- Maintained indoor humidity between 40–60% during the day
- Maintained bedroom humidity between 40–60% at night
- Used a humidifier when dryness was felt
- Became more mindful of indoor humidity levels
- Used wet towels to maintain humidity in place of a humidifier

Participants who reported behavior change in UV protection adopted the following actions:

- Added sunscreen to their morning skincare routine
- Wore hats, sunglasses, and long-sleeved clothing
- Tried to stay in shaded areas
- Thoroughly removed sunscreen and dirt at night using cleansing products

5.2.2 System Usability Evaluation

This section presents an evaluation of the user experience with the system. In the post-experiment questionnaire, we investigated whether the content and frequency of information provided by the system were appropriate.

The question regarding whether the daily and weekly advice contributed to awareness of skin quality and improvement behaviors was evaluated using a binary response format: "Yes" or "No." A "Yes" response indicates that the information provided by the system contributed to promoting awareness and behavior improvements regarding skin quality.

The question regarding the appropriateness of the frequency of information provision was rated on a three-point scale: "Too frequent," "Appropriate," or "Too infrequent." A response of "Appropriate" was considered an indication that the frequency of information provision was evaluated as suitable.

Furthermore, for the question regarding whether the system as a whole was considered useful for improving skin quality, a five-point Likert scale was used: "Strongly agree," "Agree," "Neutral," "Disagree," and "Strongly disagree."

Fig. 10 presents the survey results from the first demonstration experiment, Fig. 11 presents those from the second experiment, and Fig. 12 shows the overall results.

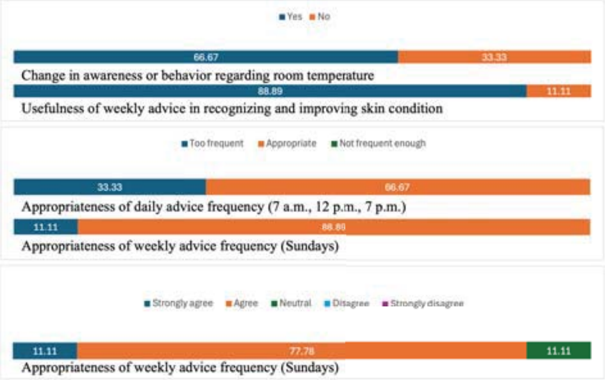


Figure 10: Answers regarding usability of the system(Experiment 1)

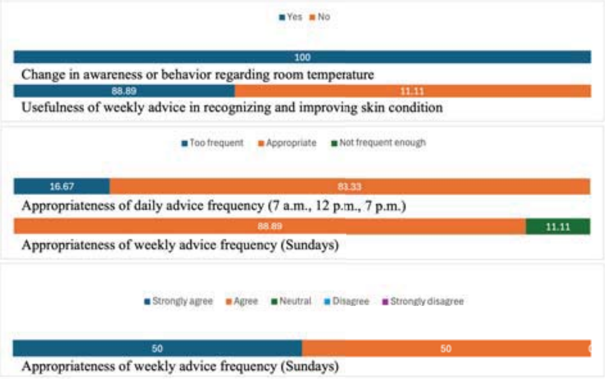


Figure 11: Answers regarding usability of the system(Experiment 2)

In the first demonstration experiment, 66.67% of participants responded that the daily advice provided each day contributed to improving their skin quality, while 88.89% responded that the weekly advice provided on Sundays was helpful for skin improvement. Regarding the frequency of information provision, 66.67% of participants indicated that the frequency of daily advice was appropriate, and 88.89% stated that the frequency of weekly advice was appropriate. In addition, 88.89% of participants responded either "Strongly agree" or "Agree" to the question regarding the usefulness of the system in improving skin quality.

In the second demonstration experiment, 100% of participants responded that the daily advice contributed to improving their skin quality, while 88.89% responded that the weekly advice was helpful for skin improvement. Regarding the frequency of information provision, 83.33% of participants indicated that the frequency of daily advice was appropriate, and 88.89% stated that the frequency of weekly advice was appropriate. In addition, 100% of participants responded either "Strongly agree" or "Agree" regarding the usefulness of the system in improving skin quality.

Overall, 88.89% of participants responded that the daily advice contributed to improving their skin quality, and 88.89% responded that the weekly advice was helpful for skin improvement. Regarding the frequency of information provision, 77.78% of participants stated that the frequency of daily advice was appropriate, and 88.89% indicated that the frequency of weekly advice was appropriate. In addition, 96.30% of participants responded either "Strongly agree" or "Agree" regarding the usefulness of the system in improving skin quality.

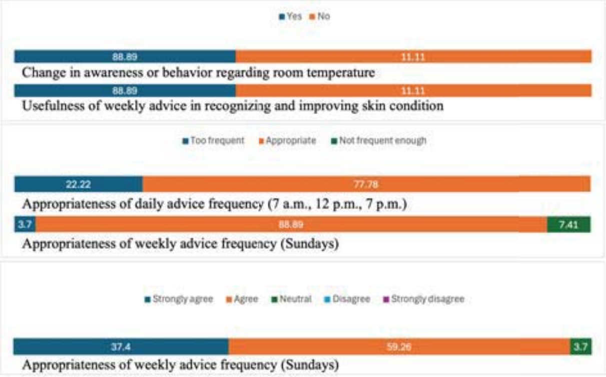


Figure 12: Answers regarding usability of the system

6 Discussion

This chapter discusses improvements in skin quality and behavior change.

6.1 Discussion on Skin Quality Improvement

This section presents a discussion on skin quality improvement using the proposed system.

First, in the first demonstration experiment, from Period A (June 17 to June 23, 2024) to Period B (June 24 to July 7, 2024), the percentage of participants with normal skin increased in 6 participants (66.67%), while it decreased in 3 participants (33.33%). This result indicates that the majority of participants showed an improvement in the proportion of normal skin, suggesting that the system had a certain effect on improving skin quality. The increase in the proportion of normal skin is considered to be due to the fact that the information and feedback provided by the system encouraged improvements in users' daily skincare practices and behavior habits. In particular, it is believed that the observed changes resulted from users being able to take appropriate actions in response to their own skin quality.

Next, in the second demonstration experiment, from Period A (November 18 to November 24, 2024) to Period B (November 25 to December 8, 2024), the percentage of participants with normal skin increased in 9 participants (50.00%), remained unchanged in 6 participants (33.33%), and decreased in 3 participants (16.67%). This result shows that, although the majority of participants experienced an increase in normal skin percentage in the second experiment as well, a notable number of participants showed no change, which differs from the results of the first experiment. This period coincides with the transition into winter, during which lower humidity and temperature levels create an environment conducive to skin dryness[4]. Therefore, compared to the first experiment, the effectiveness of the system may have been reduced. However, since a certain number of participants maintained or improved their skin quality, it is considered that the system contributed to mitigating seasonal effects.

Furthermore, when viewed as a whole, 55.56% of participants showed an increase in the proportion of normal skin, while 22.22% showed no change and 22.22% showed a decrease. These results suggest that the system demonstrated a certain effectiveness in improving skin quality. Although skin quality is affected by various complex factors and it is difficult to attribute changes solely to seasonal influences, a certain correlation between season and skin quality was observed.

Considering the influence of seasonal factors on skin quality, it is suggested that in order to maximize the effectiveness of the system, adjustments based not only on individual skin conditions but also on the season and environmental conditions are required. Specifically, by proposing care methods corresponding to changes in humidity and temperature, and by incorporating functions that compensate for environmental factors, the effectiveness of the system can be further enhanced.

6.2 Discussion on Behavior Change

This section presents a discussion on behavior change induced by the proposed system.

First, it was confirmed that all 27 participants took actions to improve their skin quality. This result suggests that the system is effective in promoting behavior change for skin quality improvement. In particular, in the area of skin moisture content, 88.89% of participants in the first experiment and 100% in the second experiment showed behavior changes, indicating that the skin moisture content-related information provided by the system had a tangible impact on users' daily behavior. The success in promoting behavior change is attributed to the presentation of specific information. The system provided users with specific information that was easy to understand and implement. For instance, by presenting concrete advice such as the timing and amount of skin moisture content required, users were able to easily modify their own behavior.

On the other hand, only 29.63% of participants showed changes in humidity control and stress management behaviors, suggesting that the information provided on these topics may have been perceived as less actionable or difficult to implement. As an improvement measure, the advice provided could include more concrete action plans and practical examples to make humidity and stress management easier to implement. For example, in the case of humidity management, suggesting how to use a humidifier at home or recommending humidity measurement methods, and for stress management, proposing quick relaxation techniques or beneficial daily habits may help make the recommended behaviors more concrete and increase the implementation rate.

In addition, regarding the evaluation of the usefulness of the daily and weekly advice, 88.89% of participants responded that the daily advice was helpful in improving their skin quality, and an equal percentage (88.89%) responded similarly for the weekly advice, indicating that both forms of advice were highly valued. In particular, in the second experiment, 100% of participants reported that the daily advice was useful, suggesting that the system effectively supported daily behavior change.

Furthermore, regarding the frequency of information provision, 77.78% of participants stated that the frequency of the daily advice was appropriate, and 88.89% responded similarly for the weekly advice. These results suggest that the system was designed to deliver information at a frequency that was not burdensome and was easily acceptable to users. Notably, the weekly advice received consistently high ratings of 88.89% in both the first and second experiments, suggesting that weekly information delivery is perceived as a useful format by users. However, since individual differences in lifestyle and sensitivity to information are considerable, allowing users to customize the frequency of information delivery may further enhance user satisfaction and promote behavior change.

7 Conclusion

In this study, we developed and evaluated a system aimed at improving skin quality and promoting behavior change. Based on skin moisture content and skin sebum content data obtained through a skin measurement device, we proposed a mechanism that provides individually tailored daily and weekly advice to promote behavior change.

In the verification experiments, we evaluated the effectiveness of skin quality improvement, behavior change, the information provided, and the overall system, based on acquired skin data and questionnaire responses from 26 participants in their twenties and one participant in their fifties. As a result, the proposed system demonstrated a certain level of effectiveness in promoting skin quality improvement and behavior change. In terms of skin improvement, 55.56% of participants showed an increase in the percentage of normal skin, and all participants engaged in improvement actions related to behavior change. Particularly regarding skin moisture content, 96.30% of participants exhibited behavior changes, confirming the effectiveness of the specific advice provided. Moreover, user satisfaction with the system was high; the frequency and content of the daily and weekly advice were well-received, contributing to users' awareness and implementation of skincare practices.

On the other hand, behavior change rates for humidity control and stress management were low at 29.63%, highlighting the lack of concrete action plans in these areas as a remaining issue. Additionally, the impact of declining humidity and temperature during winter on skin quality and behavior change was noted, suggesting the need for flexible support that considers environmental factors.

Based on these results, the proposed system was shown to be effective in promoting skin quality improvement and behavior change. However, challenges remain in addressing humidity control and stress management. In the future, it will be necessary to incorporate more concrete action plans and flexible support functions that respond to seasonal and environmental factors. This will enable the development of a practical support system that can be used daily to meet the diverse skincare needs of users.

REFERENCES

[1] S. Amagasa et al., "Evaluation of physical activity using smartphones and wearable devices in the medical and healthcare fields: current status and future prospects," Jpn. J. Public Health, vol. 68, no. 9, pp. 585–596, 2021. (in Japanese)

[2] Ministry of Health, Labour and Welfare, "Medical expenses by primary department in clinics, FY2022." [Online]. Available: https://www.mhlw.go.jp/content/001046511.pdf (accessed Jan. 21, 2025).

[3] Consumer Affairs Agency, "White Paper on Consumers 2024." [Online]. Available: https://www.caa.go.jp/policies/policy/consumer_research /white_paper/assets/consumer_research_cms201_240614 _37.pdf (accessed Jan. 21, 2025).

[4] M. Kenjo et al., "Seasonal variation in sebum composition and its relation to skin condition and skin type," J. Soc. Cosmet. Chem. Jpn., vol. 34, no. 4, pp. 365–373, 2000. (in Japanese)

[5] H. Kaneko and K. Matsui, "Design and implementation of a continuous monitoring and behavior change support system for user-led skin quality improvement," Tech. Rep. Consumer Devices and Systems (CDS), vol. 41, no. 12, pp. 1–8, 2024. (in Japanese)

[6] IPSA, "Ipsa counseling." [Online]. Available: https://www.ipsa.co.jp/shop/counseling.htmla-shop (accessed Jan. 21, 2025).

[7] S. Tsuji et al., "Design and trial of a personal workstyle advisor using behavior sensing," IPSJ Digital Practice, vol. 10, no. 1, pp. 267–282, 2019. (in Japanese)

[8] F. Alòs et al., "Effectiveness of a healthcare-based mobile intervention on sedentary patterns, physical activity, mental well-being and clinical and productivity outcomes in office employees with type 2 diabetes: Study protocol for a randomized controlled trial," BMC Public Health, vol. 22, no. 1269, 2022.

[9] Fineskin Lab, "Skin checker." [Online]. Available: https://fineskinlab.com/products/checker.html (accessed Jan. 21, 2025).

[10] H. Kumagai et al., "Development of a scientific method for skin type classification," J. Soc. Cosmet. Chem. Jpn., vol. 19, no. 1, pp. 9–19, 1985. (in Japanese)

[11] H. Iwai, "Role of skincare formulations in cosmetics," Oleoscience, vol. 1, no. 3, pp. 255–263, 2001. (in Japanese)

[12] J. Krutmann et al., "Skin hydration is significantly increased by a supplementation of curcumin extract: A randomized, placebo-controlled, double-blind study," Nutrients, vol. 11, no. 1, 2019.

[13] M. Takaoka et al., "Impact of dietary habits on skin condition," J. Jpn. Dietetic Life, vol. 19, no. 1, pp. 44–49, 2008. (in Japanese)

[14] A. Morita, "Smoking and skin aging," J. Jpn. Cosmetic Sci. Soc., vol. 32, no. 4, pp. 279–283, 2008. (in Japanese)

[15] H. Kaminaga et al., "Stress and skin — Physiological changes of skin due to overcrowded stress model —," Jpn. J. Dermatol., vol. 107, no. 5, pp. 615–620, 1997. (in Japanese)

[16] P. Oyetakin-White et al., "Does poor sleep quality affect skin ageing?," Clin. Exp. Dermatol., vol. 40, no. 1, 2015.

[17] M. Ohno et al., “Effects of temperature, humidity, season, and face washing on physiological skin functions,” Jpn. J. Dermatol., vol. 97, no. 8, pp. 953–960, 1987. (in Japanese)
[18] R. Kamide, “Dermatological significance of ultraviolet protection,” J. Soc. Cosmet. Chem. Jpn., vol. 30, no. 3, pp. 265–272, 1996. (in Japanese)
[19] Slack Technologies, “Slack.” [Online]. Available: https://slack.com/intl/ja-jp/ (accessed Jan. 21, 2025).
[20] Google, “Google Form.” [Online]. Available: https://docs.google.com/forms/u/0/ (accessed Jan. 21, 2025).

A Study on Sleep Behavior Improvement Using Wearable Device Data

Haruto Iwase[†], Ryosuke Takahashi[†], Masami Shinoda[†], and Kanae Matsui[†‡]

[†]Graduate School of System Design and Technology, Tokyo Denki University, Japan
[‡]Expolis Co., Ltd
{25amj05, 24amj24, 24amj18}@ms.dendai.ac.jp matsui@ms.mail.dendai.ac.jp

Abstract - According to the 2021 OECD survey, the average sleep duration in Japan is 7 hours and 22 minutes, the shortest among the 33 member countries. Sleep deprivation can cause various problems, including fatigue, drowsiness, emotional instability, and impaired judgment, potentially leading to accidents. This study proposes a behavior change support system that utilizes data obtained from wearable devices to address sleep issues. The system analyzes user data and provides personalized advice aimed at improving sleep habits. By encouraging behavior modification, the system seeks to support both physical and mental health, ultimately enhancing quality of life. To evaluate the effectiveness of the proposed system, an experiment was conducted with male and female participants in their early 20s. Among participants who had previously experienced insufficient sleep, two out of three showed increased total sleep time. Moreover, 80% of all participants demonstrated improvements in their Athens Insomnia Scale scores, suggesting a positive effect on sleep quality. These results indicate the potential of personalized digital interventions to promote healthy sleep behavior.

Keywords: Sleep behavior, Wearable devices, Behavioral change, Personalized support, Health promotion.

1 Introduction

In recent years, sleep deprivation among the Japanese population has become increasingly severe. According to an OECD survey report published in 2021, the average sleep duration of Japanese individuals was only 7 hours and 22 minutes, the shortest among the 33 OECD member countries[1].

Insufficient sleep has been linked to a wide range of negative outcomes, including daytime sleepiness and fatigue, increased somatic complaints such as headaches, emotional instability, reduced attention and decision-making abilities, lower productivity at work, and academic underperformance. In severe cases, sleep deprivation may even lead to accidents with serious consequences[2].

Furthermore, the 2019 National Health and Nutrition Survey in Japan reported that the proportion of individuals sleeping less than six hours per day was 39.1% among people in their 20s. Across age groups, more than 40% of individuals in their 30s to 50s reported sleeping less than six hours on average[3].

In recent years, the market for wearable devices has expanded rapidly, and their applications in personal health management have become increasingly widespread. The global wearable device market grew significantly from USD 12.98 billion in 2018 to USD 49.81 billion in 2022—an almost fourfold increase[4].

The Japanese government has been promoting the integration of information and communication technologies (ICT), including Personal Health Records (PHRs), to enable individuals to access and utilize their own health information. The use of such digital technologies is also being encouraged in the domain of sleep improvement. However, to fully realize the potential of such systems, it is essential for users to actively engage in sleep improvement behaviors. To support this, it is important to collect and analyze individual sleep data with high granularity and deliver personalized advice tailored to each user's specific needs.

This study aims to develop a system that utilizes sleep data obtained from wearable devices to provide personalized sleep improvement recommendations. The system processes and analyzes the data in accordance with the following three primary objectives.

1. Quantitative assessment and analysis of sleep patterns
2. Provision of specific recommendations for sleep improvement
3. Optimization of a personalized behavioral intervention program for better sleep

Based on these objectives, the system processes sleep-related data and provides users with appropriate information, thereby promoting improvements in individual sleep quality and overall health. In addition, the system allows for an objective evaluation of users' efforts to improve their sleep and the subsequent changes in their sleep data.

This approach aims to enable effective sleep improvement at a preclinical stage, prior to direct intervention by healthcare professionals. By enhancing the efficiency of sleep management, the system also seeks to contribute to reducing the burden on medical services.

2 Related Work

2.1 Studies on the Accuracy of Sleep Assessment Using Wearable Devices

In recent years, the widespread adoption of wearable devices has enabled the real-time collection of physical activity and biometric data. These devices are equipped with multiple sensors, allowing for detailed assessment of physical activity in daily life.

In particular, by integrating heart rate data with accelerometer data, it is possible to assess sleep conditions, including sleep states and stages, regardless of environment or location. To accurately evaluate sleep improvements, the collected data undergoes preprocessing to remove noise and irrelevant segments, followed by feature extraction using statistical techniques.

Amagasa et al.[5] provided an overview of the current state of physical activity assessment using smartphones and wrist-worn devices, and discussed future directions. The reliability of wearable devices has been evaluated by comparing them to highly accurate methods traditionally used in research, such as the doubly labeled water method for measuring energy expenditure.

Wearable devices have demonstrated high accuracy even when compared with these gold-standard techniques, while offering greater simplicity and cost-effectiveness. In particular, they have shown high reliability in measuring moderate-to-vigorous physical activity, making them a viable option for research applications.

2.2 Sleep Habit Modification Using Wearable Devices

Wearable devices have gained importance as objective tools for assessing sleep and are considered essential in advancing efforts within the field of sleep research. Ashihara et al.[6] conducted a sleep improvement program using Fitbit devices, aiming to raise awareness of sleep habits and promote better sleep practices among younger individuals.

As a result, the Pittsburgh Sleep Quality Index – Japanese version (PSQI-J)[7] showed an improvement of approximately one point in the experimental group, although this difference was not statistically significant. Furthermore, the Athens Insomnia Scale scores improved by approximately one point in the experimental group and 3.5 points in the control group. These findings suggest that while a subjective sense of insomnia improved during the intervention period, there was no substantial enhancement in overall sleep quality.

While the study by Ashihara et al. highlighted the usefulness of wearable devices as tools for self-reflection on sleep, it also indicated that achieving long-term improvements in sleep habits requires the application of behavior change techniques and continuous engagement through wearable device feedback.

2.3 Improving Sleep Behavior Using Fitbit

Numerous studies have been conducted to improve sleep behaviors using Fitbit devices. For instance, Spina et al.[15] presented sleep data obtained from wearable devices such as Fitbit to participants and provided feedback and guidance based on these data, aiming to improve subjective symptoms of insomnia and Sleep–Wake State Discrepancy. However, their intervention involved face-to-face counseling by clinical psychologists and two telephone follow-ups, relying on human support. Therefore, the need for human intervention poses challenges for long-term and large-scale interventions. In this study, considering these challenges, we aim to eliminate human involvement by completing all intervention processes within a system, thereby enabling automated interventions over extended periods.

2.4 Proposed Method in This Study

Building upon these related studies, the present research aims to promote increased awareness of sleep through sleep assessments utilizing wearable device data, and to support the development of healthier sleep habits by providing regular, sleep-related feedback.

3 Proposed System

3.1 Research Overview

In this study, we aim to promote behavior change by utilizing biometric data obtained from consumer-grade wrist-worn wearable devices to evaluate and support users' sleep. The proposed system focuses on the following four primary objectives:

1. Quantitative assessment and analysis of sleep patterns
2. Provision of specific recommendations for sleep improvement
3. Optimization of a personalized behavioral intervention program for better sleep

These objectives were selected to enable users to better understand their own sleep patterns and to facilitate a dual approach targeting both lifestyle habits and sleep environments, thereby promoting more effective sleep improvement and behavioral change. Based on these objectives, the system ultimately aims to enhance users' quality of life by reducing daytime sleepiness and fatigue, alleviating physical and mental discomfort, and improving attention and decision-making abilities.

The proposed system, aligned with these objectives, selects personalized sleep-related advice based on collected data and generates sleep data tables and visual graphs. By utilizing these functionalities and the information provided, users are expected to gain greater awareness of their sleep conditions, thereby facilitating behavioral change.

The rationale behind the selection of each objective and the expected outcomes are as follows.

1. Quantitative Assessment and Analysis of Sleep Patterns: To objectively evaluate the user's current sleep condition and identify areas for improvement.
2. Suggestions for Sleep Improvement: To demonstrate that changes in awareness and environment in daily life can lead to continuous enhancement of sleep quality.
3. Optimization of Sleep Behavior Improvement Programs: To provide feasible sleep improvement strategies tailored to individual lifestyles and personal differences in sleep habits.

3.2 System Architecture

In this study, we designed and implemented a system that collects sleep data from users via wearable devices and effectively analyzes and visualizes the data. This system enables the construction of an environment where users can utilize their sleep-related data. The system architecture was developed with reference to the work of Takahashi et al. [8]. The proposed system consists of the following three main components:

1. Sensor Module: Continuously and non-invasively collects sleep-related data from users using various sensors embedded in the wearable device.

2. Information Processing Module: Stores the collected sensing data and performs processing and analysis using threshold values based on the user's age. It also selects appropriate information to encourage sleep improvement.

3. Information Presentation Module: Presents the processed and analyzed information in formats that are easy for users to understand, such as tables, graphs, and textual advice. The amount of textual information is deliberately minimized to ensure that users can quickly and easily comprehend the content.

The overview of the system proposed in this study is shown in 1. Each component of the system is described in the following sections.

3.3 Sensing Component

In the sensor module, we used the wrist-worn wearable device Fitbit Sense 2 [9]. This device is commercially available for consumers and provides developers with extensive API documentation. It is equipped with eight types of sensors: a 3-axis accelerometer, gyroscope, altimeter, GPS/GLONASS, heart rate sensor, multi-purpose electrical sensor, skin temperature sensor, and ambient light sensor. The collected data is transmitted via Bluetooth 5.0 to an iPhone and managed on the cloud. In this study, we mainly utilized sleep-related data from the available sensor outputs.

3.4 Data Processing Module

The data processing module is implemented as a backend system centered on a server. It is responsible for acquiring and storing various types of user data obtained from the sensor module, as well as managing dedicated communication for notifications and supporting information presentation functionality.

3.4.1 Server System

The server system employs a physical server with high stability and security. This ensures a secure environment for managing and processing the collected personal sleep data.

3.4.2 Data Acquisition and Storage

The data acquisition and storage processes were implemented using Node.js. When retrieving information from the Fitbit resource server, the industry-standard authorization protocol OAuth 2.0 was utilized[10]. Sensor data were obtained by sending requests from the client to the source server. For data storage, MongoDB Atlas, a document-oriented database, was employed.

3.4.3 Dedicated Notification Communication

The dedicated notification communication function was implemented using Python 3. Slack was selected as the notification destination. Using the Slack API, dedicated channels were created for each user, enabling the delivery of feedback and informational messages related to sleep.

3.4.4 Presentation Function

The presentation function consists of two main features. First, it provides personalized feedback based on individual sensing data, including tables and graphs related to sleep status, as well as indicators for improvement and relevant advice. Second, it delivers general sleep-related information to all users. This includes common sleep issues, recommended countermeasures, and the potential benefits of improvement, all presented in an accessible and easy-to-understand format.

3.4.5 Advice Generation

The advice is provided through the notification function, with the aim of supporting individual users in improving and maintaining good sleep quality. The content of the advice is based on sleep evaluation indicators, offering information tailored to promote better sleep practices. It consists of two components: personalized improvement suggestions based on individual sleep metrics, and general educational advice applicable to all users. The design of this advice generation process was informed by the study conducted by Takahashi et al. [8].

The first type of improvement advice, specific to each sleep metric, generates information daily by comparing eight types of sleep data against standard sleep recommendation values based on the sleep evaluation indices listed in Table 2 for each age group, without considering individual differences. The second type of educational advice presents knowledge and benefits for better sleep targeting all age groups. These pieces of advice are generated utilizing Google AI Studio, leveraging the ”Sleep Guide for Health Promotion 2023” [2] and e-health net, ”Rest and Mental Health” [11], as knowledge sources.

In this study, we compared outputs generated with and without Retrieval-Augmented Generation (RAG) in terms of informational accuracy and scientific validity.

Regarding descriptions related to caffeine, the output generated with RAG stated: “Caffeine has a stimulating effect, which can make it harder to fall asleep and reduce sleep quality by making sleep shallower. Since higher daily intake increases the likelihood of such effects, it is advisable to avoid

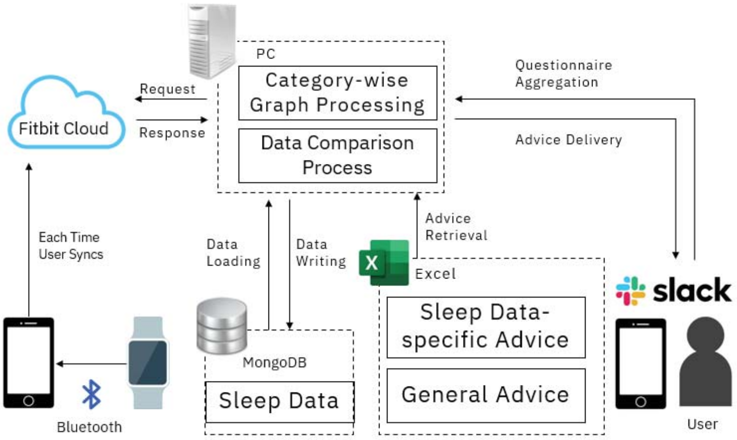


Figure 1: System Configuration Diagram

caffeine consumption particularly in the evening, and to moderate intake even during the day.” This message provided guidance considering both the timing and quantity of intake.

In contrast, the output generated without RAG stated: “Beverages containing caffeine, such as coffee, green tea, and energy drinks, should be avoided for at least four hours before bedtime. Also, it is wise to avoid alcohol before sleep as it significantly reduces sleep quality.” While this guidance may appear concrete at first glance, it overlooks individual variation in caffeine metabolism. According to the Ministry of Health, Labour and Welfare [2], the half-life of caffeine varies among individuals, typically ranging from approximately 3 to 7 hours. Therefore, for individuals with slower metabolism, caffeine consumed even four hours before bedtime may still negatively affect sleep. As such, the fixed “four-hour” recommendation lacks scientific rigor and could lead to misunderstanding.

Given that the RAG-based output presented more nuanced guidance that accounts for individual differences, it was considered to offer higher scientific validity.

Furthermore, several potentially inappropriate statements were identified in the output generated without RAG.

The first example is the following statement: “Consume foods rich in tryptophan, a precursor to the sleep hormone melatonin, such as dairy products (milk, yogurt), soy-based foods (natto, tofu), bananas, and nuts during dinner or as a light snack before bedtime.” Tryptophan is known to be metabolized into melatonin in the body and may contribute to sleep onset. However, studies have reported that eating right before bedtime can increase nocturnal awakenings and potentially impair sleep quality [16] For this reason, it is generally recommended that meals be consumed 4 to 6 hours before sleep. Therefore, the suggestion to consume a snack immediately before bedtime is considered inappropriate.

The second example is the recommendation: “After getting into bed, perform diaphragmatic breathing by slowly inhaling through the nose and exhaling through the mouth. In particular, the ‘4-7-8 breathing method’ (inhale through the nose for 4 seconds, hold the breath for 7 seconds, and exhale through the mouth for 8 seconds) is highly recommended.” Breathing techniques have been shown to potentially improve sleep quality. Tsai et al. [17], for instance, reported a significant improvement in sleep quality following the introduction of slow breathing exercises. The Ministry of Health, Labour and Welfare [2] also suggests the effectiveness of diaphragmatic breathing. However, according to the study by Vierra et al. [18], interventions using the “4-7-8 breathing method” did not demonstrate statistically significant effects among healthy young adults. In other words, although breathing techniques remain a promising intervention, clinical evidence supporting the efficacy of specific methods is still limited. Thus, recommending a particular breathing method without sufficient evidence is considered inappropriate.

These findings indicate that outputs generated without RAG, although seemingly helpful, often included scientifically unsupported or potentially misleading statements. In contrast, outputs generated with RAG were based on credible external sources and were judged to be superior in both validity and scientific consistency.

3.5 Information Presentation Section

The information presentation section is composed of presentation functions designed to support sleep improvement and the maintenance of good quality sleep for each user. Furthermore, we utilize the communication chat tool ”Slack” to provide individualized information for sleep improvement and the maintenance of good quality sleep to each user.

The presentation function plays the role of notifying the user of tables and graphs of acquired sleep data, as well as advice content. In the sleep data table, for the eight items of ”sleep efficiency,” ”total sleep time,” ”time to fall asleep,” ”number of awakenings,” ”wake time,” ”REM sleep ratio,” ”light sleep ratio,” and ”deep sleep ratio,” it shows the previous day’s sleep data and the appropriate range for the individual’s age group. Graphs are created weekly, with the X-axis representing the dates of the 7 days and the Y-axis representing each sleep data point, and the numerical values are color-coded based on sleep evaluation indices as green (appropriate), red (inappropriate), and blue (unknown). Furthermore, graphs are created for the three items of ”total sleep time,” ”sleep efficiency,” and ”time to fall asleep.” An example of a

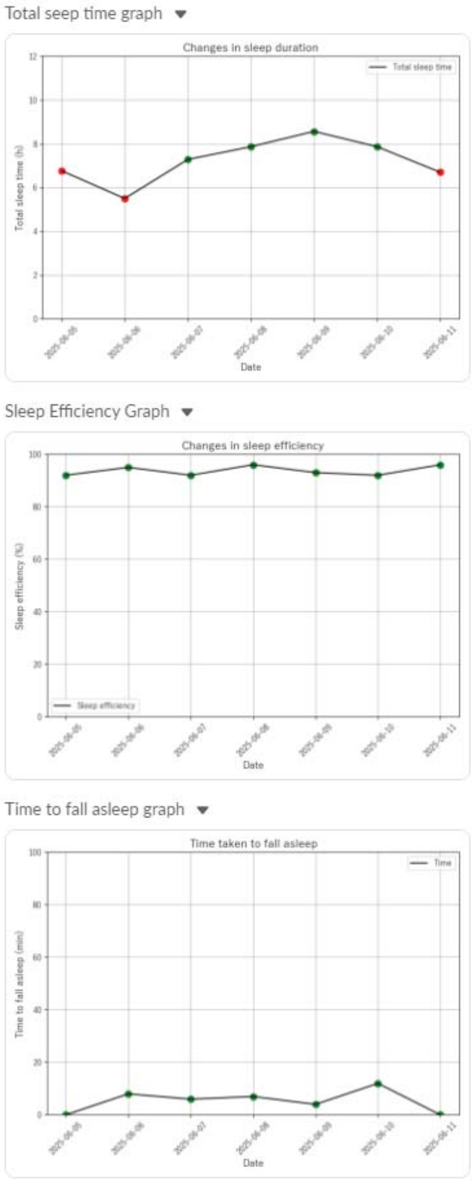
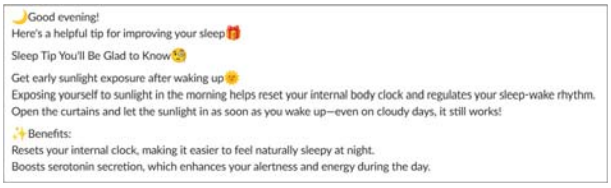


Figure 2: Example graphs to notify



🌙Good evening!
Here's a helpful tip for improving your sleep🎁

Sleep Tip You'll Be Glad to Know🧐

Get early sunlight exposure after waking up🌞
Exposing yourself to sunlight in the morning helps reset your internal body clock and regulates your sleep-wake rhythm.
Open the curtains and let the sunlight in as soon as you wake up—even on cloudy days, it still works!

✨Benefits:
Resets your internal clock, making it easier to feel naturally sleepy at night.
Boosts serotonin secretion, which enhances your alertness and energy during the day.

Figure 3: Examples of Advice

graph to be notified is shown in 2. Additionally, it plays the role of notifying a Google Forms to judge whether the advice content is appropriate. An example of the advice content to be notified is shown in 3.

3.6 Sleep Evaluation Indices

Fitbit Sense 2 performs sleep stage (Wake, REM, Light, Deep) and various sleep assessments by combining movement and heart rate patterns. The sleep assessment items used in this system are ”total sleep time,” ”sleep efficiency,” ”time to fall asleep,” ”number of awakenings lasting 5 minutes or more,” ”wake after sleep onset,” ”REM sleep ratio,” ”light sleep ratio,” and ”deep sleep ratio,” which are evaluated by age group in Sleep Health [12][13].

Total sleep time is the total duration of actual sleep. Sleep efficiency is also included in the PSQI question items, and the definition of sleep efficiency follows Arai et al. [14]. Sleep efficiency is the ratio of sleep time to the time spent in bed. In this system, sleep efficiency is calculated from the time the user was asleep divided by the total time spent in bed.

Time to fall asleep is the duration from going to bed until falling asleep. In this system, it is calculated using the time from going to bed until entering any of the ”REM sleep,” ”light sleep,” or ”deep sleep” stages. The number of awakenings lasting 5 minutes or more represents how many times the user was awake for 5 minutes or longer between going to bed and waking up. In this system, it is calculated by obtaining the wake state and counting instances where it continued for 5 minutes or longer. Wake after sleep onset indicates the total duration of wakefulness between going to bed and waking up. This system calculates it by obtaining the wake state and summing those durations.

Sleep Health is evaluated using the RAND/UCLA Appropriateness Method, classifying each sleep index into ”appropriate,” ”uncertain,” and ”inappropriate.” According to the Ministry of Health, Labour and Welfare [11], N1 sleep and N2 sleep are classified as light sleep, and N3 is classified as deep sleep. Therefore, since it is difficult to determine the appropriate values for N1 and N2, they are evaluated as light sleep. In this system, using 5%, which is the appropriate value for N1, and 80%, which is the uncertain range for N2, if the value is greater than 85%, it is judged as inappropriate.

Regarding total sleep time, the Ministry of Health, Labour and Welfare [2] also mentions it, stating that the recommended total sleep time for adults is 6 to 8 hours. However, Sleep Health [13] considers 7 to 10 hours as appropriate, showing some discrepancy. Therefore, 10 hours of sleep is significantly outside the recommendation range of the Ministry of Health, Labour and Welfare, making it difficult to judge and is treated as ”unknown.” The sleep evaluation indices for young adults in this system are shown in 1.

Furthermore, the Athens Insomnia Scale is used as a sleep evaluation index. The Athens Insomnia Scale consists of items related to sleep and items related to daytime functioning. Each question is answered on a 4-point scale from 0 to 3, and the total score determines the sleep condition. A total score of 0 to 3 indicates ”no insomnia,” 4 to 5 indicates ”possible insomnia,” and 6 or more indicates ”probable insomnia.” By comparing this score before and after using the system, it is evaluated whether there has been an improvement in sleep.

4 Demonstration Experiment

4.1 Overview

To evaluate the effectiveness of the proposed system, a user study was conducted with five participants. Prior to the start of the experiment, a questionnaire was administered to collect

Table 1: Sleep assessment index for young adults.

Index (Unit)	Appropriate	Uncertain	Inappropriate
Total Sleep Time (h)	7–9	6–7, 10–11	6, 11
Sleep Efficiency (%)	85–100	65–85	65
Time to Fall Asleep (min)	0–30	31–45	46
Number of Awakenings (times)	0–1	2–3	4
Wake After Sleep Onset (min)	0–20	21–40	41
REM Sleep Ratio (%)	N/A	0–40	41–100
N1 Sleep Ratio (%)	0–5	6–20	21–100
N2 Sleep Ratio (%)	N/A	0–80	81–100
N3 Sleep Ratio (%)	N/A	6–100	5–100

data using the Athens Insomnia Scale as well as participant background information. Specifically, the questionnaire included items related to participants' age, awareness of sleep, and their physical and mental states during the day and at bedtime.

The system validation experiment was conducted over a four-week period. During Phase A, participants wore the Fitbit Sense 2 device while maintaining synchronization with their iPhones, and continued their sleep as usual. In Phase B, the system was configured to provide feedback, including advice and graphical reports. This setup enabled an evaluation of how the feedback provided by the system influenced participants' behavior toward sleep improvement. The experimental schedule is shown in [Table IV], and participant demographics are summarized in [Table V].

4.1.1 Pre-Experiment Questionnaire

To collect data regarding participant backgrounds and the Athens Insomnia Scale (AIS), two separate questionnaires were administered via Google Forms. The first questionnaire consisted of Questions 1–13, while the AIS questionnaire comprised Questions 14–21. Questions 1–4 addressed subjective sleep assessments, including participants' age and satisfaction with their sleep. Questions 5–8 focused on current lifestyle habits. Questions 9–13 were designed to assess participants' efforts in gathering information related to sleep improvement and their awareness of sleep intervention programs.

4.1.2 Post-Experiment Questionnaire

After the experiment, a post-experiment questionnaire was administered via Google Forms to evaluate the effectiveness of the proposed system, changes in participants' sleep awareness and behavior, and insomnia symptoms based on the Athens Insomnia Scale (AIS). The questionnaire consisted of 20 items (Questions 1–20) along with an 8-item AIS survey (Questions 1–8). Questions 1–3 assessed changes in sleep patterns, Questions 4–11 evaluated the system itself, Questions 12–15 examined the influence of the system on behavior change, and Questions 16–20 focused on participants' overall assessment of the experiment.

This allows for a comprehensive evaluation of how the system influenced the user's awareness, interest in sleep, and behavioral changes. The questionnaire content for evaluating system satisfaction is shown in 4.

5 Evaluation

5.1 Evaluation Items

To evaluate the effectiveness of the proposed sleep improvement and good sleep maintenance in this study, quantitative and qualitative evaluations will be conducted.

5.1.1 Quantitative Evaluation

In the quantitative evaluation, a demonstration experiment will be conducted by setting two periods: a period where only the device was worn (Period A) and a period where the proposed content was presented in addition to wearing the device (Period B). Through a comparison of both periods, changes in sleep data and the Fitbit wearing rate will be examined to evaluate behavioral changes.

For each sleep data point, the average values during Period A and Period B will be calculated, and the difference and rate of change will be evaluated. Through these evaluations, the overall impact of the proposed content on the participants' sleep improvement will be examined. Additionally, the scores of the Athens Insomnia Scale will be calculated, and their increase or decrease will be evaluated.

In addition to user evaluations, a quantitative assessment of the system's operational status was conducted. Specifically, the system performance was examined from two perspectives: the success rate of data acquisition via sensor integration, and the success rate of notification and feedback delivery.

5.1.2 Qualitative Evaluation

In the qualitative evaluation, the appropriateness of the advice and behavioral changes will be assessed through a questionnaire survey. First, a subjective evaluation of sleep changes will be conducted. Next, regarding the effectiveness of the advice and the system, it will be evaluated whether the provided advice and information presentation methods were appropriate. Finally, using this system, it will be assessed whether changes in awareness regarding sleep and behavioral changes for sleep improvement were promoted.

Table 2: Schedule of the Demonstration Experiment

Phase	Period
Phase A	November 18, 2024 – December 1, 2024 (14 days)
Phase B	December 2, 2024 – December 15, 2024 (14 days)

Table 3: Attributes of Participants

Participant	Gender	Age Group
A	Female	20s
B	Male	20s
C	Male	20s
D	Male	20s
E	Male	20s

5.2 Evaluation Results

5.2.1 Results of Quantitative Evaluation

The results concerning the sleep indices of each participant in the demonstration experiment and the results of the Athens Insomnia Scale will be evaluated.

First, regarding total sleep time, in Period B, the average sleep time of Participant A increased by approximately 13%, and Participant C's increased by approximately 9%, indicating a notable increase in sleep duration. Conversely, Participant E's sleep time decreased by approximately 16%, and Participant D's decreased by approximately 5%. No significant change was observed in Participant B. It should be noted that Participants D and E, whose sleep times decreased, had sleep durations within the appropriate range of 7 to 9 hours in both Period A and Period B. Detailed data are shown in 5.

Next, regarding sleep efficiency, a slight increase in sleep efficiency was observed for Participants A, B, and E during Period B. Conversely, a slight decrease in sleep efficiency was observed for Participants C and D. From these results, no clear increase or decrease in sleep efficiency due to this system was confirmed. Additionally, since 4 out of 5 participants had sleep efficiency of 85% or higher in both periods, it was confirmed that there were no issues with their sleep efficiency.

However, Participant D showed an inappropriate value of 65% or less in both periods. Since an improvement in sleep is confirmed by the Athens Insomnia Scale, there is a possibility that this is due to a device malfunction or the storage condition when not worn. Detailed data are shown in 6.

Next, regarding the time to fall asleep, in Period B, the time to fall asleep for Participant B decreased by approximately 26%, Participant C by approximately 13%, and Participant D by approximately 24%, indicating an improvement in the time to fall asleep. On the other hand, Participant A showed an increase of approximately 41%, and Participant E showed an increase of 50%; however, since both values were within the appropriate range, it was confirmed that all participants were able to fall asleep appropriately. Detailed data are shown in 7.

Next, regarding the average number of awakenings lasting 5 minutes or more, in both periods, except for Participant C in Period A, the average number of awakenings for the other participants was within the range of 2 to 4 times, not reaching the recommended 2 times or less. Furthermore, since the range of 2 to 4 times is considered uncertain, it is necessary to further reduce the number of awakenings. On the other hand, Participant E showed a decrease of 1.5 times in the average number of awakenings, indicating improvement. From these results, different changes were observed for each participant, but no significant improvement was achieved. Detailed data are shown in 8.

Next, regarding the average wake time, in Period B, the average wake time for Participant B decreased by approximately 7%, and for Participant E by approximately 28%. In particular, Participant E showed a wake time of over 40 minutes in Period A, but an improvement in wake time was confirmed in Period B. On the other hand, Participant A showed an increase of approximately 3%, Participant C approximately 5%, and Participant D approximately 20%. It should be noted that for participants other than Participant E, no change in the evaluation of the indices in both periods was confirmed. Detailed data are shown in 9.

Next, regarding sleep stages, while there are no established appropriate reference values for the ratios of REM sleep, light sleep, and deep sleep used in this experiment, the ratios for all participants were categorized as uncertain in both periods, and no values considered inappropriate were observed.

Finally, regarding the Athens Insomnia Scale, in the pre-experiment questionnaire, the Athens Insomnia Scale scores of all participants were 6 points or higher, indicating "probable insomnia." In the post-experiment questionnaire, Participant A's score decreased by 11 points, and Participant B's score decreased by 5 points, showing improvement to "no insomnia." Furthermore, Participant C's score decreased by 1 point, and Participant D's score decreased by 4 points, indicating improvement to "possible insomnia." No improvement in the score was observed for Participant E. The Athens Insomnia Scale score results from the pre-experiment and post-experiment questionnaires are shown in 10.

The system was designed to automatically retrieve participants' daily sleep data using the Fitbit API. As a benchmark, successful data acquisition was expected on all scheduled usage days. In practice, data acquisition was successful on 103 out of 140 scheduled days, resulting in the following operational rate:

$$\text{Operational Rate} = \frac{103}{140} \times 100 = 73.6\%$$

The primary causes of data acquisition failures were user-related, such as participants not wearing the Fitbit device or not launching the Fitbit app. No failures due to system-side malfunctions were observed. However, this rate shows a 26% deviation from the benchmark, indicating room for improvement.

Table 4: End of the experiment Questionnaire.

No.	Question
Q1	Please provide your experiment participant number.
Q2	Compared to before using this sleep improvement program, how satisfied are you with your sleep?
Q3	Compared to before using this sleep improvement program, have you become more interested in sleep?
Q4	Compared to before using this sleep improvement program, have you deepened your knowledge about the impact of sleep on your health and daytime activities?
Q5	How often did you use this sleep improvement program?
Q6	Do you think the notification time of this sleep improvement program was appropriate?
Q7	To have a better notification time, what time of day would you have preferred?
Q8	To what extent did you utilize the Slack notification function?
Q9	How did you feel about the advice based on sleep data and the content of the summarized data table?
Q10	How did you feel about the advice content for all users?
Q11	How did you feel about the weekly advice and its graph content?
Q12	Was the information provided by this sleep improvement program easy for you to understand?
Q13	Have you started doing anything new since using this sleep improvement program?
Q14	Through this sleep improvement program, what type of importance did you become particularly interested in?
Q15	Are there any actions you stopped taking through the information and advice provided by this sleep improvement program?
Q16	Was the information and advice provided by this sleep improvement program helpful in changing your behavior?
Q17	How satisfied were you overall with this sleep improvement program?
Q18	(For those who answered ”satisfied” or ”somewhat satisfied” in Question 17) Please tell us the reasons why you were satisfied with using this sleep improvement program.
Q19	Would you like to continue using this sleep improvement program in the future?
Q20	What kind of functions would you like to see added to this sleep improvement program in the future?

Table 5: Average total sleep time over 14 days.

Participant	Period A (hours)	Period B (hours)
A	6.9	7.8
B	6.9	6.9
C	5.5	6.0
D	8.4	8.0
E	8.7	7.3

Table 6: Average Sleep efficiency over 14 days.

Participant	Period A (%)	Period B (%)
A	95.9	96.4
B	93.6	93.6
C	93.3	92.8
D	58.5	54.3
E	95.3	95.5

Table 7: Average time to fall asleep over 14 days.

Participant	Period A (minutes)	Period B (minutes)
A	7.1	10.1
B	9.7	7.1
C	4.3	3.7
D	14.2	10.8
E	12.0	18.0

Table 8: Average number of awakenings over 14 days.

Participant	Period A (times)	Period B (times)
A	3.0	2.9
B	3.6	3.2
C	1.9	2.1
D	2.3	2.9
E	3.5	2.0

As part of the intervention, periodic informational notifications were implemented, with a total of 210 notifications scheduled. Of these, 195 were successfully executed, resulting in the following success rate:

$$\text{Success Rate} = \frac{195}{210} \times 100 = 92.9\%$$

Among the 15 failures, 10 were due to system-related issues: on two notification occasions, all five intended recipients failed to receive the notification, accounting for a total of 10 failures. In addition, one notification process experienced a system malfunction that caused a delay, resulting in incomplete delivery to all five recipients, leading to another 5 failures.

These results indicate that the system operated with generally high reliability. However, the acquisition of wearable device data was found to be highly dependent on user behavior, which had a significant impact on the overall operational rate.

For future improvements, reducing user dependency is a primary challenge. Specifically, implementing a feature that automatically detects whether the Fitbit device is being worn and whether the Fitbit app is running—and sends real-time reminders accordingly—would help mitigate the risk of missed data acquisition or device non-usage in advance.

Additionally, the current system operates mainly in a Win-

Table 9: Average wake time over 14 days.

Participant	Period A (minutes)	Period B (minutes)
A	36.0	37.1
B	34.4	32.0
C	18.9	19.9
D	32.4	38.9
E	40.4	29.0

Table 10: Athens Insomnia Scale Scores Before and After the Experiment.

Participant	Pre-Experiment	Post-Experiment
A	13	2
B	6	1
C	6	5
D	8	4
E	8	8

dows environment. To improve stability, maintainability, and scalability, migration to a Linux-based system is recommended.

Furthermore, conducting repeated demonstration experiments will help establish a framework for detecting and correcting unforeseen system failures prior to actual deployment. Through these technical enhancements, the system can achieve full automation and long-term stable operation, thereby enabling more continuous and effective sleep behavior interventions.

5.2.2 Results of Qualitative Evaluation

The questionnaire focused on sleep changes (Questions 1-3), evaluation of this system (Questions 4-11), the effect and impact on behavioral change (Questions 12-15), and evaluation of this experiment (Questions 16-20).

First, regarding the evaluation results of changes in sleep satisfaction, to the question asking about the degree of satisfaction compared to before using this system, 20% answered ”satisfied” and 80% answered ”somewhat satisfied.” Furthermore, to the question asking whether they became more interested in sleep, 40% answered ”significantly more interested” and 60% answered ”somewhat more interested.” From these results, it was confirmed that many participants were satisfied with the improvement in their sleep and became more interested in it.

Regarding the evaluation results of this system, to the question about how often they used the system, 40% answered ”every day” and 40% answered ”4 to 6 times a week.” Furthermore, 20% answered ”1 to 3 times a week.” This result confirms that for some participants, the system was easily accessible enough to be used daily, but it also suggests that there were individual differences in the frequency of use.

Regarding the effect and impact on behavioral change, to the question asking whether the provided information and advice were helpful in changing behavior, all participants answered ”somewhat helpful.” Furthermore, the most common new action taken was ”started to refrain from caffeine and alcohol intake,” with 40% of participants selecting this. On the other hand, the most common action stopped was ”stopped eating high-calorie meals at night,” with 60% of participants selecting this. These results indicate that participants were motivated to change their behavior, and specific actions aimed at improving sleep were implemented.

Furthermore, as reasons for satisfaction with this system, in addition to improvements in sleep satisfaction such as ”reduced nighttime awakenings and feeling refreshed upon waking up in the morning,” improvements in daytime physical condition were reported, such as ”improved physical condition and increased daytime concentration” and ”reduced stress and improved mental well-being.” These changes are likely behaviors adopted due to the information provided by the system and increased awareness of sleep improvement. In addition, there were responses such as ”developed a habit of appropriate exercise” and ”was able to review bedding and bedroom environment,” indicating that the system’s advice effectively promotes behavioral change.

6 Discussion

In this study, we evaluated the behavioral changes related to sleep for each participant based on sensing data and questionnaire results. By using tables and graphs for information presentation, we enabled participants to visually confirm their own sleep conditions. However, the number of advice provided was insufficient, and there were repetitions in the presented content. Therefore, it is necessary to enhance the advice content further. Additionally, the quality of advice could not be improved due to the insufficient classification of user sleep attributes.

Regarding the Fitbit wearing rate, only Participant A showed an increase from 50.0% to 57.1% during the demonstration experiment period. A possible reason for this is that Participant A made an error in operating the Fitbit and could not acquire data properly, so they increased the appropriate wearing frequency in the latter half. On the other hand, Participants C and D showed a 14.3% decrease in the Fitbit wearing rate during Period B. A possible reason for this is that interest in wearing the device waned and motivation to participate in the experiment decreased as the demonstration experiment progressed.

Participant E was the only one whose Athens Insomnia Scale score did not decrease. A possible reason for this is that their Fitbit wearing rate was particularly low, with the device worn only 2 out of 14 days in Period B. These results indicate that a lack of interest in the system or sleep hinders sleep improvement, making progress difficult. Therefore, we believe that fostering interest in sleep and motivation to engage in sleep improvement behaviors are crucial for achieving sleep improvement.

The average number of awakenings lasting 5 minutes or more tended to be high overall. A possible reason for this is that this demonstration experiment was conducted on university students, who generally have more time flexibility compared to working adults. University students tend to have more leeway in the morning and can easily secure time for ”snoozing” or returning to sleep after initial awakening, which may have increased the number of awakenings. However, this high number of awakenings indicates that sufficient rest and sleep satisfaction are not being obtained. Therefore, this is an index that has the potential to be resolved as sleep improvement progresses.

Based on the results of Section 5.2, the resolution of sleep deprivation was confirmed in 4 out of 5 participants. For sleep improvement, it is crucial for users themselves to increase their interest in sleep and acquire appropriate knowledge, and further improvement is expected. Therefore, it is necessary to generate personalized advice according to individual sleep data and changes in lifestyle habits, thereby increasing user interest and enabling the acquisition of appropriate knowledge.

In this study, we evaluated the behavioral changes related to sleep for each participant based on sensing data and questionnaire results. Regarding sleep, it was suggested that advice and data visualization contributed to the improvement of sleep behavior. On the other hand, issues regarding the reliability of the device were also revealed, indicating the need for future system improvements.

7 Conclusion

In this study, we proposed a system that utilizes wearable devices and promotes the improvement of sleep behavior through the provision of information based on sleep data. In this system, participants wore wearable devices, and based on the sleep data obtained from the devices, individual areas for improvement were presented, and behavioral advice for improving sleep quality was provided.

As a result of the demonstration experiment, an improvement in sleep behavior due to this system was confirmed. In particular, among the three participants with insufficient total sleep time, an increase in total sleep time was observed in two of them. However, for the other sleep indices, individual differences were observed, and there was no consistency in the presence or absence of improvement, and no significant results were obtained. Furthermore, the Athens Insomnia Scale scores showed improvement in four out of five participants before and after the demonstration experiment.

The evaluation of the information provision system highlighted issues such as the repetition of advice for each sleep data point and the insufficient classification of user attributes. As future improvements, it is necessary to build a system that can provide sufficient and non-repetitive advice by utilizing AI for information generation. Furthermore, by classifying users according to their sleep attributes, we aim to build a system that can provide more personalized advice and enhance the effectiveness of sleep improvement.

REFERENCES

[1] OECD. Gender data portal. https://www.oecd.org/en/topics/policy-issues/gender-equality.html, (accessed Mar. 20, 2025).

[2] Ministry of Health, Labour and Welfare. Sleep Guide for Health Promotion 2023. https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/kenkou_iryou/kenkou/suimin/index.html, (accessed Mar. 20, 2025).

[3] Ministry of Health, Labour and Welfare. 2019 National Health and Nutrition Survey Report. https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/kenkou_iryou/kenkou/eiyou/r1-houkoku_00002.html, (accessed Mar. 20, 2025).

[4] Ministry of Internal Affairs and Communications. Trends and Forecasts of the Global Wearable Device Market Size. https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/r05/html/datashu.html, (accessed Mar. 20, 2025).

[5] S. Amagasa, et al., ”Evaluation of Physical Activity Using Smartphones and Wearable Devices in the Medical and Health Fields: Current Status and Future Prospects,” *Japanese Journal of Public Health*, vol. 68, no. 9, pp. 585-596, 2021.

[6] M. Ashihara, et al., ”An Attempt at a Sleep Habit Change Program Using Wearable Terminals,” *Bulletin of the Department of Junior Sports Education*, no. 6, pp. 1-8, 2018.

[7] Y. Doi, et al., ”Creation of the Japanese Version of the Pittsburgh Sleep Quality Index,” *Psychiatric Treatment*, vol. 13, no. 6, pp. 755-769, 1998.

[8] Y. Takahashi, et al., ”A Study on Activity Support Methods for Health Maintenance and Promotion Using Real-Time Data,” in *Proc. 112th Mobile Computing and New Social Systems, 83rd Ubiquitous Computing Systems, 41st Consumer Devices and Systems, and 30th Geriatric Society Design Joint Research Presentation*, vol. 2024-CDS-41, no. 13, pp. 1-8, 2024.

[9] Google Store. Fitbit Sense 2 Smartwatch. https://store.google.com/product/fitbit_sense_2?hl=ja, (accessed Mar. 20, 2025).

[10] D. Hardt, Ed., *The OAuth 2.0 Authorization Framework*. Microsoft, https://datatracker.ietf.org/doc/html/rfc6749#section-4.1, (accessed Mar. 20, 2025).

[11] Ministry of Health, Labour and Welfare. e-Health Net Rest and Mental Health. https://www.e-healthnet.mhlw.go.jp/information/heart, (accessed Mar. 20, 2025).

[12] M. Ohayon, et al., ”National Sleep Foundation’s sleep quality recommendations: first report,” *Sleep Health*, vol. 3, pp. 6-19, 2017.

[13] M. Hirshkowitz, et al., ”National Sleep Foundation’s sleep time duration recommendations: methodology and results summary,” *Sleep Health*, vol. 1, pp. 40-43, 2015.

[14] H. Arai, et al., ”Association between subjective sleep quality and physical activity and psychological adaptation,” *Japanese Journal of Psychosomatic Medicine*, vol. 46, pp. 667-676, 2006.

[15] Spina MA, Andrillon T, Quin N, Wiley JF, Rajaratnam SMW, Bei B. “Does providing feedback and guidance on sleep perceptions using sleep wearables improve insomnia? Findings from ”Novel Insomnia Treatment Experiment” : a randomized controlled trial”, *Sleep*, Vol.46, No.9, zsad167, 2023.

[16] Iao SI, Jansen E, Shedden K, O’Brien LM, Chervin RD,

Knutson KL, Dunietz GL. “Associations between bedtime eating or drinking, sleep duration and wake after sleep onset : findings from the American time use survey”. *Br J Nutr*, Vol.127, No.12, pp.1-10, 2021.

[17] Tsai HJ, Kuo TBJ, Lee GS, Yang CCH, “Efficacy of paced breathing for insomnia : Enhances vagal activity and improves sleep quality”. *Psychophysiology*, Vol.52, No.3, pp.388-396, 2015.

[18] Vierra J, Boonla O, Prasertsri P, “Effects of sleep deprivation and 4-7-8 breathing control on heart rate variability, blood pressure, blood glucose, and endothelial function in healthy young adults”. *Physiol Rep*, Vol.10, No.13, e15389, 2022.

FocusSense: Construction and Evaluation of a Multimodal Concentration Estimation Application

Noriyuki Tanaka†, Ko Watanabe‡, Shoya Ishimaru†, Andreas Dengel‡, Shingo Ata†, and Manato Fujimoto†*

†Osaka Metropolitan University, Japan
‡DFKI GmbH, Germany
* RIKEN Center for Advanced Intelligence Project AIP
sd24263r@st.omu.ac.jp, {ko.watanabe, andreas.dengel}@dfki.de, {ishimaru, ata, manato}@omu.ac.jp

Abstract - In the field of education, the spread of COVID-19 has led to a rapid increase in remote lectures, making it difficult to assess student concentration in online learning environments. This study collects multimodal sensor data from students during online lectures to estimate their levels of concentration. The data include facial orientation, gaze direction, acceleration, gyroscope data, and heart rate obtained via webcams and wearable devices. Machine learning and deep learning models are used to analyze the data and classify concentration states. We also developed a visualization application that presents the results intuitively. This tool supports students in reflecting on their learning and helps instructors adjust teaching strategies by clearly showing concentration trends. This study highlights the potential of integrating sensing and visualization technologies to enhance online learning and promote a better understanding of students' cognitive states, contributing to more effective and adaptive educational practices.

Keywords: wearable sensor, multimodal sensing, concentration detection, online learning, machine learning

1 Introduction

Since the onset of the COVID-19 pandemic, education has transitioned into an era that demands diverse and adaptable learning environments. According to a survey conducted by the Ministry of Education, Culture, Sports, Science, and Technology (MEXT), many universities adopted online lectures, particularly in 2020 in Japan [1]. Presently, 896 out of 1069 universities in Japan offer a combination of face-to-face and online lectures, solidifying online education as an integral mode of learning. Although the pandemic has subsided and face-to-face classes have resumed, many universities and educational institutions continue to offer education in a hybrid format. Consequently, it is essential to comprehensively address both offline and online learning environments to enhance educational performance.

However, a significant challenge in online education lies in accurately capturing students' cognitive states, which is inherently more complex than in face-to-face settings. This difficulty arises from the limited ability to observe nonverbal cues in online environments [2, 3]. In this context, sensing technologies have emerged as promising tools to visualize cognitive states, including attention and engagement [4, 5, 6, 7]. Among these cognitive states, concentration is particularly vital as it directly impacts learning efficiency and achievement. Accurate estimation and visualization of concentration are thus critical for improving the quality of education.

This study estimates concentration using multimodal sensing technology, integrating data from cameras (facial orientation, gaze) and wearable sensors (acceleration, gyroscope, heart rate). The developed application, *FocusSense*, provides concentration visualization to offer new insights for both instructors and students. This allows instructors to optimize content and students to enhance self-regulation by reflecting on their learning data, positioning the application as a valuable tool for a more personalized educational experience.

The main contributions (C1–C2) of this paper are summarized as follows:

- C1 **Construction of a concentration estimation model using multimodal sensors:** We built a concentration estimation system that integrates face and gaze data from a webcam, along with accelerometer, gyroscope, and heart rate data from wearable devices. The proposed model achieved a mean F1 score of 0.69 in user-dependent (UD) validation and 0.57 in user-independent (LOPO) validation.

- C2 **Development of the concentration visualization application *FocusSense*:** We developed and implemented a visualization application, *FocusSense*, which enables students and educators to utilize the estimated changes in concentration for reflection on their learning and teaching strategies.

This paper is organized as follows. Section 2 provides an overview of related work in sensor-based activity recognition and cognitive state estimation. Section 3 outlines the methodology, which includes data preprocessing, feature extraction, and the machine learning and deep learning models employed for concentration classification. Section 4 describes the experiment participants and the data collection procedure. In Section 5, the evaluation results of the proposed concentration estimation model are presented and discussed. Section 6 introduces the *FocusSense* application prototype and its user interface. Section 7 addresses the limitations of this study, and finally, Section 8 concludes the paper.

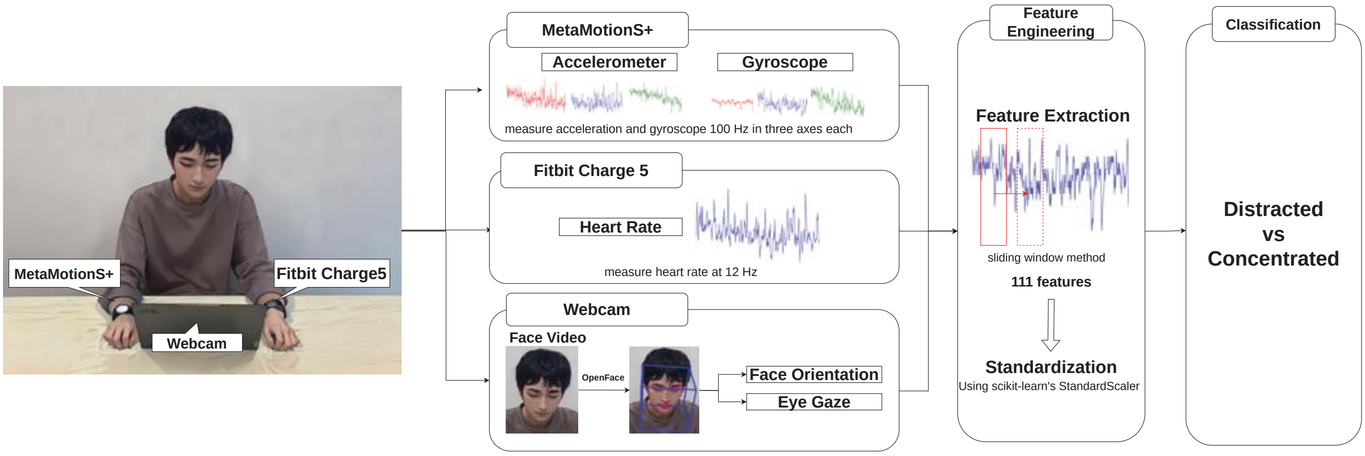


Figure 1: Overview of the data collection and analysis workflow.

2 Related Work

In this section, we organize the related work from three perspectives: "Sensor-Based Activity Recognition", "Cognitive Estimation Applications", and "Concentration Estimation". By surveying the trends and achievements in these fields, we identify the challenges and clarify the novelty of our research.

2.1 Sensor Based Activity Recognition

Previous studies on activity recognition have utilized a wide range of sensors, including wrist-worn devices [8, 9], cameras [10, 11], and microphones [12]. For instance, Bhatt et al. [10] proposed an appearance-based approach for gaze estimation using a webcam. Closer to our work, Muragi et al. [13] used hand and eye position information to recognize disruptions during assembly tasks, demonstrating the utility of multimodal sensing for understanding user states.

2.2 Cognitive Estimation Application

Applications for cognitive state estimation often involve immersive environments. For example, some systems use brain-computer interfaces and EEG sensors within VR to visualize concentration [14]. Asahi et al. [14] developed an immersive VR system that provides multimodal feedback based on brainwave and gaze data to enhance concentration. Similarly, You et al. [15] investigated effective methods for visualizing EEG-based concentration scores in VR, finding that environmental changes improved user experience. These studies highlight a trend towards using immersive feedback to modulate cognitive states.

2.3 Concentration Estimation

Research on concentration estimation has been conducted using various approaches. Tanaka et al. [16] is working on estimating the state of concentration by tracking eye movements. In this study, a machine-learning model is built using eye movement data, and a pipeline is explored for recognizing the state of concentration. Betto et al. [17] proposed a method for detecting student distraction in e-learning lectures based on student face and posture information collected from webcams. Kimura et al. [18] proposed a method for estimating intellectual concentration using pupil diameter and heart rate variability. Tanaka et al. [19] proposed a method to estimate concentration using arm acceleration, gyroscope data, heart rate, face orientation, and eye gaze. This method achieved an average accuracy of 66%. However, in that study, we validated the approach using only traditional machine learning models and did not explore the use of deep learning models.

Table 1: Lists of features extracted while applying the sliding window [19].

Function	Definition	Algorithm
$mean(s)$	Mean	$\bar{s} = \frac{1}{N}\sum_{i=1}^{N} s_i$
$max(s)$	Maximum	$\max_i(s_i)$
$min(s)$	Minimum	$\min_i(s_i)$
$std(s)$	Standard Deviation	$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(s_i - \bar{s})^2}$
$mad(s)$	Median Absolute Deviation	$\mathrm{median}_i\left(\lvert s_i - \mathrm{median}_j(s_j)\rvert\right)$
$energy(s)$	Mean Square	$\sum_{i=1}^{N} s_i^2$
$sma(s_1, s_2, s_3)$	Signal Magnitude Area	$\frac{1}{3}\sum_{i=1}^{3}\sum_{j=1}^{N}\lvert s_{i,j}\rvert$
$iqr(s)$	Interquartile Range	$Q_3(s) - Q_1(s)$
$range(s)$	Range	$\max_i(s_i) - \min_i(s_i)$
$rms(s)$	Root Mean Square	$\sqrt{\frac{1}{N}\sum_{i=1}^{N} s_i^2}$

3 Methodology

Figure 1 presents an overview of the workflow. This workflow comprises several key stages: data collection from multiple sensors, feature engineering, and, finally, classification to distinguish between concentrated and distracted states.

3.1 Preprocessing

This study used the sliding window method to extract features. We apply window sizes of 2.5, 5, 10, and 15 seconds and overlaps of 0, 0.25, 0.5, and 0.75. We extracted features using 16 different methods and verified which combination of window size and overlap yielded the best accuracy. The annotation interval is 90 seconds long, ensuring it does not reduce the user's concentration. However, because we judged that data collected at times distant from the annotation timing would have low reliability, we used only the 30 seconds immediately preceding the annotation rather than all data collected during the 90 seconds.

The ten extracted features are shown in Table 1. Note that the Signal Magnitude Area is not extracted from the line-of-sight and heart rate data because the three-axis data is required for the Signal Magnitude Area. Therefore, 111 features are extracted: 28 from the acceleration data, 28 from the gyro data, nine from the heart rate data, 28 from the facial orientation, and 18 from the line of sight. Since heart rate is measured at 12 Hz, there are cases where no value is measured within the window. In such cases, we replaced them with zeros. After extracting the features, standardization is performed to scale them. We used under-sampling and over-sampling on the dataset.

3.2 Deep Learning

In this study, we analyzed data using deep learning techniques to construct a classification model. Specifically, we employed an LSTM (Long Short-Term Memory) [20] model to enhance classification performance by effectively capturing the features of time-series data. LSTM is a type of RNN (Recurrent Neural Network) and is a learning model suitable for handling sequential data. Conventional RNNs struggle to learn long-term dependencies, but LSTM can effectively capture them by employing a three-gate structure and memory cells that control the retention and deletion of information. The architecture of the LSTM model is illustrated in Figure 2. The model structure consists of a 64-unit LSTM layer, a dropout layer, and an additional 32-unit LSTM layer. This structure enables temporal dependencies in the data while also preventing overlearning. The final result is a two-class classification using an all-coupled layer and a Softmax activation function.

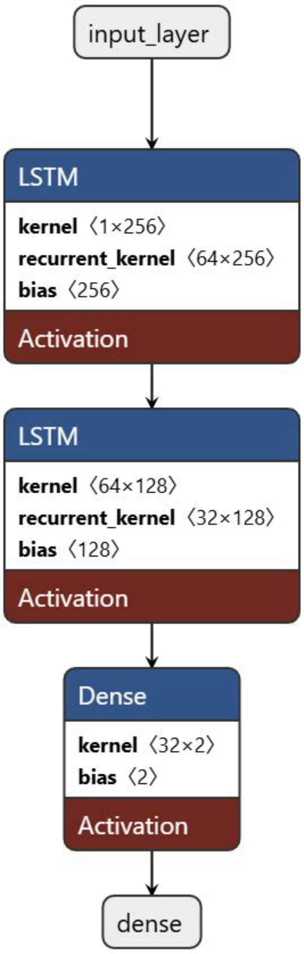


Figure 2: Model architecture

3.3 Evaluation Approach

In this study, we investigate both user-dependent and user-independent analysis for concentration estimation. Our dataset, however, exhibits an imbalance in class distribution, with “Concentrated” labels significantly outnumbering “Distracted” labels across participants. Recognizing this, we employed the macro-F1 score as our primary evaluation metric. The macro-F1 score calculates the F1 score for each class independently and then averages them, providing a more balanced assessment of the model’s performance on imbalanced datasets, as it gives equal weight to both “Concentrated” and “Distracted” states. We conduct user-dependent (UD) validation, which focuses on individual users. UD validation uses each participant’s data for training and evaluation. Since 60 labels are collected per subject, we treat six labels as one data chunk and divide it into ten segments, with one segment designated as the test data and the remaining nine as the training data. One of these is used as test data, and the remaining nine are used as training data; this procedure is repeated ten times for validation. For the validation, gradient boosting, decision trees, logistic regression, random forests, and SVM machine learning models were used for binary concentration classification. We also conduct leave-one-participant-out (LOPO) cross-validation. In this method, each participant’s data serves as test data, and the model is trained on the remaining participants’ data. For the LOPO scenario, we specifically adopted an LSTM model. This choice was made because LSTM, as a type of RNN, is particularly well-suited for handling sequential data and effectively capturing long-term dependencies, which is crucial for building a generalized model that accounts for varying temporal patterns in concentration data across different users.

4 Data Collection

In this section, we first describe the experimental design, participant details, and the multimodal data collection methodology. Subsequently, we evaluate the performance of the models constructed with this data in two scenarios, user-dependent and user-independent, and discuss the obtained results.

Table 2: Label the balance of collected concentration levels of each participant and the sum of the label counts.

User ID	U1	U2	U3	U4	U5	U6	U7	U8	U9	U10	U11	U12	U13	U14	U15	U16	U17	U18	U19	U20	U21	Sum
Concentrated	35	38	38	27	36	34	25	35	37	39	30	47	58	57	53	40	44	34	40	27	47	821
Distracted	25	22	22	33	24	26	35	25	23	21	30	13	2	3	7	20	16	26	20	33	13	439

Table 3: F1 Score for User-Dependent (UD) validation and Leave-One-Participant-Out Cross-Validation (LOPOCV) for each participant and overall.

User ID	U1	U2	U3	U4	U5	U6	U7	U8	U9	U10	U11	U12	U13	U14	U15	U16	U17	U18	U19	U20	U21	Overall
UD	0.84	0.71	0.67	0.67	0.69	0.56	0.79	0.60	0.50	0.52	0.46	0.52	0.94	0.89	0.73	0.70	0.42	0.48	0.44	0.54	0.57	0.69
LOPOCV	0.74	0.64	0.68	0.63	0.64	0.51	0.70	0.55	0.57	0.53	0.57	0.47	0.39	0.40	0.51	0.71	0.53	0.44	0.55	0.45	0.42	0.57

4.1 Participants

This study recruited 21 Japanese students (20 males and one female) from the university. Participants ranged in age from 20 to 24, with an average age of 21.9. The purpose and procedure of the experiment were explained to the participants, and their consent was obtained before participation. The explanation included the purpose of using the data obtained during the experiment, the method of wearing the sensor, and precautions to be taken during the experiment.

4.2 Data Collection Procedure

In this experiment, participants were asked to wear sensors while watching lecture videos on YouTube. The videos covered topics on Git (a version control system)[1] and SQL (Structured Query Language)[2]. The experiment lasted 90 minutes and utilized three types of sensors for data collection. The details of the sensors (S1-S3) are as follows:

- S1 MetamotionS+[3]: The sensor was attached to a participant's right wrist, recording acceleration and gyroscopic data at a 100 Hz sampling rate. The band was securely fastened to ensure sensor stability during measurements. The device also featured a button for subjective annotations. During the experiment, the sensor vibrated every 90 seconds, prompting participants to evaluate their concentration. Participants pressed the button if they were not concentrating at the time of vibration and refrained from doing so if they were focused. This protocol resulted in 60 labeled data points per participant. Data collection and vibration control were managed via Bluetooth, allowing the MetamotionS+ to connect to a PC.
- S2 Fitbit Charge 5[4]: The sensor was worn on the left wrist to monitor heart rate in real-time at a sampling rate of 12 Hz. Data was transmitted to a PC using the Fitbit API. During preprocessing, gaps in the heart rate data within the sliding window were replaced with zeros. The sensor was adjusted and firmly secured to ensure close contact with the skin for accurate measurements.
- S3 Webcam: In addition to the wrist-worn sensors, a built-in webcam on a ThinkPad X1 Yoga Gen 5 was used to capture facial data. The PC's angle and height were adjusted to ensure the participant's entire face was consistently visible to the webcam, which was recorded at 30 frames per second (fps) throughout the experiment. Following the session, facial direction and eye gaze were extracted using OpenFace [21].

5 Result & Discussion

Table 2 presents the distribution of collected concentration labels for each participant. The overall dataset is imbalanced, with a total of 821 "Concentrated" labels significantly outnumbering the 439 "Distracted" labels. Additionally, this imbalance varies among the participants. Table 3 shows the F1 score for each participant in the UD validation. The highest accuracy was achieved when the window size was 15, the overlap was zero, and the training was performed in a random forest. While P13 and P14 achieved high F1 scores, these results cannot be regarded as high due to the significant imbalance in their "Concentrated" and "Distracted" label counts, as shown in Table 2. These findings suggest that performance may improve when the model is specific to individual users. Specifically, some participants, such as P1 and P2, achieved high F1 scores, whereas others, like P18 and P11, achieved low scores. These differences may be attributed to variations in individual data characteristics, sensor data quality, or individual concentration patterns. The confusion matrix is also shown in Figure 3a. The results show that concentration is accurately estimated, while distraction has a high misclassification rate. This is likely because the number of labels is more significant for concentrated attention than for distracted attention.

In contrast to user-dependent validation, user-independent validation was conducted using the Leave-One-Participant-Out (LOPO) method. In this method, only one participant's

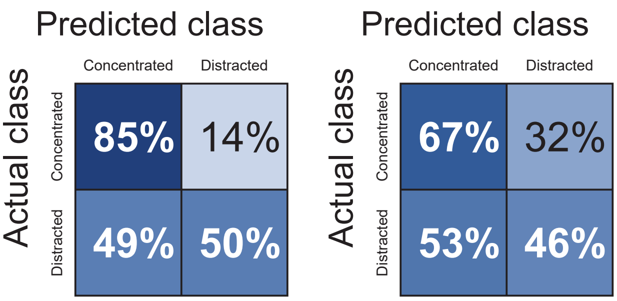
---

[1] https://youtube/WHwuNP4kalU

[2] https://youtube/v-Mb2voyTbc

[3] https://mbientlab.com/store/metamotions-p

[4] https://www.fitbit.com/global/us/products/trackers/charge5



(a) UD Validation (b) LOPOCV

Figure 3: Confusion matrices of (a) UD and (b) LOPOCV

data is used as test data, and all remaining participants' data is used as training data to train the model. The LOPO method verifies the model's generalization performance for each user since each participant is evaluated as test data. This allows us to compare the advantages of generalization in the user-independent approach with those of individual optimization in the user-dependent approach. LSTM was used for training in user-independent validation. The results of the user-independent validation are shown in Table 3. Before training, under-sampling was performed on the balanced number of labels. The Overall F1 score was 0.57, which was lower than the user-dependent result. The confusion matrix is shown in Figure 3b. The F1 score presented under the *Overall* column in both UD and LOPO validations is derived from the aggregated confusion matrix across all participants, rather than averaging individual participant F1 scores. Compared to a random chance rate for Macro F1 of approximately 0.49, our proposed method demonstrates a clear advantage in estimating concentration, particularly within the user-dependent validation scenario, highlighting its effectiveness beyond mere chance. This notable difference in F1 scores between the UD validation (0.69) and the LOPO validation (0.57) suggests that significant individual differences exist in the behavioral and physiological patterns associated with concentration. The LSTM model employed in the LOPO scenario likely struggled to generalize features learned from one participant's data to a new, unseen participant. This is particularly evident in the confusion matrix (Figure 3b), which shows a higher rate of misclassifying the "Distracted" state, suggesting that inter-subject variance and the dataset's imbalance impacted the model's generalization performance.

6 FocusSense

Based on the estimation models developed in Section 3, we designed and implemented a prototype visualization application, *FocusSense*. The primary goal of this application (C2) is to make the estimation results actionable and easy to understand, enabling students and instructors to reflect on learning patterns. The user interface of the *FocusSense* application is shown in Figure 4. The application is designed with a simple workflow: (a) The user first imports their multimodal sensor data (CSV files) and selects the sensors to use for estimation. (b) After processing, the user can visualize the raw sensor data streams (e.g., acceleration, heart rate) to check for anomalies. (c) Finally, the application presents the classified concentration states (distracted or concentrated) over the entire session, using both a detailed state transition table and an intuitive timeline bar for quick reference. This prototype serves as the foundation for the next phase of our research. As future work, we will use this tool to conduct a comprehensive user experience evaluation. This evaluation will be crucial to validate the practical effectiveness of the tool for both students and instructors and to gather qualitative feedback for further improvements.

7 Limitation and Future Work

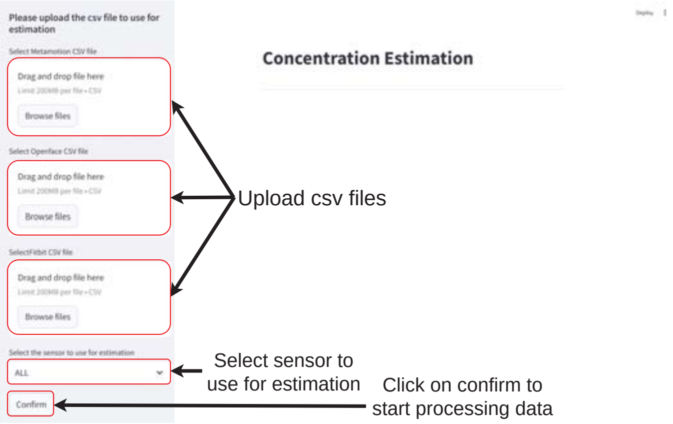
Our study employed a deep learning model to estimate the concentration of students enrolled in online classes. The results showed that the deep learning model did not significantly improve the accuracy. One reason for this is label bias. Labeling the dataset used in this study may contain a large amount of data biased toward a particular concentration state. For example, if subjects self-report too many labels as "concentrated", the model cannot accurately learn the unfocused state. This may have contributed to the reduced prediction accuracy of the model.

Additionally, there is another limitation to the annotation method. Labeling the concentration level depends on the subject's subjective judgment, and it is possible that the moment when the subject loses concentration is not accurately recorded. Such inaccuracies in annotation may have hindered the model's learning. Furthermore, considering a wider range of time series features is expected to enable the model to infer more accurately. To further enhance the model's performance, we will conduct additional experiments using Transformer-based models, which have demonstrated strong capabilities in modeling sequential data. By comparing the performance of LSTM and transformer-based architectures, we aim to evaluate the effectiveness of different deep-learning approaches for concentration estimation. In addition to these limitations, future work will involve addressing the real-world applicability of the system. Specifically, we will develop a platform for continuously collecting data to validate the system in real-world settings, such as educational or workplace environments. Additionally, we will design the system with privacy considerations to ensure the protection of users' sensitive data.

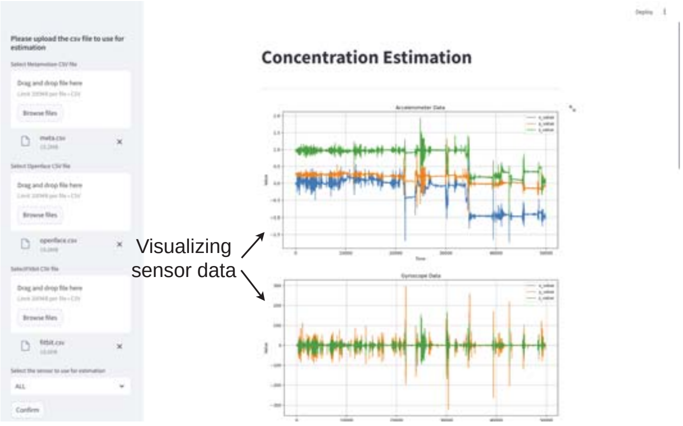
As this study focused on the construction of a prototype, future work on the application will involve (1) conducting a user experience evaluation to refine the visualization and interface, and (2) implementing and evaluating a real-time concentration feedback function.

8 Conclusion

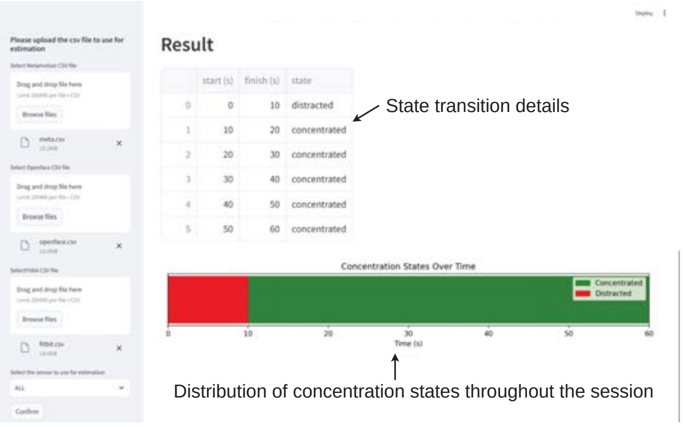
Multimodal data were collected from 21 Japanese university students while they watched a 90-minute online lecture. Data collection was conducted using wrist-worn wearable devices and a PC webcam to acquire multiple data streams,


Please upload the csv file to use for estimation
Select Metamotion CSV file
Drag and drop file here
Limit 200MB per file • CSV
Browse files
Select Openface CSV file
Drag and drop file here
Limit 200MB per file • CSV
Browse files
SelectFitbit CSV file
Drag and drop file here
Limit 200MB per file • CSV
Browse files
Select the sensor to use for estimation
ALL
Confirm
Concentration Estimation
Upload csv files
Select sensor to use for estimation
Click on confirm to start processing data

(a) First view, where the user imports a dataset.



(b) Second view, visualizing sensor data.



	start (s)	finish (s)	state
0	0	10	distracted
1	10	20	concentrated
2	20	30	concentrated
3	30	40	concentrated
4	40	50	concentrated
5	50	60	concentrated

(c) Third view showing the estimation results.

Figure 4: User interface of the *FocusSense* application.

including acceleration, gyroscope, heart rate, facial orientation, and eye gaze. Features were extracted from this data to perform a binary classification of concentration states using machine learning and deep learning models. As a result, an average F1 score of 0.69 was achieved for the UD validation and 0.57 for the LOPO validation. Furthermore, a visualization application, *FocusSense*, was developed to enable students and educators to understand the estimation results intuitively. This study demonstrated the potential of using multimodal sensors to estimate the concentration levels of students during online learning with a certain degree of accuracy. It also became clear that while visualization tools like *FocusSense* can contribute to students' self-regulated learning and instructors' teaching improvements, enhancing its practical utility requires further improvements in model accuracy and the expansion of practical feedback functions as future work.

Acknowledgements

This work was supported in part by the Japan Society for the Promotion of Science, Grants-in-Aid for Scientific Research number JP24K02934.

References

[1] Ministry of Education, Culture, Sports, Science and Technology (MEXT), "New coronavirus and distance learning measures," 2020, accessed: 2024-05-22. [Online]. Available: https://www.mext.go.jp/content/20200717-mxt_kouhou01-000004520_2.pdf

[2] K. Watanabe, T. Sathyanarayana, A. Dengel, and S. Ishimaru, "Engauge: Engagement gauge of meeting participants estimated by facial expression and deep neural network," *IEEE Access*, vol. 11, pp. 52 886–52 898, 2023.

[3] K. Watanabe, A. Dengel, and S. Ishimaru, "Metacognition-engauge: Real-time augmentation of self-and-group engagement levels understanding by gauge interface in online meetings," in *Proceedings of the Augmented Humans International Conference 2024*, ser. AHs '24. New York, NY, USA: Association for Computing Machinery, 2024, p. 301–303. [Online]. Available: https://doi.org/10.1145/3652920.3653054

[4] Y. Uema and K. Inoue, "Jins meme algorithm for estimation and tracking of concentration of users," in *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*, ser. UbiComp '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 297–300. [Online]. Available: https://doi.org/10.1145/3123024.3123189

[5] H. Yoshikawa, A. Uchiyama, Y. Nishikawa, and T. Higashino, "Combining a thermal camera and a wristband sensor for thermal comfort estimation," in *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers*, ser. UbiComp/ISWC '19 Adjunct. New York, NY, USA: Association for Computing Machinery, 2019, p. 238–241. [Online]. Available: https://doi.org/10.1145/3341162.3343813

[6] A. P. Pai, J. Santhosh, and S. Ishimaru, “Real-time feedback on reader’s engagement and emotion estimated by eye-tracking and physiological sensing,” in *Adjunct Proceedings of the 2022 ACM International Joint Conference on Pervasive and Ubiquitous Computing and the 2022 ACM International Symposium on Wearable Computers*, ser. UbiComp/ISWC ’22 Adjunct. New York, NY, USA: Association for Computing Machinery, 2023, p. 97–98. [Online]. Available: https://doi.org/10.1145/3544793.3560329

[7] R. Morita, K. Watanabe, J. Zhou, A. Dengel, and S. Ishimaru, “Genaireading: Augmenting human cognition with interactive digital textbooks using large language models and image generation models,” in *Proceedings of the Augmented Humans International Conference 2025*, ser. AHs ’25. New York, NY, USA: Association for Computing Machinery, 2025, p. 289–301. [Online]. Available: https://doi.org/10.1145/3745900.3746066

[8] P. Zolfaghari, V. F. Rey, L. Ray, H. Kim, S. Suh, and P. Lukowicz, “Sensor data augmentation from skeleton pose sequences for improving human activity recognition,” in *2024 International Conference on Activity and Behavior Computing (ABC)*, 2024, pp. 1–8.

[9] S. S. Alia and P. Lago, “Daily routine recognition from longitudinal, real-life wearable sensor data for the elderly,” in *2024 International Conference on Activity and Behavior Computing (ABC)*, 2024, pp. 1–9.

[10] A. Bhatt, K. Watanabe, A. Dengel, and S. Ishimaru, “Appearance-based gaze estimation with deep neural networks: From data collection to evaluation,” *International Journal of Activity and Behavior Computing*, vol. 2024, no. 1, pp. 1–15, 2024.

[11] R. Islam and S. W. Bae, “Pupilsense: Detection of depressive episodes through pupillary response in the wild,” in *2024 International Conference on Activity and Behavior Computing (ABC)*, 2024, pp. 01–13.

[12] K. Wakita and K. Shimada, “An utterance is enough to the gaze? gaze detection from utterance information in multiparty discussion,” in *2024 International Conference on Activity and Behavior Computing (ABC)*, 2024, pp. 1–8.

[13] T. Muragi, A. Tsuji, and K. Fujinami, “A system for assembly-work-confusion recognition based on gaze and hand positional information,” *International Journal of Activity and Behavior Computing*, vol. 2024, no. 2, pp. 1–25, 2024.

[14] H. Asahi, R. Sonoyama, C. Shoda, and N. Kotani, “The legend of holy sword: An immersive experience for concentration enhancement,” *arXiv preprint arXiv:2408.16782*, 2024.

[15] C.-W. You, H.-A. Chen, P.-C. Chen, W.-N. Lai, C. W. T. Yuan, and N. Bi, “Toward understanding the impact of visualized focus levels in virtual reality on user presence and experience,” *Proc. ACM Hum.-Comput. Interact.*, vol. 8, no. MHCI, Sep. 2024. [Online]. Available: https://doi.org/10.1145/3676527

[16] S. Tanaka, A. Tsuji, and K. Fujinami, “Eye-tracking for estimation of concentrating on reading texts,” *International Journal of Activity and Behavior Computing*, vol. 2024, no. 1, pp. 1–21, 2024.

[17] I. Betto, R. Hatano, and H. Nishiyama, “Distraction detection of lectures in e-learning using machine learning based on human facial features and postural information,” *Artificial Life and Robotics*, vol. 28, no. 1, pp. 166–174, 2023.

[18] K. Kimura, S. Kunimasa, Y. Kusakabe, H. Ishii, and H. Shimoda, “Estimation of intellectual concentration states using pupil diameter and heart rate variability.” in *CHIRA*, 2018, pp. 62–69.

[19] N. Tanaka, K. Watanabe, S. Ishimaru, A. Dengel, S. Ata, and M. Fujimoto, “Concentration estimation in online video lecture using multimodal sensors,” in *Companion of the 2024 on ACM International Joint Conference on Pervasive and Ubiquitous Computing*, ser. UbiComp ’24. New York, NY, USA: Association for Computing Machinery, 2024, p. 71–75. [Online]. Available: https://doi.org/10.1145/3675094.3677587

[20] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Comput.*, vol. 9, no. 8, p. 1735–1780, Nov. 1997. [Online]. Available: https://doi.org/10.1162/neco.1997.9.8.1735

[21] T. Baltrušaitis, P. Robinson, and L.-P. Morency, “Openface: an open source facial behavior analysis toolkit,” in *2016 IEEE winter conference on applications of computer vision (WACV)*. IEEE, 2016, pp. 1–10.

Session 2:
AI and Intelligent Systems

(Chair: Hayato Tomisu)

A Method for Classifying Kumite Techniques and Detecting Preliminary Actions during Kumite Matches

Kwangyun Kim[†], Shuhei Tsuchida[‡], Tsutomu Terada[†], and Masahiko Tsukamoto[†]

[†]Graduate School of Engineering, Kobe University, Japan
{kwangyun-kim@stu., tsutomu@eedept., tuka@}kobe-u.ac.jp
[‡]Center for Interdisciplinary AI and Data Science, Ochanomizu University, Japan
tsuchida.shuhei@ocha.ac.jp

Abstract - Kumite is a karate sparring competition in which two players fight each other using various techniques. In karate kumite matches, reducing a preliminary action (hereinafter referred to as “pre-action”) such as pulling the arms and lowering the shoulders just before an attack technique is essential. This is because pre-actions reveal the attack timing to the opponent. However, players find it difficult to grasp their pre-actions themselves. Moreover, estimating the presence or absence of pre-action is difficult with existing motion analysis methods. Because pre-action is a small action compared to punching and kicking. We previously developed a method for estimating the presence or absence of pre-actions both in static states and during kumite matches. These methods focused only on the forefist punch, which is the most basic kumite technique. We have not yet developed methods for detecting pre-actions in other kumite techniques. In addition, we could only apply the method for detecting pre-actions during a kumite match to post-match data. We therefore propose a method that can classify the types of kumite techniques and estimate the presence or absence of pre-actions for each technique.

Keywords: human motion analysis; karate; inertial sensor; sports support; classification; deep learning; convolutional neural network; long short-term memory; autocorrelation function; dynamic time warping

1 Introduction

Karate is a Japanese martial art that includes combat techniques such as punching, kicking, and blocking. There are two main types of karate competitions: Kata and Kumite. This study focuses on kumite, a form of sparring in which two karate practitioners face each other using various techniques. In kumite, a player can win a match by scoring more points than the opponent through successful attacks or defensive maneuvers. Because most kumite techniques are linear and extremely fast [1], anticipating an opponent’s movements is crucial for effective defense. To predict the opponent’s attack, anticipating an opponent’s preliminary actions (hereinafter referred to as “pre-actions”) during the attack is effective. Conversely, if the opponent cannot perceive the player’s pre-action, the attack is more likely to succeed.

A pre-action refers to the motion that occurs immediately before executing a kumite technique, such as moving the fist, lowering the arm, or raising the shoulder. Petri et al. found that punch techniques were most likely to be recognized by the opponent in the preparation phase of the attack by analyzing the “anticipatory cues” in a kumite match [2]. The anticipatory cues that take place during the attack preparation phase are the pre-actions, which inform the opponent of the timing of the attack and give the opponent time to prevent the attack.

Therefore, accurately recognizing and reducing pre-actions is important for a successful attack. However, players often have difficulty recognizing their own pre-actions, as these are frequently performed subconsciously.

We developed a method for detecting pre-actions in both static states and during kumite matches. We proposed a method to estimate the presence or absence of pre-action based on the similarity between the acceleration data of an arbitrary punch and a previously prepared dataset consisting of the acceleration data from punches performed without pre-action [3]. We refer to this method as GLID (gradually lengthening inverted-window dynamic time warping). GLID could only be applied to basic practice in the static state. It did not function effectively in kumite matches, where footwork is performed before and after kumite techniques. Here, “footwork” refers to controlling the movement of the feet to adjust position and posture in relation to the opponent. Players generally perform footwork between kumite techniques during matches.

We proposed GWA-GLID, a method for detecting the pre-action in kumite matches by combining GLID with a preprocessing step for the inertial data during kumite matches [4]. The preprocessing, named GWA (gradually window-shrinking autocorrelation-function), effectively identifies and smooths footwork segments in the inertial data. GLID and GWA-GLID focused only on the forefist punch, which is the most basic kumite technique. We have not yet developed method for detecting pre-actions in other kumite techniques. In addition, we could only apply the method for detecting pre-actions during a kumite match to post-match data.

Therefore, in this study, we propose a system that can estimate and provide real-time feedback on the presence or absence of pre-actions performed during kumite techniques in matches. We consider that this system is useful for practically and efficiently improving kumite skills.

In this paper, we propose a method for classifying kumite techniques and detecting pre-actions performed during matches to realize such a system. We refer to this method as ReGGMu (real-time GWA-GLID for multi-class).

The rest of this paper is organized as follows: Section 2 introduces related research, Section 3 explains ReGGMu, Section 4 discusses future work, and Section 5 concludes the paper.

2 Related Work

2.1 Evaluation of Karate Movements Using Sensors

Many studies have used various sensors to evaluate karate motions and support training. Motion capture was used to classify karate kicks [5, 6] and analyze kata movements [7]. Although motion capture systems can visualize and evaluate detailed karate motions, it has been pointed out that the use of numerous markers attached to the body restricts natural motion. They are also expensive, require large-scale equipment, and need ample space, making them difficult for general use.

In addition, studies analyzing karate experts' movements using EMG (electromyography) investigated the muscles that are strongly activated when punching [8], the correlation between the punch force and biomechanical parameters [9], and the effect of three months of training on the electromyographic activity during punching and kicking [10]. There are studies using computer vision to develop smart coaching systems for karate [11], to evaluate karate performance [12], and to classify kumite movements [13]. Computer vision techniques may be unable to analyze micro movements during kumite, depending on the camera angle.

In this way, there are many studies related to karate that utilize sensors. However, most focus on classifying kumite techniques or evaluating their accuracy. No studies have addressed the pre-action, one of the important elements in kumite matches. Therefore, we analyze the pre-action in karate using inertial sensors, which have recently gained widespread use in sports and martial arts research.

2.2 Motion Recognition Method Using the Inertial Sensor

Gesture classification using inertial sensors has demonstrated utility across diverse application domains. From assistive activities in healthcare and welfare [14] to driving [15] and sports skill assessment [16], as well as interaction design for education and entertainment [17], numerous studies have leveraged the ease of wear and occlusion robustness of inertial measurement to achieve reliable performance. To classify the time-series data obtained from from inertial sensors, diverse algorithms have been employed. These include DTW (dynamic time warping) [18], HMM (hidden markov model) [19], SVM (support vector machine) [20], and RNN (recurrent neural network) [21].

As a specific application, Fang et al. proposed a dynamic gesture recognition system using a glove composed of 36 inertial measurement units, applying it to construct a CNN (convolutional neural network) model for sign language recognition [22]. Teachasrisaksakul et al. presented a method that recognizes hand gestures for educational interactive games using a multilayer perceptron classifier with acceleration and angular velocity as inputs, achieving 0.93 accuracy over 15 gesture classes [17]. Alfaro et al. realized user-independent gesture classification by integrating inertial data with EMG (electromyography), attaining 0.93 accuracy with an SVM across seven gestures performed by 22 participants [23]. Sun et al. combined DTW and HMM to build a real-time model for detecting distracted-driving gestures, recognizing four hand gestures with 0.97 accuracy [15].

Table 1: Types of Kumite Techniques and Body Parts

Type of Technique	Body Part
Forefist Punch	Left Hand
Reverse Punch	Right Hand
Lunge Punch	Right Hand
Roundhouse Kick (Left)	Left Foot
Roundhouse Kick (Right)	Right Foot
Hook Kick	Left Foot

In the field of gesture classification, diverse methods have been selected according to tasks and operational conditions. More recently, advances in deep learning have driven a shift toward end-to-end models that learn features directly from sensor time series and deliver higher accuracy. This research targets kumite techniques, which exhibit diverse movement characteristics during matches. Therefore, we aim to construct a deep learning-based kumite technique classification model that simultaneously achieves real-time processing of inertial sensor time series and high recognition performance.

3 Proposed Method: ReGGMu

In this section, we describe a method for classifying kumite techniques and estimating the presence or absence of the pre-action performed during kumite matches.

3.1 Constructing a Dataset for Kumite Techniques Without the Pre-action

As a preliminary preparation for the kumite technique classification and pre-action estimation for arbitrary kumite techniques (hereinafter referred to as "input data"), we prepared a dataset consisting of inertial data from kumite techniques performed without pre-action (hereinafter referred to as "dataset without pre-action"). Figure 1 and Table 1 show the six kumite techniques examined in this study and the corresponding body parts involved in their execution. We determined the list of techniques by surveying members of the Kobe University karate club about the techniques they most frequently use during kumite matches. We constructed the dataset without pre-action using three-axis acceleration and angular velocity data collected from four karate practitioners with over three years of experience in a university karate club, each performing six different kumite techniques 10 times without pre-action.

The mounting positions of the inertial sensors on the players and the axial directions of acceleration and angular velocity of each sensor are shown in Figure 2. We placed the inertial sensors on the players' left wrist (LW), right wrist (RW), left ankle (LA), right ankle (RA), and waist (WA).

We used a compact wireless hybrid sensor Ⅱ (WAA-010) from ATR-Promotions (Kyoto, Japan) for the measurements. We used a ThinkPad X13 Gen 2 PC (OS: Windows 10 Home,

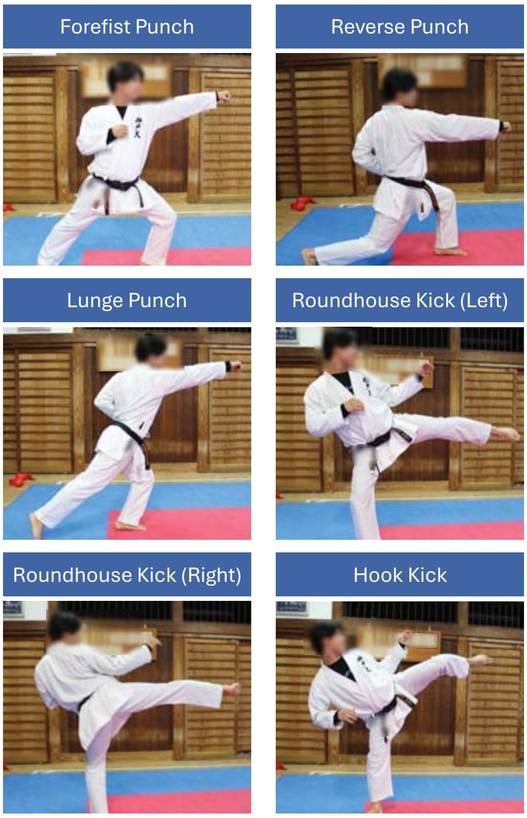


Figure 1: Types of Kumite Techniques.

CPU: 11th Gen Intel® Core™ i5-1135G7 @ 2.40 GHz, RAM: 16.0 GB) from Lenovo (Beijing, China). We obtained the inertial data using the AccelViewerHybrid-Ⅱ (ver.2.4.0) WAA-010 dedicated data receiving software at 50 Hz.

3.2 Overview of ReGGMu

An overview of ReGGMu, a method for classifying the types of kumite techniques performed during a kumite match and estimating the presence or absence of the pre-action, is shown in Figure 3.

First, we detected the "striking timing", which is when the kumite technique was performed, from the peak value in the average of the six-axis inertial data. Next, we cut out 300 samples (6 seconds) data window before and after the striking timing. Then, we detected segments in which footwork was performed within the window, and smoothed them with a moving average method. Then, we classified the smoothed inertial data into one of six kumite techniques. We used a hybrid CNN-LSTM deep learning model for classification. Finally, we estimated the presence or absence of pre-action based on the results of the DTW (dynamic time warping) calculations using the dataset. We created a dataset for each type of kumite technique in advance.

3.3 GWA: Detecting and Smoothing Method for Footwork Segment

We introduce our previously proposed GWA, which automatically detects and smooths footwork performed by players during kumite matches [4]. GWA corresponds to the third step in Figure 3.

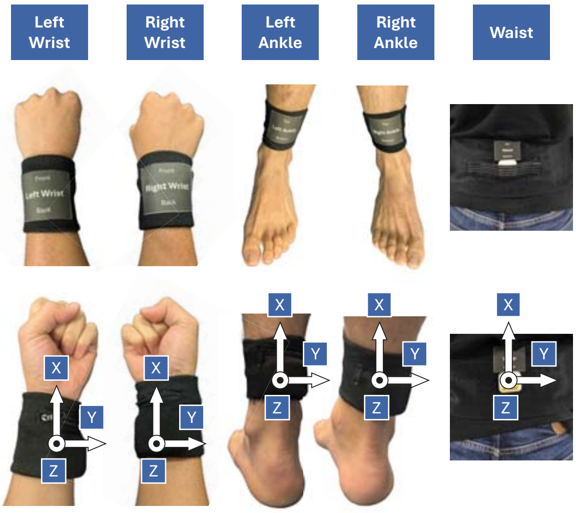


Figure 2: Mounting Position of Sensors and the Axial Direction of Inertial Data.

One of the key characteristics of kumite matches is that players perform footwork between attacks and defenses. This footwork, which involves hopping in place or moving forward, backward, or sideways using the toes, allows for smoother attacks and makes it difficult for opponents to predict the target or timing of an attack.

An overview of the GWA is shown in Figure 4. First, we classified footwork and non-footwork windows by constancy decisions based on the calculation of the ACF values in the sliding window (window size: 50 samples, sliding interval: 5 samples). We detected footwork windows by identifying peaks in the ACF values exceeding a dynamically determined threshold. The threshold for the constancy decision is defined by the following.

$$Threshold_{ACF} = \alpha \cdot (1 - k/N) \tag{1}$$

Let k be the lag when calculating the autocorrelation, and N be the window size. The value of ACF decreases linearly with increasing lag k, and the peak height of a periodic waveform such as a sine wave is $1 - k/N$. α is a scaling coefficient that should be set between 0.1 and 1.0 to maximize GWA accuracy for the data being analyzed. Following previous study [4], we set α to 0.5 in this study.

Next, we recognized the accurate boundary point between the footwork and non-footwork segments. We consider that the first window among the consecutively detected non-footwork windows consists of "footwork", "pre-action", and "kumite technique" in that order if it contains the pre-action. We gradually shrink the length of this window from its end in the time series, and calculate the value of ACF at each step, as shown in Figure 5. We defined the segment boundary point as the time when the positive peak of the ACF value was obtained by this process.

Finally, we applied a moving average method (window size: 50 samples) to the footwork segments. This smoothing process prevents the inertial data of footwork and pre-action from being recognized as similar waveforms when adapting GLID to the inertial data during kumite matches.

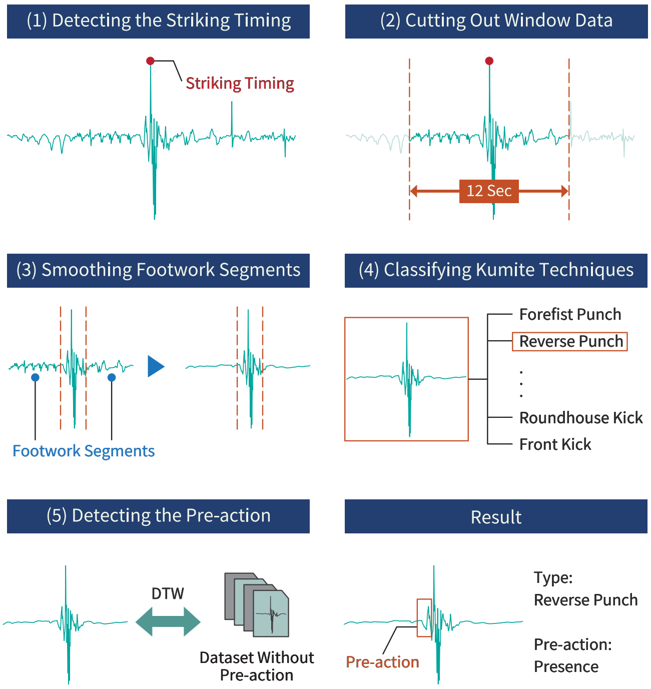

(1) Detecting the Striking Timing
Striking Timing
(2) Cutting Out Window Data
12 Sec
(3) Smoothing Footwork Segments
Footwork Segments
(4) Classifying Kumite Techniques
Forefist Punch
Reverse Punch
Roundhouse Kick
Front Kick
(5) Detecting the Pre-action
DTW
Dataset Without Pre-action
Result
Pre-action
Type: Reverse Punch
Pre-action: Presence

Figure 3: Overview of ReGGMu.

3.4 Hybrid CNN-LSTM Model for Classification of Kumite Techniques

We describe a method for classifying kumite techniques using inertial data recorded during kumite matches. This classification method corresponds to the fourth step in Figure 3.

We developed a model to classify kumite techniques, based on a hybrid CNN-LSTM deep learning model originally proposed by Supriya et al. [24] for daily activity recognition. This model, which achieved a classification accuracy of 0.98, consists of a layered architecture in which a CNN extracts key features from input data, and an LSTM predicts human activity. The layer configuration of the model used in ReGGMu is shown in Table 2.

First, the model uses stacked 1D convolutional and pooling layers to automatically extract local motion features from the input inertial data. Next, the model applies batch normalization and ReLU activation to ensure stable and efficient training. Then, the model passes the extracted features to an LSTM layer to capture temporal dependencies within the motion sequence. Finally, the model employs Dense layers and a dropout layer to refine the representation and perform classification into kumite technique types.

We employ this model to classify kumite techniques. First, we extract a 100 sample segment of the inertial data, centered on the striking timing, after footwork smoothing as described in Section 3.3, for input into the model. Then, in addition to the raw inertial data segment, we extract a set of handcrafted statistical features from the same window. These features include the mean, standard deviation, maximum value, minimum value, and corresponding vector norm mean, standard deviation, maximum value, and minimum value for each axis of acceleration and angular velocity for each of the five sensors, generating a 160-dimensional feature vector for each sample. Along with the raw inertial data, we incorporate statistical features to provide the model with both detailed sensor signals and high-level summary information, thereby enhancing the classification performance of kumite techniques.

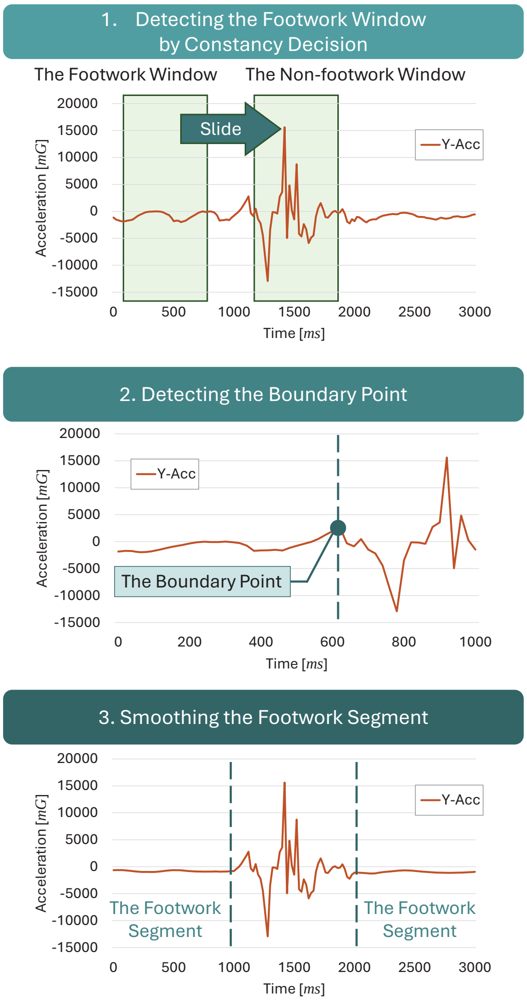


Figure 4: Overview of GWA.

The introduction of this classification model is significant in ReGGMu. Although methods for classifying only types of punches have been reported [?], a comprehensive method covering a wide range of techniques, including punches and kicks, has not yet been established. Furthermore, we enabled the extension of the conventional pre-action detection method, previously applicable only to a single kumite technique, to multiple types of kumite techniques by introducing this classification model. This represents a major novel contribution of our study.

3.5 GLID: Estimating Method for the Presence or Absence of the Pre-action

We introduce our previously proposed GLID, which is designed to estimate the presence or absence of pre-actions in kumite techniques performed in a static state. GLID is a method for estimating the presence or absence of the pre-action based on the DTW distance calculation results of the dataset without pre-action and the input data. GLID corresponds to the fifth step in Figure 3.

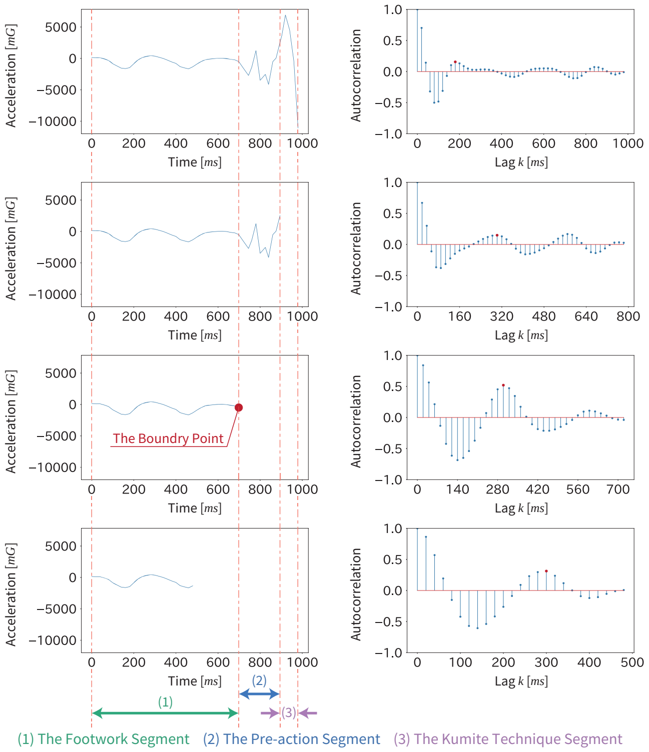


Figure 5: Detection of the Boundary Point by Gradually Shrinking Window and Calculating the Value of ACF.

An overview of GLID is shown in Figure 6. Let μ and ν be the indices of the striking timing in time series data of input data and data without pre-action in Figure 6.

First, we cut out the data from the start of the window data to the striking timing to analyze the inertial data, focusing only on the pre-action part.

Next, we reverse the time series of the cutout input data and the dataset without pre-action and calculate their DTW distance. The inertial waveforms just before the striking timing of kumite techniques are similar regardless of whether the pre-action is present. We can focus only on the pre-action waveforms that precede similar waveforms while maintaining the consistency of the number of peaks in the pre-action waveforms by setting the striking timing to the uniform starting point of the DTW and reversing the time series. We can estimate the presence or absence of the pre-action based on whether a corresponding pre-action waveform exists following this similar waveform.

Finally, we involve incrementally extending the time window of the input data and calculating the DTW distance to the dataset without pre-action sequentially. At this time, based on the classification results in Section 3.4, the corresponding kumite technique dataset without pre-action is selected and used for GLID. We calculated the DTW distance using the method of Myers et al [25]. By tracking how this distance evolves as the input data includes more of the pre-action phase, we observe that the presence of the pre-action typically causes a noticeable increase in DTW distance. This is because the waveform patterns during the pre-action differ significantly from those without the pre-action. We analyze the DTW distance graphs by identifying first local maximum and minimum, as shown in Figure 7. The gap between these points serves as a reliable indicator of the pre-action in the input data. If the gap between the local maximum and minimum exceeds the threshold, we estimate that the pre-action is present; otherwise, it is absent. We set the threshold to 13 based on the result of the preliminary investigation [3].

Table 2: Layers Configuration of CNN-LSTM Model

Layer (type)	Output Shape	Param #
conv1d (Conv1D)	(None, 100, 32)	2,912
batch_normalization (BatchNormalization)	(None, 100, 32)	128
re_lu (ReLU)	(None, 100, 32)	0
max_pooling1d (MaxPooling2D)	(None, 50, 32)	0
conv1d_1 (Conv1D)	(None, 50, 64)	6,208
batch_normalization_1 (BatchNormalization)	(None, 50, 64)	256
re_lu_1 (ReLU)	(None, 50, 64)	0
max_pooling1d_1 (MaxPooling2D)	(None, 25, 64)	0
lstm (LSTM)	(None, 64)	33,024
dense (Dense)	(None, 128)	8,320
dropout (Dropout)	(None, 128)	0
dense_1 (Dense)	(None, 8)	1,290

Total params: 52,138
Trainable params: 51,946
Non-trainable params: 192

4 Future Work

In future work, we need to address the following several challenges to develop a practical system that provides real-time feedback to players by classifying kumite techniques and detecting pre-actions during matches.

4.1 Evaluation in Match Conditions

First, we will verify the accuracy of ReGGMu under actual match conditions. Specifically, players who belonged to a university karate club will conduct kumite matches in a format equivalent to official matches while wearing inertial sensors. Using the accelerometer and gyroscope data recorded during the matches, we will evaluate technique classification accuracy and pre-action detection accuracy of ReGGMu. This evaluation will quantify robustness to dynamic characteristics of matches such as technique switching, feints, and variations in pace and thereby assess the effectiveness of ReGGMu.

4.2 Real-time Operation

Next, we will implement methods that enable ReGGMu to operate in real time. While inertial data are being measured during a match, the processing flow shown in Figure 3 will be executed automatically each time a kumite technique is performed. Specifically, when the body part with the attached sensor detects changes in acceleration or angular velocity indicative of a kumite technique, a process trigger will be activated. Subsequently, a data window of appropriate size will

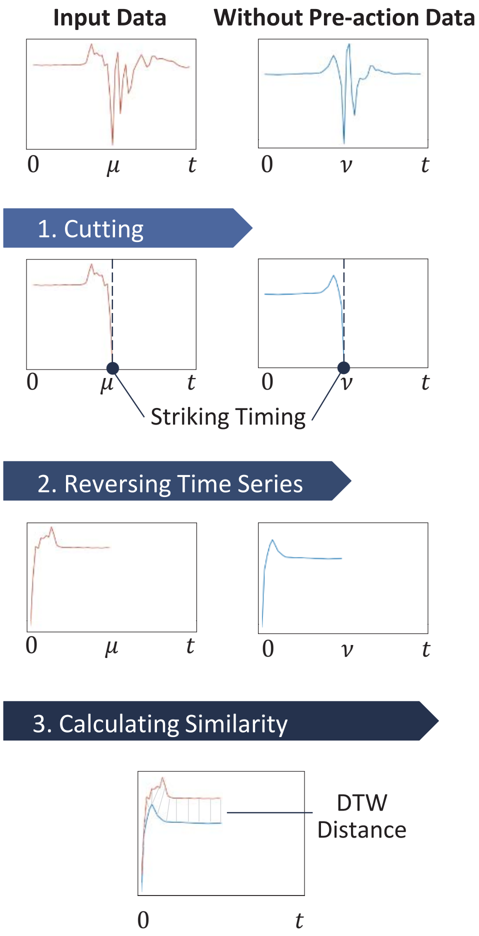


Figure 6: Overview of GLID.

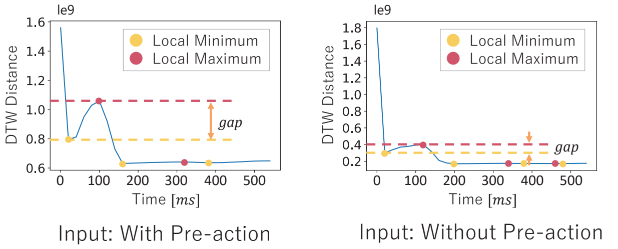
be extracted, and ReGGMu will be applied. Furthermore, to minimize the latency between trigger activation and output, we will incorporate algorithmic optimizations such as early termination within GLID.

4.3 System Design

Finally, we need to consider methods for effectively providing feedback to players based on information recognized in real-time and implement devices for this purpose. At present, we are considering the installation of LEDs inside the headgear (menho) used during matches. The system can communicate the presence or absence of pre-actions to the player by changing the light pattern or color each time a kumite technique is executed. This feedback enables real-time correction during matches and efficient skill improvement.

5 Conclusion

In this study, we proposed ReGGMu, a method for classifying kumite techniques and detecting the presence or absence of pre-actions during karate matches using inertial sensors. Previous research had focused solely on detecting pre-actions in a single technique-the forefist punch. By introducing a deep learning-based classification model and employing a dataset without pre-actions for each technique, ReGGMu enables the detection of pre-actions in six types of kumite techniques performed during matches.



Input: With Pre-action

Input: Without Pre-action

Figure 7: Comparison of DTW Distance Trend Graphs in the Presence and Absence of the Pre-action.

In future work, we aim to develop a system that facilitates the practical and efficient enhancement of kumite skills by providing real-time feedback based on classification and pre-action detection results during matches. To achieve this, we will focus on evaluating ReGGMu in actual match conditions, improving ReGGMu for real-time operation, and designing a system capable of providing effective feedback.

Acknowledgment

This work was supported by JST SPRING Grant Number JPMJSP2148, Japan.

REFERENCES

[1] J. Venkatraman, R. Manwar, and K.M. Avanaki: Development of a Punch-o-meter for Sport Karate Training, *Journal of Sensors*, Vol. 8, No. 7, p. 782 (2019).

[2] K. Petri, M. Droste, and K. Witte: Analysis of Anticipatory Cues in Karate Kumite Using an In-situ Study, *Journal of Martial Arts Research*, Vol. 3, No. 3, pp. 1–20 (2020).

[3] K. Kim, S. Tsuchida, T. Terada, and M. Tsukamoto: KARATECH: A Practice Support System Using an Accelerometer to Reduce the Preliminary Actions of Karate, *Journal of Sensors*, Vol. 24, No. 7, p. 2306 (2024).

[4] K. Kim, S. Tsuchida, T. Terada, and M. Tsukamoto: A Method for Detecting Preliminary Actions During an Actual Karate Kumite Match, *Journal of Sensors*, Vol. 25, No. 13, p. 4134 (2025).

[5] T. Hachaj, M.R. Ogiela, M. Piekarczyk, and K. Koptyra: Advanced Human Motion Analysis and Visualization: Comparison of Mawashi-Geri Kick of Two Elite Karate Athletes, *Proc. of the 2017 IEEE Symposium Series on Computational Intelligence (SSCI 2017)*, pp. 1–7 (2017).

[6] T. Hachaj, M.R. Ogiela, M. Piekarczyk, and K. Koptyra: Human Action Analysis: Templates Generation, Matching and Visualization Applied to Motion Capture of Highly-Skilled Karate Athletes, *Journal of Sensors*, Vol. 17, No. 11, p. 2590 (2017).

[7] D. Urribarri, M. Larrea, S. Castro, and E. Puppo: Visualization to Compare Karate Motion Captures, *Proc. of the XXV Congreso Argentino de Ciencias de la Computación (CACIC 2019)*, pp. 446–455 (2019).

[8] A.M. Vencesbrito, M.A.R. Ferreira, N. Cortes, O. Fernes, and P.P. Correia: Kinematic and Electromyo-

graphic Analyses of a Karate Punch, *Journal of Electromyography and Kinesiology*, Vol. 21, No. 6, pp. 1023–1029 (2011).

[9] M. Rinaldi, Y. Nasr, G. Atef, F. Bini, T. Varrecchia, C. Conte, G. Chini, A. Ranavolo, F. Draicchio, F. Pierelli, M. Amin, F. Marinozzi, and M. Serrao: Biomechanical Characterization of The Junzuki Karate Punch: Indexes of Performance, *European Journal of Sport Science*, Vol. 18, No. 7, pp. 796–805 (2018).

[10] H. Jemili, M.A. Mejri, R. Sioud, E. Bouhlel, and M. Amri: Changes in Muscle Activity During Karate Guiaku-zuki Punch and Kiza-mawashi-guiri Kick After Specific Training in Elite Athletes, *Journal of Science & Sports*, Vol. 32, No. 2, pp. 73–81 (2016).

[11] F.E. Ait-Bennacer, A. Aaroud, K. Akodadi, and B. Cherradi: Applying Deep Learning and Computer Vision Techniques for an E-Sport and Smart Coaching System Using a Multiview Dataset: Case of Shotokan Karate, *International Journal of Online and Biomedical Engineering*, Vol. 18, No. 12 pp. 35–53 (2022).

[12] G. Elkess, S. Elmoushy, and A. Atia: Karate First Kata Performance Analysis and Evaluation with Computer Vision and Machine Learning, *Proc. of the 2023 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC 2023)*, pp. 1–6 (2023).

[13] J. Echeverria and O.C. Santos: Toward Modeling Psychomotor Performance in Karate Combats Using Computer-vision Pose Estimation, *Journal of Sensors*, Vol. 21, No. 24, p. 8378 (2021).

[14] S. Khaksar, H. Pan, B. Borazjani, I. Murray, H. Agrawal, W. Liu, C. Elliott, C. Imms, A. Campbell, and C. Walmsley: Application of Inertial Measurement Units and Machine Learning Classification in Cerebral Palsy: Randomized Controlled Trial, *Journal of JMIR Rehabilitation and Assistive Technologies*, Vol. 8, No. 4, p. 29769 (2021).

[15] W. Sun, Y. Si, M. Guo, and S. Li: Driver Distraction Recognition Using Wearable IMU Sensor Data, *Journal of sustainability*, Vol. 13, No. 3, p. 1342 (2021).

[16] M.T. Worsey, H.G. Espinosa, J.B. Shepherd, and D.V. Thiel: Inertial Sensors for Performance Analysis in Combat Sports: A Systematic Review, *Journal of Sports*, Vol. 7, No. 1, p. 28 (2019).

[17] K. Teachasrisaksakul, L. Wu, G.Z. Yang, and B. Lo: Hand Gesture Recognition with Inertial Sensors, *Proc. of 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2018)*, pp. 3517–3520 (2018).

[18] B. Hartmann and N. Link: Gesture Recognition with Inertial Sensors and Optimized DTW Prototypes, *Proc. of 2010 IEEE International Conference on Systems, Man and Cybernetics (SMC 2010)*, pp. 2102–2109 (2010).

[19] K. Liu, C. Chen, R. Jafari, and N. Kehtarnavaz: Multi-HMM Classification for Hand Gesture Recognition Using Two Differing Modality Sensors, *Proc. of 2014 IEEE Dallas Circuits and Systems Conference (DCAS 2014)*, pp. 1–4 (2014).

[20] L. Zhang: Research on Human Body Movement Posture Based on Inertial Sensor, *Journal of Bioautomation*, Vol. 22, No. 2, pp. 179–186 (2018).

[21] G. Lefebvre, S. Berlemont, F. Mamalet, and C. Garcia: Inertial Gesture Recognition with BLSTM-RNN, *Proc. of International Conference on Artificial Neural Networks (ICANN 2013)*, pp. 393–410 (2013).

[22] B. Fang, Q. Lv, J. Shan, F. Sun, H. Liu, D. Guo, and Y. Zhao, Dynamic Gesture Recognition Using Inertial Sensors-based Data Gloves, *Proc. of 2019 IEEE 4th International Conference on Advanced Robotics and Mechatronics (ICARM 2019)*, pp. 390–395 (2019).

[23] J.G.C. Alfaro and A.L. Trejos: User-Independent Hand Gesture Recognition Classification Models Using Sensor Fusion, *Journal of Sensors*, Vol. 22, No. 4, p. 1321 (2022).

[24] Supriya, A. Shukla, and M. Manchanda, Inertial Sensor Based Human Activity Identification System Using CNN- LSTM Deep Learning Technique, *Proc. of 2023 10th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON 2023)*, pp. 305–310 (2023).

[25] C.S. Myers and L.R. Rabiner: A Comparative Study of Several Dynamic-time-warping Algorithms for Connected-word Recognition, *Bell System Technical Journal*, Vol. 60, No. 7, pp. 1389–1409 (1981).

Direction-of-Arrival Estimation of Impulsive Sounds Using the Spatial Audio Technology of Ambisonics

Shingo Sawazaki[†], Jinhui Chen[‡], and Takuya Yoshihiro[‡]

[†]Graduate School of Systems Engineering, Wakayama University, Japan
[‡]Faculty of Systems Engineering, Wakayama University, Japan
sawazaki.shingo@g.wakayama-u.jp, ckinki@wakayama-u.ac.jp, tac@wakayama-u.ac.jp

Abstract - In recent years, acoustic sensing has become increasingly important for robot environment recognition and search and rescue at disaster sites. However, most of the previous studies have focused on long-duration acoustic signals, and impulsive sounds such as hand-clap, which is essential for sound processing and analysis, have not been sufficiently treated. In this study, we propose a method to extract MFCCs from Ambisonics recordings and feed them into a 2-D CNN that performs regression to estimate the sound-source direction.

Keywords: Direction-of-Arrival (DOA), Ambisonics, Impulsive Sounds, Convolutional Neural Network (CNN), Deep Learning

1 INTRODUCTION

Acoustic sensing offers a contact-free way to obtain spatiotemporal environmental information, and sound-source localization is central to this capability. Sound source localization has been applied to a wide variety of fields, such as robotic environmental perception, search assistance at disaster sites, and acoustic sensor technology for use in autonomous vehicles. Against this background, research on sound source localization technology has seen rapid progress in recent years [1].

The methods for performing sound source localization using a microphone array [2]-[8] and an Ambisonics microphone [9]-[11] are outlined below. In the microphone array method, a large number of microphones are arranged in a planar, circular, or spherical configuration, and the location of the sound source is estimated based on time and intensity differences among the recorded acoustic signals. However, the large number of microphones increases the physical footprint of the array. On the other hand, the method using an Ambisonics microphone enables the recording of sound sources in all directions with a single microphone unit and can be used for sound source localization, thereby alleviating the constraints on installation space.

Sound Event Localization and Detection (SELD) is a key challenge in acoustic sensing. However, most existing SELD studies have focused on relatively long-duration sounds, such as speech and everyday environmental sounds, and the localization accuracy for impulsive sounds, such as hand-clap sounds, has not yet been thoroughly investigated. In addition, for practical applications in real-world environments, it is essential to estimate real sounds rather than sounds reproduced through loudspeakers. However, the accuracy of impulsive sound localization under such conditions has not yet been clarified.

Multiple simultaneous impulsive sounds produce overlapping directional components in both time and frequency, making separation difficult for either microphone arrays or Ambisonics microphones. For this reason, few studies have systematically evaluated the direction-of-arrival (DOA) estimation performance for multiple simultaneous impulsive sounds.

Deep learning, which has advanced rapidly in recent years, has attracted considerable attention as a promising approach to addressing these challenges. Deep learning enables efficient learning of complex acoustic signal features and achieves high-precision estimation. When combined with Ambisonics, it is expected to enable sound source localization of short impulsive sounds and facilitate practical performance evaluation. However, there have been no clear reports of studies applying Ambisonics or deep learning to short-time DOA estimation of impulsive sounds, and the potential of these approaches remains largely unexplored. Advances in impulsive sound recognition using Ambisonics microphone could potentially lead to significant progress in practical applications, such as acoustic-based rescue operations at disaster sites and anomaly detection systems. Specifically, in disaster rescue operations, estimating the direction of brief sounds—such as faint voices, coughs, and knocking noises made by victims—is expected to help streamline rescue activities and reduce the risk of secondary disasters. In addition, from a security perspective, anomaly detection systems can facilitate rapid response by identifying the direction of abnormal sounds, such as breaking glass or screams. In this way, this study represents an important step toward the advancement of acoustic sensing technology.

The aim of this study is to investigate the feasibility of estimating the direction of impulsive sounds—which has not been sufficiently explored—by using deep learning. Hand-clap sounds are used as a representative example of impulsive sound sources.

The structure of this paper is as follows. Section 2 reviews related work, and Section 3 presents the proposed method. Finally, Section 4 concludes the paper.

2 RELATED WORK

This section reviews related work on microphone array and Ambisonics-based approaches used as sensing methods for sound source localization. In a study employing the microphone array method, Tan et al. constructed a two-microphone array and estimated the DOA of impulsive sounds within a range of 0 to 180 degrees using a CNN-based regression model

(CNN-R), which integrates a convolutional neural network and a regression model [2]. However, this method has a limitation; it cannot determine the correct sound source position when the estimated direction falls within the 0–180 degree range and the source is symmetrically located with respect to the microphone array plane. Pujol et al. proposed an DOA estimation method for a single sound source using a microphone array composed of six microphones arranged in a circle and one microphone placed at the center, employing a convolutional neural network (CNN) [3]. They demonstrated that this method achieves high estimation performance even in noisy environments, outperforming conventional methods such as SRP-PHAT [4] and MUSIC [5]. However, this method is designed for speech and music with a single sound source, and its accuracy for impulsive sounds or multiple simultaneous sources remains unknown.

Diaz-Guerra et al. proposed a regression-based method for estimating the DOA of a single sound source [6]. Their approach utilizes a microphone array consisting of 12 microphones mounted on the head of a robot and constructs a model that combines a two-dimensional icosahedral CNN with a one-dimensional time-series CNN applied to SRP-PHAT maps. Huang et al. proposed a method for estimating the direction of a sound source using a spherical microphone array composed of 32 microphones arranged on a sphere with a radius of 4.2 cm [7]. They employed a CNN, using a covariance matrix as the input feature. However, these methods are designed for conversational speech from a single sound source, and their accuracy for impulsive sounds and multiple sound source scenarios remains unclear. Furthermore, He et al. proposed a multi-task deep neural network (DNN) that simultaneously performs DOA estimation and voice/non-voice classification for multiple sound sources, using a ResNet-based CNN [8]. They demonstrated that their method improves the accuracy of voice and noise discrimination. However, the accuracy of this method for impulsive sounds, as well as its ability to estimate multiple simultaneous impulsive sources, remains unclear.

In a study using the Ambisonics method, Adavanne et al. demonstrated that SELD can be performed in environments with overlapping sound sources using a convolutional recurrent neural network (CRNN) [9]. Tang et al. proposed a method for estimating the azimuth and elevation angles of a sound source using a CRNN [10]. They used four channels of First-Order Ambisonics (FOA) signals as input to the model. Krause et al. proposed a method for 3D sound event localization and detection (3D SELD), which simultaneously performs sound event detection, DOA estimation, and distance estimation [11]. They used four channels of FOA signals as input, and compared two approaches; a multi Activity-Coupled Cartesian Distance and DOA (multi-ACCDDOA) model and a multi-task model.

However, since these methods also target non-impulsive sounds, the estimation accuracy for impulsive sounds remains unclear.

3 PROPOSED METHOD

3.1 Overview of the Proposed Method

This study proposes a method for estimating the DOA of impulsive sounds recorded by an Ambisonics microphone using a deep learning model. In particular, the sound data recorded in the Ambisonics A-format is first converted into the B-format and pre-processed. Then, Mel-frequency cepstral coefficient (MFCC) features are extracted and used as input to a 2-D CNN for DOA estimation via regression or classification. This study aims to achieve high-accuracy DOA estimation by leveraging the Ambisonics format, which captures the spatial information of sound, and by training a CNN to learn time-frequency features of MFCCs.

3.2 Input Data

This study employs Ambisonics – a spatial-audio technology – as the input representation for the deep-learning model. By adopting Ambisonics, it becomes theoretically possible to estimate sound direction by decomposing pre-recorded omnidirectional sound data into directional components. Ambisonics is considered to outperform conventional stereo or multi-channel recording formats in DOA estimation. Therefore, we adopt this technology in our study.

Furthermore, MFCC features are extracted from the recorded sound after pre-processing that includes logarithmic transformation and noise removal. The raw audio waveform is converted into a two-dimensional feature map in the time-frequency domain, as using the raw audio waveform directly as input to a deep learning model tends to result in high-dimensional and complex data. MFCCs are widely used in various audio-related tasks, including speech recognition, as they can effectively compress and retain important acoustic features. Therefore, we use MFCCs as input features for DOA estimation in this study. The details of the pre-processing are described in Section **??**.

3.3 Composition of Deep Learning Model

In this study, we propose a method for estimating the DOA of impulsive sounds using deep learning. The structure of the proposed DOA estimation model is shown in Fig. 1. We repeatedly apply convolution and pooling operations four times to the MFCC features extracted from nine-channel hand-clap sounds to compress the characteristics of impulsive sounds. Thereafter, the proposed model outputs the estimated angle through three-layered fully connected layers. In the model shown in Fig. 1, the output dimension of the final layer differs depending on the task: it is 1 for regression, where the angle is estimated as a continuous value, and 24 for classification, where the angle is assigned to one of 24 discrete classes. In this study, regression is applied for the estimation of a single impulsive sound, while classification is used for overlapping impulsive sounds.

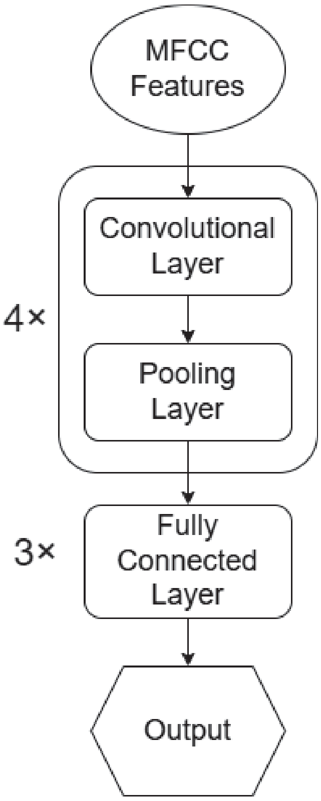


Figure 1: Overview of the DOA Estimation Model.

3.4 Convolutional Layer

In this study, the MFCC features are represented as a two-dimensional feature map. We employ the 2-D CNN to leverage the two-dimensional structure of the features. CNNs have a structure well-suited for handling two-dimensional feature maps, and are therefore effective at capturing local patterns. They are widely used in image recognition tasks, as they are well suited for extracting local features from two-dimensional data. Since MFCCs also form a two-dimensional feature map, applying convolutional layers enables effective extraction of local features in the time-frequency domain. In the convolutional layer, convolution is applied to the MFCC features to effectively compress the temporal and frequency characteristics of impulsive sounds. This convolutional processing reduces the amount of data and improves computational efficiency, while retaining important features necessary for estimating the direction of impulsive sounds. In the proposed model, the kernel sizes of the convolutional layers are set to 5×5, 5×5, 3×3, and 3×3 from the input layer onward. The purpose of this architecture is to effectively capture broad features of the input data in the initial layers and to extract more detailed local features in the later layers. The input to the convolutional layer is represented as a tensor of shape [C, T, M]. Here, C denotes the number of audio channels, T denotes the number of time frames (samples along the time axis), and M is the dimensionality of the feature vector. The ReLU function is used as the activation function. To prevent overfitting, L2 regularization was applied to the weight parameters of all convolutional layers. In particular, regularization was applied to the weight parameters responsible for transforming inputs to outputs in each layer, so that the regularization term would be taken into account during parameter updates at each layer.

3.5 Pooling Layer

Pooling layers serve to reduce the size of feature maps extracted by convolutional layers. This process reduces the size of the feature maps while preserving their spatial information. In this method, a pooling layer is placed after each convolutional layer. Max pooling is used to extract the maximum value from each region of the feature map.

3.6 Fully Connected Layer

Three fully connected layers are placed at the end. These layers serve to produce the final angle prediction by integrating features extracted from the convolutional and pooling layers. The first and second fully connected layers employ the ReLU activation function to improve the model's expressiveness by introducing non-linearity. The final fully connected layer employs the sigmoid activation function to normalize the predicted angle value to the range [0, 1]. This architecture allows the output to be appropriately scaled according to the actual range of angles. In addition, dropout was applied to the first two layers, and L2 regularization was applied to the weight parameters of all fully connected layers to prevent overfitting. This helps prevent overfitting to the training data and enhances the model's generalization ability to unseen data.

4 CONCLUSION

In this study, we developed a deep learning-based model for estimating the DOA of impulsive sounds.

Future work includes improving the estimation accuracy for real impulsive sounds and extending the method to handle multiple simultaneous sound sources.

REFERENCES

[1] Tong, K., Hu, Y., Dikic, B., Solmaz, S., Fraundorfer, F., and Watzenig, D., "Robots Saving Lives: A Literature Review About Search and Rescue (SAR) in Harsh Environments," 2024 IEEE Intelligent Vehicles Symposium (IV), Jeju Island, Korea, Republic of, 2024, pp. 953-960.

[2] Tan, T.-H., Lin, Y.-T., Chang, Y.-L., and Alkhaleefah, M., "Sound Source Localization Using a Convolutional Neural Network and Regression Model,"Sensors, vol. 21, no. 8031, pp. 1–17, Dec. 2021. doi: 10.3390/s21238031.

[3] Pujol, H., Bavu, É., and Garcia, A., "BeamLearning: an end-to-end Deep Learning approach for the angular localization of sound sources using raw multichannel acoustic pressure data," arXiv preprint, arXiv:2104.13347v1, 2021. Accessed: May 5,2025.

[4] DiBiase, J. H., "A High-Accuracy, Low-Latency Technique for Talker Localization in Reverberant Environments Using Microphone Arrays," Ph.D. dissertation, Dept. of Eng., Brown Univ., Providence, RI, USA, 2000.

[5] Schmidt, R. O., ”Multiple emitter location and signal parameter estimation,” *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, Mar. 1986, doi: 10.1109/TAP.1986.1143830.

[6] Diaz-Guerra, D., Miguel, A., and Beltran, J. R., “Direction of Arrival Estimation of Sound Sources Using Icosahedral CNNs” IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 30, pp. 3342–3355, 2022, doi: 10.1109/TASLP.2022.3224282.

[7] Huang, Q., and Fang, W., “ A Deep Learning Method for DOA Estimation with Covariance Matrices in Reverberant Environments,” Applied Sciences, vol. 12, no. 9, Art. no. 4278, Apr. 2022, doi: 10.3390/app12094278.

[8] He, W., Motlicek, P., and Odobez, J. M., “Joint Localization and Classification of Multiple Sound Sources Using a Multi-task Neural Network,” Interspeech, 2018.

[9] Adavanne, S., Politis, A., Nikunen, J., and Virtanen, T., “Sound Event Localization and Detection of Overlapping Sources Using Convolutional Recurrent Neural Networks,” arXiv preprint, arXiv:1807.00129v3, 2018. Accessed: May 5,2025.

[10] Tang, Z., Kanu, J. D., Hogan, K., and Manocha, D., “ Regression and Classification for Direction-of-Arrival Estimation with Convolutional Recurrent Neural Networks, ” Proc. Interspeech, Graz, Austria, Sep. 2019, pp. 654–658, doi: 10.21437/Interspeech.2019-1111654.

[11] Krause, D. A., Politis, A., and Mesaros, A., “ Sound Event Detection and Localization with Distance Estimation, ” arXiv preprint, arXiv:2403.11827v2, Jun. 2024. Accessed: May 12,2025.

[12] Loshchilov, I., and Hutter, F., “Decoupled Weight Decay Regularization,” in Proc. of the International Conference on Learning Representations (ICLR), 2019.

[13] Voyage Audio, “Spatial Mic - Voyage Audio”, voyageaudio, `https://voyage.audio/spatialmic/`. Accessed: May 5,2025.

[14] Anker, “Soundcore 3”, Anker, `https://www.ankerjapan.com/products/a3117`. Accessed: May 5,2025.

Exception-Oriented Test Case Generation and Evaluation using GitHub Copilot

Sogo Yano[†], Kozo Okano[‡], Shinpei Ogata*

[†]Graduate School of Science and Technology, Shinshu University, Japan
[‡]* Faculty of Engineering, Shinshu University, Japan
{25w6097a, ogata}@shinshu-u.ac.jp, okano@cs.shinshu-u.ac.jp

Abstract

Testing is indispensable to software robustness, and exception-oriented tests are particularly critical for safety and reliability. In practice, however, they remain underutilized, largely because of the oracle problem—deciding whether a thrown exception reflects the specification or a latent bug. We propose an LLM-based method for generating exception-oriented test cases and evaluate it using a quantitative oracle framework based on an Expected–Thrown vs. Designed–Undesigned context matrix to characterise exception semantics. Using GitHub Copilot, we automatically generated Java exception tests with three prompt strategies—Zero-Shot, Test-aware Prompting, and Iterative—producing 1,008 tests for fixed versions of Defects4J programs. Only about 20% of the generated tests actually executed a throw statement. Test-aware Prompting and Iterative Prompting surpassed Zero-Shot in throw-statement coverage and the number of detected Undesigned exceptions, yet incurred more compilation and oracle-mismatch errors, underscoring the intrinsic difficulty of exception-test generation.

Keywords: Exception Testing, GitHub Copilot, Test Generation, Large Language Models

1 Introduction

Software testing ensures program correctness and stability even under unexpected conditions [1]. Recent studies explore automating test generation using Large Language Models (LLMs), which can assist in producing and refining test suites [2, 3], improving efficiency and software quality [4].

However, most LLM-based approaches focus on normal behavior. The automatic generation of *exception tests*—those targeting abnormal or failure-inducing inputs—remains largely unexplored [4], even though robust exception handling is critical for preventing crashes and security flaws [5]. Traditional generators such as Randoop and EvoSuite [6, 7] broaden input coverage but ignore semantic intent, limiting their value for exception analysis.

LLM-based testing faces the classic *oracle problem* [8]: the model generates both test logic and assertions but cannot verify their correctness, especially for exceptions where expected behavior is often undocumented. Existing metrics (e.g., coverage or compile rate) assume known oracles and thus fail to evaluate exception tests properly [3, 9].

To address this, we classify exceptions along two axes—*Expected vs. Thrown* and *Designed vs. Undesigned*—to build a semantic oracle that distinguishes intended exceptions from potential bugs. Our framework extracts specifications (E_spec) and identifies Undesigned exceptions as those absent from the specification. This enables automated, context-aware evaluation of LLM-generated tests.

We use GitHub Copilot for its IDE-integrated code generation and target Java for its explicit exception system and structured specifications [10].

Based on the above, we formulate three research questions:

- RQ1: Do LLMs possess the ability to generate test cases that appropriately involve exception occurrences?
- RQ2: Does variation in prompt design—an essential feature of LLM usage—significantly affect the generated outputs?
- RQ3: Can the contextual meaning of exceptions be captured to detect Undesigned exceptions?

The remainder of this paper is organized as follows. Section 2 clarifies the terminology and definitions used throughout the study. Section 3 introduces the proposed approach, including exception specification extraction and prompt design. Section 4 explains the experimental setup and evaluation metrics, followed by Section 5, which presents the results. Section 6 offers analysis and discussion. Finally, Section 7 reviews related work.

2 Preparation

We first define key terms and concepts related to generating and evaluating exception tests. This includes classification of JUnit output messages, exception types, and exception contexts (“thrown,” “observed,” “expected”). These definitions form the basis for our experimental design and evaluation metrics.

2.1 Definition of Test Types Used in This Study

This study focuses on two types of tests: positive tests and exception tests. While software tests can be categorized in various ways, this classification reflects a practical distinction between verifying normal behavior and handling exceptional conditions.

Positive Tests. A positive test verifies that a program behaves as specified under normal input or conditions. It checks expected results via assertions such as *assertEquals()*, *assertTrue()*, or *fail()* in JUnit. If a mismatch occurs, JUnit throws an *AssertionError*.

Exception Tests. An exception test verifies that a program correctly throws or handles exceptions under abnormal conditions.

Its goal is to detect both expected and undesigned exceptions. Expected exceptions are explicitly declared in the test, and JUnit 4 supports two common patterns for such verification.

2.2 JUnit Internal Messages

JUnit outputs stable, pattern-based messages. We classify them by matching fixed *AssertionError* formats and similar deterministic patterns, which enables consistent parsing and exception-type identification.

2.3 Definition of Exception Types

In Java, all throwable objects inherit from the root class *Throwable*, which divides into two main branches: *Error* and *Exception* [11, 12]. *Error* types (e.g., *OutOfMemoryError*) represent critical system-level failures and are excluded from this study. We focus instead on two subtypes of *Exception*: checked and unchecked exceptions.

Unchecked Exceptions. These extend *RuntimeException* and do not require a *throws* clause. Their implicit behavior often leads developers to overlook them. Examples include *NullPointerException*, *IndexOutOfBoundsException*, and *ClassCastException*. Some APIs deliberately throw unchecked exceptions to enforce input validation.

Checked Exceptions. These extend *Exception* but are not *RuntimeException* subclasses. They must be declared with a *throws* clause, or compilation will fail. Typical examples are *IOException*, *SQLException*, and *ClassNotFoundException*. Because they are explicitly declared, they are rarely thrown unintentionally.

2.4 Definition of Context

This section defines five key terms—*Thrown*, *Observed*, *Expected*,and *Detection*—used to interpret exception behavior during test execution.

Thrown Exceptions. All exceptions thrown at the JVM level, including those raised via *throw* statements in application or library code. Exceptions swallowed by *try–catch* blocks without logging or output (i.e., subsumed exceptions) technically occur but may not be fully recorded, making comprehensive logging difficult.

Observed Exceptions. Exceptions that are visible in some form—test results, logs, or stack traces. Examples include error messages printed by JUnit or stack traces in log files. If an exception is caught and suppressed without output, it is *not* observed.

Expected Exceptions. Exceptions anticipated by the test case or specification in advance. If the specified exception does not occur—or a different one occurs—the test framework can judge that "the expected exception did not occur."

Detection. An undesigned exception both occurs and is observed during test execution. This indicates that the test has successfully revealed an exception that was not declared in the specification.

2.5 Definition of Designed vs. Undesigned Exceptions

As evaluation metrics for test generation, we measure whether the generated tests properly cover *designed exceptions* and whether they detect unintended *undesigned exceptions*. Their definitions are as follows:

Designed Exception. An exception that is explicitly declared in the program logic or specification, and is anticipated as part of normal behavior. Such exceptions are typically verified in tests using mechanisms like *@Test(expected=...)*, *assertThrows*, or by explicitly handling them through *try-catch* blocks.

Undesigned Exceptions. Exceptions that occur unintentionally and are not anticipated by the test code. These are primarily unchecked exceptions such as NullPointerException or IndexOutOfBoundsException, which arise from bugs or defects. If test generation fails to anticipate these, the tests may fail unexpectedly at runtime.

3 Proposed Approach

This section details the "LLM-based exception-test generation method" proposed in this study. First, we present a test-generation flow that employs three prompt strategies—Zero-Shot Prompting, Test-aware Prompting, and Iterative Prompting—and explain the underlying conceptual structure and intent [13]. Next, we introduce the "exception-context model," a simple oracle framework based on a two-axis context matrix—Expected vs. Thrown and Designed vs. Undesigned—to characterize exception semantics and evaluate the generated test cases.

3.1 Prompt-Based Test Generation Flow

We use Large Language Models (LLMs) to automate exception test creation, focusing on inputs that trigger undesigned exceptions beyond typical generator reach. Guided by the SUT–test relationship (Fig. 1), we adopt three prompt strategies:

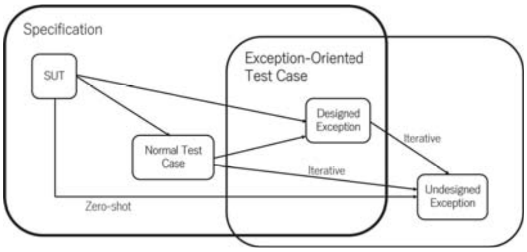


Figure 1: Conceptual relationship between SUT and test cases

- **Zero-Shot Prompting(ZSP)**: Baseline approach. Generates exception tests without additional context, relying on the model's internal knowledge.
- **Test-aware Prompting(TaP)**: Uses existing human-written tests as implicit context via GitHub Copilot, generating complementary exception tests.

- **Iterative Prompting(Iterative)**: Refines results iteratively, reusing generated tests to gradually shift from Designed to Undesigned exceptions.

Example Prompts.

"Please generate exception-handling test cases within the specified scope."

"Generate exception tests complementing existing normal and designed cases, focusing on undesigned exceptions from unexpected inputs or states."

"Identify uncovered inputs and generate broader exception test cases."

3.2 Exception-Context Model for Oracle Construction

To address the oracle problem for exceptions, we introduce an *exception-context model*. This model classifies and evaluates exceptions along the following two axes:

- **Expected exceptions vs. Thrown exceptions**
- **Designed exceptions vs. Undesigned exceptions**

If the expected exception type is thrown, the test *passes* (assertion succeeds).

If the expected and thrown types differ, the event is treated as *exceptionMismatch*.

If the expected exception is not thrown, it is treated as *exceptionNotThrown*.

If an unexpected exception occurs without being expected, it is treated as *exceptionUnexpected*.

The oracle problem here is the need to decide whether an observed exception is specification-compliant or a bug. Using the exception-context model enables the following:

If the thrown exception is an Undesigned exception, it is deemed a bug that violates the specification.

Introducing $E_{\text{spec}}(s)$ For each SUT s, we pre-collect the set of specification-defined exceptions $E_{\text{spec}}(s)$ from source code and existing tests [14]. Designed exceptions are extracted from four sources:

$$E_{\text{spec}}(s) = \{\, e \mid e \in \texttt{throws} \,\} \cup \{\, e \mid e \in \texttt{throw} \,\} \\ \cup \{\, e \mid e \in \texttt{expected=e} \,\} \cup \{\, e \mid e \in \texttt{try-catch} \,\}$$

By comparing this $E_{\text{spec}}(s)$ with the set of observed exceptions after test generation, we can quantitatively evaluate the appearance of Undesigned (unexpected) exceptions. In the rest of this paper, we use the notation E_spec to refer to this exception set, omitting (s) for brevity.

4 Evaluation Experiment

This section describes the experiment.

4.1 Experimental Procedure

This experiment evaluates the quality and characteristics of exception-oriented test cases generated by LLMs. The procedure consists of the following five steps:

1. **Prompting and Test Generation:** For each of the six SUTs, we applied the three prompt strategies introduced in Section 3—Zero-Shot Prompting (ZSP), Test-aware Prompting (TaP), and Iterative Prompting. To account for stochastic variability in LLM outputs, each strategy was executed five times, resulting in a total of 90 test generation attempts [15].
2. **Static Analysis and Coverage Preparation:** For each SUT method, we performed static analysis using Spoon [16]. First, we identified all *throw* statements and inserted logging statements immediately before them to enable exception coverage tracking. Additionally, we extracted the exception specifications (E_spec) from the SUTs and their existing test suites, which were later compared against observed runtime exceptions. We also analyzed method dependencies within the SUTs.
3. **Compilation and Execution:** The generated test files were compiled using Apache Maven (with manual fixes for compilation errors as needed) and executed using the JUnit 4 framework in a Java environment.
4. **Log Collection and Error Classification:** Execution logs and stack traces were collected. We classified error types using a combination of pattern matching, LLM-based inference, and manual validation.
5. **Metric Calculation:** Evaluation metrics were computed based on the exception-context model described in Section 3.

The following subsections provide further details on the experimental environment, the systems under test (SUT), and the definitions and collection methods of the evaluation metrics.

4.2 Experimental Environment

All experiments were conducted in a Java/JUnit 4 environment using GitHub Copilot as the LLM-based test generator.

4.3 System Under Test (SUT)

We selected target methods from the Defects4J dataset, focusing on those prone to undesigned exceptions. In total, 68 methods were analyzed across multiple open-source libraries.

4.4 Evaluation Metrics

We evaluated the generated test cases using the following metrics, based on our exception-context model.

Compilation Errors. Counts the number of test cases that failed to compile using Maven. This serves as a basic quality check for the generated code.

Exception Coverage. Measures how thoroughly exception-throwing behaviors are exercised.

Throw Statement Coverage: Adopted from Goffi et al. [17], defined as the ratio of executed *throw* statements to the total number of unique ones. Only actually executed statements are counted; *try-catch* blocks are excluded.

Throw Path Intensity: Counts all executions of *throw* statements, including repeated executions.

Runtime Errors.
Pass Rate: Tests where all assertions succeed.
Error Classification:
assertionMismatch — Failed assertions not involving exceptions.
exceptionMismatch — Thrown exception differs from expected; categorized by type pairings.
exceptionNotThrown — Expected exceptions not thrown. Further categorized as:
  pathexists — Throw reachable but not triggered.
  pathmissing — Throw absent.
exceptionUnexpected — Unanticipated exceptions, labeled as checked or runtime.
All exceptions are labeled as *Designed* or *Undesigned*.

4.5 Metric Collection Methods

Exception Coverage. We performed static analysis using *Spoon* and inserted log output statements before each *throw* statement. This allowed us to identify the *throw* statements that were executed during test execution.
Runtime Error Classification. JUnit stack traces were first parsed using regular expressions. These were then verified using a combination of GPT-4o-based inference and manual inspection.

5 Results

In this experiment, a total of 1,008 test cases were generated using three prompt strategies: ZSP, TaP, and Iterative. This section presents results on compilation errors, exception coverage, runtime error classification, and detection of Designed versus Undesigned exceptions.

5.1 Compilation Errors

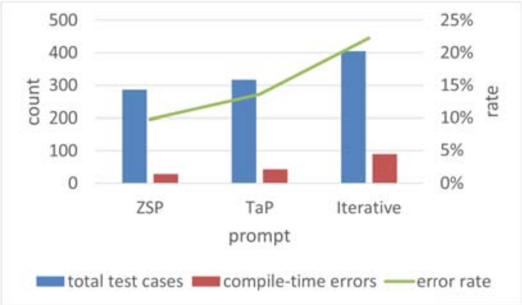


Figure 2: Number of compilation errors and total test cases per prompt strategy

Many initial compilation errors stemmed from environment inconsistencies. For example, the generated tests mixed JUnit 3/4/5 syntax, had invalid JUnit annotations, or missed required import and package statements. These issues likely arose because Defects4J programs include older tests using different JUnit versions. We resolved these problems by standardizing all tests to JUnit 4 and applying necessary fixes (using Copilot suggestions and manual edits) before recompiling.

After this standardization process, we recompiled all test cases and collected only those errors that persisted as semantically or syntactically meaningful faults. These formally recognized compilation errors—along with the total number of generated test cases per prompt strategy—are reflected in Figure 2. The types of errors included issues such as passing *null* to parameters of primitive types, mismatched or reversed argument orders and types, incompatibilities caused by version differences in external libraries (e.g., Commons CSV or CLI), inconsistent array types or confusion between values and variables, and access violations resulting from the use of non-public APIs.

As shown in Figure 2, Iterative Prompting resulted in the highest number of such compilation errors. Importantly, all of these errors were subsequently corrected, and the repaired test cases were included in the later stages of evaluation, including exception coverage and runtime error classification.

5.2 Exception Coverage

Exception coverage was measured for the existing test cases (*existing*) as well, using the same criteria as for the generated tests. This served as a reference point to assess how the LLM-generated tests compare to manually written tests in terms of exception path coverage.

We excluded the createNumber method (from Lang) because its existing tests achieve 100% exception coverage by explicitly targeting every throw (likely due to many historical bugs). This extreme manual targeting made it an outlier and an unfair comparison for our automatically generated tests.

As a result, the denominator in the exception coverage metric was reduced by one SUT. Nevertheless, the overall exception coverage achieved by the existing tests remained relatively low.

5.2.1 Unique Coverage

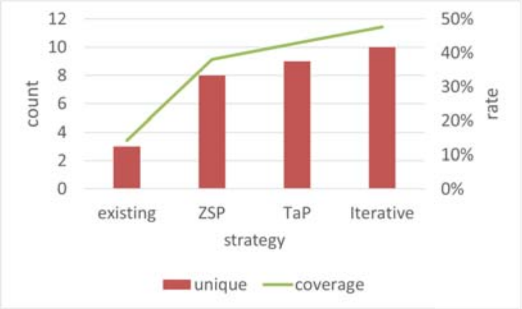


Figure 3: Unique throw statement coverage

Unique throw statement coverage improved in the order of existing $<$ ZSP $<$ TaP $<$ Iterative, as shown in Figure 3.

5.2.2 Total Coverage

For total (non-unique) throw coverage, both TaP and Iterative showed similarly high ratios, confirming that these strategies generated many exception-focused tests4.

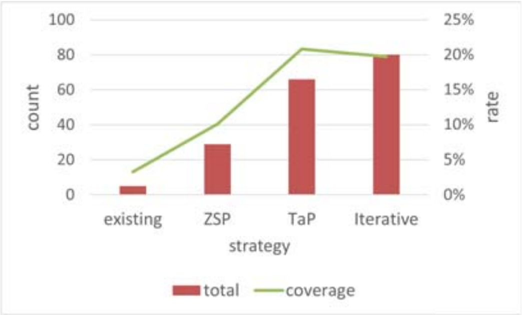


Figure 4: Total throw statement coverage (including duplicates)

5.3 Runtime Errors

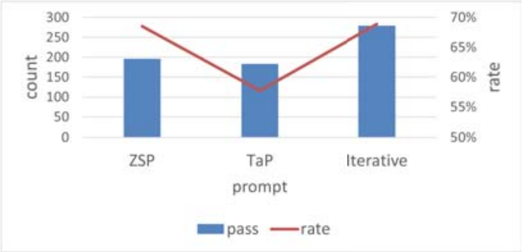


Figure 5: Pass rate of test cases per strategy

In terms of runtime results, TaP exhibited the lowest *pass rate* among the strategies5. Here, “pass” refers to a test case that completed without an *error* or *failure*.

5.3.1 Four Major Runtime Error Categories

Runtime errors were classified into the following four categories and summarized in an Expect–Thrown matrix in Table 1. Note that *Normal (N)* refers to *non-exceptional behavior*. A pass refers to a case where *Expected* and *Thrown* match exactly; we did not record the breakdown of these counts. For cases where an exception was expected but a normal outcome occurred, we did not distinguish between Checked and Runtime and instead counted them together. The numerical values in each cell denote the number of test cases corresponding to that Expected–Thrown combination, where “Value mismatch” indicates differing outputs under normal execution, and “Type mismatch” refers to mismatched exception types.

AssertionMismatch. Occurs in the *Normal→Normal* pattern, where the output value does not match the expectation and an *AssertionError* is raised. These are classified as non-exception mismatches.

ExceptionMismatch. Occurs in exception–exception matrix cells where the thrown exception differs from the expected one. We further subdivided these mismatches based on whether the expected and actual types were *RuntimeException* or *CheckedException*. Overall, such mismatches were rare.

ExceptionNotThrown. Refers to matrix cells where an exception was expected but a normal result was observed. This pattern was common in Iterative but less frequent in TaP. We further categorized these as *pathexists* (a feasible path to the exception exists) or *pathmissing* (no feasible path exists—likely a hallucination).

ExceptionUnexpected. Occurs when an exception is thrown despite a *Normal* expectation.

5.3.2 Undesigned Exceptions

We counted the number of undesigned exceptions by calculating the difference between the observed exception types and the predefined set of designed exceptions E_spec for each SUT. The number of detections of undesigned exceptions was computed by counting the number of unique exception types per SUT and summing them across all SUTs. All undesigned exceptions observed were of type RuntimeException.

The number of detected undesigned exceptions becomes 3 for ZSP, 9 for TaP, and 8 for Iterative.

These results indicate that TaP and Iterative detected many undesigned exceptions.

6 Discussion

This section answers each research question.

6.1 RQ1: Can LLMs generate tests that appropriately involve exceptions?

When generating exception-handling tests, the LLM exhibited a higher rate of compilation and runtime errors, reflecting the structural complexity of exception control flow. Using throw path intensity as an indicator, only about 20% of generated tests reached code paths with explicit *throw* statements. This result suggests that exception-oriented generation remains challenging, partly due to the limited representation of exception-handling patterns in LLM training data [5, 18]. Among observed cases, *exceptionNotThrown* (expected but not raised) was the most frequent, showing that the model often failed to produce inputs satisfying exception-triggering conditions, whereas true type mismatches were relatively rare.

6.2 RQ2: Does prompt design materially change the outputs?

Variation in prompt design significantly influenced the quality and coverage of generated tests. Test-aware Prompting (TaP) improved assertion accuracy by referencing existing tests, reducing *assertionMismatch* cases, yet suffered from lower overall pass rates due to increased *exceptionUnexpected* and oracle-related errors. Iterative Prompting, while prone to more compilation errors, achieved the highest unique *throw*-statement coverage and improved undesigned-exception detection through successive refinement. Overall, both TaP and Iterative outperformed the Zero-Shot baseline in exploring exception paths, indicating that structured prompt design enhances exception-oriented reasoning but also amplifies syntactic and contextual risks.

6.3 RQ3: Can exception context modeling enable detection of Undesigned exceptions?

A limited number of undesigned exceptions were detected, most of them *RuntimeException*s representing unforeseen bugs

Table 1: Expect–Thrown Matrix

(a) ZSP

Expected Thrown	**Normal (N)**	**Checked (C)**	**Runtime (R)**
Normal (N)	✓ NN-Pass (Value match) Value mismatch: 46		34
Checked (C)	2	✓ CC-Pass (Type match) Type mismatch: 0	1
Runtime (R)	8	0	✓ UU-Pass (Type match) Type mismatch: 0

(b) TaP

Expected Thrown	**Normal (N)**	**Checked (C)**	**Runtime (R)**
Normal (N)	✓ NN-Pass (Value match) Value mismatch: 28		49
Checked (C)	0	✓ CC-Pass (Type match) Type mismatch: 0	1
Runtime (R)	38	0	✓ UU-Pass (Type match) Type mismatch: 12

(c) Iterative

Expected Thrown	**Normal (N)**	**Checked (C)**	**Runtime (R)**
Normal (N)	✓ NN-Pass (Value match) Value mismatch: 54		48
Checked (C)	2	✓ CC-Pass (Type match) Type mismatch: 2	0
Runtime (R)	20	0	✓ UU-Pass (Type match) Type mismatch: 0

outside the specification. GitHub Copilot could identify potential exception types for each method but often failed to generate the specific inputs needed to trigger those paths. This indicates that while LLMs possess some awareness of exception contexts, their ability to reproduce undesigned exceptions through valid input generation remains limited.

6.4 Threats to Validity

This study is subject to several threats to validity.

Learning-data contamination. GitHub Copilot's training set may include source code or tests from Defects4J, so prior knowledge could have biased generation and evaluation.

Extent of contextual influence. The precise scope of Copilot's internal context—prompt history and code snippets—cannot be inspected or controlled, making it impossible to determine exactly how far iterative prompts propagate their influence.

Ambiguity in error classification. Relying solely on JUnit logs makes it difficult to distinguish whether an exception stems from the SUT or from faults in the test code itself, so both kinds of errors may be conflated. Additionally, exceptions suppressed inside try-catch blocks (”subsumed”) are not observable in logs, making them unmeasurable and potentially misclassified as unthrown.

Accuracy of E_spec extraction. Because Defects4J does not provide an explicit mapping between each method and its corresponding design-time tests, static extraction of designed exceptions (E_spec) may contain inaccuracies.

Evaluation limited to JUnit 4. For consistency we used JUnit 4, but differences in specifications and exception-message patterns in other versions (e.g., JUnit 5) were not considered, which may restrict the generalizability of our classification logic.

Prompt-design fragility. Because GitHub Copilot is not fully robust, minor variations in natural-language prompts can greatly affect the generated code, demanding meticulous prompt design and reproducible prompt configurations [19].

7 Related Work

Prior studies on exception testing can be grouped into three areas. First, Zhang et al. [20] proposed EXLONG, an LLM-based framework for generating exceptional behavior tests. Unlike EXLONG, our method classifies exceptions by context and evaluates their semantic validity.

Second, TOGA [21] uses neural models to generate exception oracles from documentation, but lacks code-level reasoning. In contrast, we leverage LLMs with full method access for context-aware exception handling.

Finally, Barr et al. [8] highlighted the oracle problem's complexity in software testing. Our work extends this discussion by redefining exception oracles through designed/undesigned classification.

ACKNOWLEDGEMENTS

The research is being partially conducted as Grant-in- Aid for Scientific Research C (21K11826). The authors wish to express their profound gratitude to Professor Shin Nakashima of the National Institute of Informatics for his insightful guidance and steadfast support, which have been instrumental to the successful completion of this research.

REFERENCES

[1] Glenford J. Myers, Corey Sandler, and Tom Badgett. *The Art of Software Testing*. John Wiley & Sons, Hoboken, NJ, 2 edition, 2011. 256 pages.

[2] Zhiqiang Yuan, Yiling Lou, Mingwei Liu, Shiji Ding, Kaixin Wang, Yixuan Chen, and Xin Peng. No more manual tests? evaluating and improving chatgpt for unit test generation. *Proceedings of the ACM on Software Engineering*, 1(FSE):1703–1726, 2024.

[3] Khalid El Haji, Carolin Brandt, and Andy Zaidman. Using github copilot for test generation in python: An empirical study. In *Proceedings of the 5th ACM/IEEE International Conference on Automation of Software Test (AST)*, pages 45–55, 2024.

[4] Angela Fan, Beliz Gokkaya, Mark Harman, Mitya Lyubarskiy, Shubho Sengupta, Shin Yoo, and Jie M. Zhang. Large language models for software engineering: Survey and open problems. *arXiv preprint*, 2023.

[5] Francisco Dalton, Márcio Ribeiro, Gustavo Pinto, Leo Fernandes, Rohit Gheyi, and Baldoino Fonseca. Is exceptional behavior testing an exception? an empirical assessment using java automated tests. In *Proceedings of the 25th International Conference on Evaluation and Assessment in Software Engineering (EASE)*, pages 170–179, 2020.

[6] Randoop Development Team. Randoop manual. `https://randoop.github.io/randoop/manual/`, 2024. Accessed: 2024-06-11.

[7] Sebastian Schweikl, Gordon Fraser, and Andrea Arcuri. Evosuite at the sbst 2022 tool competition. In *International Workshop on Search-Based Software Testing (SBST)*, pages 33–34, 2022.

[8] Earl T. Barr, Mark Harman, Phil McMinn, Muzammil Shahbaz, and Shin Yoo. The oracle problem in software testing: A survey. *IEEE Transactions on Software Engineering*, 41(5):507–525, 2015.

[9] Shuo Ren, Daya Guo, Shuai Lu, Long Zhou, Shujie Liu, Duyu Tang, Neel Sundaresan, Ming Zhou, Ambrosio Blanco, and Shuai Ma. Codebleu: A method for automatic evaluation of code synthesis. In *Proceedings of the 35th AAAI Conference on Artificial Intelligence (AAAI)*, pages 4503–4510, 2021.

[10] Mohammed Latif Siddiq, Joanna C. S. Santos, Ridwanul Hasan Tanvir, Noshin Ulfat, Fahmid Al Rifat, and Vinicius Carvalho Lopes. Using large language models to generate JUnit tests: An empirical study. In *Proceedings of the 28th International Conference on Evaluation and Assessment in Software Engineering (EASE)*, pages 313–322, 2024.

[11] Oracle. Java language specification, java se 21 edition, chapter 11: Exceptions. `https://docs.oracle.com/javase/specs/jls/se21/html/jls-11.html`, 2025. Accessed: 2025-06-08.

[12] Oracle. Java platform se 8 api specification: Package `java.lang`. `https://docs.oracle.com/javase/8/docs/api/java/lang/package-tree.html`, 2025. Accessed: 2025-06-08.

[13] Jason Wei, Maarten Bosma, Vincent Y. Zhao, Kelvin Guu, Adams Wei Yu, Brian Lester, Nan Du, Andrew M. Dai, and Quoc V. Le. Finetuned language models are zero-shot learners. In *Proceedings of the 10th International Conference on Learning Representations (ICLR)*, 2022.

[14] Martin P. Robillard and Gail C. Murphy. Analyzing exception flow in java programs. *ACM SIGSOFT Software Engineering Notes*, 24(6):322–337, 1999.

[15] Laura Plein, Wendkûuni C. Ouédraogo, Jacques Klein, and Tegawendé F. Bissyandé. Automatic generation of test cases based on bug reports: A feasibility study with large language models. In *Proceedings of the 2024 IEEE/ACM 46th International Conference on Software Engineering: Companion Proceedings (ICSE Companion)*, pages 360–361, 2024.

[16] Renaud Pawlak, Martin Monperrus, Nicolas Petitprez, Carlos Noguera, and Lionel Seinturier. Spoon: A library for implementing analyses and transformations of java source code. *Software: Practice and Experience*, 46(9):1155–1179, 2015.

[17] Alberto Goffi, Alessandra Gorla, Michael D. Ernst, and Mauro Pezzè. Automatic generation of oracles for exceptional behaviors. In *Proceedings of the 25th International Symposium on Software Testing and Analysis (ISSTA)*, pages 213–224, 2016.

[18] Diego Marcilio and Carlo A. Furia. How Java programmers test exceptional behavior. In *Proceedings of the 18th IEEE/ACM International Conference on Mining Software Repositories (MSR)*, pages 207–218, 2021.

[19] Antonio Mastropaolo, Luca Pascarella, Emanuela Guglielmi, Matteo Ciniselli, Simone Scalabrino, Rocco Oliveto, and Gabriele Bavota. On the robustness of code generation techniques: An empirical study on github copilot. In *Proceedings of the 45th IEEE/ACM International Conference on Software Engineering (ICSE)*, 2023.

[20] Jiyang Zhang, Yu Liu, Pengyu Nie, Junyi Jessy Li, and Milos Gligoric. exlong: Generating exceptional behavior tests with large language models. *arXiv preprint*, 2024.

[21] Elizabeth Dinella, Gabriel Ryan, Todd Mytkowicz, and Shuvendu K. Lahiri. Toga: A neural method for test oracle generation. In *Proceedings of the 44th International Conference on Software Engineering (ICSE)*, pages 2130–2141, 2022.

Keynote Speech 1:
Mr. Shingo Kinoshita
(Senior Vice President of Research and Development Planning Department, NTT)
(Chair: Tetsuya Yokotani)

(No distribution of materials)

Session 3:
User Behavior and Interaction Design

(Chair: Kei Hiroi)

A Cycling Route Recommendation System based on AI-Predicted Narrative Scores

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

[†]Graduate School of Data Science, Shiga University, Japan
[‡]University Library, The University of Osaka, Japan
tomisu@ieee.org, s6024150@st.shiga-u.ac.jp, kai.n.lib@osaka-u.ac.jp, yoshihisa@biwako.shiga-u.ac.jp

Abstract - Existing cycle tourism initiatives often overlook casual tourists by focusing on physically demanding, sport-oriented routes. However, existing initiatives and routing tools tend to prioritize sport-oriented, physically demanding courses and overlook casual tourists and the narrative value of the experience. This study proposes an automated system that generates personalized cycling routes by evaluating the narrative quality of the tourism experience. We introduce the Narrative Factor Score (NS). This novel metric quantifies a route's thematic coherence, story depth, and emotional connection using a generative AI to analyze and enhance data from points of interest.

On a real-world route (Biwaichi Minami), persona-tuned routes achieved higher NSs than the baseline (0.687), reaching 0.710–0.798 with only +1.30–+2.60% additional distance; by contrast, a control persona with low narrative weights reduced NS to 0.389, confirming controllability. These findings validate our narrative-based framework as a practical approach for creating high-quality, personalized cycling routes, successfully connects quantitative adaptation with qualitative experience design.

Keywords: Route suggestion, Tourism informatics, Generative AI, Cycle tourism, Bicycle

1 Introduction

In recent years, cycle tourism initiatives have been implemented worldwide as a strategy for regional revitalization. In Europe, the EuroVelo project consists of 19 routes spanning 42 countries with a total length of approximately 90,000 [km] [1]. Taiwan's 990 [km] round-island route, known as the Huandao route[2], has become popular for long-distance cycling tours. Following these international examples in Japan, infrastructure developments such as the 60 [km] Shimanami Kaido and the 200 [km] Lake Biwa circuit route have drawn attention as tourism models centered around cycling[3].

However, these long-distance courses and sport-oriented cycling routes often fail to accommodate general tourists or individuals with limited physical capabilities who rely on rental bicycles[4]. Given these challenges, a moderate approach to cycling that balances appropriate physical activity with diverse tourist experiences has been inadequately explored despite its potential significance for tourism development.

To address this limitation, our study[1] proposes an automated cycling route generation system incorporating novel metrics derived from quantitative online map data and activity-based elements. The system analyzes terrain and distance information alongside descriptions and reviews of tourist locations using a generative AI agent to establish a comprehensive reward structure.

Through this methodology, cycle tourism can become more accessible to casual tourists rather than being limited primarily to cycling enthusiasts. By balancing sightseeing elements with the experiential consumption aspects inherent to cycling itself, the system aims to enhance overall traveler satisfaction.

This paper is organized as follows. We begin by reviewing related studies and case examples to identify challenges in existing long-distance course design and activity-focused routing approaches. We then present our route-generation algorithm using a generative AI agent, followed by the design and results of our evaluation experiments. Following this, we discuss the results and the potential implications of this research for cycle tourism and the broader tourism industry. Finally, the paper concludes by summarizing our findings and outlining avenues for future work.

2 Related Work

2.1 Diversity in Regional Characteristics and User Needs in Tourism Cycling

Previous research on cycling route design in tourist destinations has consistently emphasized that route attributes, such as safety, traffic volume, and scenic quality, significantly influence the preferences of tourist cyclists. Ritchie demonstrated that tourism cyclists in New Zealand tend to prefer scenic circular routes[5], while Downward and Lumsdon revealed that clear signage and well-maintained rest facilities contribute to route attractiveness in England's Staffordshire Moorlands[6].

Also, user needs in tourism cycling exhibit significant variation based on individual attributes such as gender, age, and health status. Research by Emond et al.[7] and Heesch et al.[8] demonstrates that motivations and constraints in bicycle use for both recreational and commuting purposes differ by gender, highlighting the inadequacy of a singular user model for tourism route design.

These findings collectively indicate that cycling route design for tourism requires a flexible approach that addresses regional geographic and cultural characteristics, the composition of tourism resources, and the diverse preferences and constraints of travelers. Uniform and standardized designs risk failing to capitalize on region-specific attractions and user

---

[1] This work was supported by JSPS KAKENHI Grant Numbers 25K15077, 23K24955, 23K24843, 23K21660, also supported by the Public Interest Incorporated Foundation Yamaha Motor Foundation for Sports; and by the Tateisi Science and Technology Foundation 2025 Research Grant (C).

needs.

2.2 Evaluation Perspectives and Tourism Style Selection

While infrastructure provision and route maintenance have dominated the evaluation and design of tourism cycling from a supply-side perspective, Ritchie emphasized that the sustainable development of cycle tourism requires the systematic incorporation of demand-side perspectives[5].

Demand-side evaluation methodologies have emerged in different forms. Mundet et al. demonstrated the potential to leverage professional cyclists' advanced route evaluation capabilities for developing road cycling tourist destinations[9]. In complementing this approach, Carvalhinho et al. identified that recreational cycling route assessment requires evaluators with tourism expertise in western Portugal[10]. These contrasting findings emphasize how evaluation criteria and necessary expertise vary substantially depending on the intended style of tourism.

2.3 Route Generation Technologies and Experience Design

Recent advancements in route generation technology have introduced adaptation methods based on quantifiable metrics. Černá et al.[4] and Giovannini et al.[11] proposed approaches using attractiveness and convenience indicators such as tourist spot placement, scenic views, and facility availability. These studies employ mixed-integer programming models to enhance tourism experiences within budget and time constraints, providing engineering frameworks applicable to administrative route planning.

Beyond technical adaptation, recent research has recognized the importance of enhancing the qualitative experience through narrative elements. Mossberg et al.[12] and Blichfeldt and Halkier[13] identified storytelling as crucial in forming emotional and cultural connections between tourists and destinations. Bassano et al. suggested that systematically managing place storytelling in digital environments can lead to competitive advantages and brand enhancement for tourist destinations[14].

3 Proposed Method

This study proposes a quantitative framework to enhance cycle tourism by evaluating routes based on their narrative quality, moving beyond purely physical metrics. Our approach is centered on a metric, the Narrative Factor Score (NS) score[15], and a novel framework for personalizing it. The method consists of three core components: (i) AI-driven narrative feature extraction from Points of Interest (POIs), (ii) the formulation of the NS to quantify qualities, and (iii) a weight controlled framework on the user personas. The following subsections provide a detailed description of this framework, including the functionality of each component.

3.1 System Overview

Figure 1 illustrates the workflow of our proposed method. The system begins by collecting POI data from cycling route. A generative AI then enriches this data by analyzing reviews and descriptions to extract narrative features. From this enriched data, our method formulates the NS. This NS is used for our weight controlled framework on the user personas. The resulting score, NS, serves as the primary criterion for evaluating and ranking potential cycling routes.

3.2 POI Data Collection and Analysis

The system obtains tourist spot information in the target area using map services. The system preserves these data at this initial stage without assigning any original attractiveness scores.

The system uses a generative AI to analyze the names and descriptions and then reviews the obtained results. This analysis extracts information such as historical or cultural background, main appeal points, and relevant episodes associated with each location. The system organizes the extracted information into a structured data table for each spot as POIs. Through this process, the system generates a set of POIs enriched with comprehensive post-analysis metadata.

3.3 Calculation of NS

The NS introduced in this study is an evolution of the concept we previously proposed to quantify the narrative quality of a tourism experience. While the original work defined NS as a simple sum of its components, this study extends it to a weighted average to accommodate diverse user values.

The dynamic learning mechanism for these parameters is a key contribution of this study. The parameters α, β, and γ are adjustable according to user interests and objectives. For example, if a user emphasizes historical and cultural experiences, β can be assigned a higher value. Alternatively, if a user prioritizes social media appeal, γ can be assigned a larger value.

$$\mathrm{NS} = \alpha \cdot \mathrm{TC} + \beta \cdot \mathrm{SD} + \gamma \cdot \mathrm{EC} \tag{1}$$

where:

- **Thematic Coherence (TC):** The system analyzes the textual similarity of metadata related to keywords and background information to evaluate the consistency of the overall theme. Routes visiting heritage sites from a single cultural domain or spots related to a particular historical figure yield higher scores.
- **Story Depth (SD):** Through generative AI-based text analysis, each POI's description is evaluated for the extent to which it includes historical and episodic content. This evaluation is initially rated on a scale and subsequently normalized. This measure quantifies the depth of the narrative offered to the user.
- **Emotional Connection (EC):** The system employs generative AI to analyze social media posts and review

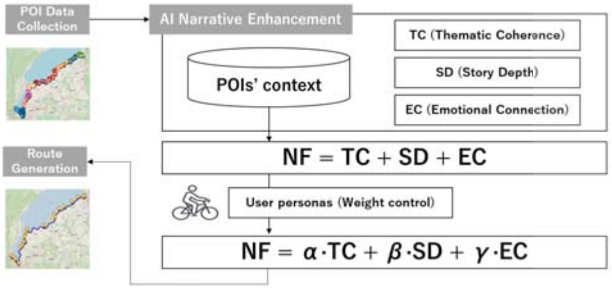


Figure 1: System Overview based on AI-Predicted NSs

texts for positive sentiment and enthusiasm, assigning a score to each POI. The system then normalizes this score by calculating an average or summation across the entire route.

3.4 User-Adaptive Personalization

This study extends the basic route generation algorithm by proposing a weight controlled framework on the user personas. This framework adapts route recommendations based on individual preferences.

We personalize NS primarily by learning the weights (α, β, γ). For persona shaping, we tune ζ and μ as correction hyperparameters in the route search (including prompt focus and thematic repetition control). The revisitation penalty δ, is conceptually available but disabled in this paper's evaluation ($\delta = 0$) to model first-time usage. In this paper, ζ and μ are used to bias POI texts and to discourage trivial thematic repetition within the search, but they are not part of the NS formula itself.

To achieve robust and fine-grained personalization, we introduce a dual-model approach that switches learning strategies based on the user's interaction frequency.

a) Individual-level Personalization for High-Frequency Users: This model targets high-frequency users, such as Hobbyist Cyclists, who may use the system weekly. For these users, a sufficient amount of feedback data is accumulated to learn their specific preferences. The system maintains a dedicated set of parameters $(\alpha_u, \beta_u, \gamma_u)$ for each user u.

To enable this personalization, the system learns from the user's explicit feedback. The fundamental principle is to minimize the error e_i between a user's given rating r_i for a chosen route i, and the NS predicted by the system. We formulate this as an adaptation problem and employ a gradient-descent-based approach to iteratively adjust the parameters.

First, we define the error e_i as:

$$e_i = r_i - \mathrm{NS}(i; \alpha, \beta, \gamma) \tag{2}$$

This difference is minimized by adjusting the NS weights (α, β, γ). The search-time hyperparameters (ζ, μ) may be tuned separately, while δ is fixed to 0 in this paper's evaluation. We employ a gradient-descent-based learning approach.

We define a cost function L_i based on the squared error:

$$L_i(\alpha, \beta, \gamma) = \frac{1}{2}\big(r_i - \mathrm{NS}(i; \alpha, \beta, \gamma)\big)^2 \tag{3}$$

We then update the weights with their learning rates:

$$\alpha \leftarrow \alpha - \eta_\alpha \frac{\partial L_i}{\partial \alpha} \tag{4}$$

$$\beta \leftarrow \beta - \eta_\beta \frac{\partial L_i}{\partial \beta} \tag{5}$$

$$\gamma \leftarrow \gamma - \eta_\gamma \frac{\partial L_i}{\partial \gamma} \tag{6}$$

Search-time hyperparameters (ζ, μ) are tuned outside the NS formula; $\delta = 0$ throughout our experiments. Through this iterative learning, the system progressively acquires more accurate estimates of the user's individual preferences.

b) Persona-level Personalization for Low-Frequency Users: For low-frequency users, such as Tourist Users, who may use the system once every few months, amassing individual data is challenging. To address this, our framework pre-defines several virtual user personas, such as Cultural Heritage Seeker or Nature Explorer. Each persona p has its own set of parameters $(\delta_p, \zeta_p, \mu_p)$. A new or infrequent user is assigned to the most fitting persona based on an initial questionnaire or their first-choice POIs. The user's feedback is then used to update the parameters of their assigned persona, not their profile. This approach enables robust personalization by aggregating sparse data from multiple users within the same persona.

4 Implementation

The system proposed in this study was implemented using Python 3.10.11 and several external APIs. Figure 2 illustrates the overall implementation architecture and data flow of our system. The implementation followed a sequential pipeline approach where each subsystem processed and enriched the data before passing it to the next stage.

4.1 POI Collection and Analysis

The foundation for our analysis was baseline route data from GPX files corresponding to real-world cycling courses. The system first parsed the GPX file to extract the route's geometry and calculated its primary physical characteristics, such as total distance and elevation.

Next, to systematically gather new POIs along this baseline route, we established sampling points at 1 [km] intervals. At each point, the Google Maps Places API was queried to retrieve a comprehensive set of nearby POIs within a 1 km radius. The collected POIs were then merged with the baseline set, and duplicates were removed based on their unique identifiers to form a comprehensive candidate pool.

Finally, this candidate pool was subjected to a quality-based filtering process. We selected high-quality, relevant POIs for the subsequent narrative enhancement phase by choosing those with user ratings of 3.0 or higher out of 5.0 and 20 reviews or more.

4.2 Narrative Enhancement

For the narrative enhancement of the filtered POIs, we integrated the DeepSeek V3. We calibrated the model's temperature parameter to 0.7 to strike a balance between factual

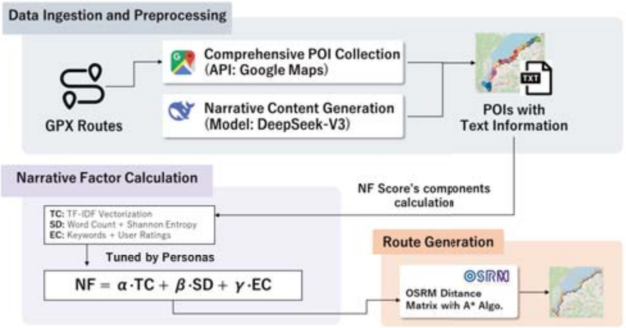


Figure 2: Overall implementation architecture showing the main subsystems and data flow

accuracy and creative expression and set a maximum token limit of 1000 to ensure comprehensive yet concise outputs.

The main user prompt instructed the AI to analyze a POI's information provided by Google Maps and return a JSON object containing five key narrative elements:

- historical_background: The historical context or origin of the location.
- cultural_significance: Its cultural, atmospheric, or landscape-related importance.
- appealing_episodes: Noteworthy features or episodes associated with the spot.
- emotional_connection: A list of 3-5 keywords describing potential visitor emotions.
- categories: An expanded list of up to three detailed categories.

To facilitate personalization, this base prompt was dynamically augmented with persona_specific_additions.

4.3 NS Calculation

The NS, our core metric for quantifying the narrative quality of a route, was implemented using the formula 1. The implementation of each component is as follows:

- **TC**: This measured the thematic consistency among POIs on a route. We implemented this by vectorizing the narrative texts of each POI (including categories, historical background, and cultural significance) using TF-IDF (Term Frequency-Inverse Document Frequency). The final score represents the average pairwise cosine similarity between all POIs on the route.
- **SD**: This quantified the richness of each POI's narrative. It was implemented by calculating the total word count of the AI-generated historical_background, cultural_significance, and appealing_episodes fields, which was then normalized to a 0-1 scale.
- **EC**: This evaluated the route's potential for emotional engagement. The score was a composite of the average of AI-generated emotional keywords and the POI's average user rating from Google Maps, both of which have been normalized.

The final NS was calculated as a weighted sum of these components. The initial parameters were set as follows: $\alpha = 0.3$, $\beta = 0.4$, and $\gamma = 0.3$.

4.4 Route Generation

The route generation process was designed to create an adaptive sequence of POIs that enhanced the narrative quality, as measured by the NS while adhering to practical cycling constraints. We integrated the OSRM, which leverages the OpenStreetMap road network to provide accurate, real-world cycling distances. For the core adaptation task, we implemented the A* search algorithm. The algorithm explores potential routes by prioritizing paths that balance the accumulated NS with the distance traveled.

5 Performance Evaluation

To validate the effectiveness of our proposed framework, we conducted an evaluation of the experiential quality of the recommended routes for our persona-based personalization. This study aimed to verify that our proposed NS generates routes with a higher experiential value compared to traditional baseline routes.

5.1 Procedure

For this evaluation, we selected governmental official cycling routes in Japan: the southern half of the Lake Biwa circuit ("Biwaichi Minami") in Shiga. This route was designated by local governments and was promoted through official guidebooks and websites.

First, we obtained the official GPX data for the course. From the associated guidebooks and websites, we identified an initial set of ten POIs to establish a baseline for each route. In parallel, we defined six user personas corresponding to common cycling tourism preferences, as detailed in Table 1. In all experiments, we fixed the revisitation penalty at $\delta = 0$ and tuned only (α, β, γ) and (ζ, μ) according to each persona; thus, revisitation effects were deactivated during evaluation. These personas are implemented through two primary mechanisms. The first is the adjustment of the NS weighting parameters. The second mechanism is prompt focus, which refers to an extra instruction appended to the system prompt that biases the generative AI's output toward a persona's theme. This mechanism allows the system to tune the narrative vectors of POIs without changing the core weighting parameters.

The chosen personas were not derived from a large-scale quantitative survey but were established as representative examples based on qualitative insights gathered from preliminary discussions with experienced cyclists and findings from our prior user interviews. This mechanism tunes the narrative vectors without changing the underlying weighting parameters.

Next, our system generated routes tailored to each persona. During this route generation process, the system calculated and recorded the resulting NS, the parameters used for adaptation, and the increase in the total distance required to visit the recommended POIs.

Table 1: Persona Definitions with Actual Configurations
δ is disabled ($= 0$) for first-time usage and thus omitted.

Persona	Actual Configuration				
	α	β	γ	ζ	μ
Story-focused	0.300	0.500	0.200	0.150	0.300
History-focused	0.400	0.450	0.150	0.250	0.250
Food-focused	0.250	0.350	0.400	0.100	0.400
Nature-focused	0.400	0.300	0.300	0.200	0.300
Variety-focused	0.200	0.300	0.500	-0.200	0.800
Performance-focused	0.150	0.200	0.150	0.000	0.600

5.2 Result

The quantitative characteristics of the routes generated by our system for each persona, alongside the baseline routes, are detailed in Table 2. A primary observation from this table was that the NSs for all persona-specific routes without the "Performance-focused" persona were consistently higher than the baseline in the Biwaichi Minami dataset. Conversely, the "Performance-focused" route, which was intentionally designed with low NS weights to prioritize physical aspects over narrative quality, showed a significantly lower NS. This result confirmed that our system could easily control the narrative emphasis of a route by the specified persona parameters, either enhancing or diminishing it as required.

6 Discussion

6.1 Quantitative Discussion of NS

This quantitative examination confirms interplay between the NS component weights (α, β, γ) and the corrective parameters (ζ, μ) allow the system to generate routes with measurably different characteristics, validating the effectiveness and flexibility of the proposed personalization model. A detailed quantitative analysis of the generated routes, as presented in Table 2, validates the efficacy of our persona-based parameterization in steering the route adaptation process.

For instance, the Story-focused persona is configured with a high weight for SD ($\beta = 0.50$). This configuration directs the A* search algorithm with POI to prioritize routes with high SD values. This is empirically confirmed by the results for the Biwaichi Minami dataset, where the Story-focused route achieved an SD score of 0.967, which is significantly higher than the baseline's 0.831. This demonstrates that the high β weight effectively translated the abstract preference for "story" into a quantifiable route characteristic. Conversely, the Variety-focused persona is designed to find diverse experiences, which is implemented through a high weight for EC ($\gamma = 0.50$) and, more critically, a negative correction for thematic repetition ($\zeta = -0.20$). The effect of this negative ζ is evident in the TC scores.

Another illustrative case is the Performance-focused persona, which is characterized by uniformly low NS weights. This parameterization deliberately diminishes the importance of narrative elements in route adaptation. Consequently, the resulting NSs for this persona are the lowest across all personas. By reducing the narrative weights in the Performance focused persona, the A* search greedily favors the shortest detours. Consequently, generated routes adhere closely to the baseline geometry, which better satisfies athletes who mainly value physical metrics.

6.2 Limitation

The proposed framework, while demonstrating promising results, is subject to several limitations that warrant discussion.

A primary limitation lies in the foundation of the system: the POI data and its subsequent narrative evaluation. The system's NS is entirely dependent on POI data collected from the Google Maps API. This reliance introduces a potential data skew, as the API and the filtering process, which prioritizes locations with a high number of reviews, may favor well-established commercial venues and major tourist sites over lesser-known locations that could be highly relevant to specific narrative themes. Closely related to this is the quality and objectivity of the AI-driven narrative evaluation. The narrative enhancement process relies on a generative AI to interpret and structure information. The resulting output is inherently constrained by the AI's training data and algorithmic biases. Furthermore, core metrics such as SD, which word count quantifies, and TC, measured by TF-IDF-based cosine similarity, are quantitative proxies that may not holistically capture the qualitative richness and contextual depth of a truly compelling narrative experience.

Further limitations exist within the personalization framework itself. For low-frequency users, the system employs a set of predefined personas to tailor recommendations. However, the current set of personas was not designed to be mutually exclusive or collectively exhaustive (MECE), meaning it may not adequately serve users with hybrid interests or niche preferences that fall outside the defined categories. This issue directly contributes to the classic cold-start problem for new users. If this initial data is sparse or ambiguous, an inaccurate persona assignment may occur, leading to a cascade of poorly aligned route recommendations that could fundamentally degrade the user's satisfaction. Additionally, the current NS metric is calculated from static POI attributes and does not account for their time-series component. This is a notable constraint because the experiential quality of a tourist destination is often highly time-dependent.

7 Conclusion

This study proposed and developed an automated system for generating personalized cycling routes based on a metric, the NS. Our goal was to address a gap in current cycle tourism offerings, which often cater to enthusiasts while neglecting the needs of casual tourists seeking more moderate, experience-rich journeys. By leveraging a generative AI to enrich points of interest with narrative content and by personalizing route generation through user personas, our system successfully creates routes that balance physical activity with qualitative tourism experiences. The outcomes can serve not only as an operational system for tourists but also as a founda-

Table 2: Quantitative Characteristics of Generated Routes (Dist. = Distance, Incr. = Increased)

Route Type	Persona	NS	TC	SD	EC	Dist. [km]	Dist. Incr. [%]
Biwaichi Minami	Baseline	0.687	0.610	0.831	0.621	46.2	0.00%
	Story-focused	0.798	0.600	0.967	0.672	47.1	+1.95%
	History-focused	0.784	0.633	0.967	0.635	46.8	+1.30%
	Food-focused	0.773	0.667	0.963	0.674	47.1	+1.95%
	Nature-focused	0.710	0.567	0.951	0.658	46.8	+1.30%
	Variety-focused	0.737	0.533	0.953	0.689	47.4	+2.60%
	Performance-focused	0.389	0.667	0.959	0.646	46.9	+1.52%

tional methodology for local governments and tourism operators aiming to design more inclusive cycle tourism policies.

REFERENCES

[1] A. Bodor, E. Lancaster, B. McEldowney, and J. Freire, EuroVelo, the European Cycle Route Network: Press Pack, European Cyclists' Federation, Brussels, Belgium, (2016).

[2] Transportation Tourism Bureau, Ministry of Transportation and Communications, Cycling Route No. 1: A Bicycle Touring Guide Around Taiwan, 1st ed., Ministry of Transportation and Communications, Taipei, Taiwan, (2017).

[3] Ministry of Land, Infrastructure, Transport and Tourism, National Cycle Route, [Online]. Available: `https://www.mlit.go.jp/road/bicycleuse/good-cycle-japan/national_cycle_route/`, [Accessed: Mar. 20, 2025].

[4] A. Černá, J. Černý, F. Malucelli, M. Nonato, L. Polena, and A. Giovannini, Designing Optimal Routes for Cycle-Tourists, Transportation Research Procedia, Vol. 3, pp. 856–865 (2014).

[5] B. W. Ritchie, Bicycle Tourism in the South Island of New Zealand: Planning and Management Issues, Tour. Manag., Vol. 19, No. 6, pp. 567–582 (1998).

[6] P. Downward and L. Lumsdon, The Development of Recreational Cycle Routes: An Evaluation of User Needs, Managing Leisure, Vol. 6, pp. 50–60 (2001).

[7] C. R. Emond, W. Tang, and S. L. Handy, Explaining Gender Difference in Bicycling Behavior, Transp. Res. Rec., Vol. 2125, No. 1, pp. 16–25 (2009).

[8] K. C. Heesch, S. Sahlqvist, and J. Garrard, Gender Differences in Recreational and Transport Cycling: A Cross-Sectional Mixed-Methods Comparison of Cycling Patterns, Motivators, and Constraints, Int. J. Behav. Nutr. Phys. Act., Vol. 9, No. 1, p. 106 (2012).

[9] L. Mundet, J. Marin, and A. Figueroa, How to Develop a Road Cycling Tourism Destination: Girona as a Case Study, J. Outdoor Recreat. Tour., Vol. 39, No. 100566, p. 100566 (2022).

[10] L. Carvalhinho, E. Pereira, R. Durao, and P. Rosa, Assessment of Recreational and Tourist Cycling Routes: Case Study in the West Region of Portugal, J. Outdoor Recreat. Tour., Vol. 45, p. 100729 (2024).

[11] A. Giovannini, F. Malucelli, and M. Nonato, Cycle-Tourist Network Design, Transp. Res. Procedia, Vol. 22, pp. 154–163 (2017).

[12] L. Mossberg, A. Therkelsen, E. H. Huijbens, P. Björk, and A.-K. Olsson, Storytelling and Destination Development, Nordic Innovation Centre, Oslo, Norway, pp. 1–65 (2010).

[13] B. S. Blichfeldt and H. Halkier, Mussels, Tourism and Community Development: A Case Study of Place Branding Through Food Festivals in Rural North Jutland, Denmark, Eur. Plan. Stud., Vol. 22, No. 8, pp. 1587–1603 (2014).

[14] C. Bassano, S. Barile, P. Piciocchi, J. C. Spohrer, F. Iandolo, and R. Fisk, Storytelling About Places: Tourism Marketing in the Digital Age, Cities, Vol. 87, pp. 10–20 (2019).

[15] H. Tomisu, S. Morita, N. Kai, and T. Yoshihisa, Narrative-Aware Cycling Route Design Using Generative AI, Proc. IEEE COMPSAC, 9357 (2025).

Designing Traces as Social Cues to Support Visitor Exploration

Ayaka Negishi*, Hiroki Echigo** and Minoru Kobayashi***

*Graduate School of Advanced Mathematical Sciences, Meiji University, Japan
**Institute for Advanced Study of Mathematical Sciences, Meiji University, Japan
***Faculty of Interdisciplinary Mathematical Sciences, Meiji University, Japan
ayaka.negishi@koblab.org, hirokiechigo.info@gmail.com, minoru@acm.org

Abstract -In unfamiliar spaces such as museums or exhibitions, visitors often struggle to decide where to go and what to view. While existing navigational aids like maps or signs guide visitors based on the designer's intent, they may limit spontaneous discovery and personal agency. In contrast, social cues—such as footprints—represent traces of others' activity and offer indirect hints. These cues invite interpretation: "Why did someone stop here?" or "What's ahead?" This process can trigger self-directed exploration, like choosing to visit an unfamiliar area. Thus, traces do not prescribe a path but instead invite interpretation by visitors, potentially encouraging more self-directed exploration. This study investigates the potential of visually presenting such traces to support decision-making and broaden exploration in exhibition spaces. Before conducting the main study, we focus on the design of trace-based visualizations and examine which forms best convey the presence and actions of others in an intuitive and engaging way.

Keywords: social cues, trace, social awareness, footprints, interface

1 INTRODUCTION

Human behavioral traces can function as social cues that make visible the presence or activity of others in a space [1][2][3][4]. For example, footprints may offer hints about where people have previously gone and how a space has been used [5][6]. These cues can trigger interest or actions, such as thinking, “Maybe this spot is valuable,” or “Others paid attention here, so I should take a look too.”

Since traces show how a space has been used at a glance, they can also help people decide where to look or what might be important. This is especially helpful in unfamiliar environments or when there are too many options to choose from. While traditional tools like maps and signs provide navigation, they are typically based on the designer’s intent, which may limit spontaneous exploration or personal discovery [7][8][9].

In contrast, traces do not explicitly direct behavior. Instead, they invite interpretation, such as wondering “Why are there footprints here?” This ambiguity can encourage people to explore based on their own curiosity or interest. Traces also offer another advantage: they are implicit and instantly understandable through visual perception, without requiring complex interaction or explanation [10][11].

This study focuses on these characteristics of traces and aims to investigate how they influence free exploration behavior. Specifically, we examine whether showing the footprints of others in a space with multiple exhibits can function as social cues and support self-directed exploration. We created a virtual museum that visualizes such traces and compared participants' behavior in two conditions: one with footprints (intervention) and one without (control).

The results showed that footprints helped them decide where to focus and where to stand to see better.

2 RELATED WORK

2.1 Studies on Visualizing Traces of Human Behavior

Several studies have explored the presentation of human behavioral traces in physical or digital spaces. Monastero et al. [10] conducted a six-week field study in a university lobby, where real-time walking traces were visualized. Their findings suggest that such traces can enhance social awareness in the space through passive visual perception alone, without requiring active interaction. Similarly, Hirsch et al. reported that the visualization of traces in virtual environments can enhance social presence, even among asynchronous participants [12][13].

These studies demonstrate that behavioral traces function as implicit social cues, making others’ presence or activity in a space visible and enhancing social awareness and presence. Building on these insights, the present study investigates how such traces are interpreted in relation to their surrounding environment and how they influence exploratory behavior.

Albarrak et al. [11] examined the impact of behavioral traces on path selection by presenting footprints on only one side of a branching corridor. Their results revealed that participants did not always choose the side with traces, indicating that such cues do not necessarily dictate user decisions. While their study was conducted in a narrowly constrained environment (a binary corridor), the present study analyzes the influence of traces in a more open and unconstrained exploratory setting.

Shirai et al. [14][15] also explored non-verbal traces such as poster residue and thermal marks, demonstrating that these subtle cues can serve as effective aids for information-seeking tasks. In digital contexts, other studies have visualized users’ behavioral history—such as browsing paths on websites—to support social navigation and information retrieval [16][17][18]. These studies have shown that traces can be valuable in task-oriented scenarios where users have clear goals.

In contrast, the present study focuses on environments without predefined tasks or objectives. We aim to examine how behavioral traces influence user behavior in free exploration, particularly in relation to the initial formation of interest or attention.

Drawing on these prior studies, our research explores how traces function in large-scale environments with multiple options for exploration.

2.2 Design of Traces

Albarrak et al. [11] compared six types of visual representations for behavioral traces—such as heatmaps, dots, lines, and footsteps—and found that footsteps were the most intuitive for conveying human behavior. Based on this, the present study adopts the footprint format as a simple and recognizable method of presenting traces.

Wong et al. [16], in their study on web navigation history, proposed highlighting only salient events instead of the entire behavioral history in order to reduce visual complexity. Inspired by this approach, our study also limits visualized traces to stay locations and durations, allowing for a concise and focused display. This helps reduce visual load while still providing meaningful information for users to interpret.

Shirai et al. [4] proposed a three-level framework for progressive disclosure of interaction history, consisting of:

- Level 1: Indicating the presence of information
- Level 2: Specifying the content of the information
- Level 3: Presenting detailed information

In our study, footprints function as a Level 1 cue—an abstract indicator that suggests where others have stayed. These traces are shown before users approach the content itself, serving as a subtle prompt to help them quickly grasp the overall use of the space and decide where to focus their attention. This design aims to encourage users to develop their own interest and engage in self-directed exploration.

2.3 Influence of Others' Behavior in Exhibition Spaces

The behavior of others in exhibition spaces has a significant impact on individual viewing experiences. Previous studies by Goffman, Heath, and others [19][20][21] have shown that the attention and actions of people nearby can shape how exhibits are encountered and interpreted.

In this study, we focus on traces left by previous visitors to the space. By presenting these traces, we aim to enable emergent collaboration [22] between asynchronous visitors—where users gain hints or new discoveries by indirectly interacting with the behavior of others.

Based on these insights, our study examines how traces of past behavior can influence exploratory actions in a free-viewing environment.

3 STUDY

3.1 Research Purpose and Hypotheses

The goal of this study is to understand how traces of others' behavior can be used in free exploration settings.

Specifically, we examine the following hypotheses:

- **Hypothesis 1**: Traces of others serve as social cues that help visitors decide which exhibits to focus on before directly engaging with the content.

3.2 Experimental Design

In this study, we built a VR museum that replicates the structure and experience of a real museum. This virtual environment allowed participants to freely explore a wide area, enabling us to compare their behaviors under different conditions regarding the presence or absence of traces.

Experimental Conditions: We used a between-subjects design, with a single independent variable: whether behavioral traces were presented or not. To generate the trace data, we first asked one participant (P1) to explore the museum freely, and we recorded P1's behavior. All other participants were randomly and evenly assigned to one of the following two conditions:

- Control condition (no traces): Participants explored the VR museum freely without any visual traces of others on the floor. (Figure 3 - (a))
- Intervention condition (with traces): Participants explored the same museum, but with footprints displayed on the floor, based on the behavioral data of P1 recorded in the control condition. (Figure 3 - (b))

Trace Design (Figure 4): Based on the findings of Albarrak et al. [11], we used footprints to represent the behavioral traces of P1. To avoid clutter and ensure readability, only locations where P1 stayed for more than 1 second were visualized. P1's behavior was recorded as follows:

- Walking only: green line
- Stay less than 1 second: gray dots
- Stay between 1–10 seconds: yellow dots
- Stay between 10–20 seconds: orange dots
- Stay 20 seconds or more: red dots

Among these, only stay points of **1 second or longer** (yellow, orange, and red) were visualized as **footprints**. In addition, the longer the dwell time, the **darker the footprint color**, allowing viewers to intuitively understand where P1 spent more time. The orientation of each footprint was determined based on the camera angle where P1 spent the longest time facing while standing still (from recorded video). If no stable view direction was available during a pause, we visualized a wider foot angle to indicate the general facing range. Any unintentional long pauses, such as those caused by operational difficulties, were excluded from trace visualization.

Participants: A total of 11 participants (5 female, 6 male), aged between 20 and 24 years (M = 21.9), were recruited from our laboratory. The sample included both undergraduate and graduate students. All participants were unfamiliar with the virtual museum used in the study. According to responses to the pre-questionnaire, the average rating of interest in Buddhist art was 1.6 in the control group and 1.8 in the intervention group (on a 5-point Likert scale). This suggests that participants in both conditions did not have a strong prior interest in Buddhist art.

3.3 Experimental Environment

In this study, participants experienced the VR museum on a desktop PC monitor (dimensions: 332.1 mm × 567.7 mm; resolution: 2560 × 1440) (figure 1). This experimental setup was chosen because it is difficult to control environmental factors such as exhibit layout and lighting in a real museum, and detailed behavior tracking is also challenging. Additionally, using a VR headset (HMD) may hinder participants' ability to read the exhibition texts clearly—especially since the exhibits in this study included detailed explanations—which could interfere with in-depth viewing behavior.

VR Museum (Contents): The virtual museum was developed using Blender and Unity, and its theme was Buddhist art. This theme was selected for several reasons: First, most participants were assumed to have limited prior knowledge of Buddhist art. Second, Statues such as Buddhas and Bodhisattvas contain symbolic meanings in their posture and decorations, making them suitable for 360-degree viewing and open to diverse interpretations. Third, exhibits such as scrolls, sutras, and Buddhist paintings require not only visual inspection but also deep engagement with contextual and cultural information through reading, allowing us to examine both superficial and in-depth viewing behaviors. The museum included the following four exhibit areas (figure 2):

- **Area 1: Scrolls, Buddhist paintings, and scriptures:** These were displayed along the outer walls with explanatory texts. A sutra exhibit was placed at the center of the area. Understanding these exhibits requires reading the detailed texts and observing fine visual elements, as they are not easily recognized from a distance.
- **Area 2: Buddha statues (Nyorai):** Large and visually striking statues that are easy to recognize from afar.
- **Area 3: Bodhisattva statues** Similar to Area 2, these are large, decorative statues that are visually noticeable from a distance.
- **Area 4: Large mural paintings** This area is located deep within the museum space, requiring participants to move further in to access it, encouraging extended exploration.

An online version of the simplified experimental environment is available at the following URL: https://vrmuseumothertrace.netlify.app

Due to file size constraints, only Area 1 and Area 4 are included in the demo, and the visual quality is reduced compared to that used in the actual experiment.

Operation: Participants navigated using the arrow keys or WASD keys for movement and changed their viewing direction using the mouse.

3.4 Procedure

1. **Operation Check**
   Participants first launched the VR application, received instructions on the controls, and practiced the basic operations.
2. **Instructions**
   All participants (in both control and intervention conditions) were told: *"Please explore the VR museum freely. There is no time limit. You may end the session whenever you like. When you are finished, please let us know."* Additionally, for participants in the intervention condition, the following explanation was provided before exploration: *"The footprints shown on the floor represent the walking paths of a previous participant. Only the locations where the person stopped are displayed. The darkness of each footprint indicates the length of time they stayed at that spot."*
3. **Exploration Phase**
   Participants in the control condition explored the VR museum without footprints, while those in the intervention condition explored it with footprints displayed. Participants were free to move and view the exhibits at their own pace. No time limit was imposed.
4. **Informed Consent**
   Before the experiment, participants were informed that questionnaire responses would be collected and analyzed anonymously. They were also told that the results might be used for academic presentations or publications. Consent was obtained accordingly.
5. **Post-Exploration Questionnaire**
   Participants in the intervention condition completed a questionnaire (Table 1), focusing on their awareness and use of the presented footprints.

3.5 Data Collection

This study compared two experimental groups: a control condition with no visual traces, and an intervention condition in which P1's footprints were visualized. To investigate how traces were interpreted and utilized during free exploration, qualitative evaluations were conducted using a post-experiment questionnaire. A questionnaire (Table 1) was administered only to participants in the intervention condition. It focused on whether participants noticed the traces, whether the traces influenced their behavior, whether the traces affected their impressions of exhibits, and whether they used the traces as cues for deciding where to pay attention.

Figure 1: Scenes from the Experiment.

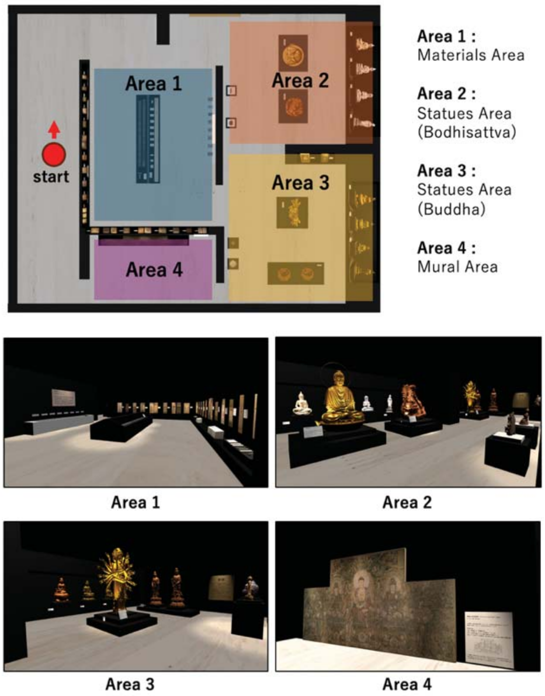
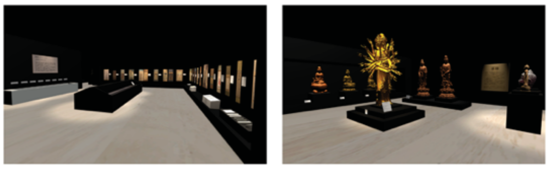
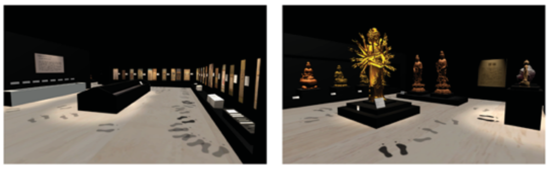


Figure 2: Area map.

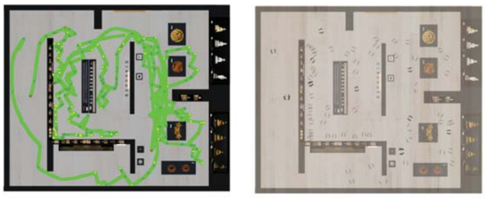


(a) Control condition (no traces)



(b) Intervention condition (with traces)

Figure 3: Experimental environments under each condition (Top: No Traces, Bottom: With Traces).



(a) Movement path data of P1 (dots represent stop locations: ~1s = gray, 1–10s = yellow, 10–20s = orange, 20s and above = red)

(b) Trace design based on (a)

Figure 4: Trace design.

Table 1: Questionnaire for Intervention condition.

	Question	Answer Format
Q1	Did you ever feel unsure about which exhibit to view next during your exploration?	5-point scale (1: Never unsure – 5: Always unsure)
Q2	Please explain the reason for the answer to Q2?	Open-ended
Q3	Were you aware of the traces on the floor during your exploration?	5-point scale (1: Not at all aware – 5: Constantly aware)
Q4	Did the traces influence or change your behavior during the exploration?	5-point scale (1: Not at all – 5: Very frequently)
Q5	If the traces affected your behavior or how you viewed the exhibits, please describe in as much detail as possible.	Open-ended
Q6	Do you feel that the traces affected your impression of the exhibits?	5-point scale (1: Not at all – 5: Very strongly)
Q7	(For those who answered that there was an influence in Q) Please describe specifically what kind of influence you experienced.	Open-ended
Q8	Did the traces help you decide where to focus your attention?	5-point scale (1: Not at all – 5: Very much so)
Q9	(For those who answered positively in Q6) Please explain how the traces helped you decide where to focus.	Open-ended
Q10	Please describe any ways in which the traces negatively affected your exploration.	Open-ended
Q11	If there were any types of information you wish the traces had included, please describe them.	Open-ended

4 RESULT

Table 3 summarizes the questionnaire scores related to the influence of footprints in the intervention group.

According to the results of Q1 and participants' free responses, few participants interpreted the footprints as a clear "path" to follow through the exhibits. Therefore, the footprints did not function as navigational guidance that suggests a fixed viewing order (as noted in P7 and P8's comments). However, many participants perceived the footprints' location and density as indicators that "something is there" or "others have paid attention here," which led them to focus on those spots. This suggests that the footprints were not used as a navigational guide but rather as a *cue for deciding where to pay attention*.

4.1 Intrinsic Motivation and Curiosity Triggered by Traces

This study observed that the presence and density of traces not only served as visual information but also prompted participants to ask themselves, "Why are there traces here?" or "Why are there no traces there?", which in turn sparked interest and intrinsic motivation toward the exhibits.

Participant P10 stated, "The presence and density of footprints increased my expectations of the exhibit, and I began to think that the exhibits with footprints might be more valuable." Similarly, P7 commented, "Even in areas with many exhibits, I wondered why there were few traces." P9 mentioned, "Exhibits with more footprints felt like the main highlights of the exhibition" and "I thought exhibits with dark traces might be important, so I read the captions carefully." These responses suggest that traces can stimulate curiosity and foster deeper engagement by encouraging visitors to interpret their meanings. In this way, traces served as a trigger for intrinsic motivation and curiosity, shaping participants' attention and exploration strategies through questions such as "Why is this here?" or "Is this an important exhibit?"

Table 3: Questionnaire Scores by Item (Intervention Group)

	Question	Answer (P7-P11)
Q1	Did you ever feel unsure about which exhibit to view next during your exploration	4,4,1,4,2 (M = 3.0)
Q3	Were you aware of the traces on the floor during your exploration?	4,3,4,5,4 (M = 4.0)
Q4	Did the traces influence or change your behavior during the exploration?	4,4,3,4,4 (M = 3.8)
Q6	Do you feel that the traces affected your impression of the exhibits?	2,2,3,5,2 (M = 2.8)
Q8	Did the traces help you decide where to focus your attention?	5,3,3,3,4 (M = 3.6)

4.2 Utilization of Traces

The results of Q3 and Q4 indicate that most participants noticed the traces and adjusted their behavior accordingly. 4 out of 5 participants rated Q4 with a score of 4, suggesting that traces had a notable influence on behavior.

From Q8 and Q9, it is evident that most participants used traces as cues to decide where to focus. Several free responses showed that participants inferred the importance of an exhibit based on the number and density of traces. For example:

- "I tried to look more carefully at areas where the footprints were darker." (P7, P11)
- "There were a lot of footprints, so I thought that was an important spot to look at. I read the captions carefully and examined the artwork closely." (P8)
- "The presence of traces gave me an idea of which exhibits were drawing attention. I felt that those with more and darker footprints might be important." (P9)

In addition, some participants noted:

- "Since many of the dark traces were in front of captions, I began to assume that traces indicated the presence of explanations." (P9)
- "Standing on the footprints made it easier to read the captions, so I began following the traces to view the exhibits more comfortably." (P8)

These responses suggest that traces influenced not only "what to look at" but also "where to look from" and "where to stand," affecting participants' viewpoint selection.

4.3 Limitations of the Trace Design

Some participants expressed uncertainty about the meaning and source of the traces. Questions such as "Are these traces from one person or overlapping traces from many?" (P9) and "Are they really related to important exhibits?" (P8) were raised. These concerns suggest that while the ambiguity of traces can encourage flexible interpretation, it may also lead to doubts about the reliability of the information and the stability of its meaning. This highlights a limitation of the current trace design and indicates a need for refinement in future implementations.

5 DISCUSSION

5.1 Do Traces Help Visitors Decide Where to Focus?

This study suggests that visual traces—specifically, footprints—may influence visitors' initial decisions about where to focus their attention in exploratory environments.

Several participants reported paying closer attention to exhibits marked by dense footprints, suggesting that the *intensity of traces* may function as a proxy for the level of attention the exhibit deserves. One participant noted that they realized interpretive signage was easier to read from the locations of the footprints, and thus used the footprints to determine not only *what* to view, but also *from where* to view it. These observations suggest that footprints supported not only content selection but also viewpoint selection, pointing to new design possibilities for trace-based information presentation in spatial environments.

5.2 Intrinsic Motivation and Curiosity Triggered by Footprints

In addition to guiding attention, traces appeared to stimulate *intrinsic motivation* and *curiosity*. Responses from both the questionnaire revealed that participants interpreted footprints as indicators of value or popularity—especially when footprints were dense or located in otherwise inconspicuous areas. For example, P10 reported that footprints raised expectations about the exhibit, while P9 inferred that denser footprint marked more significant displays.

This form of engagement was not driven by instruction, but rather by a self-initiated process of meaning-making: participants asked themselves, “Why are there traces here?” and proceeded to explore out of personal interest.

These results indicate that the absence of traces can itself serve as meaningful information. Rather than merely directing behavior, traces prompted reflective questions that activated self-directed exploration—suggesting their value as *cognitive triggers* in UX and spatial design.

5.3 Design Limitations of Trace Presentation

In this study, footprints were presented based on stay location and duration, with color intensity reflecting the length of time spent at each location. This design effectively conveyed depth of attention, and participants used the intensity as a guide for deciding how carefully to view an exhibit.

However, the trace design also introduced interpretive ambiguity. Some participants questioned the origin or nature of the footprints, asking, “Are these from one person or multiple people?” or “Do they really indicate important exhibits?” Such comments suggest that a lack of contextual metadata—such as who left the traces or when—can lead to confusion or reduced trust in the cues provided.

While ambiguity can encourage flexible interpretation, it may also compromise clarity and reliability for certain users. Future designs should consider offering optional supplementary information (e.g., source, number of contributors, time of visit) to those who seek it.

5.4 Limitation of the Experiment

This study has several limitations. First, since the number of participants was limited to ten, the generalizability of the results is restricted. Future studies should include a larger and more diverse group of participants to obtain more robust and broadly applicable findings.

Second, the footprints used in this study were manually created, which limits their accuracy. Although we attempted to reconstruct the orientation of the footprints using video recordings, visualizing dynamic behaviors—such as head turning or changes in walking direction—remains a challenge. While we adjusted the color intensity of the footprints based on dwell time, prolonged stays may have included moments of hesitation or environmental scanning, which do not necessarily indicate focused visual attention. Some participants also expressed doubts about the reliability of the traces. To address these concerns, future systems should consider generating footprints automatically and in real time based on actual user behavior. This would help viewers intuitively recognize that the footprints were left by real visitors, thereby enhancing both the credibility and interpretability of the traces.

Third, the experiment was conducted in a virtual environment, which may differ from real-world contexts in important ways. Factors such as field of view, level of immersion, and movement control in VR may influence how traces are perceived and interpreted. Future work should compare the effectiveness of trace-based cues in both virtual and physical settings to assess their applicability in real-world environments.

Finally, this study did not compare footprint-based cues with conventional navigation aids commonly used in museums, such as maps or signage. Consequently, it remains unclear how footprints function differently or more effectively as behavioral cues. Future research should incorporate such comparisons to better position the role of footprints in spatial navigation and visitor engagement.

5.5 Contributions

This study demonstrates that presenting others’ behavioral traces—specifically footprints—can function not only as a guide for attention ("where to look"), but also as a *social cue* that fosters intrinsic interest and self-directed exploration. Unlike explicit navigational aids, traces invite interpretation by prompting questions such as “Why are traces here?” or “Why are they absent?”, thereby encouraging deeper engagement.

Additionally, traces influenced not only *what* to view, but also *how* to view—participants used them to determine the best *vantage point* or position for observing the exhibits. This highlights the potential of footprints as design elements that support viewpoint orientation and enhance spatial experience quality.

These insights suggest that behavioral traces can serve as subtle, implicit cues that empower users’ agency in exploration. As such, they hold promise for UX and spatial design strategies that prioritize user autonomy, curiosity, and interpretive engagement over prescriptive guidance.

6 CONCLUSION

This study investigated how others’ behavioral traces—specifically footprints—are used in a large-scale free exploration environment with multiple exhibits. By comparing a virtual museum with and without visible traces, we examined how these traces function as social cues and influence visitor behavior and perception.

The results showed that footprints served as useful references for deciding “where to focus” and “where to stand to see better.” This was especially true for document-based exhibits that required close viewing. In addition, participants also visited areas without any footprints, suggesting that the traces did not enforce a fixed path but instead helped expand the range of choices and support self-directed exploration.

Footprints also influenced not only what to see but how to see it—for example, which angle to view from or where to stand. This indicates that traces can function as guidance for viewpoint selection and enhance the quality of the experience.

Overall, we suggest that others' behavioral traces act as implicit social cues that do not require explicit instructions or complex interactions. By leaving room for interpretation, these traces help foster personal interest and intrinsic motivation in exploratory settings.

REFERENCES

[1] Luluah Albarrak, Oussama Metatla, and Anne Roudaut. 2021. (Don't) Mind the Step:Investigating the Ef-fect of Digital Social Cues on Navigation Decisions. Proc. ACM Hum.-Comput. Interact. 5, ISS, Article 492(November 2021), 18 pages. https://doi.org/10.1145/3488537

[2] Ruth C. Dalton, Christoph Holscher, and Daniel R. Montello. 2019. Wayfinding as a SocialActivity. Frontiers in Psychology 10 (2019), 142. https://doi.org/10.3389/fpsyg.2019.00142

[3] Ruth Conroy Dalton, Renato Troffa, John Zacharias, and Christoph Hoelscher. 2011. VisualJohn Zacharias. 2001. Path choice and visual stimuli: signs of human activity and architecture.Journal of environmental psychology 21, 4 (2001), 341–352.

[4] Yoshinari Shirai, Kumiyo Nakakoji, Yasuhiro Yamamoto, Elisa Giaccardi: A framework for presentation and use of everyday interaction histories, 1st Korea-Japan Joint Workshop on Ubiquitous Computing and Networking Systems(ubiCNS2005), pp. 257-261, 2005.

[5] Dieberger, A., Dourish, P., Hook, K., Resnick, P. and Wexelblat,A.:Social navigation: Tech-niques for building more usable systems, inter-actions, Vol.7, No.6, pp.36-45(2000).

[6] Wexelblat, A. and Maes, P.: Footprint:History-Rich Tools for Information Foraging, CHI99, pp.270-277(1999).

[7] Stephen Bitgood. 2002. Environmental psychology in museums, zoos, and other exhibition centers. In In Handbook of environmental. Citeseer.

[8] Daniel R Montello and Corina Sas. 2006. Human factors of wayfinding in navigation. (2006).

[9] Jan M Wiener, Simon J Büchner, and Christoph Hölscher. 2009. Taxonomy of human wayfinding tasks: A knowledge-based approach. Spatial Cognition & Computation 9, 2 (2009), 152–165.

[10] Beatrice Monastero and David K McGookin. 2018. Traces: Studying a public reactive floor-projection of walking trajectories to support social awareness. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. ACM, 487.

[11] Albarrak, L., Metatla, O. and Roudaut, A.: Exploring the Design of History-Enriched Floor Interfaces for Asynchronous Navigation Support, Proceedings of the 2020 ACM Designing Interactive Systems Conference, 2020, p. 1391-1403.

[12] L. Hirsch, C. George and A. Butz, "Traces in Virtual Environments: A Framework and Exploration to Conceptualize the Design of Social Virtual Environments," in IEEE Transactions on Visualization and Computer Graphics, vol. 28, no. 11, pp. 3874-3884, Nov. 2022, doi: 10.1109/TVCG.2022.3203092.

[13] L. Hirsch, A. Haller, A. Butz and C. George, ""What a Mess!": Traces of Use to Increase Asynchronous Social Presence in Shared Virtual Environments," 2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), Christchurch, New Zealand, 2022, pp. 598-599, doi: 10.1109/VRW55335.2022.00150.

[14] Yoshinari Shirai, Tatsuo Owada, Koji Kamei, Kazuhiro Kuwabara: Optical Stain: Amplifying vestiges of a real environment by light projection, HCI International2003, 2003.

[15] Shirai, Yoshinari & Nakakoji, Kumiyo & Yamamoto, Yasuhiro. (2007). Interacting with Interaction Histories in a History-Enriched Environment. 10.4018/978-1-59904-693-8.ch005.

[16] Yuet Ling Wong, Jieqiong Zhao & Niklas Elmqvist (2015) Evaluating Social Navigation Visualization in Online Geographic Maps, International Journal of Human-Computer Interaction, 31:2, 118-127, DOI: 10.1080/10447318.2014.959106

[17] Hill, W. C., Hollan, J. D., Wroblewski, D., & McCandless, T. (1992). Edit wear and read wear. *Proceedings of the ACM Conference on Human Factors in Computing Systems*, 3–9.

[18] Alexander, J., Cockburn, A., Fitchett, S., Gutwin, C., & Greenberg, S. (2009). Revisiting read wear: Analysis, design, and evaluation of a footprints scroll-bar. *Proceedings of the ACM Conference on Human Factors in Computing Systems*, 1665–1674.

[19] Vom Lehn, Dirk. (2006). Embodying experience: A video-based examination of visitors' conduct and interaction in museums. European Journal of Marketing. 40. 1340-1359. 10.1108/03090560610702849.

[20] Vom Lehn, Dirk & Heath, Christian & Hindmarsh, Jon. (2001). Exhibiting Interaction: Conduct and Collaboration in Museums and Galleries. Symbolic Interaction. 24. 189-216. 10.1525/si.2001.24.2.189.

[21] Vom Lehn, Dirk & Heath, Christian & Hindmarsh, Jon. (2005). Examining exhibits: Interaction in museums and galleries. Communication and Cognition. 38. 229-247.

[22] Terveen, L., & Hill, W. (1998). Evaluating emergent collaboration on the web. *Proceedings of ACM Conference on Computer-Supported Cooperative Work*, 355–362.

Analysis of Gaze Direction and Turn-Taking in the Multitasking Videoconference[1]

Taketo Imagawa*, Atsuto Kurokochi*, Koki Yanagii*, Kazuyuki Iso**, Masayuki Ihara***,
and Minoru Kobayashi****

*Graduate School of Advanced Mathematical Sciences, Meiji University, Japan
**Tokyo Information Design Professional University, Japan
***RIKEN Center for Integrative Medical Sciences, Data Science and Design Special Team, Japan
****Faculty of Interdisciplinary Mathematical Sciences, Meiji University, Japan
{taketo.imagawa, atsuto.kurokochi, koki.yanagii}@koblab.org
iso@tid.ac.jp
{ihara, minoru}@acm.org

Abstract - Recent videoconferencing tools require participants to juggle multiple tasks, such as viewing shared materials, responding to chat messages, and taking notes, at the same time. This multitasking environment can reduce awareness of nonverbal signals like gaze and facial expressions, making it harder to identify who is speaking or being addressed. As a result, interruptions and pauses often disrupt the natural flow of conversation. To examine these issues, we analyzed participants' gaze directions and speech immediately before smooth turn-giving using data collected from wearable eye trackers during a decision-making meeting that involved chat interactions and document browsing. We report the gaze analysis results and discuss how multitasking shapes gaze direction and turn coordination in videoconferences.

Keywords: Gaze, Turn-Taking, Videoconference, Multitasking

1 INTRODUCTION

Videoconferencing platforms such as Zoom [1] and Microsoft Teams [2] have made remote collaborative work [3] commonplace. However, during online meetings, participants are typically required to manage several activities at once. For example, reviewing shared materials, replying to chat messages, taking notes, and observing others' behaviors. Such multitasking situations can reduce sensitivity to nonverbal signals, including facial expressions, gestures, and gaze, which are essential for maintaining smooth interpersonal communication.

This study aims to identify the nonverbal cues that support seamless turn-taking in multitasking videoconferences, where participants' attention is divided among multiple visual and communicative channels, such as video feeds, document browsing, and chat interactions. To this end, we focus on analyzing participants' gaze directions immediately before smooth turn-giving to reveal where they tend to look prior to successful turn transitions under multitasking conditions.

2 RELATED WORK

2.1 Effects of Eye-Gaze and Multitasking on Turn-Taking and Video Meetings

Kendon [4] demonstrated a strong association between gaze behavior and turn-taking. For instance, a current speaker may direct their gaze toward the next speaker as a signal of yielding the floor, while a listener may avert their gaze to plan or formulate their response. Similarly, Jokinen et al. [5] found that gaze direction can help infer whether the current speaker intends to continue or which participant is preparing to take the next turn. These studies highlight the crucial role of gaze cues in coordinating conversational turns.

However, in videoconferencing, gaze information is often distorted or partially lost due to camera placement and limited transmission of visual cues, making it difficult for participants to perceive others' gaze behavior accurately. Moreover, multitasking activities such as responding to chat messages, switching between windows, or managing other concurrent tasks can increase cognitive load and stress during meetings [6, 7, 8, 9, 10]. Together, these factors make fluent turn-taking even more challenging in online communication settings.

2.2 Our Previous Study

To address the issues discussed in Section 2.1, we previously conducted both in-person and video-based meetings involving a single task in which participants were required to browse supplementary materials. We then analyzed their gaze directions immediately before turn-taking and turn-giving [11]. The results indicated that participants often fixated on the shared document just prior to these conversational transitions. This finding suggests that, in multitasking videoconferences, smooth turn-taking may be supported not only by gazing at the speaker's face but also by establishing joint attention on a common visual target.

---

[1] This work was supported by JSPS KAKENHI Grant Numbers 22H03635, 23K24891.

However, our earlier study had several limitations, including the absence of a non-document condition, a restricted range of multitasking situations, hesitation of making eye-contact, and potential inaccuracies in gaze tracking. To address these issues, we designed a new experiment and conducted a more detailed analysis of gaze behavior.

3 DEFINITIONS

This study investigates cues that enable smooth turn taking in multitasking videoconferences by analyzing participants' gaze behavior. We define speaker roles and conversational categories used in the analysis.

Previous Speaker. The participant who spoke immediately before the current utterance.

Current Speaker. The participant who is speaking at the present moment.

Next Speaker. The participant who takes the floor next, following the previous or current speaker.

Non-Speaker. A participant who is neither speaking nor involved in the ongoing turn transition.

Turn-Giving. When the current speaker intentionally yields the floor to the next speaker.

Turn-Taking. When the next speaker begins speaking after the previous speaker has finished.

The image of time interval for turn-giving and turn-taking is shown in Figure 1.

Interruption. When an interrupter begins speaking within 200 ms (milliseconds) after the current speaker's onset.

Silence. When a gap longer than 1700 ms occurs between the end of one speaker's turn and the beginning of the next.

Overlap. When another participant begins speaking more than 200 ms after the current speaker's onset while the current speaker is still talking.

Turn-continuation. When the same speaker resumes speaking without transferring the floor.

A turn changes without interruption, overlap, silence, or turn continuation is classified as **smooth turn giving** (or smooth turn taking where applicable; in this paper we analyze smooth turn giving).

All utterance events were annotated on a per-participant timeline; for each event we recorded timing, the identities of current / next speakers, and the duration between the end of one utterance and the onset of the next. For interruptions, we examined non-speakers' timelines to determine who started first and whether the onset difference between the current speaker and the interrupter was within 200 ms.

4 EXPERIMENT

We designed a videoconferencing experiment that incorporated multitasking conditions, specifically, document browsing and chat message responses to identify the factors that contribute to smooth turn changes. An overview of the meeting setup is presented in Figure 2. During the session, we recorded both gaze and utterance data from all participants.

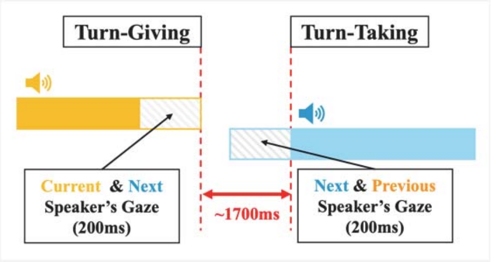

Turn-Giving
Turn-Taking
Current & Next Speaker's Gaze (200ms)
~1700ms
Next & Previous Speaker's Gaze (200ms)

Figure 1: Image of time interval for turn-changing

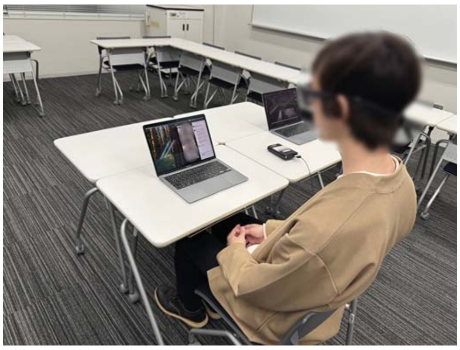


Figure 2 : Image of the video meeting in the experiment

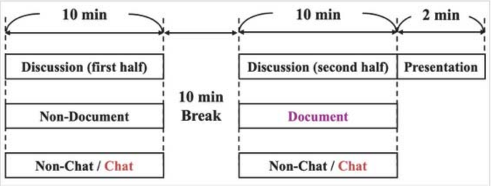


Figure 3: Flow of the discussion in the experiment

4.1 Participants and Apparatus

Participants. Four students (two undergraduates and two graduates; one male and three females) from the same laboratory as the authors participated in the experiment. All participants were acquainted with each other and had prior experience using Zoom for videoconferencing.

Apparatus. Each participant was provided with a 13-inch laptop for operating Zoom, browsing documents, and responding to chat messages. During the meeting, the participants' video windows were displayed in a vertical column showing live video feeds. Participants were informed that their audio, video, and gaze data would be recorded using an iPhone and a wearable eye tracker (Tobii Pro Glasses 2 and 3 [12, 13]), and their consent was obtained prior to the experiment.

4.2 Design of the Videoconference

Procedure. Figures 2 and 3 illustrate the meeting setup and discussion flow. Each meeting lasted 20 minutes, divided into two 10-minute sessions: the first under the *Non-Document* condition (without document browsing) and the second under the *Document* condition (with browsing). Both halves also included two sub-conditions: *Non-Chat* (no replies) and *Chat* (at least one reply). After the second session, one participant, chosen freely by the group, gave a brief two-minute presentation summarizing the discussion outcome.

Theme. The discussion topic was "Decide on three ideal conditions for a laboratory." This theme was selected because (1) it encourages balanced participation, (2) it is suitable with or without shared materials, and (3) it is familiar and engaging for all participants.

Document. In the second half, a PDF summarizing interview responses from other students (Figure 4) was shared via Zoom chat. Participants referred to it when needed. The purpose of this document was to guide discussion and to distribute visual attention across multiple areas. In the *Non-Document* condition, only the chat and video interfaces were displayed.

Chat. Throughout both halves, the experimenter sent one direct message per participant every minute via Zoom chat (see Table 1). Messages were short questionnaire-style prompts used as external interruptions rather than discussion aids. Sessions with no replies were categorized as *Non-Chat*, while those with one or more replies were *Chat* conditions. Participants could also use the chat to take notes during discussion.

4.3 Setup for Gaze Analysis

Gaze data were collected using Tobii Pro Glasses 2 and 3, while participants' video and audio were recorded with iPhones and Zoom. Gaze points were sampled every 10 ms. Areas of Interest (AOIs) were defined to identify what participants were looking at; when automatic classification was inaccurate, gaze positions were manually corrected based on the original recordings. The eight AOIs were: previous speaker, current speaker, next speaker, non-speaker, self, chat, document, and others. We analyzed gaze patterns during the 200 ms preceding turn-giving (shaded in Figure 1). The most frequently viewed AOI was identified using Tobii Pro Lab [14] and Python, with a 100 ms buffer applied to compensate for Zoom's transmission delay [15].

5 RESULTS

5.1 Eye-Gaze before Smooth Turn-Giving

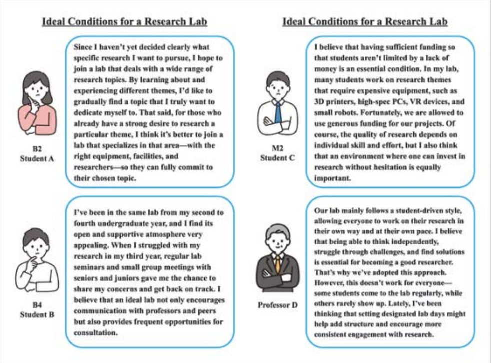
Figure 5 summarizes the most frequently viewed AOIs by the current speaker during 200 ms before smooth turn-giving in the videoconference.

Non-Chat & Non-Document condition. Excluding 15 cases categorized as "others", the non-speaker was most frequently viewed (11 times), followed by the next speaker (9 times).

Chat & Non-Document condition. Excluding 7 "others", the non-speaker and next speaker were both viewed most frequently (3 times each), followed by chat (2 times).

Non-Chat & Document condition. The most frequent AOI was the chat (7 times), followed by the non-speaker, self, and document (3 times each).

Chat & Document condition. Excluding 7 "others", the non-speaker was most frequently viewed (9 times), followed by the next speaker, chat, and document (5 times each).



Ideal Conditions for a Research Lab

B2
Student A

Since I haven't yet decided clearly what specific research I want to pursue, I hope to join a lab that deals with a wide range of research topics. By learning about and experiencing different themes, I'd like to gradually find a topic that I truly want to dedicate myself to. That said, for those who already have a strong desire to research a particular theme, I think it's better to join a lab that specializes in that area—with the right equipment, facilities, and researchers—so they can fully commit to their chosen topic.

B4
Student B

I've been in the same lab from my second to fourth undergraduate year, and I find its open and supportive atmosphere very appealing. When I struggled with my research in my third year, regular lab seminars and small group meetings with seniors and juniors gave me the chance to share my concerns and get back on track. I believe that an ideal lab not only encourages communication with professors and peers but also provides frequent opportunities for consultation.

Ideal Conditions for a Research Lab

M2
Student C

I believe that having sufficient funding so that students aren't limited by a lack of money is an essential condition. In my lab, many students work on research themes that require expensive equipment, such as 3D printers, high-spec PCs, VR devices, and small robots. Fortunately, we are allowed to use generous funding for our projects. Of course, the quality of research depends on individual skill and effort, but I also think that an environment where one can invest in research without hesitation is equally important.

Professor D

Our lab mainly follows a student-driven style, allowing everyone to work on their research in their own way and at their own pace. I believe that being able to think independently, struggle through challenges, and find solutions is essential for becoming a good researcher. That's why we've adopted this approach. However, this doesn't work for everyone—some students come to the lab regularly, while others rarely show up. Lately, I've been thinking that setting designated lab days might help add structure and encourage more consistent engagement with research.

Figure 4: Example of the document shared in the experiment

Table 1: Example of the chat message sent in the experiment

Please keep the content of this chat strictly confidential.	
Please respond to the following questions using a 7-point Likert scale (1 = Not at all, 7 = Extremely). Example of response: Q1: 3, Q2: 3, Q3: 3	
Q1	I feel anxious about interrupting others when I try to speak.
Q2	I feel anxious about uncomfortable silences occurring.
Q3	I feel stressed about replying to this chat.

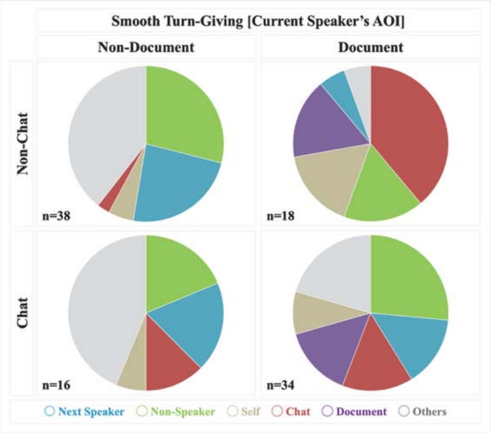


Figure 5: Visualization of the current speaker's AOI 200ms before smooth turn-giving in the videoconference

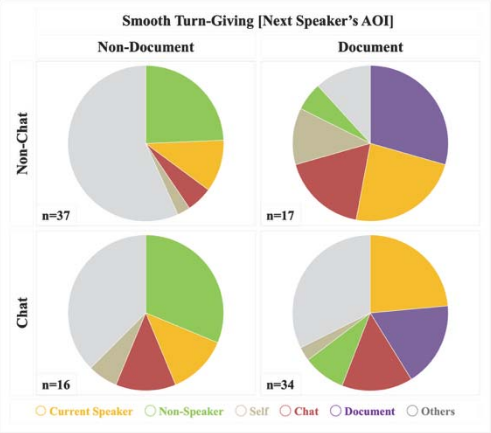


Figure 6: Visualization of the next speaker's AOI 200ms before smooth turn-giving in the videoconference

Figure 6 summarizes the most frequently viewed AOIs by the next speaker during the 200 ms before smooth turn-giving in the videoconference.

Non-Chat & Non-Document condition. Excluding 21 "others", the non-speaker was most frequently viewed (9 times), followed by the current speaker (4 times).

Chat & Non-Document condition. Excluding 6 "others", the non-speaker was most frequently viewed (5 times), followed by the current speaker and chat (2 times each).

Non-Chat & Document condition. The most frequent AOI was the document (5 times), followed by the current speaker (4 times).

Chat & Document condition. Excluding 11 "others", the previous speaker was most frequently viewed (8 times), followed by the document (6 times).

6 DISCUSSION

As shown in Figures 5 and 6, except in the Non-Chat & Document condition, the current speaker tended to gaze at the non-speaker or next speaker, and the next speaker tended to gaze at the non-speaker or current speaker during the 200 ms preceding smooth turn-giving. Previous studies have extensively discussed the close relationship between mutual gaze and turn-taking in in-person communication [16, 17, 18, 19]. The results of this study suggest that mutual gaze may also occur during turn-changings, even when the current and next speakers are not co-located.

However, because it is difficult to accurately perceive and convey gaze direction in videoconferencing, it remains unclear whether such mutual gaze is truly achieved. This implies that simply looking at the other's face video feeds may not directly contribute to smooth turn-taking, and that participants may in fact be overly reliant on face videos feeds during turn-giving.

While excessive fixation on others' face videos might sometimes reflect overreliance, it could also serve as a nonverbal signal indicating readiness to yield the floor. In multitasking videoconferences where verbal cues are limited, directing one's gaze toward another participant may function as a compensatory cue to coordinate turn transitions. Furthermore, the alignment of gaze between speakers, an implicit form of joint attention, may help establish shared awareness and timing, thereby contributing to smoother turn-giving.

These findings highlight that gaze behavior, even when imperfectly conveyed through video, may still play a constructive role in coordinating conversational flow under multitasking conditions.

7 LIMITATIONS AND FUTURE WORK

7.1 Constraints of Analysis

This study focused primarily on analyzing gaze behavior immediately before smooth turn-giving in multitasking videoconferences. However, future research should also examine turn-taking behavior to provide a more comprehensive understanding of how gaze contributes to the coordination of conversational turns. Further insights could be gained by comparing gaze patterns across in-person, audio-only, and video meetings, as well as by analyzing gaze behavior preceding non-smooth turn changes such as speech contention and silence. Comparing audio-only and video meetings may help isolate the effect of eye gaze in remote communication, while examining the differences between smooth and non-smooth turn transitions would strengthen the validity of the findings. Finally, because the experimental videoconference was conducted only once, the results should be interpreted within that limitation. Additional sessions with a larger dataset are required for more robust analysis.

7.2 Technical Limitations in Data Alignment

As described in Sections 3 and 4.3, utterances were measured from the perspective of individual participants, and a 100 ms buffer was applied to account for network delay. However, the temporal alignment of data may still differ across participants, as latency can vary depending on network conditions and hardware performance. Consequently, the perceived timing of speech events may differ between participants, leading to potential misclassification of utterances. For instance, whether a brief pause was identified as silence or which participant was labeled as the interrupter. This possibility of temporal mislabeling should be carefully considered when interpreting the results.

In addition, portions of the gaze data required manual correction, which limited the scope of quantitative analysis. Improving the accuracy of gaze tracking and synchronization between multimodal data sources is essential for future work. Developing automated analysis pipelines capable of efficiently processing large-scale gaze and speech datasets would also enhance the reproducibility and scalability of similar studies.

7.3 Conditions of the Experiment

This study incorporated two types of multitasking during the videoconference: document browsing and chat replying. However, some participants responded to chat messages with noticeable delays, which may have reduced the intended level of multitasking difficulty. Future experiments should therefore increase the temporal urgency or cognitive demand of chat responses to better simulate real multitasking conditions.

In addition, browsing a document does not always constitute multitasking or necessarily induce distraction. To elicit stronger cognitive engagement, future studies could employ documents containing figures or tables that require active interpretation and information integration. Such materials would better reflect realistic scenarios in which participants must balance multiple visual and cognitive tasks simultaneously.

Moreover, gaze behavior can be influenced by various contextual factors, including interpersonal relationships and display conditions. In our previous experiment [11], participants who had not met before tended to avoid eye contact, possibly due to unfamiliarity. In contrast, the present study recruited acquaintances to promote natural interaction. Nevertheless, subtle relational differences such as age hierarchy or social familiarity, may still have affected their gaze patterns. Future research should consider recruiting participants of similar age and familiarity levels to ensure equal and open communication dynamics.

Finally, screen size may also affect gaze behavior. To minimize distortion of visual attention caused by small video windows, future experiments should use external displays rather than built-in laptop screens, thereby providing a more consistent visual environment across participants.

7.4 Future Work

Based on the findings and limitations of this study, we plan to conduct further analyses to determine whether joint attention contributes to smooth turn changes under various meeting formats and enhanced multitasking conditions. Future experiments should examine how shared gaze coordination facilitates both turn-giving and turn-taking processes, as these are complementary mechanisms essential for natural conversational flow.

Additionally, analyzing non-smooth turn changes such as speech contention and silence, could provide further evidence for the role of gaze in regulating conversational dynamics. Comparing gaze patterns across in-person, audio-only, and video meetings would also help clarify how visual cues interact with cognitive load and task complexity.

We further aim to increase the number of experimental sessions to improve the statistical reliability and generalizability of our conclusions. Expanding the participant pool to include diverse communication styles and technical environments could also offer broader insights into how multitasking affects turn management.

Ultimately, our long-term goal is to develop design guidelines and assistive systems that promote smoother and more inclusive turn-taking in multitasking videoconferencing environments involving simultaneous document browsing and chat interactions.

8 CONCLUSION

We analyzed gaze direction immediately before smooth turn-giving in multitasking videoconferences involving document browsing and chat replying, with the aim of identifying cues for smooth turn-taking. The results suggest that joint attention, where the current (or previous) and next speakers simultaneously focus on the same area, such as a shared document, may contribute to smoother turn coordination. Our findings also indicate that speakers sometimes become overly focused on other participants' video feeds, implying a complex balance between attention sharing and overreliance on visual cues. In addition, participants engaged in chat responses appeared less involved in turn-taking or joint attention, suggesting that multitasking can fragment conversational engagement.

As discussed in Section 7, future studies will seek to generalize these findings and clarify the specific role of joint attention in facilitating turn-taking. We plan to examine gaze patterns preceding non-smooth turn changes, such as interruptions and silences, to further validate our interpretation. Enhancing the precision of gaze data and expanding the dataset will also allow for more reliable conclusions. Ultimately, based on these insights, we aim to develop design principles and support systems that help participants especially those engaged in multitasking environments achieve smoother, more natural turn-taking in videoconferencing contexts.

REFERENCES

[1] Zoom, https://www.zoom.com, last accessed 2025/06/02.

[2] Microsoft Teams, https://www.microsoft.com/en/microsoft-teams/group-chat-software, last accessed 2025/06/02.

[3] Choudhury, P., Foroughi, C., and Larson, B., Work-from-anywhere: The Productivity Effects of Geographic Flexibility, Strategic Management Journal, Vol.42, pp.655-683 (2020).

[4] Kendon, A., Some functions of gaze-direction in social interaction, Acta Psycho logica, Vol.26, pp.22-63 (1967).

[5] Jokinen, L., Nishida, M., and Yamamoto, S, On eye-gaze and turn-taking, Eye gaze in intelligent human machine interaction (EGIHMI'10), pp.118-123 (2010).

[6] Riedl, R., On the stress potential of videoconferencing: definition and root causes of Zoom fatigue, Electron Markets, Vol.32, pp.153–177 (2022).

[7] Ansah, A. A., Xing, A., Kamaraj, A. V., Tosca, D., Boyle, L., Iqbal, S., Kun, A. L., Lee, J. D., Pahud, M., and Shaer, O., "I need to respond to this" – Contributions to group creativity in remote meetings with distractions, CHIWORK'22, pp.1-12 (2022).

[8] Cao, H., Lee, C. J., Iqbal, S., Czerwinski, M., Wong, P. N, Rintel, S., Hecht, B., Teevan, J., and Yang, L., Large scale analysis of multitasking behavior during remote meetings, CHI Conference on Human Factors in Computing Systems, pp. 1-13 (2021).

[9] Lee, M., Park, W., Lee, S. and Lee, S., Distracting Moments in Videoconferencing: A Look Back at the Pandemic Period, CHI Conference on Human Factors in Computing Systems, pp.1-21 (2022).

[10] Reinecke, L., Aufenanger, S., Beutel, M. E., Dreier, M., Quiring, O., Stark, B., Wölfling, K., Müller, K. W., Digital Stress over the Life Span: The Effects of Communication Load and Internet Multitasking on Perceived Stress and Psychological Health Impairments in a German Probability Sample, Media Psychology, Vol.20, pp.90-115 (2017).

[11] Ohnaka, K., Imagawa, T., Iso, K., Ihara, M. and Kobayashi, M., Quantitative Observation to Explore the Turn-Changing Mechanisms of Conversations in Remote Meetings Accompanying Supplemental Materials, CollabTech, Vol.14890, pp.161-176 (2024).

[12] Tobii Pro Glasses 2, https://www.tobii.com/en/products/discontinued/tobii-pro-glasses-2, last accessed 2025/06/02.

[13] Tobii Pro Glasses 3, https://www.tobii.com/en/products/eye-trackers/wearables/tobii-pro-glasses-3, last accessed 2025/06/02.

[14] Tobii Pro Lab, https://www.tobii.com/products/software/behavior-research-software/tobii-pro-lab, last accessed 2025/07/18.

[15] Chang, H., Varvello, H., Hao, F., and Mukherjee, S., Can You See Me Now? A Measurement Study of Zoom, Webex, and Meet, In Proceedings of the 21st ACM Internet Measurement Conference, pp.216-228 (2021).

[16] Rutter, D. R., Stephenson, G. M., and Dewey, M. E., Visual communication and the content and style of conversation, British Journal of Social and Clinical Psychology, Vol.20, pp.41-52 (1981).

[17] Rossano, F., Gaze in Conversation, The handbook of conversation Analysis, Wiley-Blackwell, pp.308-329 (2012).

[18] Jokinen, K., Furukawa, H., Nishida, M., and Yamamoto, S., Gaze and turn-taking behavior in casual conversational interactions, ACM Transactions on Interactive Intelligent Systems (TiiS), Vol.3, Article. 12, pp.1-30 (2013).

[19] Kendrick, K. H., Holler, J., and Levinson, S. C., Turn-taking in human face-to-face interaction is multimodal: gaze direction and manual gestures aid the coordination of turn transitions, Philosophical Transactions of the Royal Society B, Vol.378, pp.1-17 (2023).

Session 4:
Multimedia and Sensor Systems

(Chair: Manato Fujimoto)

Design and Control Olfactory Displays based on Diffusion and Flow of Aroma Gas

Koichi ONUKI[*], Motofumi HATTORI[*], Yohei SETA[*], and Yuichi BANNAI[**]

[*] Faculty of Information Technology, Kanagawa Institute of Technology
1030 Shimo-ogino, Atsugi city, Kanagawa prefecture, Zip 243-0203, Japan
Telephone +81-46-291-3084
hattori@ic.kanagawa-it.ac.jp

[**]Faculty of Sociology, Urawa University, Japan
yu.bannai@urawa.ac.jp

Abstract - In order to offer scent in a virtual space, the authors are developing hardware and software of olfactory displays. These olfactory displays are controlled by the game engine UNITY. A user who wears a Head Mounted Display can see 3D Computer Graphics and smell scent which is associated with the 3DCG scene. To analyze the specifications of an olfactory display, the authors are simulating flow and diffusion of aroma gas in olfactory displays. In these phenomena that aroma gas is transported by air flow, a huge number (amount) of aroma gas molecules are transported by colliding with a huge number (amount) of air molecules. Although it is impossible to compute these molecular motions, we will coarse-grain these phenomena by hydrodynamics limit theory of statistical physics and we can compute them approximately by Moving Particle Simulation.

Keywords: VR, Olfactory display, Aroma gas, diffusion, flow, Moving Particle Simulation, Hydrodynamics limit

1 Simulation to Design Olfactory Display

When a user wears a Head Mounted Display, one can immerse oneself in a Virtual Space. This virtual space is generated to one's eyes by 3D Computer Graphics. If some scent is offered to one's nose, one feels reality more in the virtual space.

Thus the authors are developing olfactory displays which offer some kinds of scent to a user's nose, as shown in Figure 1. Figure 3 shows the mechanism of olfactory displays by the authors. [1] [2] [3]

In order to analyze the specifications of an olfactory display, the authors simulated how to the air flow transports the aroma gas , in the authors' previous research [3]. In the results of these simulations, the aroma gas diffused into the air flow by the effect of the difference between the amount density of the aroma gas molecules and the amount density of the air molecules. But in the physical experiments, the aroma gas diffuses into the air when there is no difference between the amount density of the aroma gas molecules and the amount density of the air molecules, since all the molecules collide frequently by molecules' thermal motion. In this presentation, the authors will simulate the phenomena that the aroma gas diffuse into tha air flow by the effect of the thermal motion of all molecules.

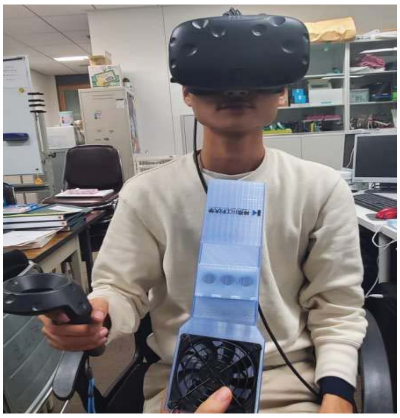


Figure 1 : A Head Mounted Display and an olfactory display

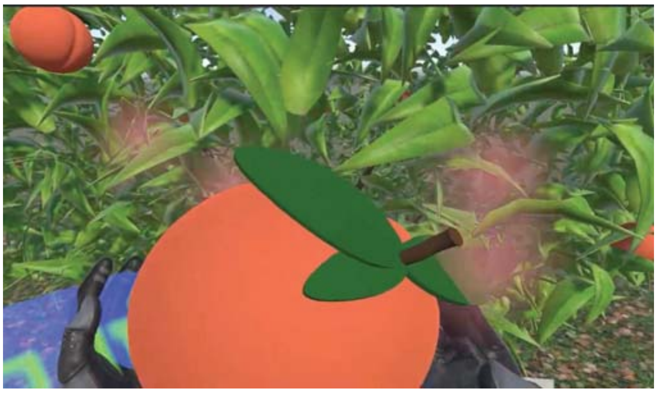


Figure 2 : VR space by scent and 3DCG graphics

2 Olfactory Display controlled by the game engine Unity

The authors control the olfactory display by the game engine Unity. Since the Head Mounted Display and the olfactory display are controlled by the same Unity scene, the olfactory display offers appropriate scent associated with the 3DCG scene to the user's nose.

The user who wears a Head Mounted Display (HMD) can see 3D Computer Graphics which is rendered by the Unity scene. As shown in Figure 2, when the HMD user grasp a fruit, the scent of the fruit is offered to the HMD user's nose.

These VR systems with scent and graphics will be developed and be reported by the authors' group (a doctor course student and master course students) at future International Workshops on Informatics (IWINs).

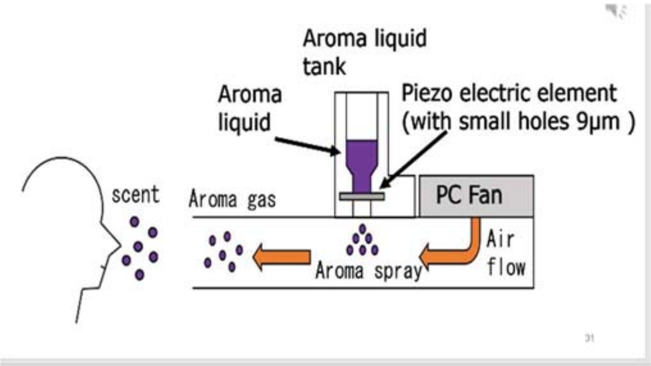


Figure 3 : Mechanism of olfactory displays

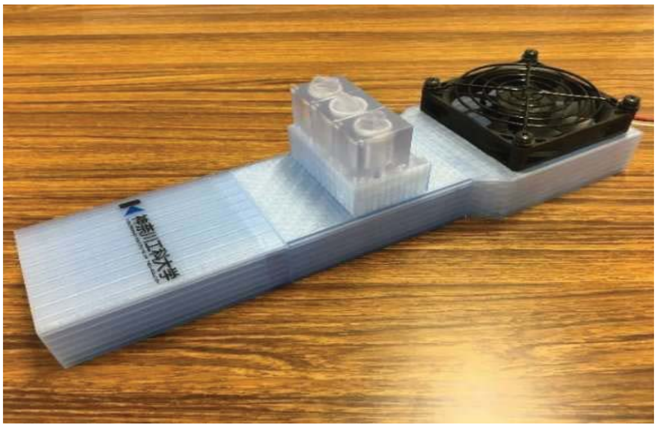


Figure 4 : Photograph of an olfactory display

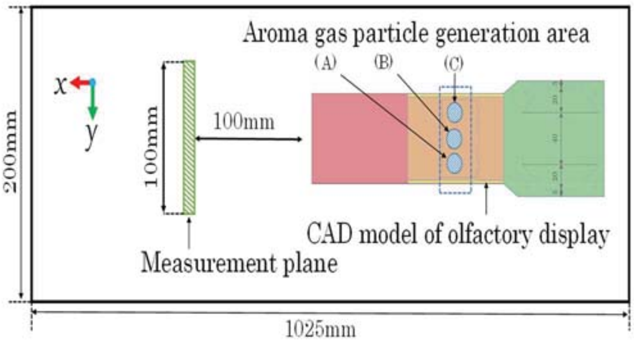

Aroma gas particle generation area
(A)
(B)
(C)
x
y
100mm
100mm
200mm
CAD model of olfactory display
Measurement plane
1025mm

Figure 5 : Top view of the whole range in Fluid Simulation

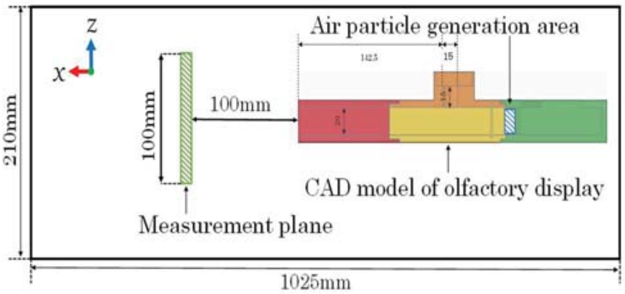


Figure 6 : Side view of the whole range in Fluid Simulation

3 Hardware of Olfactory Display

The mechanism of olfactory display is shown in Figure 3 and Figure 4. The detailed hardware of olfactory displays was reported in [11].

In order to offer appropriate scent to a user's nose according to an object in a VR space, the authors had developed olfactory displays. as shown in Figure 4. When the user immerses a VR space as shown in Figure 1, appropriate scent is offered from an olfactory display as shown in Figure 2. When a peach appears in the VR space, the scent of peach is offered from the olfactory display. When a lemon appears in the VR space, the scent of lemon is offered from the olfactory display.

Figure 3 shows the structure of olfactory displays which the authors had developed. Figure 4 shows an example of olfactory displays which the authors had developed. As Figure 3 shows, aroma liquid tank ejects aroma spray into air flow in a flow tunnel of the olfactory display. Aroma spray is vaporized into aroma gas. Aroma gas is transported to a user's nose by the air flow in the flow tunnel of the olfactory display. Since there are 3 aroma liquid tanks on the olfactory display which is shown in Figure 4, three kinds of scent are offered from 3 aroma liquid tanks on this olfactory display. A photograph of an aroma tank is shown in Figure 7.

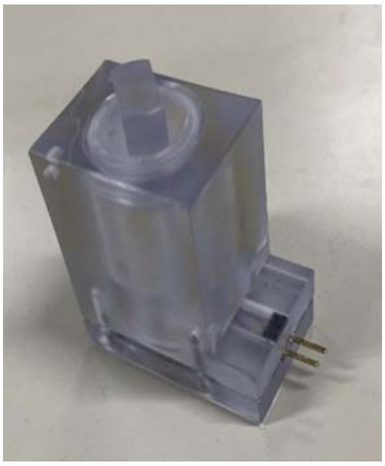


Figure 7 : An aroma liquid tank of olfactory display

There is a piezo-electric element in the bottom of the aroma liquid tank. Aroma spray is dripped from the bottom of the aroma liquid tank into the air flow tunnel of the olfactory display during electric voltage is applied to the piezo-electric element. Aroma spray is dripped from many holes whose diameters are 9 μm (9 micro meters) into the air flow tunnel of the olfactory display. Since these holes are very small, aroma spray is never dripped when we do not apply electric voltage to the piezo-electric element. By the effect of intermolecular force between aroma liquid and the piezo-electric element, aroma spray is never dripped from the above small holes when we do not apply electric voltage to the piezo-electric element.

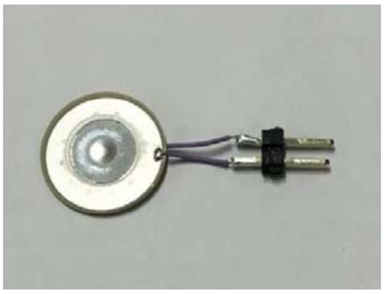


Figure 8 : Piezo-electric element which has many small holes.

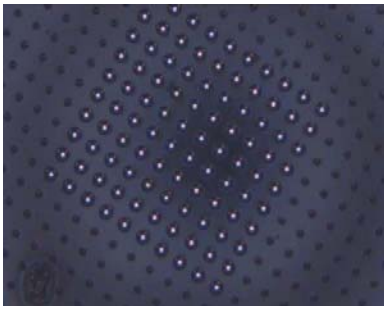


Figure 9 : Piezo-electric element with many small holes.

In order to control the electric voltage which is applied to the piezo-electric element as shown in Figure 8, the authors had developed a Digital/Analog converter as shown in Figure 10. By controlling time length of applied voltage, the strength of scent is controlled. If we apply electric voltage to the piezo-electric element long time, scent becomes strong. If we apply electric voltage to the piezo-electric element short time, scent becomes weak. Such time length is controlled from the Unity C# script of the Unity project which makes a VR space through Head Mounted Displays (HMDs). The authors use Vive Pro Eye (by HTC Corporation) as the Head Mount-ed Display which makes VR spaces.

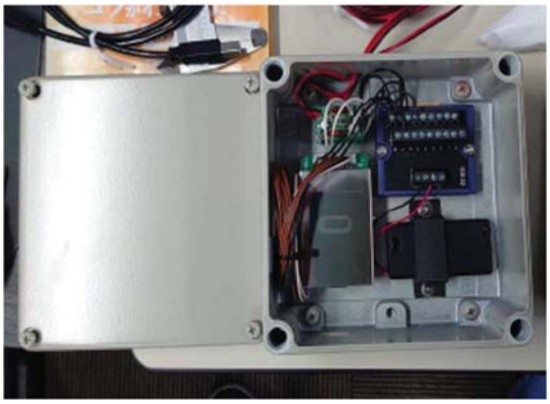


Figure 10 : D/A converter controls the piezo-electric elements

4 Design Shape and Size of Flow Tunnel of Olfactory Display

In order to optimize the shape and size of the olfactory display, the authors are simulating the flow of aroma gas which is transported by the air flow which was caused by PC fan (See the right part of Figure 3). In the authors' previous research [2] [3], the flows of aroma gas were simulated and analyzed. These simulations [2][3] could compute dynamic diffusion of aroma gas into air. But these previous researches [2] [3] did not consider the static diffusion of aroma gas into air.

When the PC fan stops and the air flow stops, the aroma gas diffuses to the air irreversibly as Boltzmann's statistical physics teaches [7]. In this presentation, the authors simulate and analyze these static diffusion of the aroma gas (limonene) to the air (nitrogen and oxygen).

5 Diffusion and Transport in Flow of Air and Aroma gas

In the real physics phenomena about the olfactory display, the motion of a huge number (amount) of air molecules (nitrogen molecules and oxygen molecules) is generated as an air flow by the PC fan, forcibly as shown in Figure 3. The huge number (amount) of molecules of the aroma gas (limonene molecules) are transported by colliding a huge number (amount) of air molecules (nitrogen molecules and oxygen molecules). During this transport, the huge number (amount) of aroma gas molecules (limonene molecules) diffuse into the huge number (amount) of air molecules (nitrogen molecules and oxygen molecules).

If the authors could compute the motions of these huge number (amount) of molecules numerically, the very accurate phenomena would be analyzed numerically. But this very accurate numerical computation is impossible as statistical physics shows [4].

Instead of the above ideal dream (the very accurate computation), Moving Particle Simulation can compute the above molecular motions approximately [1], as the section 6 and the section 7 of this proceeding explain.

Statistical physics shows that the motion of a huge number (amount) of molecules can be computed as the motion of many number (amount) of fluid particles approximately by coarse-graining [4] [5].

6 Computational Discrete Fluid Dynamics can compute Molecular Dynamics approximately

In computational discrete fluid dynamics, the motion of a huge number of molecules $\boldsymbol{q} = \left(q_x,\ q_y, q_z \right)$ is coarse-grained to the motion of many number of fluid particles $\boldsymbol{R} = \left(R_x,\ R_y, R_z \right)$. Although each molecule $\boldsymbol{q} = \left(q_x,\ q_y, q_z \right)$ is a point in the 3 dimensional Euclidean O-XYZ space, each fluid particle $\boldsymbol{R} = \left(R_x,\ R_y, R_z \right)$ is a small volume in the 3 dimensional Euclidean O-XYZ space.

Since each fluid particle $\boldsymbol{R} = \left(R_x,\ R_y, R_z \right)$ contains many molecules, we can define mass density ρ, momentum density M, pressure P, temperature T, of each fluid particle $\boldsymbol{R} = \left(R_x,\ R_y, R_z \right)$. The velocity v(t , R) of a fluid particle $\boldsymbol{R} = \left(R_x,\ R_y, R_z \right)$ is a ratio of momentum density M with respect to mass density ρ, i.e.

$$v(t, R) = M(t, R) \;/\; \rho(t, R)$$
$$M(t, R) = \rho(t, R)\ \ v(t, R)$$

at time t [4] [6].

Let N be the huge number which express the number (amount) of both air molecules (nitrogen molecules and oxygen molecules) and aroma gas molecules (limonene molecules). The order of the huge number N is Avogadro's number 6.02×10^{23} .

Let $\boldsymbol{q_i} = \left(q_{i\,x},\, q_{i\,y}, q_{i\,z} \right)$ be an air molecule (a nitrogen molecule or an oxygen molecule) or an aroma gas molecule (a limonene molecule) for i = 1, 2, ・・・, N. Let $\boldsymbol{p_i} = \left(p_{i\,x},\, p_{i\,y}, p_{i\,z} \right)$ be a momentum of the molecule $\boldsymbol{q_i} = \left(q_{i\,x},\, q_{i\,y}, q_{i\,z} \right)$ for i = 1, 2, ・・・, N. Let m_i be a mass of the molecule $\boldsymbol{q_i} = \left(q_{i\,x},\, q_{i\,y}, q_{i\,z} \right)$ for i = 1, 2, ・・・, N. Each m_i is a mass of a nitrogen molecule or an oxygen molecule or a limonene molecule for i = 1, 2, ・・・, N. If $\boldsymbol{q_i} = \left(q_{i\,x},\, q_{i\,y}, q_{i\,z} \right)$ is a nitrogen molecule, m_i = 28 [g / mol]. If $\boldsymbol{q_i} = \left(q_{i\,x},\, q_{i\,y}, q_{i\,z} \right)$ is an oxygen molecule, m_i = 18 [g / mol]. If $\boldsymbol{q_i} = \left(q_{i\,x},\, q_{i\,y}, q_{i\,z} \right)$ is a limonene molecule, m_i = 136 [g / mol].

Let $U(\boldsymbol{q_i} , q_j)$ be the Lenard-Jones potential which computes the intermolecular force between the molecule $\boldsymbol{q_i}$ and the molecule q_j for i , j = 1, 2, ・・・, N [4].

Let $q[N] = (q_1 ,\ q_2 ,\ \cdots,\ q_N)$ be a position vector of N molecules. Let $p[N] = (p_1 ,\ p_2 ,\ \cdots,\ p_N)$ be a momentum vector of N molecules.

Then the Hamiltonian H(q[N] , p[N]) of N molecules dynamics is defined as follows.

$$H(q[N]\,,\, p[N]) = \qquad (2)$$

$$\sum_{i=1}^{N}\frac{1}{2m\_i}\langle p_i, p_i\rangle + \frac{1}{2}\sum_{i=1}^{N}\sum_{i\neq j}U\left(q_i - q_j\right)$$

where $\left\langle p_i \,,\, p_j \right\rangle$ is an inner product of 3 dimensional vectors p_i and p_j .

We obtain the Hamilton dynamics of N molecules as follows

$$\frac{dq_i(\tau)}{d\tau} = \frac{\partial H}{\partial p_i}\left(q[N](\tau)\ ,\ p[N](\tau) \right)$$

$$\boldsymbol{\frac{dp_i(\tau)}{d\tau} = (-1)\, \frac{\partial H}{\partial q_i}\left(q[N](\tau)\ ,\ p[N](\tau) \right)}$$

(3) (4)

for i = 1, 2, ・・・, N and time τ .

Here consider small positive real number ε > 0 such that

$$\epsilon^3 = \frac{1}{N} \qquad (5)$$

and Dirac's delta function δ (r) on the 3 dimensional Euclidean O-XYZ space.

The mass field of N molecules becomes

$$\mathrm{Rho}[N](t,r) = \frac{1}{N}\sum_{i=1}^{N} m_i\ \delta\left(r - \epsilon\, q_i\left(\frac{t}{\epsilon}\right)\right) \qquad (6)$$

The momentum field of N molecules becomes

$$M[N](t,r) = \frac{1}{N}\sum_{i=1}^{N} p_i\left(\frac{t}{\epsilon}\right)\ \delta\left(r - \epsilon\, q_i\left(\frac{t}{\epsilon}\right)\right) \qquad (7)$$

The energy field of N molecules becomes

$$E[N](t,r) = \frac{1}{N}\sum_{i=1}^{N} e_i\left(\frac{t}{\epsilon}\right)\ \delta\left(r - \epsilon\, q_i\left(\frac{t}{\epsilon}\right)\right) \qquad (8)$$

where Hamiltonian e_i of the molecule q_i is defined by

$$e_i(t) = \frac{1}{2m_i}\,\langle p_i(t),\, p_i(t)\rangle + \frac{1}{2}\sum_{j\neq i} U\left(q_i(t) - q_j(t)\right) \qquad (9)$$

The total volume Ω is the whole range of Figure 5 and Figure 6. The total volume Ω consists of the gas flow tunnel of the olfactory display and the space between the user's nose and the olfactory display in Figure 3.

In computational discrete fluid dynamics, the total volume Ω is divided into 100,000 fluid particles R_k, i.e.

$$\Omega = R_1 \cup R_2 \cup R_3 \cup \cdots \cup R_{100{,}000} \qquad (10)$$

where

$$R_k \cap R_l = \emptyset \qquad (11)$$

for $k \neq l$.

Based on the hydrodynamics limit theory [4] [5] of statistical physics, the fields (6) (7) (8) of N molecules are approximated by the fields (12) (13) (14) on each fluid particles R_k for k = 1, 2, 3, ・・・, 100000.

The mass field Rho[N](t , r) of molecules is approximated by the mass density ρ (t , r) on each fluid particle $\boldsymbol{R} = \left(R_x\,,\, R_y\,, R_z \right)$ as

$$\lim_{N\to\infty}\int_{r\in R} \mathrm{Rho}[N](t,r)\,dr = \int_{r\in R}\rho(t,r)\,dr \qquad (12).$$

The momentum field M[N](t , r) of molecules is approximated by the momentum density ρ (t , r) v(t , r) on each fluid particle $\boldsymbol{R} = \left(R_x\,,\, R_y\,, R_z \right)$ as

$$\lim_{N\to\infty}\int_{r\in R} M[N](t,r)\,dr = \int_{r\in R}\rho(t,r)\ v(t,r)\,dr \qquad (13).$$

The energy field E[N] (t , r) of molecules is approximated by the energy density e(t , r) on each fluid particle $\boldsymbol{R} = \left(R_x\,,\, R_y\,, R_z \right)$ as

$$\lim_{N\to\infty}\int_{r\in R} E[N](t,r)\,dr = \int_{r\in R} e(t,r)\,dr \qquad (14).$$

By the theory of hydrodynamics limit theory [4] [5] of statistical physics, the mass density ρ(t , r) and the momentum density ρ(t , r) v(t , r) satisfy the following equations (15) and (16) of continuity i.e. Equation (15) of conserving mass and Equation (16) of conserving momentum.

Mass is conserved as

$$0 = \frac{\partial\rho}{\partial t} + \frac{\partial}{\partial \boldsymbol{r}}\cdot(\boldsymbol{\rho v}) = \frac{\boldsymbol{D\rho}}{\boldsymbol{Dt}} + \boldsymbol{\rho}\,\frac{\partial}{\partial \boldsymbol{r}}\cdot \boldsymbol{v} \qquad (15).$$

Momentum is conserved as

$$\boldsymbol{0} = \frac{D(\rho\boldsymbol{v})}{Dt} + \frac{\partial P}{\partial \boldsymbol{r}} \qquad (16).$$

where P(t , r) is a pressure around a point $\boldsymbol{r} = \left(r_x\,,\, r_y\,, r_z \right)$ which is defined by Equation (22) in the end of this section 6.

In computational discrete fluid dynamics, each fluid particle R_k moves as the time t goes on. Equations (10) and (11) are rewritten as

$$\Omega \;=\; R_1(t) \;\cup\; R_2(t) \;\cup\; R_3(t) \;\cup\; \cdots \;\cup\; R_{100{,}000}(t) \qquad (\ 17\ )$$

where

$$R_k\ (t) \;\cap\; R_l(t) \;=\; \emptyset \qquad (\ 18\ )$$

for $k \neq l.$

By Equation (12), the mass Mass_$k(t)$ of a fluid particle $R_k\ (t)$ is defined as

$$\begin{aligned}\mathrm{Mass\_}k(t) \;&=\; \int_{r \in R\_k(t)} \rho(t,r)\, dr \\ &=\; \lim_{N\to\infty} \int_{r \in R\_k(t)} \mathrm{Rho}[N](t,r)\, dr\end{aligned} \qquad (\ 19\ ).$$

which is a sum of all masses of all molecules (nitrogen and oxygen and limonene) in a fluid particle $R_k\ (t)$ for k = 1, 2, 3, ・・・, 100000.

By Equation (13), the momentum Mmt_$k(t)$ of a fluid particle $R_k\ (t)$ is defined as

$$\begin{aligned}\mathrm{Mmt\_}k(t) \;&=\; \int_{r \in R\_k(t)} \rho(t,r)\ v(t,r)\, dr \\ &=\; \lim_{N\to\infty} \int_{r \in R\_k(t)} M[N](t,r)\, dr\end{aligned} \qquad (\ 20\ ).$$

which is a sum of all momentums of all molecules (nitrogen and oxygen and limonene) in a fluid particle $R_k\ (t)$ for k = 1, 2, 3, ・・・, 100000.

The velocity $V_k\ (t)$ of a fluid particle $R_k\ (t)$ is defined as

$$V_k\ (t) \;=\; \frac{\mathrm{Mmt\_}k(t)}{\mathrm{Mass\_}k(t)} \qquad (\ 21\ ).$$

for k = 1, 2, 3, ・・・, 100000. Equations (19) (20) (21) are derived based on the hydrodynamics limit theory [4] [6].

To formulate Moving Particle Simulation based on Equations (25) and (26) (Lagrange type Euler equation), **let** $\boldsymbol{r} = \left(r_x\,,\ r_y\,, r_z \right)$ be a center point of a fluid particle $\boldsymbol{R} = \left(R_x\,,\ R_y\,, R_z \right)$ (22).

7 Moving Particle Simulation computes Transport and Diffusion of aroma gas

About 1.8[m/s] is the velocity of gas flow (flow of nitrogen and oxygen and limonene) which is controlled by the PC fan in Figure 3 and Figure 4. Since the velocity of the gas flow is always about 1.8[m/s] and this value is sufficiently smaller than the velocity 340[m/s] of sound wave in air, the gas flow is approximated by incompressible flow [8]. Although the mass density ρ(t , r) of the gas (nitrogen and oxygen and limonene) changes as the time t goes on, the change of ρ(t , r) is very small. The mass density ρ(t , r) changes near the positive constant real number ρ_0 (> 0).

$\rho(\,t\,, r\,)$ nealy equals ρ_0 (23).

Equation (23) yields that the pressure P(t , r) changes lineally as

$$P(t,r) \;-\; P_0 \;=\; \mathrm{Sound}^2\ \{\,\rho(\,t\,,r\,) \;-\; \rho_0\ \} \qquad (\ 24\ ).$$

where Sound (> 0) is a velocity 340[m/s] of the sound wave in air and P_0 is a pressure of air at mass density ρ_0 [8].

In Moving Particle Simulation, the mass density $\rho(\,t\,, R\,)$ of the fluid particle $\boldsymbol{R} = \left(R_x\,,\ R_y\,, R_z \right)$ is computed from the particle number density (PND) of the fluid particle $R(t)$ at time t [9].

Equations (16) (22) (23) yields the Lagrange type fluid dynamics equation (Euler equation)

$$\frac{D\boldsymbol{r}}{Dt} = \boldsymbol{v}$$

$$\rho_0\ \frac{D\boldsymbol{v}}{Dt} \;=\; (-1)\frac{\partial P}{\partial \boldsymbol{r}} \qquad (\ 25\ )(\ 26\ )$$

By Equations (24) (25) (26), the authors compute each fluid particle $R_k\ (t)$, each velocity $V_k\ (t)$ and each pressure $P($ t, R_k) as the time *t* goes on for *k* = 1, 2, 3, ・・・, 100000.

When the aroma gas is transported by the air flow, the aroma gas diffuses into the air. Since the mass of the aroma gas in the fluid particle $\boldsymbol{R} = \left(R_x\,,\ R_y\,, R_z \right)$ increases as the time t goes on, Let $\beta($ t , r) be the purity (concentration) of the aroma gas in the air.

$$0 \;\le\; \beta(\,t\,, r\,) \;\le\; 1 \qquad (\ 27\ )$$

The purity (concentration) $\beta($ t , r) is the ratio of the number (amount) of the aroma gas molecules with respect to the total number (amount) of molecules in the fluid particle $\boldsymbol{R} = \left(R_x\,,\ R_y\,, R_z \right)$ [6] [7].

When $\beta($ t , r) = 0, the fluid particle $\boldsymbol{R} = \left(R_x\,,\ R_y\,, R_z \right)$ contains no molecule of the aroma gas. All molecules in fluid particle $\boldsymbol{R} = \left(R_x\,,\ R_y\,, R_z \right)$ are the air.

When $\beta($ t , r) = 1, the fluid particle $\boldsymbol{R} = \left(R_x\,,\ R_y\,, R_z \right)$ contains no molecule of the air. All molecules in fluid particle $\boldsymbol{R} = \left(R_x\,,\ R_y\,, R_z \right)$ are the aroma gas.

When $0 \;<\; \beta(\ t\ , r\ ) \;<\; 1$, the fluid particle $\boldsymbol{R} = \left(R_x\,,\ R_y\,, R_z \right)$ contains both aroma gas molecules and air molecules.

The diffusion of the aroma gas into the air is modeled by the following diffusion equation (28) [7]

$$\frac{\partial \beta}{\partial t} \;=\; \mathrm{Diffuse}\ \frac{\partial}{\partial \boldsymbol{r}} \cdot \frac{\partial}{\partial \boldsymbol{r}}\ \beta(t, \boldsymbol{r}) \qquad (\ 28\ )$$

where Diffuse (>0) is a diffusion coefficient.

The pressure (24), the Lagrange type fluid dynamics (25) (26), and the diffusion equation (28) of the purity (concentration) are solved simultaneously by Moving Particle Simulation [9]. The authors compute Moving Particle Simulation by developing SDK source codes to the software

Particleworks which is supported by Prometech Software Inc. [10]

The computed results was presented and discussed on August 31 Sunday – September 3 Wednesday 2025 at 19 th International Workshop on Informatics (IWIN 2025) in Sendai-city Miyagi-prefecture Japan.

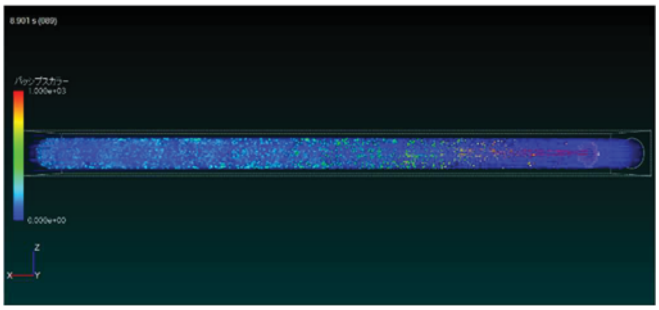


Figure 11 : Diffusion is simulated by Purity in MPS

8 Conclusions and Future Works

The authors discuss the diffusion and flow of the aroma gas to the air in detail. (The same simulation method will analyze the diffusion and flow of the aroma gas to the laboratory room.)

For the above subjects, the authors confirmed that molecular dynamics of gas is approximated by computational discrete fluid dynamics which includes Moving Particle Simulation based on the hydrodynamics limit theory in statistical physics.

In order to compute the phenomena that the aroma gas diffuses to the air, the authors paid attention to the purity (concentration) $\beta(t , r)$ of the aroma gas in the air, and implemented the diffusion equation about $\beta(t , r)$ to Moving Particle Simulation.

The computed results will be presented and analyzed on September 1 Monday 2025 at 19 th International Workshop on Informatics (IWIN 2025) in Sendai-city Miyagi-prefecture Japan.

The authors will improve olfactory displays and develop VR systems with scent and graphics.

In order to design and optimize flow tunnels of olfactory displays, the second author (Motofumi Hattori) is learning and studying molecular dynamics about gas and liquid in the physics group of Open University (Prof. Jun Kishine, Prof. Kei Iida, and Prof. Tetsuo Matsui).

Informatics theories and technologies are becoming more and more important in statistical physics. By simulating natural laws in VR worlds, the natural laws are comprehended deeply both by logic and by mind. The academic discussions between VR research and statistical physics research will be reported by the second author (Motofumi Hattori) at future International Workshops on Informatics (IWINs).

Acknowledgement

The authors would like to express their thanks to Dr. Tomoki Sawada, Mr. Hiromori Takakura, and Mr. Yosuke Tanigawa (Prometech Software Inc. [7]) for teaching them the source code which computes the purity (concentration) in the fluid particles.

Also the second author (Motofumi Hattori) would like to express his thanks to Prof. Takuya Matsunaga (The University of Tokyo) who introduced him to statistical physics which guarantees Moving Particle Simulation by showing the book [12].

References

[1] Motofumi Hattori, Yohei Seta, Yuichi Bannai
“The Meaning of Moving Particle Simulation for Gas Flow in Olfactory Display”
Abstract of The 15th World Congress on Computational Mechanics (WCCM 2022) 0709 (2003) 2022/7/31

[2] Yohei Seta, Mitsunori Makino, Yuichi Bannai, Motofumi Hattori
“Air Flow Channel Planning for Droplet-Spray-type Olfactory Displays Using a Small Wind Tunnel”
Inter national Journal of Informatics Society (IJIS) Volume 15, Number 2, pp.55-65, 2023/10/23

[3] Yuichi Bannai, Yohei Seta, Motofumi Hattori,
“3DCG Visualization to analyze aroma gas flow in olfactory displays by MPS”, Journal of the Virtual Reality Society of Japan, 29(1) pp. 18-19, 2024/04/10.

[4] P. M. Chaikin and T. C. Lubensky
"Principles of Condensed Matter Physics"
Chapter 8 Fluid Dynamics,
Cambridge University Press 1995

[5] Kohei Uchiyama and Tadahisa Funaki,
“ From Micro To Macro Vplume 2 Hydrodynamics Limit on Lattice Gas ”, Maruzen publish 2012

[6] L.D. Landou and E.M. Lifshitz, “Fluid Mechanics”, Chapter 6 Diffusion, Pergamon Press, 1987

[7] Kazuo Kitahara, “Nonequilibrium statistical physics”, Chapter 4 Diffusion Phenomena, Iwanami Basic Physics Series Volume 8, 1997/10/28

[8] Alexander L. Fetter and John Dirk Walecka
“Theoretical Mechanics of Particles and Continua”
Chapter 9 Sound Waves in Fluids, Dover 2003

[9] Seiichi Koshizuka, Kazuya Shibata, Masahiro Kondo, and Takuya Matsunaga, “Moving Particle Semi-implicit Method (A meshfree Particle Method for Fluid Dynamics)”, Academic Press 2018

[10] Particleworks supported by Prometech Software Inc.
https://particleworks.com/

[11] Noriko Takimoto, Koichi Onuki, Yohei Seta, Yuichi Bannai, Motofumi Hattori,
“A Head Mounted Display attachment Olfactory Display to offer Multimodal Sensory Information for Advertisements in Metaverse”, Virtual, Augmented and Mixed Reality 15th International Conference, VAMR 2023, Held as Part of the 25th HCI International Conference, HCII 2023, Copenhagen, Denmark, July 23--28, 2023, Proceedings pp. 278-294. 2023/07/23

[12] William Graham Hoover,
“Smooth Particle Applied Mechanics”
World Scientific Publishing Corporation, 2006

An Empirical Study of Broadcaster-following POV Heatmap based on MR technologies in 360-degree Internet Live Broadcasting

Yoshia Saito*, Junichiro Suto*

*Graduate School of Software Information Science, Iwate Prefectural University, Japan
y-saito@iwate-pu.ac.jp

Abstract - In 360-degree internet live broadcasting using omnidirectional cameras, broadcasters can communicate with viewers in real time, and viewers can freely view images of the area surrounding the broadcaster in 360 degrees. On the other hand, there is an issue in 360-degree internet live broadcasting where broadcasters cannot determine the Point Of View (POV) which is viewing direction of viewers. This leads to communication errors between broadcasters and viewers. To address this issue, we previously proposed a system using Mixed Reality (MR) technologies and POV heatmaps, which partially resolved the problem. However, since the display position of the POV heatmap is fixed, the burden on broadcasters remains unchanged. In this paper, we propose an MR-based broadcaster-following POV heatmap system to further reduce the broadcaster's burden in 360-degree internet live broadcasting. By displaying the POV heatmap in the real space using MR and having it follow the broadcaster, we expect to reduce the broadcaster's burden of confirming the viewers' POV. This paper reports on the implementation and evaluation of a prototype system.

Keywords: Mixed Reality, 360-degree Internet Live Broadcasting, POV, Heatmap

1 INTRODUCTION

In recent years, 360-degree videos, which use omnidirectional cameras to capture images of the entire 360-degree surroundings, have been gaining attention. As a result, 360-degree internet live broadcasting, which combines 360-degree video with internet live broadcasting, has emerged. In 360-degree internet live broadcasting, broadcasters can communicate in real time with viewers, who can enjoy a 360-degree view of the broadcaster's surroundings. Viewers can freely change their viewing direction. As a result, compared to traditional webcam-based broadcasting, broadcasters find it difficult to track viewers' viewing directions, leading to challenges in communication. The viewer's viewing direction in 360-degree internet live broadcasting will be referred to as Point Of View (POV) hereafter in this paper.

To address this issue, we implemented a system using heatmap in Mixed Reality (MR) devices, as a previous study [1], to place POV heatmaps in the MR space and reduce the broadcaster's workload. While the previous study successfully reduced the broadcaster's workload to some extent, the fixed position of the heatmap remained a challenge, as it still imposed a significant burden on the broadcaster. Therefore, a more effective method to further reduce the broadcaster's workload is necessary.

In this study, we propose a method to reduce the broadcaster's workload in 360-degree internet live broadcasting by using MR devices to display POV heatmaps that cover the broadcaster. MR, also known as mixed reality, is a technology that creates a space where the real and virtual worlds interact with each other. MR devices are transparent head-mounted displays that project the MR space. By displaying the POV heatmap in real space and tracking it in real time, this approach will address the issues.

The rest of this paper is organized as follows. Section 2 describes related work and our previous work of the MR spherical POV heatmap. Section 3 describes an overview of the proposed system and the implementation of the prototype system. It also describes a preliminary experiment for performance verification of the prototype system and system improvements to resolve issues identified in the preliminary experiment. Section 4 shows evaluation results using the improved system and discusses the evaluation results. Section 5 summarizes this study.

2 RELATED WORK

2.1 Information sharing using MR

Benko [2] implemented a system called VITA (Visual Interaction Tool for Archaeology) to visualize archaeological excavation sites. Using VITA, data can be easily obtained and spatial and temporal characteristics can be visually represented, making it highly effective for analysis. This result suggests that MR can significantly reduce the time required for data classification and analysis.

Gun [3] sought to enhance the collaborative experience of live panoramas using MR technology. To this end, they implemented SharedSphere, a wearable MR remote collaboration system. This system reduces work time and errors, enabling efficient overlapping work with virtual objects.

Based on the results of these studies, it is considered that the use of MR in 360-degree Internet live broadcasting can reduce the effort and burden on broadcasters and shorten response times.

2.2 Heatmap display

Maurus [4] proposed a method for generating and visualizing realistic heatmaps of real-time 3D gaze data. Unlike existing methods, this method projects the gaze of tracked individuals onto each fragment of the scene and uses a shadow mapping algorithm to account for occlusions, thereby achieving accurate visualization of the scene. This enables accurate visualization of the perceived scene regardless of the complexity of the underlying geometry.

Blignaut [5] proposed an algorithm that allows users to set the range of peripheral vision to be considered and adjust the measurement method and weight of fixations. This algorithm allows users to set the decay of gaze in a linear scale, Gaussian function, or no decay, and allows adjustment of the red threshold and transparency of the heatmap. Additionally, a method for visualizing areas of equal gaze concentration by adding contour lines has been demonstrated.

By visualizing viewers' POV using the heatmap, areas with concentrated POV can be shown as high heat levels (e.g., red), while areas with fewer POVs can be shown as low heat levels (e.g., green or yellow). This enables broadcasters to intuitively grasp viewers' POV.

Li [6, 7] demonstrated that spherical displays are superior to 2D flat displays in remote collaboration. Specifically, it was reported that spherical displays have less distortion, improve 3D spatial recognition, and offer advantages in terms of directional relationships and telepresence. Additionally, they demonstrated that spherical displays mitigate visibility challenges in remote observation and provide intuitive and effective support for spatial recognition and gaze guidance.

Oyekoya [8] demonstrated that spherical displays are visible from all directions compared to flat displays, and that they enhance telepresence by accurately conveying the gaze direction of remote participants. In particular, the display methods in Inflated mode and Normal mode were experimentally confirmed to have high gaze tracking accuracy and are effective for remote collaboration.

These characteristics are also considered effective for live broadcasting using omnidirectional cameras.

2.3 Our previous study

We discovered that in 360-degree internet live broadcasting, broadcasters are unable to determine the viewing direction of viewers, resulting in communication errors between broadcasters and viewers and preventing smooth communication. It is assumed that the area where viewers are directing their POV contains objects of interest to them. The viewers' POV represents information that encompasses their interests. In communication, the direction in which the communication partner is looking—i.e., where their gaze is directed—clearly indicates their interests, concerns, and the current focus of the conversation.

To address this issue, previous research [1] implemented a system that visualizes viewers' POVs in MR space using heatmap. This system overlays a heatmap on a spherical image captured by an omnidirectional camera, enabling the visualization of POV distribution. By using an MR device with the POV heatmap application installed, a spherical heatmap like Figure 1 is displayed in the MR space. Broadcasters can intuitively grasp the viewers' POV by confirming this heatmap while conducting a 360-degree internet live broadcasting.

We conducted an evaluation experiment comparing two systems: an MR-based POV heatmap system and an AR-based POV heatmap system proposed by Takada [9]. The purpose of the evaluation experiment was to determine whether the MR-based POV heatmap system could reduce the broadcaster's workload. The expected effects were: "reduction in the time required to complete pronouns or omitted words used in comments sent by viewers, and improvement in accuracy," and smoother communication using POV." To confirm these effects, the evaluation measured response times in communication between broadcasters and viewers, analyzed communication content, and conducted a subjective evaluation of comment transmission accuracy through a questionnaire. The experiment involved casual conversation about objects prepared in advance in a room. The results showed that the system using MR-based POV heatmap reduced the time it took for broadcasters to respond, thereby reducing the burden on broadcasters.

However, in the previous study, the display position of the POV heatmap was fixed, which made it difficult for broadcasters to check the heatmap. Since the position was fixed, broadcasters had to move themselves to check the heatmap. Additionally, it was necessary to compare the heatmap with the target object, which added to the burden on broadcasters. Based on the above results, it is necessary to develop methods to reduce the burden on broadcasters in 360-degree internet live broadcasting.

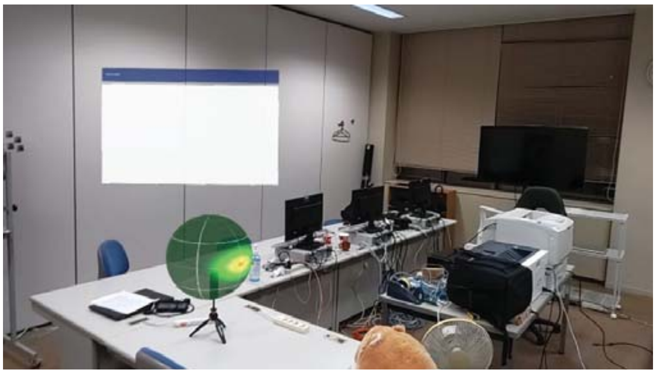


Figure 1: Previous system

3 PROPOSED SYSTEM

In the previous system, the broadcaster was burdened with the task of checking the viewers' POV. This is because the display position of the POV heatmap implemented using MR is fixed, requiring the broadcaster to move to check it. In addition, the POV heatmap and the object in real space must be compared with each other. To solve this problem, we propose a system that displays a broadcaster-following POV heatmap in MR space using an MR device so that it covers the broadcaster. This system reduces the burden of checking the POV heatmap, which has been an issue in previous studies.

3.1 System Architecture

In this study, we used Microsoft's HMD-type HoloLens 2 as the MR device and RICOH THETA V as the omnidirectional camera. For the broadcasting system, an existing 360-degree internet live broadcasting system [1] was utilized and operated on a web browser. During broadcasting, the RICOH THETA V is connected to a notebook computer in live mode to stream 360-degree video. Similarly, the viewer client is also an existing system, and the POV of viewers watching the 360-degree live stream on the web server is obtained, and this POV information is sent to the application running on the HoloLens 2. Heatmap is displayed at corresponding locations on the sphere, and the broadcaster confirms them through the HoloLens 2. Figure 2 shows the system architecture. The red frames and words indicate the new parts implemented in this study.

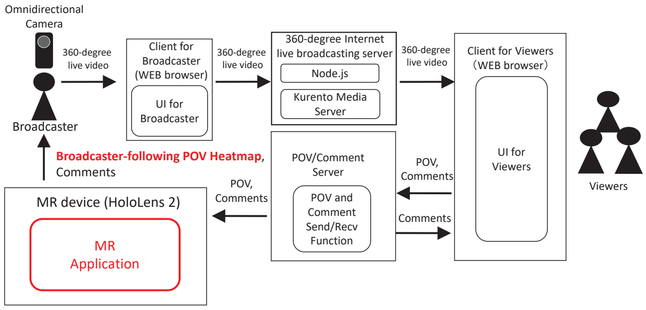


Figure 2: System architecture of the proposed system

There are two issues in implementing this system. The first is that it is necessary to update the position of the heat map displayed on the sphere depending on the position of the broadcaster. The second is that when viewers face the broadcaster, the broadcaster cannot detect this using only the POV heatmap.

To solve the first problem, we implemented a solution using Raycast. Raycast is a function that shoots a transparent ray from a specified location and obtains information about objects that the ray hits. In this system, we first convert the coordinates of the viewer's location from two-dimensional coordinates to three-dimensional coordinates, then shoot a Raycast from the origin toward those coordinates to obtain the collision point. Next, a second Raycast is fired from the collision point toward the broadcaster to detect the collision point with the sphere displayed around the broadcaster. Finally, the second collision point is converted from 3D coordinates to 2D coordinates to determine the display position of the heatmap. This allows the heatmap's display position to be updated in real-time by continuously referencing the broadcaster's position. Figure 3 shows the implemented POV heatmap. It can be confirmed that the viewer is looking at a hamster stuffed toy.

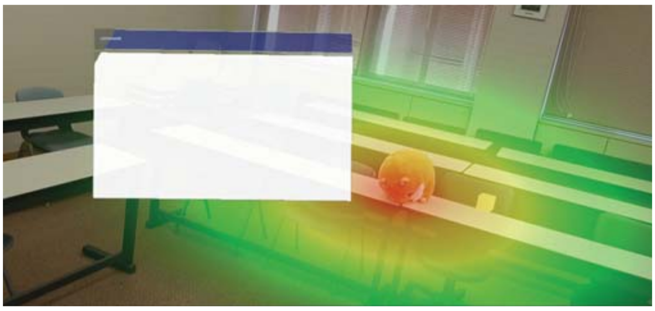


Figure 3: POV heatmap of the prototype system

To solve the second problem, we added an arrow function. This displays an arrow pointing toward the broadcaster when the Raycast hits the broadcaster. This allows the broadcaster to determine whether the viewing direction is pointed toward themselves by checking the presence or absence of the arrow when the viewing direction is directed toward themselves. To implement the arrow function, it was necessary to recognize the broadcaster in the MR space. However, due to the specifications of HoloLens 2, it was not possible to distinguish between the broadcaster and other objects. Therefore, we added an object for determining the broadcaster's position at the broadcaster's location. When the first Raycast, based on the viewer's viewing direction, hits this object, an arrow is displayed. This effectively causes the arrow to appear when the broadcaster is being viewed. At this point, the color of the arrow changes from green to yellow to red depending on the percentage of viewers facing the broadcaster. This allows the broadcaster's attention level to be judged by color, similar to a heatmap. Additionally, there is a possibility that the arrow may be displayed unintentionally when the broadcaster moves and accidentally hits the Raycast. To address this, we display the heatmap at the position where the Raycast would have hit if the broadcaster had not been there, simultaneously with the arrow. The implemented arrow is shown in Figure 4. This figure shows a broadcast with three viewers. Since two viewers are watching the broadcaster, yellow arrows are displayed.

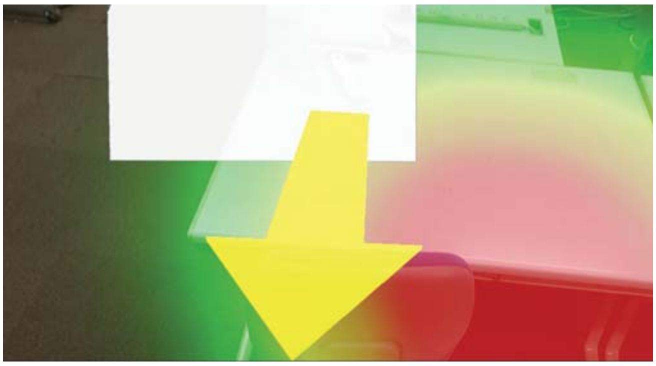


Figure 4: Arrow function of the prototype system

3.2 Preliminary Experiment

A preliminary experiment was conducted using three systems: the system in the previous study, the proposed system with the arrow function, and the proposed system without the arrow function. The purpose of the experiment was to confirm whether the system implemented in this study reduced the burden on broadcasters in understanding the viewers' POV compared to the system from the previous study. In addition, the experiment aimed to confirm whether the arrow function was useful when it was directed toward the broadcaster by the viewers.

Each experiment was conducted with one broadcaster and three viewers, for a total of six experiments. In the experiment, the broadcaster and viewers performed tasks in separate rooms. The equipment used in the experiment consisted of one notebook PC for the broadcaster, three notebook PCs for the viewers, one notebook PC for the server, a RICOH THETA V omnidirectional camera, and HoloLens 2 as an MR device. The broadcast content was casual conversation about objects prepared in advance and placed in the room. The objects included three types—“laptop PC, game console, and book”—with varying conditions such as color and purpose. Three of each type were prepared and placed in different locations so that objects of the same type were not visible simultaneously. Additionally, six types of stuffed animals were placed in the room.

The viewers were asked to send comments containing pronouns every 2 minutes for a total of 4 times as communication from viewers to broadcasters. Viewers sent comments containing the type of object and pronouns, such as “What is this game console?”. At this point, the comments to be sent and the objects to be viewed are specified in the procedure manual, and the timing of transmission is indicated in the chat by sending the instruction “Please send the first comment.” The broadcaster determines which object the comment refers to and responds to the viewer.

As communication from the broadcaster to the viewers, point out an object every two minutes, twice. The broadcaster points to the designated object and instructs the viewers, “Everyone, please look at this.” At this point, the broadcaster engages in casual conversation about the object that received focused attention when the viewers' POVs converged. When the broadcaster is aware of the viewers' viewing direction, it is expected that casual conversation will begin at the same time that the viewers' POVs converge. Furthermore, the broadcaster also asks viewers to pay attention to the broadcasters themselves every two minutes at twice changing standing position to check the arrow function

3.3 Results and Issues

As a result of the experiment, it was found that the system in the previous study was easier to understand in terms of the clarity of the heatmap. This is thought to be because, in the proposed system, the heatmap is displayed over a wide area, including areas outside the broadcaster's field of view, making it difficult to find the heatmap. Furthermore, it was found that circles of the heatmap occupy a large portion of the field of view in the proposal system, which affects the broadcaster's awareness of the surrounding situation and increases their workload. In addition, the large size of the circles of the heatmap made it difficult for the broadcaster to identify the objects the viewers were looking at.

Regarding the effectiveness of the arrow function in the proposal system, it was found that the broadcaster can easily understand that viewers are looking in their direction. On the other hand, since the arrows were displayed in the direction of the omnidirectional camera, it was found that the broadcaster could not notice when facing other directions.

3.4 System Improvements

Directional indicators for eye guidance

One issue with the proposal system was that it was burdensome for broadcasters to detect circles of heatmap outside their field of view.

Bork [10] demonstrated that in an MR environment, using gaze guidance technology for navigation to a target enabled users to quickly detect objects outside their field of view and reduce the time required to collect them. Burigat [11] compared visualization methods for objects outside the field of view on mobile devices and demonstrated arrow-based directional indicators.

Based on these studies, we added directional indicators as an additional feature to the proposed system. The directional indicators guide the broadcaster's gaze to the heatmap displayed outside the field of view. Figure 5 shows the directional indicators that were additionally implemented. When the heatmap is displayed outside the broadcaster’s field of view, directional indicators are displayed pointing toward the heatmap, and when it is within the field of view, they are hidden.

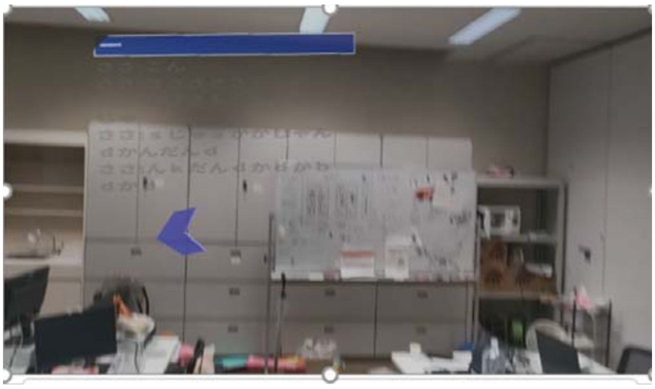


Figure 5: Directional indicators

Adjustment of the circle size of heatmap for accurate recognition of objects

The results of the preliminary experiment revealed that the heatmap was displayed too large and directed toward multiple objects, making it difficult to determine which object the heatmap was indicating. To address this issue, adjustments were made to the size of the heatmap.

We conducted an experiment assuming distances of 3 m and 1 m between the broadcaster and the target object. The size of the heatmap in the prototype system of the preliminary experiment was measured to be approximately 33 cm in diameter. In the experiment, four systems were used, with the size of the heatmap in the prototype system set to 1, and the sizes of the other systems set to 1, 0.75, 0.5, and 0.25. To eliminate differences in familiarity, the order in which the systems were used was changed for each broadcaster.

As a result of the experiment, we found that the appropriate heat map size is 0.25 (approximately 8 cm in diameter), so we changed the size of the heatmap in the prototype system. Figure 6 shows a comparison between the original size and the changed size.

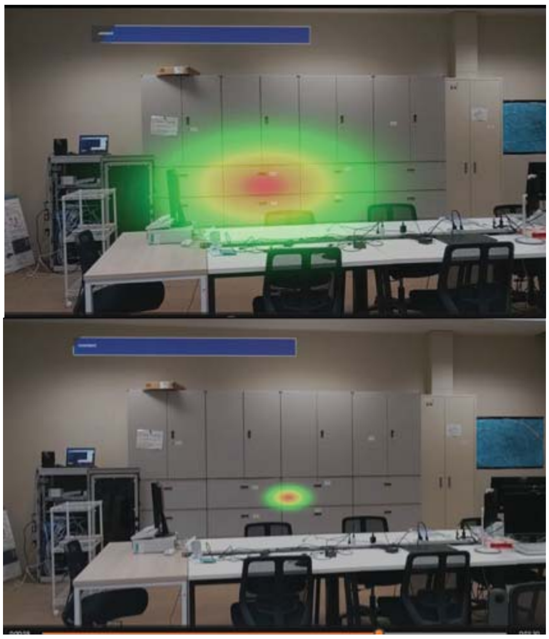


Figure 6: Circle size of heatmap (The upper figure is the original, and the lower figure is the adjusted version.)

- **Adjustment the arrow display position**

We made changes to always display the arrow within the field of view. In the prototype system, the arrow is displayed toward the broadcaster when the camera is facing the broadcaster. Therefore, the broadcaster must look at the camera to confirm the arrow, and if the broadcaster is not looking at the camera, they may not notice when the arrow is displayed toward them. To address this issue, we improved the system so that the arrow is always displayed within the broadcaster's field of view. This allows broadcasters to immediately notice the arrow display even when not looking at the camera, enabling more flexible broadcasting. For implementation, we modified the existing arrows to always follow the field of view on the HoloLens. Additionally, we adjusted the direction of the arrows so that their tips point toward the broadcaster. This adjustment enables broadcasters to notice the audience's gaze even when sitting in a chair, for example.

4 EVALUATION

We conducted evaluation experiments using three systems: the previous system, the proposed system before improvement (pre-improvement system), and the proposed system after improvement (post-improvement system). The purpose of the experiment was to confirm whether the post-improvement system reduced the burden on broadcasters in understanding viewers' POV compared to the pre-improvement system and the previous system.

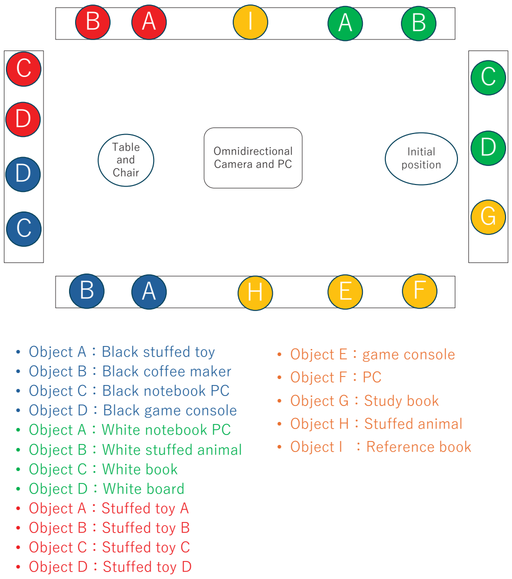

B
A
I
A
B
C
D
D
C
Table and Chair
Omnidirectional Camera and PC
Initial position
C
D
G
B
A
H
E
F
• Object A：Black stuffed toy
• Object B：Black coffee maker
• Object C：Black notebook PC
• Object D：Black game console
• Object A：White notebook PC
• Object B：White stuffed animal
• Object C：White book
• Object D：White board
• Object A：Stuffed toy A
• Object B：Stuffed toy B
• Object C：Stuffed toy C
• Object D：Stuffed toy D
• Object E：game console
• Object F：PC
• Object G：Study book
• Object H：Stuffed animal
• Object I ：Reference book

Figure 7: Evaluation environment

4.1 Environment

Each experiment was conducted with one broadcaster and three viewers, for a total of six experiments. The equipment used in the experiment consisted of one notebook PC for the broadcaster, three notebook PCs for the viewers, one notebook PC for the server, a RICOH THETA V as an omnidirectional camera, and HoloLens 2 as an MR device. The broadcast content was casual conversation about objects prepared in advance and placed in the room. The objects were prepared in three categories—“black objects,” “stuffed animals,” and “white objects”—with four objects of each category sharing the same conditions. These objects were arranged in such a way that objects of the same category were visible simultaneously within the field of view. Additionally, six other objects were placed in the room. The layout of the room with the objects is shown in Figure 7.

4.2 Tasks

In the evaluation experiment, broadcasters were asked to perform two tasks. Task 1 involved tasks related to communication from viewers to broadcasters, while Tasks 2 involved tasks related to broadcasters communicating with viewers. A total of three tasks were performed in a single broadcast, and this was repeated once for each of the three systems. To eliminate differences in familiarity with the systems, the order in which the systems were used was changed for each broadcaster.

Before starting the tasks, broadcasters were given an initial broadcast operation confirmation period to familiarize themselves with the system. Once the broadcasters were confirmed to be familiar with the system, the actual broadcast began. As preparation before the start of the tasks, broadcasters were asked to move to their initial positions and face forward. They were then instructed to remain in place until a specified comment was received.

In Task 1, viewers were asked to send comments containing pronouns indicating the type of object being referred to every two minutes, four times in total. As communication from viewers to the broadcaster, the viewers send comments containing the type of object and pronouns, such as “Instruction: What is this black thing?”. The viewers send comments with “Instruction:” at the beginning so that it is clear that the comment is for the task. The comments to be sent and the objects to be faced are specified in the procedure manual, and the timing of sending is indicated in the chat with “Please send the first comment.”. Broadcasters determine which object the comment refers to and respond to the viewer. If the broadcaster understands what the viewer is commenting on, the response time is expected to be shorter. The time taken for the broadcaster to provide the correct response to the comment is measured, and the broadcaster is asked to provide a subjective evaluation via a survey regarding how well the comment was understood or perceived. In this study, the time it takes for the broadcaster to provide a correct response to a comment is referred to as “consideration time”. If consideration time increases, the response to the viewer's comment is delayed, which may lead to a decrease in viewer satisfaction. Conversely, if consideration time is short, the response to the viewer is faster, which is expected to improve viewer satisfaction. Additionally, by comparing consideration time across different systems, it is possible to determine whether the broadcaster's workload has been reduced.

In Task 2, the broadcaster sits in a designated chair and asks the viewers to focus their attention on him or her. he broadcaster then asks the viewers to draw a picture related to the theme on a worksheet in four minutes. To help the broadcaster concentrate on the task, he or she is instructed to ask the viewers to grade the completed drawings and to aim for the highest possible score. In addition, the broadcaster is instructed to check the viewers' gaze from time to time while working and to ask them to focus their attention again if anyone is not looking. The viewers are instructed to avert their gaze from the broadcaster every 60 seconds after the task begins. One viewer averts their gaze after 60 seconds, two after 120 seconds, and three after 180 seconds. At this point, the time it takes for viewers who have averted their gaze to notice that the broadcaster is not looking at them is measured.

4.3 Results of Task 1

The results of the evaluation for Task 1 are shown in Table 1. Focusing on the average time, the previous system took 19.8 seconds, the pre-improvement system took 36.8 seconds, and the post-improvement system took 22.9 seconds. The Tukey-Kramer's test showed no significant difference. One possible reason for this result is that the post-improvement system has a smaller heatmap, so even if the heatmap is displayed slightly out of position, it can be difficult to make a judgment.

The results of the error count for Task 1 show that the error count was 2 in the previous system, and 5 in both the pre-improvement and post-improvement systems. The results of the Tukey-Kramer's test for the error count showed no significant difference.

The results of the questionnaire regarding ease of understanding were as follows: the previous system scored 3.7 points, the pre-improvement system scored 2.9 points, and the post-improvement system scored 4.3 points. The Tukey-Kramer's test showed a significant difference between the pre-improvement system and the post-improved system (***p*** = 0.002). These results indicate that the post-improvement system is easier to understand than the pre-improvement system due to the reduction in the size of the heatmap, and that it is as easy to understand as the previous system.

Table 1: The results for Task 1

	Consideration time (avg.)	Error count (total)	Questionnaire score (avg.)
Previous	19.8 sec	2	3.7
Pre-improvement	36.9 sec	5	2.9
Post-improvement	22.9 sec	5	4.3

4.4 Results of Task 2

The results for Task 2 are shown in Table 2. The results of the waiting time for the broadcaster's own attention instruction were 14.5 seconds for the previous system, 8.8 seconds for the pre-improvement system, and 8.6 seconds for the post-improvement system. The Tukey-Kramer's test showed no significant difference. This is likely because when giving attention instruction to oneself, it is necessary to look at the camera, so there was no difference between the systems.

The results of the awareness time when viewers diverted their gaze from the broadcaster during the task until the broadcaster noticed it were 39.2 seconds for the previous system, 35.5 seconds for the pre-improvement system, and 24.2 seconds for the post-improvement system. Note that if the broadcaster did not notice that the viewer had looked away during the task, the time was recorded as 60 seconds. The Tukey-Kramer's test revealed significant differences between the previous system and the post-improvement system (***p*** = 0.007) and between the pre-improvement and post-improvement systems (***p*** = 0.007). These results suggest that the post-improvement system is more suitable for confirming the viewer's gaze during the task. One factor contributing to these results is that in the improved system, the broadcaster can always confirm whether the viewer's gaze is directed toward them by an arrow displayed on the screen. In contrast, in the pre-improvement system, the arrow was displayed pointing from the broadcaster toward the camera.

Therefore, it is difficult to track the viewer's gaze during the task. Furthermore, in the previous system, as in the pre-improvement system, when checking the viewer's gaze during the task, it is necessary to look at the camera and confirm the spherical heatmap. Additionally, due to the specifications of the heatmap, when the number of people in the heatmap changes from three to two, it is difficult to notice the change in the size of the heatmap. Furthermore, in the previous system, when the heatmap is displayed on the back side from the broadcaster's perspective, the broadcaster must move to the back side to confirm it. For these reasons, it is difficult to track the viewers' gaze during the task even in the previous system. Therefore, it is considered that such results were obtained.

As a questionnaire related to Task 2, we asked broadcasters to rate on a 5-point scale how easy it was to understand that viewers were watching them while they were working, and to give reasons for their answers. The results were 3.3 points for the previous system, 3.3 points for the pre-improvement system, and 4.7 points for the post-improvement system. The Tukey-Kramer's test showed no significant difference. Regarding the reasons for their answers, many respondents said that it was troublesome to check the camera to determine where the viewers were looking in the previous system. Regarding the pre-improvement system, many respondents said that it was easy to determine the number of viewers by the color of the arrow. On the other hand, some respondents said that the arrow was displayed a little too low and that they forgot to pay attention to the arrow when concentrating on their work. Finally, regarding the post-improvement system, there were opinions that the arrow being within the field of view made it easier to notice color changes during tasks. Additionally, there were opinions that displaying the indicators in a different direction made it easier to notice changes in the viewers' POV.

Table 2: The results for Task 3

	Waiting time at starting time (avg.)	Awareness time (avg.)	Questionnaire score (avg.)
Previous	14.5 sec	39.2 sec	3.3
Pre-improvement	8.8 sec	35.5 sec	3.3
Post-improvement	8.6 sec	24.2 sec	4.7

5 CONCLUSION

In this study, we proposed a method to reduce the burden on broadcasters in 360-degree internet live broadcasting. As the proposed system, we proposed a system that displays a broadcaster-following POV heatmap in MR space and displays an arrow when the viewer's gaze is directed toward the broadcaster. By comparing this system with previous study, we clarified its advantages and challenges. The evaluation experiment revealed two major findings. First, a function that allows broadcasters to recognize when viewers are looking in their direction is necessary for the broadcaster-following POV heatmap. The arrow function implemented in this study was found to be useful in alerting broadcasters when viewers are looking in their direction. Second, the broadcaster-following POV heatmap needs to be improved. Compared to the previous system, the proposed system made the POV heatmap less easy to understand, which increased the burden on the broadcaster. A future challenge is to improve the positional accuracy of the heatmap display.

REFERENCES

[1] Yoshia Saito and Junichiro Suto: A Spherical POV Heatmap using Mixed Reality in 360-degree Internet Live Broadcasting, International Workshop on Informatics (IWIN), pp. 215-218 (2023).

[2] Hrvoje Benko, Edward W. Ishak, Steven Feiner: Collaborative Mixed Reality Visualization of an Archaeological Excavation, Proceedings of the 3rd IEEE/ACM International Symposium on Mixed and Augmented Reality, pp.132-140 (2004).

[3] Gun A. Lee, Theophilus Teo, Seungwon Kim, Mark Billinghurst: Mixed reality collaboration through sharing a live panorama, SIGGRAPH Asia 2017 Mobile Graphics & Interactive Applications, No. 14, pp. 1-4 (2017).

[4] Michael Maurus, Jan Hendrik Hammer, Jürgen Beyerer: Realistic heatmap visualization for interactive analysis of 3D gaze data, ETRA '14 : Proceedings of the Symposium on Eye Tracking Research and Applications, pp. 295-298 (2014).

[5] Pieter Blignaut: Visual span and other parameters for the generation of heatmaps, Proceedings of the Symposium on Eye-Tracking Research & Applications 2010, pp. 125-128 (2010).

[6] Zhengqing Li, Shio Miyafuji, Toshiki Sato, Hideki Koike, Naomi Yamashita, Hideaki Kuzuoka: How Display Shapes Affect 360-Degree Panoramic Video Communication, Proceedings of the 2018 Designing Interactive Systems Conference, pp. 845-856 (2018).

[7] Zhengqing Li, Shio Miyafuji, Erwin Wu, Hideaki Kuzuoka, Naomi Yamashita, Hideki Koike: Omni-Globe: An Interactive I/O System For Symmetric 360-Degree Video Communication, DIS '19: Proceedings of the 2019 on Designing Interactive Systems Conference, pp. 1427-1438 (2019).

[8] Oyewole Oyekoya, William Steptoe, Anthony Steed: SphereAvatar: a situated display to represent a remote collaborator, Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 2551-2560 (2012).

[9] Masaya Takada, Dai Nishioka and Yoshia Saito: Proposal of the Viewer's "POV Heat Map" in 360-degree Internet Live Broadcasting, IPSJ Journal, Vol.62, No.1, pp. 145–159 (2021) (in Japanese).

[10] Felix Bork, Christian Schnelzer, Ulrich Eck, and Nassir Navab: Towards Efficient Visual Guidance in Limited Field-of-View Head-Mounted Displays, IEEE transactions on visualization and computer graphics, Vol.24, No.11, pp. 2983-2992 (2018).

[11] S. Burigat, L. Chittaro, and S. Gabrielli: Visualizing locations of off-screen objects on mobile devices: A comparative evaluation of three approaches, In Proceedings of the 8th Conference on Human-computer Interaction with Mobile Devices and Services, MobileHCI '06, pp. 239-246 (2006).

Session 5:
Smart Cities and Transportation

(Chair: Masashi Saito)

Real-Time Position Estimation of Non-Line-of-Sight Obstacles Using Reflected Images on Reflective Surfaces

BaiLi Sheng *, Yusuke Takatori **

* Graduate School of Engineering, Course of Electrical and Electronic Engineering
Faculty of Engineering, Japan
** Department of Electrical and Electronic, Kanagawa Institute of Technology, Japan
sei24@ele.kanagawa-it.ac.jp, takatori@ele.kanagawa-it.ac.jp

Abstract- This paper presents a real-time system for estimating the position of non-line-of-sight (NLOS) obstacles using reflected images. Based on previous work with stereo cameras and reflective surfaces such as vehicle panels and roadside windows, a complete system was implemented and evaluated on both mirror and painted surfaces. The method estimates NLOS vehicle positions from reflected image displacement. Experiments achieved an average processing time of 56.7 ms for mirrors and 100 ms for painted surfaces, satisfying real-time requirements. The average X-axis errors were 32 mm and 61 mm, respectively, confirming practical feasibility even under low reflectivity.

Keywords: ITS; obstacle of non-line of sight; obstacle position estimation; real time

1 INTRODUCTION

Intelligent Transportation Systems (ITS) enhance traffic safety and mobility through information and communication technologies. Obstacle detection using sensors such as LiDAR, millimeter-wave radar, and stereo cameras has been widely studied, but these methods are limited to detecting only Line-of-Sight (LOS) obstacles.

To overcome this, Vehicle Information Sharing (VIS) using V2V and V2R communication has been studied. V2R shares obstacle data via Roadside Units (RSUs) [1], while V2V enables direct vehicle-to-vehicle exchange. However, these methods are constrained by RSU range and communication unit penetration [2][3].

Besides communication-based methods, several sensing approaches have also been explored to detect NLOS obstacles. Millimeter-wave radar detects preceding vehicles but suffers from limited angular resolution [4]. Doppler radar has been applied to blind-spot human detection [5][6][7], and mirror-based reflection methods have been tested without position estimation [8]. Stereo cameras, by contrast, can measure distance and recognize object types [9]. Although conventional cameras only capture LOS information, reflective surfaces—such as vehicle panels or roadside windows—can reveal NLOS obstacles [10][11][12].

This paper presents a system for estimating the position of NLOS obstacles by analyzing their virtual images reflected on mirror-like or painted surfaces using stereo vision. The proposed method employs a virtual "folding" technique based on reflective surface geometry and supports both real-time and image-based processing. This study focuses on flat reflective surfaces to clarify basic properties.

Section 2 introduces the NLOS estimation concept, Section 3 describes the real-time prototype, Sections 4 and 6 present evaluations using mirror and painted surfaces, and Section 7 concludes the paper.

2 NLOS OBSTACLE POSITION ESTIMATION FROM REFLECTED IMAGE[12]

2.1 Stereo Camera-Based Position Estimation

A stereo camera consists of two cameras placed side by side at a fixed distance to simultaneously capture two images. Due to the difference in the cameras' positions, there is a displacement in the position of the subject between the two images, which is referred to as "disparity." Disparity is calculated through stereo matching and serves as essential information for estimating the distance to an object. Fig.1 illustrates the process of obtaining disparity using a stereo camera.

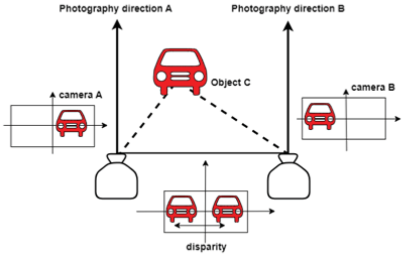


Figure 1: Position estimation using stereo cameras

2.2 Position Estimation Using Reflected Images on Reflective Surfaces

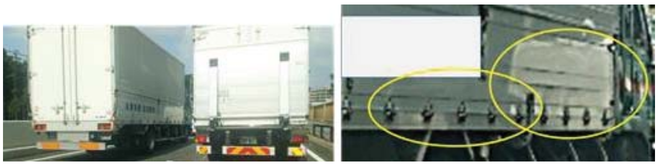


Figure 2: Position estimation for non-line-of-sight obstacles

In this study, we address the estimation of the position of NLOS obstacles by utilizing their reflections on surfaces such as the side panels of vehicles in adjacent lanes.

A method has been proposed using an on-board stereo camera, which estimates the position of obstacles based on their reflected images on mirror-like surfaces, including vehicle sides and roadside glass panels. As illustrated in Fig. 2, the position is estimated by analyzing the geometric relationship between the reflective surface and the reflected obstacle.

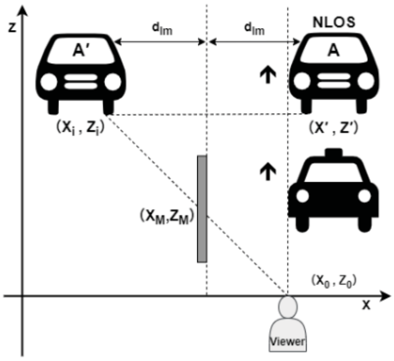


Figure 3: Position estimation of obstacles reflected on a reflective surface

As illustrated in Fig. 3, the position of the actual object was geometrically derived from its reflected image appearing on the side surface of a nearby vehicle.

In this model, the coordinates of the actual object (X',Z') are expressed as a linear function of the reflected image point (X_i,Z_i) based on the mirror-symmetry geometry, allowing direct estimation of the object's real-world position from stereo camera images.

3 CONSTRUCTION OF A BASIC REAL-TIME NLOS SYSTEM

In this study, before discussing the painted surface, this section focuses on the position estimation system using reflections on mirror-like surfaces.

The overall process of the proposed system is illustrated in Fig. 4. The input image is first converted to grayscale and flattened to reduce brightness variation. Then, the image data is transferred to the GPU, where Semi-Global Matching (SGM) [13] is performed to generate a disparity map. The processed data is subsequently sent back to the CPU, converted into 3D coordinates, and used to determine the position of the target object.

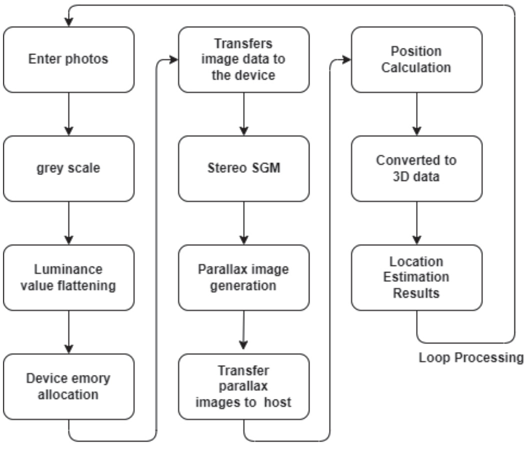

Enter photos
grey scale
Luminance value flattening
Device emory allocation
Transfers image data to the device
Stereo SGM
Parallax image generation
Transfer parallax images to host
Position Calculation
Converted to 3D data
Location Estimation Results
Loop Processing

Figure 4: Overview of the process flow

Many stereo matching methods have been proposed, including representative approaches such as Semi-Global Block Matching, Block Matching (BM), and Semi-Global Matching (SGM) [14]. In this study, SGM implemented in OpenCV is adopted as the matching method. SGM offers stable disparity estimation even in flat regions and effectively preserves object contours. Moreover, compared to BM, it is more robust against external factors such as lighting, allowing for more reliable matching performance.

4 EVALUATION OF REAL-TIME NLOS SYSTEM USING MIRROR SURFACES

4.1 Experimental Setup And Performance Evaluation of The Real-Time System

To evaluate the performance of the developed real-time position estimation system, an experiment was conducted in which the target object was continuously moved within a defined experimental range.

A schematic diagram of the assumed road environment is shown in Fig.5 In this scenario, two vehicles are present ahead of the ego vehicle, and the vehicle located two vehicles ahead is in a NLOS position. The system estimates the position of this NLOS vehicle by utilizing its reflection on the side panel of a vehicle in the adjacent lane. The direct distance to the target vehicle was assumed to be approximately 20 m, and the lateral distance about 2.5 m. In the experiment, measurements were conducted at 1/6 scale within the detectable range of the system.

For evaluation, the position estimation accuracy was verified by comparing three types of positions:

- **Direct estimated position**: the position of the Target Vehicle directly captured by the camera
- **Virtual estimated position**: the position of the Virtual Vehicle as it appears in the reflected image
- **Reflected estimated position**: the reflected image's position mirrored over the reflective surface to estimate the real position of the actual object

Specifically, the average and standard deviation of the positional differences between the Reflected Estimated Position and the Direct Estimated Position were evaluated. Furthermore, to assess processing performance, the average and standard deviation of the processing time required for position estimation were used as evaluation metrics.

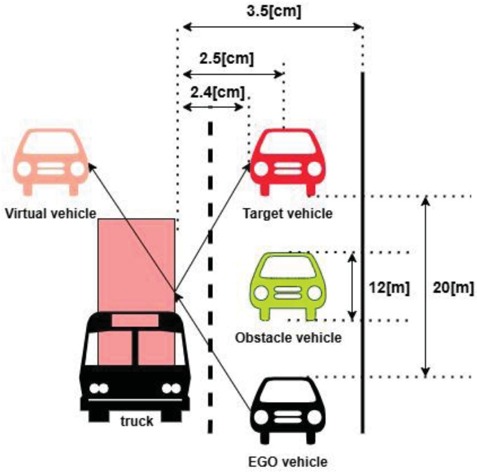


Figure 5: Assumed Road Environment

4.2 Processing Time Analysis

As the experimental method, a flat panel with a rear-view image of a vehicle affixed to it is used as the target object. To mount the panel, an electric model vehicle with adjustable speed is selected as the base. This model vehicle is capable of running at a constant speed along a rail. The panel is continuously captured by the stereo camera, and in each frame, the four corners of the panel equipped with LED light sources are extracted. The centroid of these points is then calculated. The position estimation is performed based on the coordinates of this centroid.

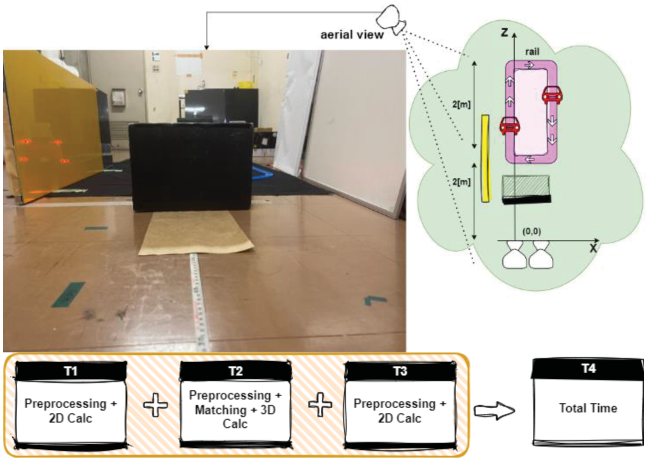


Figure 6: Experimental scene (top) and time for each section (bottom)

The rail on which the model runs is constructed as a loop track with both straight and curved sections, as shown in Fig. 6 (top). The straight section is placed at a depth of approximately 1 to 5 meters from the camera. During actual measurements, real-time position estimation is continuously performed as the model travels from 2 to 4 meters along this straight section (defined as one trip). The processing time for each estimation is measured during this interval. As shown in the image in Fig. 6 (bottom).

The total processing time (T_4) is divided into the following three stages:

- T_1: 2D image preprocessing and calculation of the object's center point
- T_2: 3D image preprocessing and calculation of spatial coordinates
- T_3: Position estimation through 3D data transformation

4.3 Evaluation Methodology For Position Estimation Accuracy

To evaluate the proposed system, experiments were conducted using mirror surfaces. The estimated position of the reflected object (assuming NLOS) was compared with the estimated position of the object directly in front (assuming LOS). To capture the virtual image reflected from the mirror surface, an electric model track was installed as in Section 4.1, and the mirror surface was positioned 0.4 meters to the left from the direction of the camera.

When estimating the position using only the NLOS vehicle, an obstacle was placed in front of the camera to block the direct view of the real object. One run was defined as the movement along the straight section of the rail from a distance of 2 meters to the end point at 4 meters. Ten measurements were recorded for both the direct estimated position, the reflected estimated position and virtual estimated position from the mirror surface. The experimental setup is illustrated in Fig. 7.

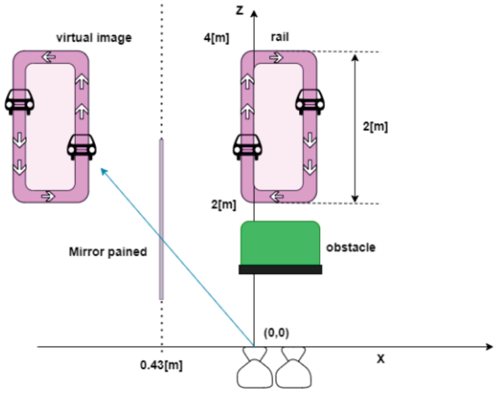


Figure. 7: Arrangement during real-time position estimation.

4.4 Device and PC specifications

In this study, the ZED 2i Stereo Camera developed by Stereo labs is used as the stereo camera device. The specifications of the ZED 2i Stereo Camera are shown in Table 1.

Table 1: Specifications of the ZED 2i Stereo Camera

Item	Specification
Baseline Distance	120 mm
Output Image Size	1280 × 720 (WXGA)
Focal Length	2.1 mm
Maximum Field of View	110° (H) × 70° (V) × 120° (D)

The PC used for this system is a laptop equipped with an Intel Core i7-12700H CPU and an NVIDIA GeForce RTX 3060 GPU (with 7.9 GB of dedicated memory).

4.5 Evaluation Results of Average Processing Speed

The average processing time for 10 runs was measured on the mirror surface using WXGA resolution. As shown in Table 2, the processing time per frame (T_4) is approximately 56.7 ms. According to Fig. 8, the processing speed is concentrated between 3 [ms] and 4 [ms], and within 97% cumulative relative frequency, the processing time remains stable within 4 [ms].

Table 2: Results of evaluation of average processing times and standard deviations

Results of evaluation		T1[ms]	T2[ms]	T3[ms]	T4[ms]＝T1+T2+T3
Mirror	Ave	10.65	42.08	4.02	56.7
	Stedv	0.352	0.233	0.337	0.712

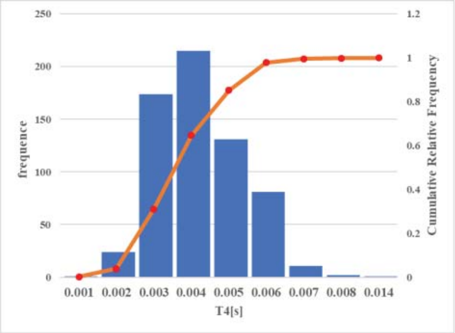


Figure.8: Frequency Distribution of Real-time Processing Time (T4) for Mirror Surface at WXGA Resolution

4.6 Evaluation Results of Real-Time Position Estimation Accuracy

In Fig.9, the virtual estimated positions derived from the reflected images appear near X = 800 mm, consistent with the expected mirror symmetry across the reflective surface placed at X = 400 mm.The direct estimated positions, which represent the real object's location observed without reflection, are distributed around X = 0 mm, indicating accurate stereo estimation. This symmetrical relationship confirms that the virtual positions are correctly mirrored with respect to the reflective surface, validating the system's geometric consistency.

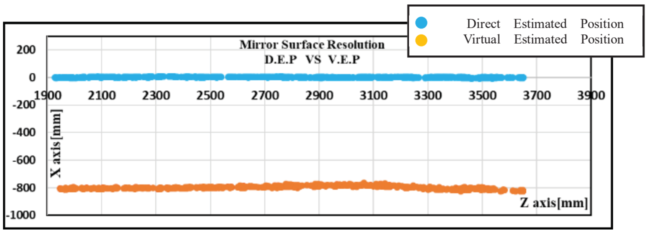


Figure 9: Mirror Surface-Direct vs. Virtual Estimated Position (Z-axis trajectory)

Fig.10 shows the relationship between the X-axis difference (ΔX) and the Z-axis distance for both the direct estimated positions and the reflected estimated positions over 10 trials. The results consistently exhibit the same trend across all trials, suggesting that the observed variations stem from differences in image appearance and matching accuracy rather than systematic error. Although slight fluctuations are observed at longer distances, the overall deviation remains small. The X-axis error (ΔX) is approximately 0.019 m on average, indicating high stability and robustness of the position estimation even in reflected scenarios.

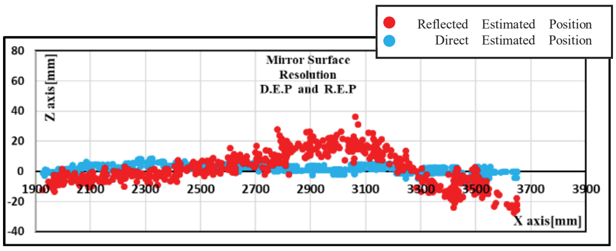


Figure 10: Mirror Surface-Direct vs. Reflected Estimated Position (X-axis fluctuation)

5 CONSTRUCTION OF A PAINTED SURFACE-COMPATIBLE SYSTEM

5.1 Process Flow of The Real-time System for Refrected Image for Painted Surface

This section, discussing the yellow painted surface, a real-time position estimation of NLOS obstacles reflected on painted surfaces around the vehicle is performed using an in-vehicle stereo camera. The painted surfaces referred to here are assumed to be the side panels of large vehicles such as trucks or passenger vehicles.

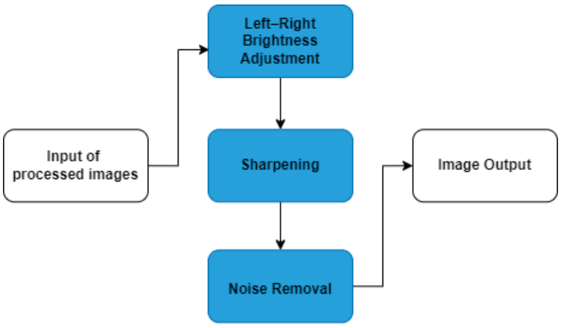


Figure. 12: Processing flow for images reflected on the reflective surface.

We propose a novel method for estimating the position of NLOS obstacles reflected on a painted surface. In the proposed approach, position estimation of the reflected image on the painted surface is performed by applying image enhancement processing after histogram equalization, as shown in Fig. 12. The overall processing flow is based on the flow already shown in Fig. 4. The method includes adjusting the brightness difference caused by the camera's position, sharpening the image, removing noise, and performing stereo matching.

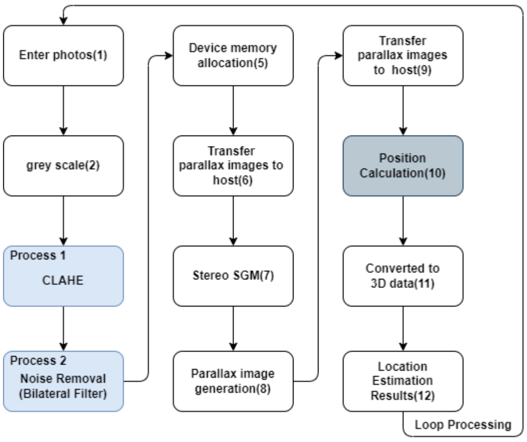

Enter photos(1)
grey scale(2)
Process 1
CLAHE
Process 2
Noise Removal (Bilateral Filter)
Device memory allocation(5)
Transfer parallax images to host(6)
Stereo SGM(7)
Parallax image generation(8)
Transfer parallax images to host(9)
Position Calculation(10)
Converted to 3D data(11)
Location Estimation Results(12)
Loop Processing

Figure 13: Overview of Image Processing on Painted Surfaces

Fig.13 shows an overview of the image processing on the painted surface. Based on the model configuration described in Section 4.1, the image processing following the grayscale conversion of the input image was implemented with reference to the method shown in Fig. 12. Due to the difference in reflection angles between the left and right cameras, the reflected images captured by each camera often exhibit brightness inconsistencies. To suppress the effect of these brightness differences, Contrast Limited Adaptive Histogram Equalization (CLAHE) [15] is applied to both the left and right grayscale images. This allows for more stable stereo matching performance under varying lighting conditions. By applying adaptive histogram equalization (CDF), the brightness and contrast of the image can be improved. This method especially enhances details in both dark and bright regions, resulting in a more visually balanced overall contrast in the image. Furthermore, for noise removal as shown in Fig.13, a bilateral filter is implemented. This filter was chosen because it can remove noise while preserving detailed features, and a GPU-compatible library is available in OpenCV. Examples of the image quality improvement achieved by each processing step are shown in Fig. 14.

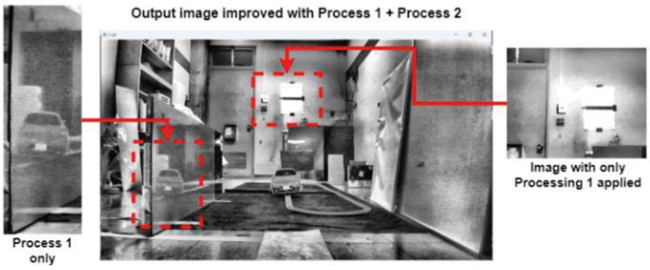

Output image improved with Process 1 + Process 2
Image with only Processing 1 applied
Process 1 only

Figure 14: Comparison of image processing before and after improvements

According to the model described in Section 4.2, unlike mirror-like surfaces, painted surfaces can exhibit significant outliers in disparity values during disparity map generation. To address this issue, the proposed system extracts the target measurement point from the captured image and, based on the disparity map obtained through matching, uses a 5×5 region centered around the target point. Within this region, the system removes outliers using the interquartile range (IQR) method and outlier detection techniques. It then calculates the average of the remaining disparity values and estimates the 3D coordinates (X, Y, and Z) of the pixel position based on the average disparity value.

6 EVALUATION OF REAL-TIME NLOS SYSTEM USING PAINTED SURFACES

The experiment was conducted under the same conditions as described in Section 4, with the mirror surface replaced by a painted surface. According to Table 3, the processing time per frame (T4) is approximately 100 [ms]. As shown in Fig. 15, the processing speed is concentrated between 0.1 and 0.105 seconds. Within a cumulative relative frequency of 95%, the processing time remains stable within 0.105 seconds.

Table 3: Results of evaluation of average processing times and standard deviations

Results of evaluation		T1[ms]	T2[ms]	T3[ms]	T4[ms]= T1+T2+T3
Painted	Ave	44.46	46.98	8.76	100.2
	Stedv	2.64	0.96	0.23	2.05

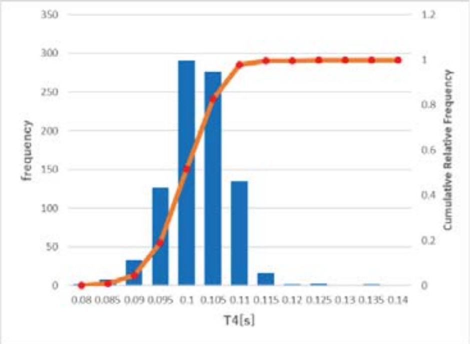


Figure.15: Real-time processing the (T4) with Painted surface WXGA resolution

6.1 Evaluation Results of Real-Time Position Estimation Accuracy

Fig.16 shows the spatial distribution of the direct estimated positions and virtual estimated positions derived from reflections on a painted surface. The direct positions are concentrated near X = 0 mm, while the virtual positions are symmetrically distributed around X = 800 mm, relative to the reflective plane at X = 400 mm. This demonstrates that the proposed method maintains geometric consistency even on painted surfaces with lower reflectivity, confirming that the reflection-based position estimation is theoretically sound.

However, compared to the mirror surface case (Fig.9), the data in Fig.16 shows greater spread and positional variance. This is primarily attributed to uneven disparity distributions caused by appearance inconsistencies during stereo matching on the painted surface. (In particular, some image regions fail to produce valid disparity values and are output as zeros, leading to local errors.)Although an interquartile range (IQR)-based filtering method was used to remove outliers, it may also exclude correctly matched pixels, further increasing the overall variance in position estimates.

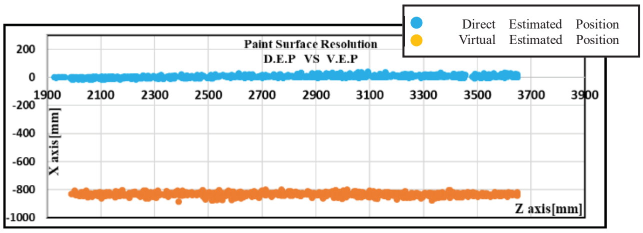


Figure 16: Painted Surface-Symmetry Analysis of Virtual vs. Direct Estimated Position (X-Axis)

Fig.17 shows the horizontal difference (ΔX) between the direct and reflected estimated positions over 10 trials on the painted surface. Compared to mirror reflections, the reflected positions on the painted surface exhibit larger and more scattered deviations due to irregularities in reflected appearance and lower stereo matching reliability. A quartile-based filtering method was applied to reduce the impact of outliers. However, this process also introduced minor inconsistencies in the direct estimates, as valid data may also be inadvertently removed. Despite these effects, the average ΔX error remained around 0.0352 m, demonstrating that reliable horizontal estimation is still achievable under low-reflectivity conditions.

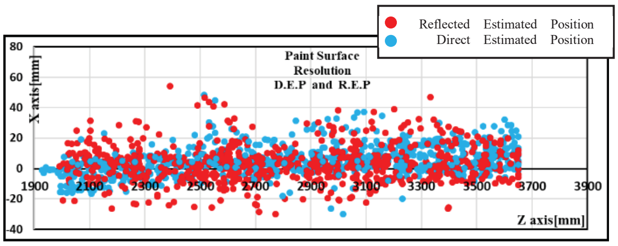


Figure 17: Painted Surface-X-Axis Deviation of Reflected vs. Direct Position

7 CONCLUSION

In this study, we developed a real-time system for estimating the position of non-line-of-sight (NLOS) objects using reflections from mirrors and painted surfaces. The system achieves real-time performance, processing stereo images about at 18 fps with an average time of 56.7 ms for mirrors and 100 ms for painted surfaces. The achieved processing speed meets or exceeds the 100 ms data acquisition cycle of typical in-vehicle sensors such as LiDAR.

The experimental results showed that the average X-axis error was 32 mm (SD: 20 mm) for mirrors and 61 mm (SD: 42 mm) for painted surfaces, indicating comparable accuracy despite differences in surface reflectivity. The Z-direction error remained stable within approximately 40 mm across distances between 1.7 m and 3.0 m.

Although flat reflective surfaces such as vehicle panels were assumed in this study as a baseline condition, prior research [16][17] has shown that curved or non-planar surfaces can introduce additional ranging errors due to geometric distortions. In practice, vehicle surfaces are rarely perfectly flat or stable in orientation, and these factors may affect estimation accuracy. Future work will therefore include LiDAR-assisted surface modeling or data-driven correction methods to mitigate these effects.

REFERENCES

[1] M. Tamura, S. Takahashi, S. Yasuhara, M. Kojima, K. Minegishi, “Development of intersection safety support systems using vehicle-to-vehicle communication”, 12th ITS World Congress 2005, CD-ROM (7 pages)

[2] Yusuke Takatori, Takaaki Hasegawa, “Quantitative Performance Evaluation of Predictive Collision Warning System based on Inter-Vehicle Communication” , International Journal of ITS Research, Vol.4, No.1, December 2006,

[3] Yusuke Takatori, Hideya Takeo, “Analysis of Vehicle Information Sharing Performance of an Intersection Collision Warning System”, IEICE TRANS. FUNDAMENTALS, VOL.E100–A, NO.2 FEBRUARY 2017.

[4] Noriko Shimomura, JP Patent 2005-182137, 2005.7.7.

[5] Sora Hayashi, Kenshi Saho,Hiroaki Okinaka,Masao Masugi, “Detection of human in a blind area using micro-Doppler radar”, IEICE Tech. Rep., vol. 119, no. 102, ITS2019-4, pp. 21-25, July 2019.

[6] Nicolas Scheiner, Florian Kraus, Fangyin Wei, Buu Phan, Fahim Mannan, Nils Appenrodt, Werner Ritter, Jürgen Dickmann, Klaus Dietmayer, Bernhard Sick, Felix Heide,"Seeing Around Street Corners: Non-Line-of-Sight Detection and Tracking In-the-Wild Using Doppler Radar," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2068-2077,2020.

[7] Yuanchao Feng, Shintaro Ono, Yoshihiro Suda, Noriaki Itagaki,” Recognition of Risky Events at Blind Spots on Roads using On-vehicle Camera and Road Safety Mirror,” IEICE Tech. Rep., vol. 120, no. 94, ITS2020-2, pp. 7-12, July 2020.

[8] Sora Hayashi, Kenshi Saho,Masao Masugi, “Classification of pedestrians existing in visible or blind areas using Doppler radar –An Approach Using Long Short-Term Memory --”, IEICE Tech. Rep., vol. 120, no. 291, ITS2020-25, pp. 121-125, Dec. 2020.

[9] K. Saneyoshi,” Drive assist system using stereo image recognition”, Proc. Of the 1996 IEEE Intelligent Vehicles Symposium, pp.230-235, 1996.

[10] D. B. Lindell, G. Wetzstein and V. Koltun, "Acoustic Non-Line-Of-Sight Imaging," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 2019, pp. 6773-6782, doi: 10.1109/CVPR.2019.00694.

[11] Felix Heide, Matthew O’Toole, Kai Zang, David B Lindell,Steven Diamond, and Gordon Wetzstein. Non-line-of-sightimaging with partial occluders and surface normals. ACMTransactions on Graphics, 38(3):1–10, 2019. 1, 2

[12] Y. Takatori, "NLOS Obstacle Position Estimation from Reflected Image," 2020 IEEE Intelligent Vehicles Symposium (IV), Las Vegas, NV, USA, 2020, pp. 1265-1270, doi: 10.1109/IV47402.2020.9304553.

[13] Hirschmüller, H.: "Stereo Processing by Semiglobal Matching and Mutual Information", IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, pp. 328-341.

[14] Hirschmüller, H. “Accurate and efficient stereo processing by semi-global matching and mutual information.” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2005.

[15] Y. Chang, C. Jung, P. Ke, H. Song and J. Hwang, "Automatic Contrast-Limited Adaptive Histogram Equalization With Dual Gamma Correction," in IEEE Access, vol. 6, pp. 11782-11792, 2018, doi: 10.1109/ACCESS.2018.2797872.

[16] Koch, Rainer & May, Stefan & Nüchter, A. (2017). DETECTION AND PURGING OF SPECULAR REFLECTIVE AND TRANSPARENT OBJECT INFLUENCES IN 3D RANGE MEASUREMENTS. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. XLII-2/W3. 377-384. 10.5194/isprs-archives-XLII-2-W3-377-2017.

[17] Liu Y, Ge Z, Yuan Y, et al. Study of the error caused by camera movement for the stereo-vision system [J]. Applied Sciences, 2021, 11(20): 9384.

Predicting Unexpected Traffic Congestion in Urban Scenarios Using Planned Route Information

Shota Okazaki †and Takuya Yoshihiro ‡

†Graduate school of Systems Engineering, Wakayama University, Japan
‡Faculty of Systems Engineering, Wakayama University, Japan
{okazaki.shota@g.wakayama-u.jp, tac@wakayama-u.ac.jp}

Abstract - Predicting the expansion of traffic congestion in advance is one of the crucial issues in modern societies because traffic congestion causes wasting time for drivers and leading to significant economic losses. Considerable part of congestion in urban areas is an unexpected congestion, which is hard to predict because it occurs depending on a coincidental vehicle interaction such as growth of the queue of right-turn vehicles. However, few studies treat this kind of unexpected congestion. In this study, we proposed to predict the unexpected congestion that will occur in a few minutes ahead using two deep learning models that incorporate vehicles' planned routes. Our proposed method predicts unexpected congestion by using all vehicles' planned routes in the road network.

Keywords: Predicting Congestion, Deep Learning, Vehicle's Planned Route, Urban Traffic, ITS

1 INTRODUCTION

Traffic congestion is recognized as a serious issue, and its impact is growing larger with increasing traffic demand. Traffic congestion causes wasting time for drivers and leading to economic losses. So, preventing expansion of traffic congestion in advance by predicting it is one of the important issues in modern societies.

Research on predicting traffic congestion has been continuing for a long time because of its importance. Predicting traffic congestion is categorized into long-term prediction and short-term prediction. Long-term prediction, which predicts it hours or days ahead, is used for urban planning or wide-area navigation systems. On the other hand, short-term prediction, which predicts it a few minutes ahead, is useful to real-time traffic control, driver assistance, and automatic driving in urban areas where traffic conditions change rapidly.

On the other hand, traffic congestion is categorized into two types: periodic congestion, which depends on the day of the week or the time of day, and unexpected congestion, which is caused by traffic accidents, road construction, etc. [1]. Especially, as the latter cases in urban areas, traffic congestion may occur depending on the intersection's status, including nearby vehicles' positions, speed, and their mutual interactions [2]. Here, note that most studies in the literature predict the periodic congestion, based on periodic characteristics that depend on the day of the week or the time of day. Only a few studies focus on the unexpected congestion, which occurs depending on the intersection's status, typically the queue length of the right-turn vehicles.

The target of this study is to predict unexpected congestion by catching up with the signs of congestions in urban areas using deep-learning models, by means of utilizing the vehicles' planned route information obtained by the route navigation systems. In this study, we assume future cities in which all vehicles are connected to the network, and we can use all vehicles' data such as vehicles' positions and planned route information. To predict unexpected congestion, we use a conbination of two deep-learning models. The first is the vehicle movement prediction model (VMP model), which predicts traffic volumes on each road a few minutes ahead based on all vehicles' planned route information. The second is the congestion prediction model (CP model), which predicts unexpected congestion that will occur in a few minutes using the output of the vehicle movement prediction model. With those two models, we try to predict short-term unexpected congestion that will occur in a few minutes using vehicles' planned route information.

The structure of this paper is as follows. The related studies are described in Section 2. The proposed method is explained in Section 3., and the study is summarized in Section 4.

2 RELATED STUDIES

Academic studies on predicting traffic congestion or traffic volumes can be classified into two types: traffic prediction on highways and in urban areas. The latter is generally more difficult because it includes the effect of intersections and traffic signals.

Many studies predicting traffic congestion focus on the highway congestion [3]–[6]. Lei et al. used AGCRN (Adaptive Graph Convolutional Recurrent Network), which is based on GCN (Graph Convolutional Neural Network), to predict traffic volume one-hour later [5]. It incorporates NAPL (Node Adaptive Parameter Learning), which generates unique parameters for nodes on the graph, and DAGG (Data Adaptive Graph Generation), which generates road network graphs from traffic volume data. Qi et al. used DSGCN (Deep Spatial and Temporal Graph Convolutional Network), which combines GCN and LSTM (Long Short-Term Memory), to predict traffic congestion levels 15-45 minutes later [6]. These studies predicted highway traffic congestion based on periodic characteristics. On the other hand, our study aims to predict unexpected congestion in the urban area, which is essentially different from these studies.

Many studies predicting urban traffic congestion focus on periodic congestion[7], [8]. Comert et al. used EGM and EGVM, which correct the Gray Model –a mathematical model capable of predicting from incomplete information– using Fourier series to predict the vehicle queue length of one-second later[7]. Shirakami et al. used QTNet, which combines STGCN (Spatio-Temporal Graph Neural Networks) and queuing theory to predict vehicle queue length one-hour ahead [8]. STGCN is a deep-learning model to predict it based on temporal and spatial correlations. Zhang et al. used ConvLSTM (Convolutional Long Short-Term Memory), which analyzes traffic patterns dependent on the day of the week or the time of day, to predict average vehicle speeds up to two hours ahead. These studies predict periodic traffic congestion that depends on the day of the week or the time of day. Therefore, it is difficult to predict unexpected traffic congestion that doesn't depend on periodic traffic patterns, which is the target of this study.

Few studies predict the urban area traffic congestion without the periodic characteristics [11], [12]. Archana et al. focused on weather data to improve the prediction accuracy of traffic congestion [11]. They predict traffic volume at multiple locations from 5 minutes to 1 hour later using traffic volume data and weather historical series (precipitation, snowfall amount, and visibility) at multiple locations. However, the target congestion in their study doesn't include the unexpected congestion targeted in this study.

As the only method close to our study, there is a study by Rui et al., which predicts traffic volume using vehicle planned route information [12]. They proposed a deep-learning model called H-STGCN (Hybrid Spatio-Temporal Graph Convolutional Network) that uses vehicle navigation system information to predict. This model predicts the average travel time of roads, i.e., the time to pass through the road, from 5 minutes to 1 hour later based on the ideal future volume, which is calculated based on vehicles' planned routes obtained from vehicles' navigation systems. However, they predict the congestion in the road network including many highways, and don't focus on unexpected congestion occurring in urban areas.

3 PROPOSED METHOD

3.1 Preliminary

These days, many drivers use navigation systems to search for optimal routes to their destinations and get real-time traffic information. In addition, with the development of ITS (Intelligent Transport Systems), it is expected that a society in which all vehicles are connected to the Internet will arrive. In such a society, it will be possible to collect information on the location and driving states of all vehicles on the server and the server proposes the optimal route for drivers as needed. In this study, we assume all drivers set the planned route on the navigation systems. Accordingly, traffic control servers can get the planned routes of all vehicles and use these data for various information processing as well as traffic control.

3.2 Deep Learning Models

3.2.1 Overview

In this study, we use two deep-learning models, the vehicle movement prediction (VMP) model and the congestion prediction (CP) model, to predict unexpected congestion approximately 10 minutes ahead based on the planned routes of the vehicles and traffic volume data. The VMP model predicts the location of a vehicle k minutes later based on the vehicle's planned route information. By predicting the location of all vehicles k minutes later, we can obtain a future traffic map for the near future, that we call the near-future complementary (NFC) data. The CP model predicts unexpected congestion a few minutes ahead based on the past traffic (PT) data and the NFC data.

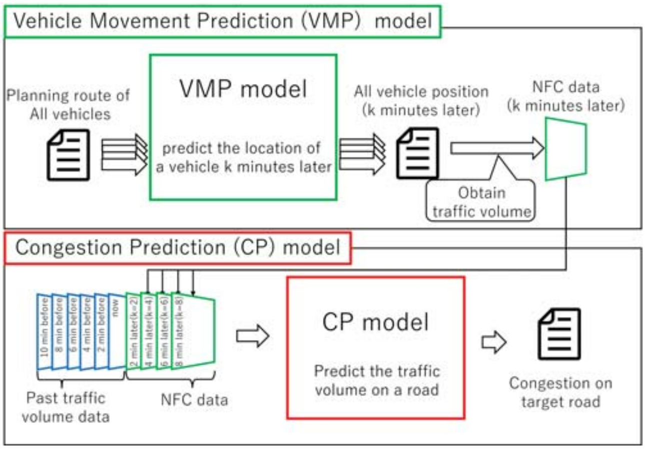

Vehicle Movement Prediction (VMP) model
Planning route of All vehicles
VMP model
predict the location of a vehicle k minutes later
All vehicle position (k minutes later)
NFC data (k minutes later)
Obtain traffic volume
Congestion Prediction (CP) model
10 min before
8 min before
6 min before
4 min before
2 min before
now
2 min later(k=2)
4 min later(k=4)
6 min later(k=6)
8 min later(k=8)
Past traffic volume data
NFC data
CP model
Predict the traffic volume on a road
Congestion on target road

Figure 1: Structure of our Proposed Method

Figure 1 is the structure of our proposed method. First, we predict all vehicle positions for k minutes later using the VMP model. As mentioned above, we can obtain the NFC data as a result. Here, we must obtain future location of all vehicles corresponding to every value k. Next, we predict traffic congestion n minutes later ($k < n$) from the actual past traffic volume and the NFC data. In Fig. 1, we show an example of those models, which predict the traffic volume 10 minutes ahead based on the actual past traffic volumes of 0 to 10 minutes before and the NFC data of $k = 2$ to 8 minutes future.

The reason for using the VMP model is the difficulty to obtain the location and speed of vehicles even for the near future. Due to the complexity of the relationship among vehicles that interact with one another, results from simple calculations are prone to error. The VMP model is expected to predict the location of vehicles considering complex interactions among vehicles from the vehicle information including planned routes of nearby vehicles. By using deep learning model, we expect to obtain more accurate locations, of near-future vehicles than simple calculative prediction such as [12] by taking vehicle interactions into account.

3.2.2 The Vehicle Movement Prediction (VMP) Model

The VMP model is a deep-learning model to predict a vehicle position k minutes later based on the planned route of the vehicle and the traffic state of near the vehicle. Vehicle location k minutes later is determined based on the pass-time of each road in the planned route of the vehicle. And, the road pass-time transits under the influence of traffic conditions of the nearby roads. Thus, predicting it using a deep-learning

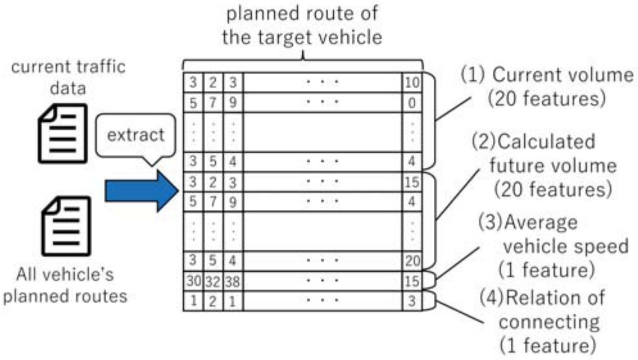


Figure 2: Creating the Input Vector of VMP Model

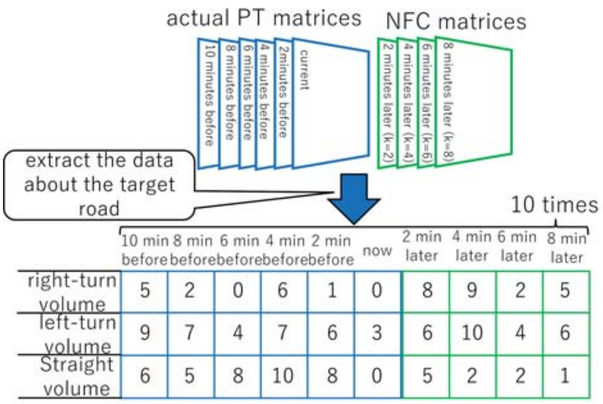
model is better than a simple calculation based on historical data. The model predicts the accuracy that the vehicle exists on each road k minutes later, based on the following four kinds of features: (1) each road's current volume, (2) each road's future volume by a simple calculation, (3) the current average vehicle speed of each road, and (4) relation between a road and the next road in the planned route of the vehicle. By predicting all vehicle future locations using this model, we can obtain the NFC data for k minutes later.

Figure 2 illustrates the method of creating the input vector for the VMP model. Each road in the vehicle's planned route has 42 values for constructing the input vector; 20 values for each of (1) and (2), and 1 value for each of (3) and (4). Because the vector includes initial 15 roads in the planned route of the vehicle, the vector has 630 ($= 15 \times 42$) values.

Traffic volumes (1) and (2) consist of four kinds of road volumes: all vehicles on the road, and the classified volumes of the straight, right-turn, and left-turn traffic on the road. Therefore, (1) and (2) include those traffic volumes of the nearby 5 roads related a road included in the planned route, because the target vehicle's movement is affected by the state of the roads that are not included in the plannned route of the vehicle. For example, a left-turn vehicle can't move on, when the left-turn next road has a congestion. Also, right-turn vehicles can't move on, in case straight vehicles exist on the oncoming road. These vehicles interfere with the following vehicles on the same road and cause congestion. By inputing traffic features of the potential interfering roads, the VMP model is expected to predict the near-future location of the target vehicle more accurately than the simple calculation.

The (2) each road's future volume is the traffic volume on each road k minutes later, obtained with an simple calculation based on all vehicle's planned routes. By using (2) the future volume as a part of the input vector, the model can predict vehicle location k minutes later considering all vehicle's planned routes and the traffic volume of each road in the future.

We describe the method of obtaining (2) the future volumes. First, we assume in the calculation that the vehicle speed is kept unchanged and there are no interaction among vehicles. Then, given the planned route of the vehicle, the position of each vehicle k minutes later is easily calclated. From the position of all vehicles k minutes later, we construct (2) the simply-estimated future volume of each road as a part of the input of the VMP model.



	10 min before	8 min before	6 min before	4 min before	2 min before	now	2 min later	4 min later	6 min later	8 min later
right-turn volume	5	2	0	6	1	0	8	9	2	5
left-turn volume	9	7	4	7	6	3	6	10	4	6
Straight volume	6	5	8	10	8	0	5	2	2	1

Figure 3: Creating Input Vector of CP Model

The NFC data,the input of the CP model, is a matrix consisting of the expected numbers of the sum of the confidences corresponding to the road in the output of the VMP model for every vehicle. This matrix represents the predicted traffic volume after k minutes.

3.2.3 The Congestion Prediction (CP) Model

The CP model predicts the traffic volume of the road n minutes later based on the traffic volume matrix of the road, which is constructed by the PT data (i.e., matrix) representing past traffic state and the NFC matrix representing future prediction of traffic state. The model uses only traffic volumes, which is the count of vehicles on each road. In this study, congestion is identified by the traffic volume, i.e., if the traffic volume on a road is larger than the congestion threshold, we regard that the road is congestion. We determined the threshold by reference to the average number of vehicles that can pass an intersection in one cycle of the traffic signal.

The traffic volumes used in the model are moving averages of actual traffic volumes in which the window size is the signal period intending to eliminate the influence of the the period of traffic signal over the traffic volume.

The CP model predicts the traffic volume 10 minutes later ($n = 10$) using the PT matrix at $p = 0, 2, 4, 6, 8, 10$ minutes before, and the NFC matrix for 2, 4, 6, 8 minutes later ($k = 2, 4, 6, 8$). Because the PT and NFC matrices are in the same format, it can be input seamlessly into the CP model. Figure 3 illustrates the method of creating the input vector of this model. The vector has 3 kinds of traffic volumes, i.e., count of straight, right-turn, and left-turn vehicles. As a result, the vector has 30 values because each of the 6 PT matrices and the 4 NFC matrices (in total 10 matrices) has 3 kinds of volume.

The CP model is a 4 layers Feedforward Neural Network. The input of the CP model is the vector that has 30 values, as mentioned above. The output of the CP model is the traffic volume 10 minutes later. The CP model has 3 middle layers. All middle layers have the ReLU function as an activation function. Because the output of the CP model is a predicted traffic volume and impossible to be negative, if the output value is under 0, the output value is set 0.

4 CONCLUSION

In this study, we proposed a method to predict unexpected congestion using deep learning models, which utilizes vehicles' planned routes.

Our proposed method predicts the unexpected congestion n minutes ahead, using two deep learning models: the VMP model and the CP model. The VMP model predicts the future location of a vehicle k minutes later based on the vehicles' planned routes. The CP model predicts the future traffic volume of a road n minutes later using the NFC matrix, which is constructed based on the future location of all vehicles.

ACKNOWLEDGMENT

This work is supported by JSPS KAKENHI Grant Number 23H03385.

REFERENCES

[1] Cambridge Systematics, Inc. et al.;SHRP 2 Report S2-L03-RR-1; TRANSPORTATION RESEARCH BOARD, Washington, D.C. (2013) `https://transops.s3.amazonaws.com/uploaded_files/SHRP2_S2-L03-RR-1.pdf` (Referenced May 21, 2025)

[2] Płaczek, Bartłomiej; A self-organizing system for urban traffic control based on predictive interval microscopic model; Engineering Applications of Artificial Intelligence; pp.75–84 (2014)

[3] Williams B. M., Hoel L. A.; Modeling and Forecasting Vehicular Traffic Flow as a Seasonal ARIMA Process: Theoretical Basis and Empirical Results; Journal of Transportation Engineering; pp.664-672 (2003)

[4] Polson N. G., Sokolov V. O.; Deep Learning for Short-Term Traffic Flow Prediction; Transportation Research Part C: Emerging Technologies; pp.1-17 (2017)

[5] Bai L., Yao L., Li C., Wang X., Wang C.; Adaptive Graph Convolutional Recurrent Network for Traffic Forecasting; Advances in Neural Information Processing Systems; pp.17804-17815 (2020)

[6] Qi Y., Cheng Z.; Research on Traffic Congestion Forecast Based on Deep Learning; Information Vol:14(2); 108 (2023)

[7] Comert G., Khan Z., Rahman M., Chowdhury M.; Grey Models for Short-Term Queue Length Predictions for Adaptive Traffic Signal Control; Expert Systems with Applications Vol:185; 115618 (2021)

[8] Shirakami R., Kitahara T., Takeuchi K., Kashima H.; QTNet: Theory-Based Queue Length Prediction for Urban Traffic; KDD '23: Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 4832-4841 (2023)

[9] Zhang K., Chu Z., Xing J., Zhang H., Cheng Q.; Urban Traffic Flow Congestion Prediction Based on a Data-Driven Model; Mathematics Vol:11(19); 4075 (2023).

[10] Wu X., Huang H., Zhou T., Tian Y., Wang S., Wang J.; An Urban Road Traffic Flow Prediction Method Based on Multi-Information Fusion; Scientific Reports Vol:15(1); 5568 (2025).

[11] Nigam A., Srivastava S.; Hybrid deep learning models for traffic stream variables prediction during rainfall; Multimodal Transportation Vol:2(1); 100052 (2023)

[12] Dai, R., Xu, S., Gu, Q., Ji, C., Liu, K.; Hybrid Spatio-Temporal Graph Convolutional Network: Improving Traffic Prediction with Navigation Data; Proceedings of the 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pp.1025-1033 (2020).

Road Congestion Prediction Method using Multimodal Macine Learning

Mikiko Sode Tanaka
Department of Electrical Engineering
and Information Science
National Institute of Technology
(KOSEN), Niihama College
Niihama, Japan
m.sode@niihama-nct.ac.jp

***Abstract*— Research is being actively conducted into systems that can simulate city road traffic on a digital twin based on information sensed in the real world and plan congestion relief measures and travel route optimization. A traffic digital twin is a computer reproduction of real-world road traffic. By using a traffic digital twin, it is possible to grasp road traffic conditions in real time and predict future traffic conditions. In order to reproduce real-world traffic flows using a digital twin and predict future traffic flows, traffic data with fine temporal and spatial granularity is required. For this reason, research and development is being conducted based on 5G and Beyond 5G, which enable high-volume communication. However, this method requires continuous collection of large amounts of data, making it difficult to use in regional cities due to the cost. Therefore, in this study, we propose a system that performs multimodal machine learning based on the travel speed of public transportation buses and the congestion level displayed on a digital map at that time to predict congestion levels. The proposed method makes it possible to create a digital twin with little data, making it a technology that can be applied even in areas where it is difficult to collect large amounts of data, such as regional cities.**

Keywords—Transportation, Digital Twin, Bus, traffic jam, Optimal Route, Traffic flow control

I. INTRODUCTION

A traffic digital twin is a computerized reproduction of real-world road traffic. By using a traffic digital twin, it is possible to grasp the road traffic situation in real time and predict future traffic conditions [1-5]. Since it will be possible to predict traffic congestion, the range of uses is wide, including studying ways to alleviate congestion and providing route guidance to avoid traffic jams. In order to reproduce real-world traffic flow using a digital twin and predict future traffic flow, traffic data with fine temporal and spatial granularity is required. Therefore, research and development based on 5G and Beyond 5G, which enable high-volume communication, is being carried out.

The population continues to decline in Japan's regional cities, making it difficult to maintain public transportation. The number of buses and trains in operation is steadily decreasing, and it is not uncommon for buses to be discontinued or to operate only two or three times a day. How to maintain local public transportation is an important issue. On the other hand, when public transportation is discontinued, the only means of transportation available is private cars. As a result, traffic congestion occurs in regional cities during morning and evening commute hours. As such, there are many traffic issues in the region. In response to these issues, various efforts have been made in recent years, such as shared buses, on-demand buses, and self-driving buses. To support these new forms of transportation, a new type of transportation digital twin system that can solve issues unique to rural areas is needed.

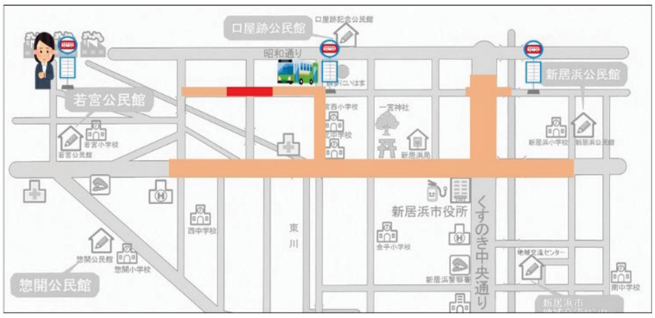


Fig. 1. Displaying road congestion using digital maps.

The biggest problem in rural areas is the limited funds available to invest in the system. Ideally, sensors would be installed in every corner of the city to collect data and build a system, but this is difficult in rural cities. We propose building a highly accurate traffic digital twin system using a small amount of data. In particular, this paper describes a system that predicts road congestion using small amounts of information, such as information that can be collected from sensors such as cameras installed in limited locations and driving data from public transportation.

It is not easy to properly manage traffic in response to unexpected road closures due to car accidents and to alleviate congestion caused by a sudden increase in urban traffic demand. Since congestion length suddenly becomes a spike shape when traffic demand exceeds the capacity of the road, it is very difficult to predict from the congestion length history, and existing methods perform poorly. The currently widely used map-based congestion display systems are very vulnerable to unexpected accidents and other unexpected occurrences, and often fail to provide accurate information. This is a problem that needs to be solved.

There is a strong correlation between the statistics of time series data of average travel speed, traffic volume, and congestion length in two adjacent road sections, which shows temporal dependence in the data and spatial dependence in the data. QTNet has been proposed as a method that focuses on this point [6]. QTNet combines constraints based on queuing

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

theory with a data-driven deep learning model. QTNet uses STGNN to predict rates and flows, and outputs queue lengths through the QT layer based on constraints derived from theoretical assumptions. Although this method is highly accurate, however, this method requires a large amount of detailed data to be feasible.

In recent years, many services have been provided that display the degree of road congestion on digital maps (Fig. 1). Although the accuracy is sufficient for daily use, it may take time to reflect the actual situation in the case of unexpected situations such as road construction, accidents, and festivals. In this paper, we propose a method that can quickly reflect the situation in the event of an unexpected event.

II. BASIC CONCEPT OF THE PROPOSED METHOD

A transportation digital twin is a technology that virtually reproduces actual transportation systems and infrastructure, and simulates, monitors, and optimizes their operation. This allows for a deeper understanding of various transportation-related issues and enables more efficient and safer traffic management. We aim to build a transportation digital twin for regional cities. Figure 2 is an image of the transportation digital twin we aim to build. In the physical space, road conditions are obtained from IoT devices, and that information is sent to cyberspace. In cyberspace, the degree of road congestion is calculated based on the data obtained, and the degree of road congestion in the physical space is returned. Our goal is to create a system that can move between cyberspace and physical space, easing congestion and facilitating mobility.

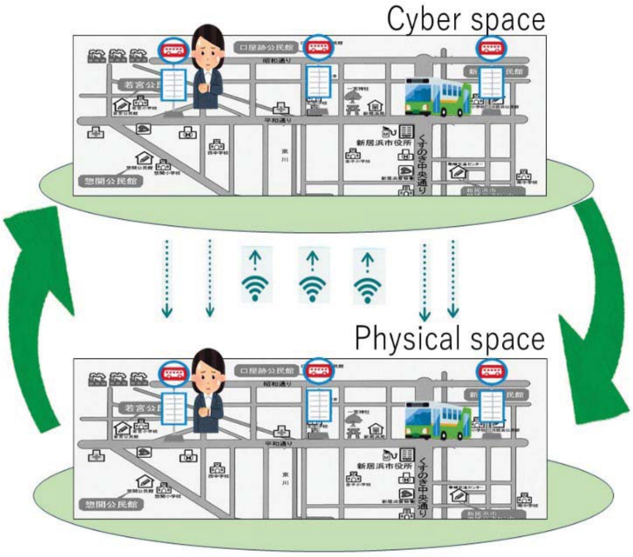


Fig. 2. Diagram of the proposed transportation digital twin configuration.

In order to build a digital twin for transportation systems in rural areas, it is necessary to minimize the number of additional IoT devices required and to effectively utilize existing assets. In urban areas, it is common to add IoT devices to key locations on roads to build a digital twin for transportation systems, but this is difficult in rural areas. Therefore, we decided to use buses, which are a form of public transportation. Although there are few buses running, they run through key locations in the city where traffic is heavy, and there are enough buses in the morning and evening when traffic is heavy. In addition, Niihama City has an application that allows users to check the running status of buses, and they can know the exact location of the bus [7]. Therefore, we decided to use the running status of buses as an IoT device. Figure 3 is a route map of buses running within the city, and Figure 4 is an application that displays the locations of buses running within Niihama City. In areas where the number of bus services is low, there is a possibility that information will be scarce and inaccurate; however, such areas can also be said to be areas where there is little demand for travel, so we decided to use a method of making predictions based on trends in the main parts of transportation.

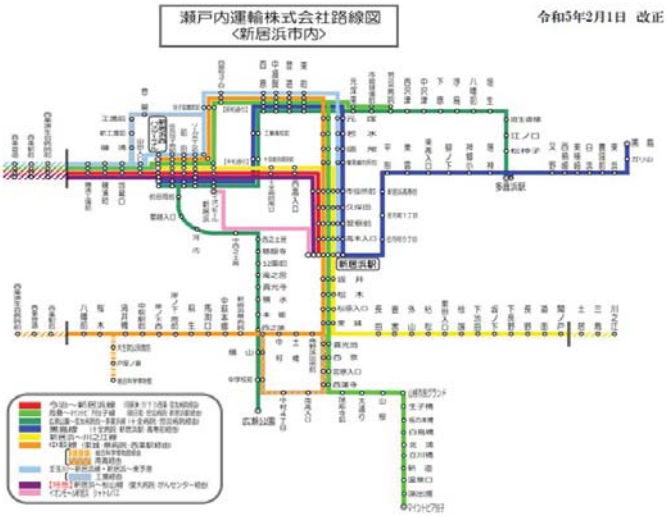


Fig. 3. Route map of buses running in Niihama City.

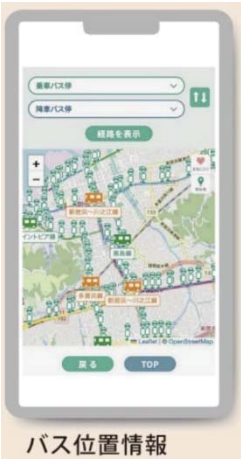


Fig. 4. An application that notifies the current location of route buses running in Niihama City [7].

III. CONGESTION PREDICTION USING MULTIMODAL AI

Multimodal AI is an AI that combines and processes multiple different types of data, such as text, images, audio, video, and sensor data. It is said that by combining information from different modalities, it is possible to process richer information and provide deeper understanding and insights. In particular, Stable Diffusion is capable of generating images from text data by adopting a latent diffusion model [8-9]. The training procedure for Stable Diffusion is shown in Figure 5. Images and text are trained as a set. Figure 6 shows the procedure for generating an image from text in Stable Diffusion.

The proposed method uses multimodal AI to predict road congestion. The speed of each bus gives a rough idea of the congestion level at that time. The results can be displayed as congestion levels on a map, and much of the data can be obtained from congestion level displays using map apps that are currently widely used. Therefore, we propose a method to train these two as a set, input the speed of each bus, and derive a map that displays the congestion level.

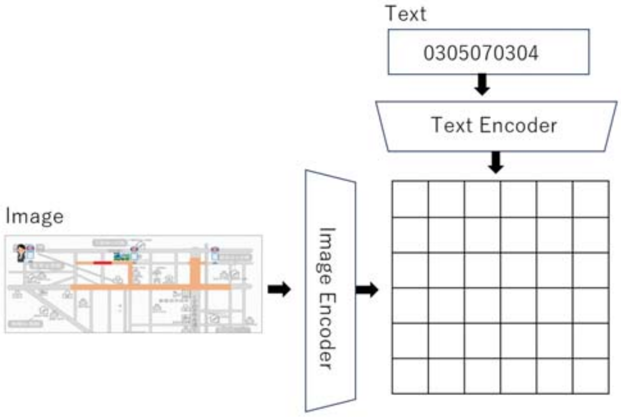


Fig. 5. Training procedure for Stable Diffusion [8-9].

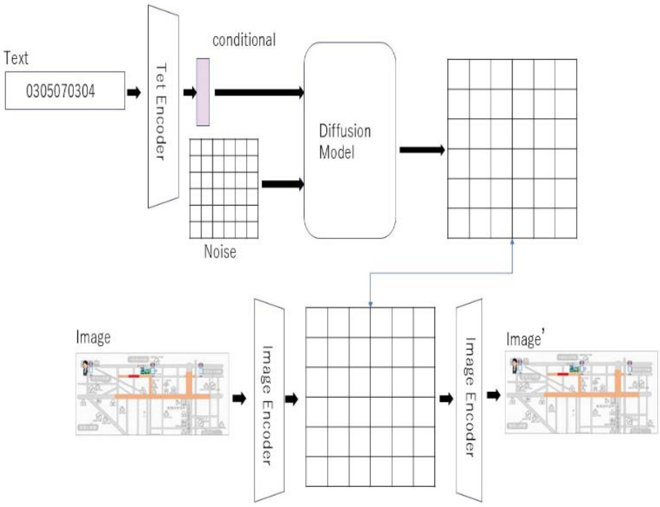


Fig. 6. Steps for generating images from text in Stable Diffusion [8-9].

IV. ROAD CONGESTION CALCULATION METHOD USING MULTIMODAL AI

We will now explain how to create the learning data. The bus speed is calculated at a certain time interval. This calculation takes into account factors such as the time the bus stops. Images of the road congestion at that time interval are created. This is done based on a commonly used web application that displays congestion levels on a map. Many sets of images showing the bus speed and congestion levels are prepared and used as input data for learning. An image of the learning data is shown in Figure 7.

The format of the text to be entered is explained below. The string is entered in the order of the bus location number. The bus location information is divided into 15-minute intervals, and the bus speed is entered for each time. Speed should be entered in three digits. For example, if the bus is traveling at 50 km/h, enter 050. If there is no bus traveling at that time and location, insert the value -10. Figure 8 shows an example of bus position information. The target area is divided into meshes and numbered. Figure 9 shows the time format for expressing bus travel times.

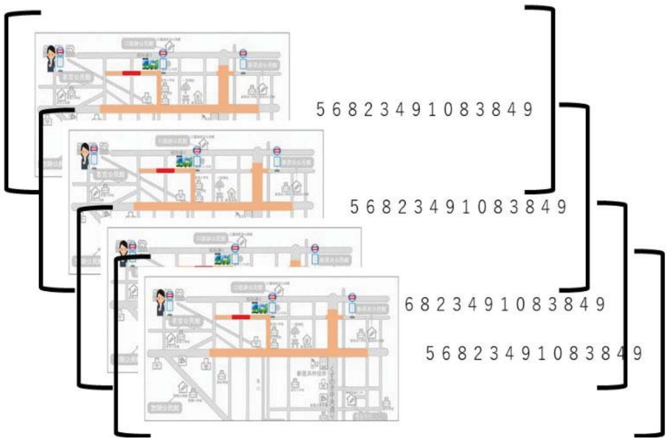


Fig. 7. Image of the learning dataset for calculating road congestion levels.

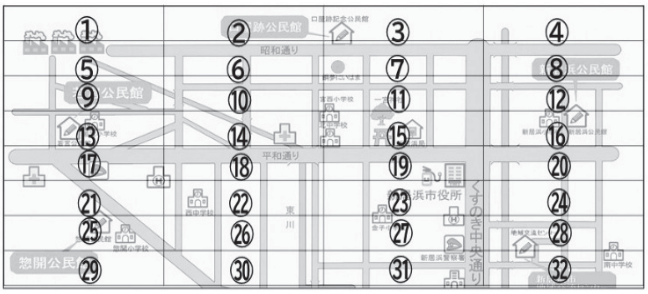
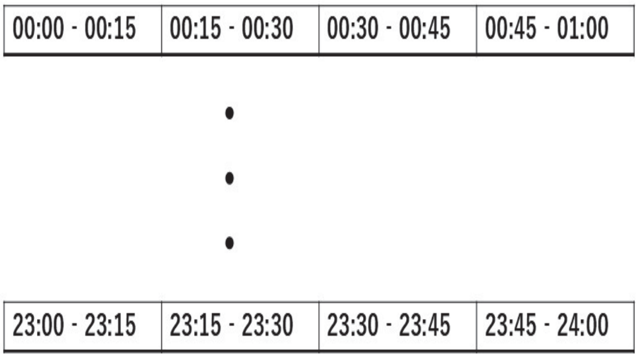


Fig. 8. Example of bus location displayed on a map.



00:00 - 00:15	00:15 - 00:30	00:30 - 00:45	00:45 - 01:00

•

•

•

23:00 - 23:15	23:15 - 23:30	23:30 - 23:45	23:45 - 24:00

Fig. 9. The time format for bus travel times.

Using the trained library, an image showing the degree of road congestion is generated. For input, text is generated according to the text format used in training and input. The resulting image shows the degree of road congestion. Figure 10 shows the flow of generating an image from text.

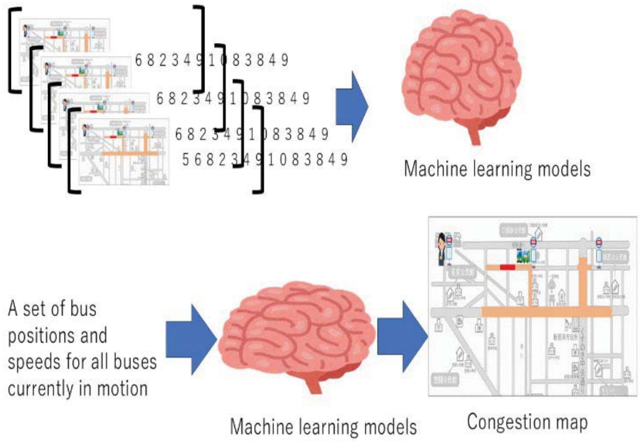


Fig. 10. Flow of generating an image from text.

V. EXPERIMENTAL RESULTS

We conducted experiments to verify the effectiveness of the proposed method. As a first step, we fixed the time and set the target area to a small area of 3 × 3. Figure. 11 shows an example in which three roads are set horizontally in a 3×3 area, and the degree of congestion on the roads is displayed in color. We used five colors to indicate the degree of congestion, and assigned bus speeds of 40km, 35km, 30km, 20km, 10km, and 0km. We created data for all of these combinations and created input data.

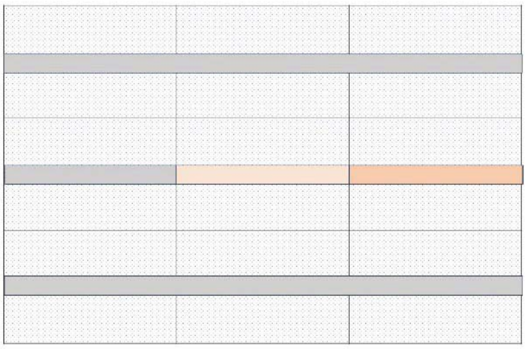


Fig. 11. An example of setting three roads horizontally in a 3x3 area and displaying the degree of congestion on the roads in color.

VI. CONCLUSION

Research is being actively conducted into systems that can simulate urban road traffic on a digital twin based on information sensed in the real world and plan congestion relief measures and optimized travel routes. The aim is to use a traffic digital twin to grasp road traffic conditions in real time and predict future traffic conditions. Conventional methods require the continuous collection of large amounts of data, which is applicable in large cities, but it is difficult to use in regional cities from a cost perspective. Therefore, in this study, we propose a system that predicts congestion levels by performing multimodal machine learning based on the travel speed of public transportation buses and the congestion levels displayed on digital maps at that time. A simplified version of the experiment was conducted to verify the effectiveness of the proposed method.

Acknowledgment

This work was supported by JSPS KAKENHI Grant Number 24K14939.

REFERENCES

[1] Miho Fujishima , Masaru Takagi, Masato Yokoya, Ryota Nakata, “Utilizing digital twins for traffic streamlining,” https://www.rd.ntt/research/JN202302_20942.html, *(Access 2024.12.15)*

[2] Takuya Nishimura, Takashi Kodama, Daiki Kawamori, Yuta Ojima, Ryo Nakata, Jun Tanabe Hiroshi Kiyotake, Satoshi Fukuda, “Proposal of Traffic Digital Twin for situation-based real-time traffic management,” Proceedings of the 43rd Traffic Engineering Research Conference, May 12, 2023, No. 100.

[3] Takashi Kodama, Yasuyuki Iwasato, Ryo Nakata, Kenichi Takashima, Takuya Nishimura, Daiki Kawamori, “Concept of transportation demand management using digital twin and consideration for social implementation,” 21st ITS Symposium 2023.

[4] Ministry of Land, Infrastructure, Transport and Tourism Regional Transportation Division, Policy Planning Bureau, “Current situation and issues surrounding regional transportation,” https://www.mlit.go.jp/policy/shingikai/content/001311082.pdf, (Access 2024.12.15).

[5] Tatsuro Sakai, Tomohide Ichikawa, Asami Shiro, “Current status and issues of regional public transport in depopulated areas,” Ministry of Land, Infrastructure, https://www.soumu.go.jp/main_content/000569916.pdf, (Access2024.12.15).

[6] Ryu Shirakami, Toshiya Kitahara, Koh Takeuchi, Hisashi Kashima, “QTNet: Theory-based Queue Length Prediction for Urban Traffic,” KDD '23: Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining.

[7] Bus location system "LAC Bus", https://heartnetwork.jp/ict_service/?page_id=180 (Access 2025.06.22).

[8] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, Björn Ommer, “High-Resolution Image Synthesis with Latent Diffusion Models,” Computer Vision and Pattern Recognition 2022.

[9] Aditya Ramesh, Mikhail Pavlov, Gabriel Goh, Scott Gray, Chelsea Voss, Alec Radford, Mark Chen, Ilya Sutskever, “Zero-Shot Text-to-Image Generation,” Computer Vision and Pattern Recognition 2021.

[10] Mikiko Sode Tanaka, "Study on Digital Twin Computing for Predicting General Road Traffic Volume," 2025 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), Fukuoka, Japan, 2025, pp. 0520-0522, doi: 10.1109/ICAIIC64266.2025.10920764.

[11] Mikiko Sode Tanaka, Nito Yamada“Research on the performance required for CPS databases to realize smart cities,” 2025 IEEE 14th International Conference on Consumer Electronics - Berlin (ICCE-Berlin).

Propose of Visit Promotion System Based on Visit Prediction and Common Topics

Kohei Hayashi[†], and Katsuhiko Kaji[†]

[†]Graduate School of Business Administration and Computer Science, Aichi Institute of Technology
{b25720bb, kaji}@aitech.ac.jp

Abstract - The recent spread of diverse work styles has led to a decline in opportunities for face-to-face communication. In particular, the loss of informal communication—which serves as a foundation for spontaneous information exchange and relationship building—has become a critical issue. This study proposes a visit support system designed to promote face-to-face interactions among members who share common topics, targeting small-scale communities such as university laboratories. The proposed approach extends an existing occupancy management system by incorporating a function to predict future room presence based on past stay records. By combining visit prediction with topic information, the system identifies time slots when members who share the same topic are likely to be present. This aims to naturally trigger face-to-face communication and revitalize the community. Although the evaluation experiment did not reveal clear behavioral changes among members, the study identified several points for improvement and remaining challenges.

Keywords: Behavior Modification, Face-to-Face Communication, Common Topics, Visit Prediction

1 Introduction

hogehoge

In recent years, remote work and flexible working arrangements have become increasingly widespread. As a result, opportunities for face-to-face communication have been significantly reduced. Some studies have indicated that remote work can hinder communication[1][2]. It tends to make interactions more difficult and less frequent. Moreover, the decline of traditional co-located interactions has led to a decrease in informal communication opportunities, such as spontaneous encounters and casual conversations. The importance of informal communication—hereafter referred to simply as informal communication—and its association with face-to-face interactions has been highlighted in prior studies [3] [4]. Additionally, face-to-face communication has been reported to positively influence relationship building and information sharing [5].

For smooth face-to-face communication, information regarding member's presence plays a crucial role. Knowing when and where someone is present enables others to approach them efficiently, reducing wasted effort. Furthermore, it is generally believed that the psychological barrier to initiating interaction is lower when more people are present. This makes visits and spontaneous communication more likely to occur. Therefore, understanding visit status is not merely useful for monitoring individual behavior but also contributes to enhancing community engagement as a whole.

Various systems have been proposed to manage the stay status of members. One such system developed and operated in our laboratory is Stay Watch [6]. Human behavior is known to follow recurring patterns, such as preferred places and times of stay. These tendencies, referred to as work rhythms, are shaped by individual work styles and environmental factors [7]. They often vary across individuals and days of the week. In recent years, several methods have been proposed to leverage work rhythms for predicting future stay patterns. However, most existing research has focused solely on prediction. There have been few attempts to utilize these predictions to support behavioral change or facilitate communication.

Informal face-to-face communication plays a vital role in facilitating the exchange of contingent information and fostering interpersonal relationships. Such communication typically arises spontaneously in contexts unrelated to formal tasks, often triggered by shared interests or casual consultations. Members who share common topics are more likely to experience a sense of affinity, thereby lowering the threshold for initiating conversation and facilitating face-to-face interaction. For example, small talk among individuals with shared interests or brief consultations on research themes can serve as entry points for everyday face-to-face communication.

However, face-to-face communication inherently assumes the physical co-presence of the parties involved. In environments such as offices adopting telework or university laboratories with flexible attendance policies, member visit frequencies may decline, severely limiting opportunities for in-person interaction. Consequently, incidental encounters and relationship-building opportunities are reduced, raising concerns about a decline in community vitality. In such settings, there is a need for a system that fosters and encourages face-to-face communication opportunities.

This study aims to increase visit frequency and encourage small-group activities based on shared topics within the community. As an approach, we propose a visit promotion system that combines information on common topics with visit prediction. the proposed system provides a mechanism to facilitate face-to-face communication based on the availability of each member in the future and the common topics shared among the members. Specifically, the approach combines visit prediction with common topics to identify situations where members related to a particular topic are likely to gather at the same time of day. The aim is to increase their motivation to visit. This is expected to result in natural face-to-face communication based on a common topic. To enable the prediction of future visit status, we extended the existing stay management system by adding a new prediction function. This function learns each member's visiting and returning tendencies based on their past stay history. This extension enables the system to flexibly capture the stay patterns of each member and provide predictions that are in line with actual visitation behavior.

2 Related works

2.1 Research on Prediction of Human Behavior

There is a wide range of research on predicting human behavior, much of which is based on the analysis of past behavioral history. Based on the size of the data used, these studies can be broadly categorized into two groups: those that utilize large-scale data and those that rely on small-scale data.

In the former group, various methods have been proposed to accurately predict people’s movement trends and stay distributions across entire cities or large-scale facilities. These methods rely on various large-scale datasets, including GPS logs and Wi-Fi connections. From these, they extract behavioral patterns over time and space for individuals or groups. In particular, recent advances in deep learning have enabled the use of millions of behavioral logs to learn high-dimensional features and estimate future actions with high accuracy [8][9][10]. However, such approaches require large amounts of data and computational resources. They are therefore difficult to apply in environments where historical data are scarce or in small-scale contexts with limited operational capacity.

In contrast, studies using small-scale data typically focus on the behavioral patterns of individuals within constrained communities, such as university laboratories or small offices. Due to the limited data available, these studies often incorporate explicit contextual information to improve prediction accuracy. For example, some approaches estimate visit times and durations by matching shared calendar schedules with activity history [11][12]. Others improve accuracy by referencing the stay status of members who frequently share activities [13]. However, many of these methods require manual schedule input or prior knowledge of interpersonal relationships, increasing the operational burden on users. Furthermore, the designs are often tailored to specific environments or applications, limiting the reusability of prediction results and integration with other systems.

As described above, large-scale data methods face challenges in real-world deployment due to environmental constraints. Meanwhile, small-scale data methods often place a higher burden on users and lack flexibility. Therefore, this study aims to develop a prediction function that is suitable for small-data environments, minimizes user effort, and allows seamless integration with external systems.

2.2 Research on Communication Support

Some studies on communication support have proposed systems that intentionally create opportunities for incidental encounters, thereby promoting informal communication. These systems often aim to visualize the presence of others in shared spaces, such as offices, and encourage spontaneous interactions by presenting relevant topics or contextual information.

For instance, several mechanisms have been developed to visually convey the activities of others from a distance, serving as triggers for chance conversations [14]. Other approaches use mediators such as food, beverages, or physical artifacts to promote movement and conversation within a space [15]. In addition, systems that display shared interests or current topics in common areas—prompting users to engage in dialogue—have also been proposed [16][17]. These approaches primarily focus on facilitating interaction among people who are already co-present.

In contrast, this study aims to proactively create situations where multiple members are likely to be in the same space at the same time, which is essential for face-to-face communication. Specifically, the system analyzes historical entry and exit logs to identify time slots during which members with shared topics are likely to be co-present. It then uses these predictions to promote visits. This approach is designed not merely to support incidental communication, but to increase the likelihood of meaningful dialogue and activity among users who share specific goals or interests.

3 Visit Prediction Function Based on Past History

Figure 1 shows an overview of the functions. This chapter describes the requirements, implementation, and evaluation of a function for predicting user's future visits and departures, which serves as a prerequisite for the visit promotion system proposed in this study.

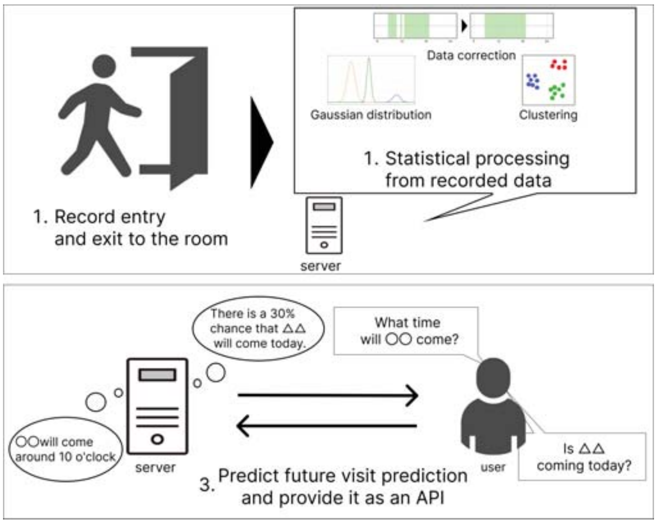

Data correction
Gaussian distribution
Clustering
1. Statistical processing from recorded data
1. Record entry and exit to the room
server
There is a 30% chance that △△ will come today.
What time will ○○ come?
○○will come around 10 o'clock
server
user
Is △△ coming today?
3. Predict future visit prediction and provide it as an API

Figure 1: Overview of the Visit Prediction Function

3.1 Requirement Definition

This function probabilistically estimates the likelihood of a user's future visit or departure based on their past entry and exit history. To realize this, it is necessary to implement a sys-

Table 1: API endpoint URL

URL	Description
{BASE URL}/api/v1/prediction/probability:action	Probability of user behavior occurrence at any given time
{BASE URL}/api/v1/prediction/time:action	Predicted time of user behavior occurrence

Table 2: Request Parameters (User Activity Probability Relative to an Arbitrary Time)

Path parameter	Value	Description
action	string	visit: Probability of a visit departure:Probability of going home
Query parameter	**Value**	**Description**
user-id	int	Required. User ID. Multiple Select.
weekday	int	Prediction Day of the Week (Monday: 0 - Sunday: 6). Default: 0.
time	string	Prediction Time. Format: HH:MM. Default: 24:00.
isForward	boolean	true: Calculate probabilities before the specified time. false: Calculate probabilities after the specified time. Default: true.

Table 3: Request Parameters (Predicted Time of User Activity Occurrence)

Path parameter	Value	Description
action	string	visit: Probability of a visit departure:Probability of going home
Query parameter	**Value**	**Description**
user-id	int	Required. User ID. Multiple Select.
weekday	int	Prediction Day of the Week (Monday: 0 - Sunday: 6). Default: 0.

Table 4: Response Field (User Activity Probability)

Field	Description
weekday	Day of the week
time	Time for prediction
isForward	Whether to predict before the specified time
result.userId	User ID
result.probability	Probability of visit

Table 5: Response Field (Predicted Time of User Activity Occurrence)

Field	Description
weekday	Day of the week
result. userId	User ID
result. predictionTime	Predicted time

tem that collects data unobtrusively, predicts visit prediction trends, and ensures interoperability with external systems.

To predict future visits, it is essential to understand user's staying tendencies based on their past behavioral patterns. For this purpose, the continuous collection of reliable behavioral data is crucial. In this study, to minimize the burden on users during data collection, the prediction function was implemented as an extension of an existing occupancy management system. The system utilizes entry and exit logs that are automatically recorded by IC cards, beacons, and similar technologies. This approach enables the system to obtain visit histories through user's routine activities, without requiring any special actions on their part.

The visit status consists of multiple pieces of information. For example, it includes data such as who is currently in the room, how many people are present, and who visited the room and when. Therefore, in order to predict future visit trends, it is necessary to obtain information from multiple perspectives corresponding to these aspects.

The two types of forecasts addressed in this study are the probability of a user's behavior occurring at an arbitrary time and the predicted time at which the behavior will occur. In this study, we define the probability of user behavior at any given time as the likelihood of one of four possible events. These include the user having already visited, visiting later, having already departed, or departing after the specified time. The predicted time of user behavior refers to the time at which a visit or departure is expected to occur with the highest probability.

These two types of information serve as valuable indicators for visit prediction. Combining these probabilities enables diverse predictions from both temporal and contextual perspectives. For instance, the system can estimate whether a specific user will visit during the day, or whether another user is likely to be present upon arrival. This information also provides a foundation for supporting individualized forecasts targeting specific users. As a result, it is well-suited for flexible and highly personalized prediction in small communities.

In designing a visit prediction function, it is important to ensure interoperability with external systems. In particular, the visit promotion system proposed in this study assumes that the prediction function will be used as an external module. This design allows the prediction component to be reused or replaced independently. Therefore, the prediction function must be implemented in a generic manner that does not depend on any specific use case or application. Furthermore, this prediction function is expected to be used not only for visit promotion. It also has potential applications in areas such as congestion visualization, facility reservation support, and documentation of community activities. Therefore, its output format must be designed to facilitate integration with external systems.

3.2 Implementation of Predictive Functions

In this study, to probabilistically predict visit prediction, we first extract individual user behavior patterns from the access logs stored in the occupancy management system. As the first step, we extracted the daily visit and departure times for each user. Here, visit refers to the first entry time of the day, and departure refers to the last exit time. Note that leaving the room in the middle of the day does not necessarily indicate that the person has returned home; it may also represent a temporary absence, such as going out for lunch or attending a meeting. Even if there are multiple entries and exits between the first entry and the last exit of the day, they are regarded as part of a continuous stay. Exceptional cases, such as stays spanning multiple days or missing exit records, were preprocessed as no exit.

The extracted daily visit and return times were clustered by day of the week to clarify behavioral tendencies specific to each day. Since users may show diverse behavioral patterns on the same weekday, we performed the analysis under the assumption of multiple clusters per day.

The clustering method used in this study was the Gaussian Mixture Model (GMM). GMM assumes that the data consists of a mixture of multiple Gaussian distributions and probabilistically estimates the mean and variance of each distribution corresponding to a cluster. Among various clustering methods such as k-means and DBSCAN, GMM was chosen because this study assumes that member behavior patterns can be represented as a combination of multiple-peaked normal distributions. GMM is particularly suitable for modeling overlapping distributions and capturing their stochastic characteristics.

The number of clusters was determined using the Bayesian Information Criterion (BIC). After setting an upper limit, we selected the cluster count with the lowest BIC value. To reduce the impact of minor time differences on the clustering results, the time data were rounded to the nearest 30 minutes before applying GMM.

Figure 2 shows the number of times a user departed from the lab at each time of day on Mondays, along with its corresponding Gaussian function. The histogram and the Gaussian function do not align well in the original data. However, the clustering results indicate that each cluster can be approximately modeled by a Gaussian distribution.

The clustering results include the mean time, standard deviation, and the number of data points for each cluster. The mean time represents the typical timing of the action within the cluster, while the standard deviation serves as an indicator of the consistency of the action and the uncertainty of the prediction. The number of data points in a cluster serves as a feature indicating the frequency of occurrence of the corresponding behavior pattern.

Since each obtained cluster is assumed to follow a normal distribution, a user's behavioral history is represented as a mixture of multiple probability distributions to estimate the future probability of user visits. Since the data collection period includes non-visit days, it is necessary to adjust the weight of each cluster accordingly. In this study, weights were assigned to each cluster to reflect the visit frequency over the entire recording period.

The probability of a user action occurring at a given time is calculated by computing the cumulative distribution function (CDF) for each cluster, applying the corresponding weight, and summing the results. This allows the probability that a user has visited by a specified time to be estimated based on actual behavioral trends and their frequency. The predicted time of occurrence of user behavior was derived as a weighted average of the mean times of the clusters, using their corresponding weights. Since the mean and median coincide in a normal distribution, this weighted average can be directly interpreted as the predicted time of occurrence of the user behavior.

The resulting visit prediction results are provided as a RESTful Web API, enabling flexible and broad utilization independent of specific systems or platforms. Users and external systems can access the forecast results via the Internet, allowing the prediction models developed in this study to be integrated into various environments.

The API provided by this function is designed to support a wide range of use cases that leverage visit prediction. For example, it is possible to specify target members for prediction, retrieve action probabilities at arbitrary times, and switch between action conditions such as visit or stay.

The API endpoint (request URL) is shown in Table1. This API assumes simple requests via URL parameters and is designed to use the HTTP GET method. The required request parameters and response formats are shown in Table2, Table3, Table4, and Table5. If an invalid request is received or an error occurs during server-side processing, the request is aborted, and an appropriate HTTP error code along with an error message is returned.

3.3 Accuracy Verification of Visit Prediction Prediction

To confirm the effectiveness of the proposed visit prediction function, we conducted a verification experiment on its prediction accuracy. Here, we focus on the probability of a visit by a given time and evaluate the prediction accuracy by comparing the predicted results with the actual visitation data.

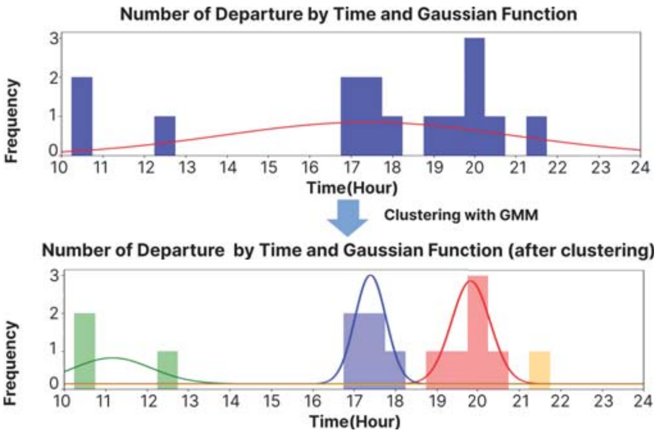


Figure 2: Discovering Patterns of Behavior Through Clustering

3.3.1 Experimental Setup

The data used in this study consist of entry and exit logs of 32 undergraduate and graduate students recorded by Stay Watch. The recording period spanned from each user's registration date to January 9, 2025, and the most recent 10 weeks were used for evaluation. The visitation probabilities were estimated using historical data accumulated during the period outside the evaluation window.

The verification procedure is as follows. First, for each day of the week, the prediction model calculates the earliest time t at which the probability of a visit by a given time exceeds a predefined threshold (50% in this experiment). It is then verified whether the user had actually visited by time t. Then, the actual visit frequency by time t during the verification period (measured value) is calculated, and the difference from the predicted probability is evaluated. If t could not be obtained (i.e., the visit probability never exceeded the threshold), the case was excluded from the evaluation.

3.3.2 Overall Performance Evaluation

Aggregate validation results for all users are summarized in Table 6. The estimated probabilities and the averages of the measured values generally agreed, confirming that the proposed method captured visitation trends without significant overall bias.

On the other hand, the mean absolute error was approximately 18 percentage points, indicating variability in predictions both across individuals and between different days of the week. This was especially pronounced for users with limited historical data and during periods featuring atypical behaviors, such as the year-end and New Year holidays.

3.3.3 Individual Examples and Discussion

Individual case results for member A and member B are presented in Table 7 and Table 8, respectively. Even when the visitation probability was around 50%, cases were observed in which the actual measured values deviated significantly from the predicted probability.

Factors contributing to such errors include individual circumstances such as health issues or internships, as well as changes in behavioral patterns during specific periods such as year-end and New Year holidays, long vacations, and the start of a new fiscal year. During these periods, work rhythms themselves tend to shift, which may reduce the relative effectiveness of past visit patterns for predicting near-term behavior. In addition, irregular factors such as holidays specific to the target community and holidays that fall on weekdays can also cause significant fluctuations in visit behavior. The current system cannot adequately handle such fluctuations, which contributes to a decline in prediction accuracy. Therefore, future work should focus not only on enhancing robustness against individual temporary factors but also on developing methods to dynamically adjust the reliability of past visit records according to the circumstances.

Table 6: Overall Summary of Visit Probability Prediction

Item	Value
Mean Predicted Probability	50.16%
Mean Actual Value	49.77%
Mean Error	0.39 points
Mean Absolute Error	18.29 points
Mean Absolute Percentage Error	0.646

4 Visit Promotion System Based on Visit Prediction and Common Topics

This chapter describes the requirements for the system designed to achieve the goals of this study, the architecture of the system based on those requirements, and an investigation into the effects of the system on community members. An overview of the system is shown in Figure3.

4.1 Target Users of the System

The proposed system targets members of informal ”ad hoc groups” that naturally form within small communities, such as university laboratories or offices, based on shared topics or interests. These groups emerge organically through casual interactions and are not based on formal structures or explicit agreements. Group participation is voluntary and flexible, often depending on mutual interest rather than formal membership.

Examples include students who share research themes, use similar lab equipment, or enjoy the same recreational activities. Since such activities typically occur when members happen to be co-present, the system aims to increase opportunities for spontaneous face-to-face interactions by predicting overlapping presence based on shared topics.

In environments with flexible attendance—like research labs —meeting opportunities can be limited. This is especially problematic for users with low visit frequency, as they may gradually lose connection with the group. Conversely, highly active users may not require system support, and inactive users may need different approaches.

Therefore, the system primarily targets users whose visit frequency and presence duration could benefit from improvement. By encouraging simultaneous presence among members with common topics, the system aims to revitalize informal group communication and foster engagement.

4.2 System Requirements

This study aims to increase opportunities for face-to-face communication and promote more active engagement within ad hoc groups composed of members with shared topics or interests. To achieve this, the system must predict time slots when group activities are likely to occur and deliver information proactively while minimizing user burden.

For predicting likely activity time slots, we utilize the prediction API proposed in the previous chapter. In ad hoc groups, whether an activity takes place often depends on whether enough members with shared interests are present at the same time. Therefore, predicting each member's room presence is essential for estimating when activities are likely to occur. How-

Table 7: Validation Results for Member A (Excerpt)

Day of Week	Time t	Predicted Probability (%)	Number of Visits	Actual Value (%)
Tue	09:47	50.12	2	20
Wed	15:07	50.00	5	50
Thu	12:42	50.12	6	60
Fri	13:05	50.12	6	60

Table 8: Validation Results for Member B (Excerpt)

Day of Week	Time t	Predicted Probability (%)	Number of Visits	Actual Value (%)
Tue	11:30	50.00	3	30
Wed	12:07	50.17	4	40
Thu	18:00	53.57	6	60
Fri	12:38	50.12	3	30

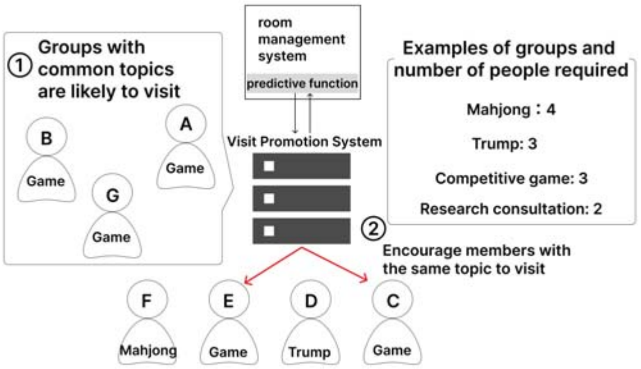


Figure 3: Overview of Visit Promotion System Based on Visit Prediction and Common Topics

ever, building a new system from scratch for occupancy management and prediction may impose a significant burden on users, including changes in usage environments and the need to provide additional data. To address this, the proposed system extends an already-deployed occupancy management system by adding a prediction feature, enabling future room presence prediction through a Web API using existing entry and exit logs.

Since participation in ad hoc group activities is based on user initiative, the means of information delivery must also be carefully considered. If users must actively retrieve information themselves, they may miss participation opportunities or face psychological barriers. This system was implemented to operate entirely within existing chat tools or web platforms, avoiding the need for new installations or complex configurations and thereby reducing user burden. Furthermore、 the system proactively sends notifications to users, including those with low visit frequency or those who may have missed opportunities, encouraging their participation in group activities.

4.3 Implementation of the Proposed System

To meet the objectives and requirements of this study, the system was implemented as a bot-type application extension that runs within an existing chat tool. Chat tools are likely already in use in many communities, making them suitable platforms for seamless function integration. A bot-type application running on a chat tool enables proactive engagement through notifications. Given its compatibility and ease of interaction via bot-based notifications, this study adopts Slack as the target platform.

The system follows several stages before delivering notifications to users. These stages include: (1) registration of ad hoc groups and users, (2) prediction of user visits and their room presence durations, (3) extraction of time slots when ad hoc groups are likely to gather, and (4) notification delivery to relevant members.

First, as a preparatory step, users and ad hoc groups are registered. Here, ad hoc groups refer to relationships formed based on user's shared interests or activities (e.g., Mahjong, seminars, board games) and do not require explicit awareness by users. To represent groups, the system uses tags as a substitute, allowing users to freely register topics of interest. When registering tags, users specify a tag name and the required number of participants for the activity (e.g., 4 for Mahjong). To make registration simple, this can be done directly within the Slack application. Additionally, the system provides initial registration functionality to link user's occupancy management system IDs with their Slack IDs.

Next, the system predicts which users will visit and their room presence durations. It retrieves each member's likelihood of visiting by 23:59 using the prediction API and extracts users whose probability exceeds a specified threshold as expected visitors. This threshold is set based on past data to represent a high probability of visiting on a given weekday. In this study, the threshold was set at 50Then, for each expected visitor, their estimated stay duration is calculated based on predicted time points obtained from the API.

Using the predicted room stay duration, the system calculates overlapping time periods among users sharing the same tag and extracts time slots during which people are likely to gather. As shown in Figure4, the overlapping presence of members A, B, and C indicates time periods when an activity is likely to occur. If the number of users expected to be present during such overlaps meets or exceeds the required number for that tag, a notification is sent. If the number of tagged users is below the required minimum, the system sends a notification regardless of the predicted attendance to encourage contact opportunities.

Finally, the system sends notification to members of rele-

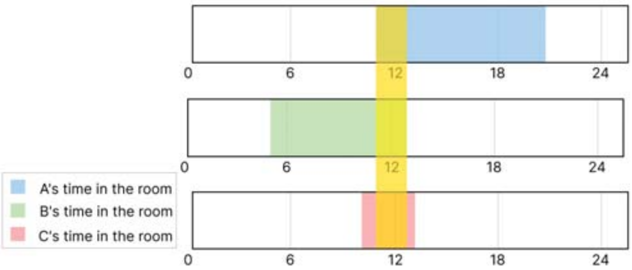
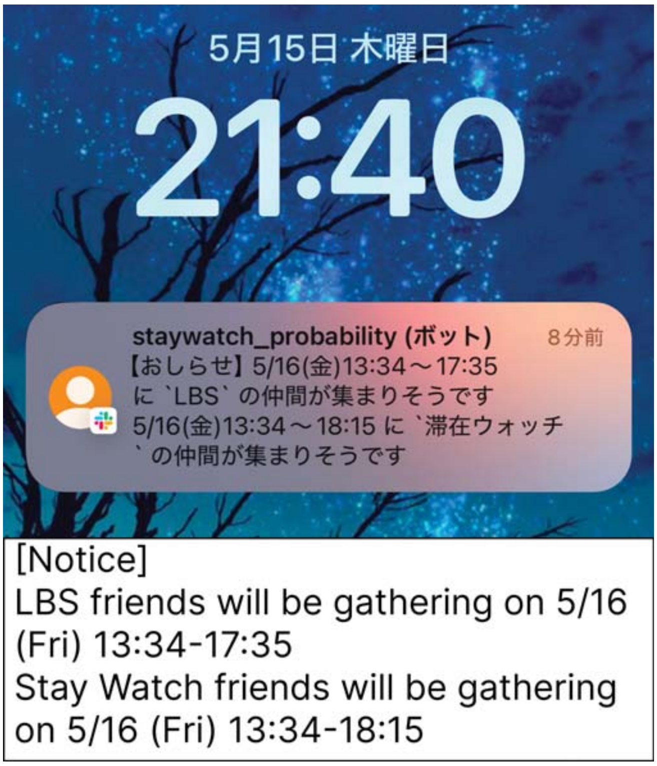


Figure 4: Comparison of User's Room Stay Duration



5月15日 木曜日

21:40

staywatch_probability (ボット) 8分前
【おしらせ】5/16(金)13:34～17:35 に `LBS` の仲間が集まりそうです
5/16(金)13:34～18:15 に `滞在ウォッチ` の仲間が集まりそうです

[Notice]
LBS friends will be gathering on 5/16 (Fri) 13:34-17:35
Stay Watch friends will be gathering on 5/16 (Fri) 13:34-18:15

Figure 5: Example Notification from the Proposed System

vant ad hoc groups. The bot-type application sends a notification once per day on Slack. The notification follow the format shown in Figure5, explicitly stating the tag and the suitable time period for activity, encouraging users to consider visiting the next day. The notification time was set to 9:00 PM so that it could influence user's planning before going to bed.

4.4 Evaluation of the Visit Promotion System

An evaluation experiment was conducted to verify the effectiveness of the proposed visit promotion system. The aim of this experiment was to determine whether the system contributed to increasing face-to-face communication opportunities within the laboratory. Specifically, we analyzed how the notifications influenced user's visit frequency and room presence durations.

4.4.1 Experimental Settings

The evaluation involved 14 undergraduate and graduate students belonging to the same laboratory. The system ran for 16 business days from April 15 to May 12, 2025, and a control period was set from March 1 to March 31, during which the system was not deployed.

The system was operated as a bot-type application on Slack. Prior to deployment, users selected and registered tags corresponding to their topics of interest to join groups. The topics were classified into the following two categories, with the required number of participants specified for each:

- Recreational/Social Topics: Super Smash Bros.(3), Catan(3), Jenga(3)
- Research/Consultation Topics: Project work, programming languages, lab equipment, etc.(2–3 people each)

The evaluation was based on three components: First, we used entry and exit logs recorded by a stay monitoring system to calculate each user's visit frequency and average stay duration for each weekday. We then compared the changes before and after the system was deployed. Second, to confirm whether activities actually occurred, a camera was installed in the lab for continuous recording. We visually verified whether observable activities (e.g., playing Super Smash Bros.) took place. However, due to limitations in visually confirming certain activity types, this method was sed as a supplementary measure. Lastly, we conducted a questionnaire to gather user's impressions of the notifications and their influence on visit motivation. The survey also covered topics they participated in, the impact on scheduling, future willingness to use the system, and included open-ended responses.

4.4.2 Analysis of Changes in Visit Frequency and Stay Duration

Table 9 shows the change in average visit frequency by weekday for all 14 participants. Following the introduction of the visit-promotion system, visit frequency increased across all weekdays. Notably, Wednesdays (+42.9 percentage points) and Thursdays (+34.5 points) exhibited the most significant improvements. These results suggest that the system was particularly effective in encouraging visits during the middle of the week.

Although this trend may indicate the system ' s effectiveness, it is also possible that variations in class schedules across weekdays influenced visit frequency. To address this potential confounding factor, future studies should investigate the relationship between class schedules and visit behavior. Such analysis would help more clearly isolate the effects of the system.

One possible explanation for the midweek increase is that students tend to have more flexible schedules on Wednesdays and Thursdays. In contrast, Mondays are often used for weekly planning, while Fridays may be constrained by weekend plans. As a result, notifications suggesting potential group activities may have acted as stronger triggers for midweek visits. Conversely, the smaller increases observed on Mondays and Fridays imply that users' schedules on those days were less flexible, diminishing the impact of the notifications.

Table 9: Change in Visit Frequency by Day of the Week (Overall Average, n=14)

Day	Before Introduction	After Introduction	Change
Monday	47.1%	64.3%	+17.2 pts
Tuesday	55.4%	71.4%	+16.0 pts
Wednesday	23.2%	66.1%	+42.9 pts
Thursday	25.0%	59.5%	+34.5 pts
Friday	35.7%	52.4%	+16.7 pts

Table 10: Average Stay Duration by Month (Overall Average)

Period	Average Stay Duration (minutes)	Change (minutes)
Before Introduction	363.7	-
After Introduction	424.7	+61.1

As for the average stay duration, as shown in Table 10, it slightly increased from 363.7 minutes in March to 424.7 minutes in April (+61.1 minutes). This may indicate that visitors tended to stay for longer periods rather than for brief visits.

Looking at individual trends, some users showed noticeable increases in both visit frequency and stay duration after the system was introduced, while others showed a decrease in stay duration. These differences may be attributed to individual characteristics and responsiveness, such as varying levels of interest in the suggested topics and the motivational effects of the notifications.

4.4.3 Activity Verification Using Camera Footage

To supplementally verify whether activities were carried out, we manually reviewed footage from cameras installed in the lab to check if activities related to the topic groups were actually conducted during the target period. The verification was limited to leisure and social interaction topics—specifically, Smash Bros., Catan, and Jenga. These activities involved physical game tools, making them relatively easy to identify from video. On the other hand, research- or consultation-related topics were excluded, as it was difficult to determine from footage whether such activities were taking place.

From the perspective of visit promotion through notifications, no clear emergence of new behavior triggered by the system was observed during the evaluation period. Most of the observed activities were conducted by members who already frequently visited the lab. Nevertheless, topics with a high number of activity instances indicated strong interest within the community and may serve as a starting point for engaging more members or promoting future actions.

The results also suggest that activity occurrence may be more influenced by temporal factors—such as the ease of conducting activities at certain times—than by who is present. Most activities took place around similar times, and participants were often already present before the activities began. Thus, it is likely that not only the simultaneous presence of members but also the suitability of the time slot and personal rhythms influence the likelihood of activity occurrence.

4.4.4 Analysis and Discussion of the Questionnaire Results

A questionnaire was conducted after the system introduction period. The purpose was to understand user's subjective evaluations of the system's usefulness and the notification feature, as well as their intentions for future use. The questionnaire consisted of 5-point scale ratings, binary (yes/no) questions, and open-ended responses. The 5-point scale questions asked whether users felt their visit frequency had increased and whether interactions with related members had grown due to notifications. The binary questions asked whether they noticed the notifications, changed their plans because of them, and whether they would like to continue receiving notifications.

Regarding visit frequency, 12 respondents reported no change, while 2 indicated a significant increase. The main reasons for the lack of change were that some respondents did not know how to act even when informed that others interested in the same topic would be present, some were already usually in the lab at the suggested times, and others only visited when they had a specific reason. On the other hand, those who reported an increase in visits attributed it to personal circumstances. One respondent mentioned being unable to come in March due to personal matters but actively visiting in April. Another stated that a paper deadline was approaching, which naturally led to more frequent visits. hese responses suggest that the system itself did not directly influence user behavior.

To foster interaction via notifications, user's voluntary actions play a crucial role. At the same time, a lack of visibility into interpersonal relationships was found to inhibit such actions. Regarding changes in interaction, 9 respondents expressed disagreement, 3 reported no change, and 2 agreed. Those who saw no increase in interaction mentioned that they were unsure who the relevant members were or felt that there was no clear reason to gather, even when notified. In contrast, the respondents who felt their interactions had increased noted that they had more opportunities to engage in activities related to the suggested topics or that the notifications provided them with a reason to invite others. These findings indicate that not all users take proactive steps to interact with others, and that support mechanisms are needed to reduce psychological barriers to action. To fully leverage the effect of notifications, it is important to implement features that make visible who shares the same topics of interest and is likely to be present at the same time. Furthermore, it is essential to design support that facilitates natural interaction regardless of user's level of initiative, such as by proactively suggesting actions that can serve as social triggers.

If the content of the notifications overlaps with user's routine presence patterns, their impact may be diminished. In fact, all respondents reported noticing the notifications, indicating that awareness itself was not an issue. However, none of the respondents reported changing their plans based on the notifications. One commonly mentioned reason was that the same topics and time slots were repeatedly suggested, which made the notifications feel ordinary rather than special.

User's reactions to the notifications indicated a high level of interest in the provided information. However, the effectiveness of the notifications was influenced by each user's stay patterns and the nature of the topic. When asked whether they would like to continue receiving notifications, eight participants responded positively. Their reasons included the low psychological burden of the notifications and a general curiosity about who might be present and when, as well as the fact that the notifications were not perceived as disruptive. In contrast, those who preferred not to receive notifications gave reasons such as already being present in the lab regularly, making the notifications unnecessary, or relying on natural conversation for consultation-related topics, which made advance knowledge of other's presence irrelevant to their behavior. The system is not intended for users who are frequently present, but rather for those who tend to visit infrequently. Therefore, the lack of perceived value among constantly present users is consistent with the system's original design. Nonetheless, depending on the nature of the topic, prediction-based notifications alone may not be sufficient. To support time-sensitive communication needs, an additional mechanism for real-time topic sharing may be beneficial.

Throughout the evaluation, several participants commented that the suggested activity windows were too wide. This was because the system treated all time slots in which the necessary number of members might be present as equally viable for activities. Consequently, the notifications included overly broad time frames that did not align well with the actual times when activities typically occurred, potentially diminishing their perceived accuracy and usefulness. This issue stems from the system's limitation: it predicts only entry and exit patterns and lacks access to historical records of actual activity times. As a result, the system was unable to identify the time slots that were most conducive to real activities and instead notified users of all periods that merely satisfied the minimum co-presence condition.

In summary, the proposed system did not induce a clear behavior change in terms of increasing visits. However, it demonstrated some effectiveness in prompting activity initiation among those already present. Thus, strategies for promoting visits and strategies for facilitating activities during presence should be treated as distinct objectives and considered separately in system design.

5 Conclusion

This study aimed to promote face-to-face communication in small-scale communities, such as university laboratories. To achieve this, we proposed and implemented a visit-promotion system that combines predictions of future room presence with shared topics of interest. The system analyzes past entry and exit logs to estimate when members are likely to visit, identifies overlapping time windows among users with shared interests, and sends notifications to encourage co-presence and interaction.

To evaluate its effectiveness, we analyzed changes in visit frequency and duration, monitored group activity via video recordings, and collected post-experiment feedback through questionnaires. While no clear behavioral changes were directly attributed to the notifications, some participants reported that the notification served as useful triggers for initiating activities. However, concerns emerged regarding uncertainty about who shared the same topics and a lack of clarity about what to do once gathered. These issues point to the need for visualizing social connections and clarifying group activity objectives.

Our findings also indicate that the triggers for visiting the lab and those for participating in activities are not always aligned. For visit promotion, personalized notifications that match individual schedules and interests are crucial. In contrast, activity facilitation requires timely and proactive suggestions during the stay to reduce users' hesitation and encourage spontaneous engagement. Future work should clearly separate these two support functions and explore methods tailored to each.

Regarding the topics selected for evaluation, we found that recreational themes—such as games and shared casual interests—were effective in fostering informal communication. They required no preparation and were familiar to most users, lowering the barrier to participation. However, some users showed limited engagement due to a lack of interest in the suggested topics or a preference for goal-oriented interactions. These findings suggest that while recreational topics were suitable for our system's purpose, more serious or task-focused themes may require additional support, such as clearer goals or role assignments.

In addition, mechanisms that visualize topic-sharing relationships among members are necessary. Some participants expressed anxiety due to not knowing who shared their interests. Making such connections visible could enhance users' psychological comfort and awareness of community ties.

It is also important to support real-time topic sharing and spontaneous communication. The questionnaire revealed the presence of time-sensitive topics that are not effectively addressed through predictive notifications. Integrating tools such as chat, voice, or screen sharing could enable more immediate and responsive interactions.

Finally, the notification time windows should be more precisely tuned. When the suggested time ranges were too broad, users found them less actionable. By incorporating real activity data, the system can narrow down effective time slots and improve practical value. Future development should aim for a flexible system that adapts notification content, timing, and support methods to each user's context and needs.

REFERENCES

[1] L. L. Martins, L. L. Gilson, and M. T. Maynard, “Virtual Teams: What Do We Know and Where Do We Go From Here?,” *Journal of management*, vol. 30, no. 6, pp. 805–835, 2004.

[2] M. Mortensen and P. J. Hinds, “Conflict and shared identity in geographically distributed teams,” *International Journal of Conflict Management*, vol. 12, no. 3, pp. 212–238, 2001.

[3] R. E. Kraut, R. S. Fish, R. W. Root, B. L. Chalfonte, I. S. Oskamp, and S. Spacapan, “Informal Communication in Organizations: Form, Function, and Technology,” *People's Reactions to Technology*, pp. 145–199, 1990.

[4] A. Mehrabian and M. Wiener, “Decoding of inconsistent communications,” *Journal of personality and social psychology*, vol. 6, no. 1, pp. 109–114, 1967.

[5] S. A. Al Saifi, S. Dillon, and R. McQueen, “The relationship between face to face social networks and

knowledge sharing: an exploratory study of manufacturing firms," *Journal of knowledge management*, vol. 20, no. 2, pp. 308–326, 2016.

[6] F. Naruse and K. Kaji, "Estimation of Person Existence in Room Using BLE Beacon and Its Platform," *Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICST*, vol. 240, pp. 251–257, 2018.

[7] J. B. Begole, J. C. Tang, R. B. Smith, and N. Yankelovich, "Work Rhythms: Analyzing Visualizations of Awareness Histories of Distributed Groups," *Proceedings of the 2002 ACM Conference on Computer Supported Cooperative Work*, pp. 334–343, 2002.

[8] A. Al-Molegi, M. Jabreel, and B. Ghaleb, "STF-RNN: Space Time Features-based Recurrent Neural Network for predicting people next location," *2016 IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 1–7, 2016.

[9] D. Kong and F. Wu, "HST-LSTM: A hierarchical spatial-temporal long-short term memory network for location prediction.," *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*, vol. 18, no. 7, pp. 2341–2347, 2018.

[10] H. Terashima, N. Tamura, K. Shoji, S. Katayama, K. Urano, T. Yonezawa, and N. Kawaguchi, "Human Mobility Prediction Challenge: Next Location Prediction using Spatiotemporal BERT," *Proceedings of the 1st International Workshop on the Human Mobility Prediction Challenge*, pp. 1–6, 2023.

[11] Tullio, Joe and Goecks, Jeremy and Mynatt, Elizabeth D. and Nguyen, David H., "Augmenting shared personal calendars," *Proceedings of the 15th Annual ACM Symposium on User Interface Software and Technology*, vol. 4, no. 2, pp. 11–20, 2002.

[12] Tanaka, Yuto and Fukushima, Taku and Yoshino, Takashi, "Docoitter: A Presence Display System Capable of Predicting Future In-the-room Information(in Japanese)," *Journal of Information Processing Society of Japan*, vol. 54, no. 9, pp. 2265–2275, 2013.

[13] A. Sugiyama, H. Egi, I. Takata, and K.-i. Okada, "Proposal of the method to prospect individual behavior considering conformity behavior(in japanese)," *Proceedings of the GN Workshop 2007*, vol. 2007, pp. 31–36, 2007.

[14] A. Obata and K. Sasaki, "OfficeWalker: a virtual visiting system based on proxemics," *Proceedings of the 1998 ACM Conference on Computer Supported Cooperative Work*, pp. 1–10, 1998.

[15] Nakano *et al.*, "The Traveling Café: A Communication Encouraging System For Partitioned Offices," *CHI EA '06*, pp. 1139–1144, 2006.

[16] Y. Chiba and K. Nishimoto, "An Intrablog-Based Informal Communication Encouraging System that Seamlessly Links On-Line Communications to Off-Line Ones," *IEICE TRANSACTIONS on Information*, vol. E90-D, no. 10, pp. 1501–1508, 2007.

[17] T. Matsubara, K. Sugiyama, and K. Nishimoto, "Raison D'etre object: A cyber-hearth that catalyzes face-to-face informal communication," *International Conference on Engineering and Employment of Cooperative Information Systems*, pp. 537–546, 2002.

Session 6:
Human-Computer Interaction and Education

(Chair: Kanae Matsui)

Rhythm Rally: Entertainment Sport Combining Table Tennis and Rhythm Game

Koki Fuseya†, and Katsuhiko Kaji†

†Graduate School of Business Administration and Computer Science, Aichi Institute of Technology
{b25722bb,kaji}@aitech.ac.jp

Abstract - This is an entertainment sport that combines table tennis and rhythm games using smartphones. The system uses accelerometers in smartphones attached to rackets to detect hits, sending the timing data to a host device. The host compares the timing with the music beats and provides feedback by changing the audio pitch-stable for “GOOD” and lowered for “MISS.” This rhythm-based interaction encourages consistent rallies and adds a cooperative gaming element. While the system is accessible to beginners for entertainment, its effectiveness as a training tool may require basic rally skills. Experimental results showed strong entertainment potential, suggesting the value of combining physical activity with musical feedback for engaging user experiences.

Keywords: Human-Computer Interaction, Ubiquitous Computing, Consumer Devices and Systems, Entertainment Computing

1 Introduction

In recent years, the rapid spread of smartphones has led to an increase in a variety of services and applications. Among them, there are those that make use of the sensors built into smartphones. For example, there are healthcare services using acceleration sensors and navigation services using GPS. As smartphones become more popular, more sensors are being installed and their performance is improving.

Recently, we have many opportunities to see IoT utilized in various fields due to the advancement of technology. IoT stands for Internet of Things and refers to the technology that connects various objects to the Internet. The IoT includes the means to extend convenience and functionality by attaching sensors not only to computers and other devices, but also to various other things in the world, and making use of the data obtained.

There are some cases in which the IoT is utilized in the field of sports. For example, in tennis, there is the Smart Tennis Sensor. The sensor is attached to the tip of the racket grip and senses practice. The sensing enables determination of swing speed and which part of the racket surface is hit by the ball. However, a dedicated device needs to be built to attach the sensor. In addition, most of the feedback is provided by a terminal that is separate from the tool. There is a limit to the amount of feedback that can be provided by simply installing sensors.

There is a new culture that combines sports and entertainment. There is slipper ping-pong, in which table tennis is played with slippers that we usually use, cap-throwing baseball using plastic bottle caps, and balloon volleyball, in which volleyball is played with balloons [1] [2] [3]. These approaches to adding changes to the tools used in sports are thought to lead to entertainment that entertains people.

As part of our approach, we are developing a smartphone table tennis that extends the entertainment value of table tennis by integrating a smartphone and a table tennis racket [4]. This is an entertainment sport that uses the sensors in the smartphone to make swing and hit judgments during a rally, and adds some element from the smartphone side according to the judgments to extend the entertainment of ping-pong. Such entertainment sports are designed so that even inexperienced players can enjoy them. Proposing a sport that enhances the entertainment value of table tennis and improves table tennis skills. Performing sensing with a smartphone and playing rhythm games using the information obtained from sensing.

The purpose of this research is to research an entertaining sport that can be enjoyed even by table tennis beginners, and to support practice that can also aim at improving technical skills to perform rallies at regular intervals. As an approach, we create an application that provides real-time feedback based on sensor data from a smartphone.

2 Related research

There are a wide range of examples of IoT applications. In the primary industry, efforts are underway to address issues such as the lack of successors due to the aging of the workforce and to improve work efficiency. In the fishing industry, Blue Ocean Gear has developed smart buoys equipped with GPS and sensors that can obtain location information in real time for the purpose of tracking the location of fishing gear and preventing its loss, enabling remote management of fishing gear. [5] Furthermore, research on extending the convenience and functionality of everyday tools is also progressing [6]. Prohaska et al. proposed SPICE, a projection-type interface that projects recipes and operation guides onto the work surface and tracks the user’s movements in real time, with the aim of improving efficiency and intuitive operability during cooking [7]. The system recognizes the user’s hand movements and the placement of utensils, and displays appropriate information according to the cooking process, thereby reducing cooking time and improving immersion in the task. In this research, we employ a smartphone as a sensor to be incorporated into the IoT and propose a method to provide feedback by taking advantage of its multifunctionality. This will enable a terminal that integrates sensing and feedback and open up new possibilities for the IoT.

There is existing research on sensing technologies for table tennis. Peter et al. attached inertial sensors to table tennis rackets and collected data on eight basic strokes from ten amateur and professional players [8]. This research demonstrated high stroke detection and classification accuracy. It showed potential applications in training, match analysis, match statistics display, and player training progress support. There is also research using piezoelectric sensors to classify the impact location on the racket surface [9]. This enables the player to develop the ability to instantly judge the course to be hit and to improve his or her ability to respond in real games. In this research, among the sensors installed in smartphones, the acceleration sensor is used for swing sensing in table tennis.

There is research that extends sports competitions to support physical activity and entertainment sports [10]. Arzehgar et al. used IMUs and fusion sensors to sense athletes' movements and physical conditions, and constructed a real-time model for injury prevention and recovery support [11]. This enables intervention and feedback to correct movements that are difficult to notice during competition.

There is research aimed at changing the trajectory of ping pong balls. Zhou et al. are developing a smart table tennis racket with controllable stiffness using anisotropic electrofluidic elastomers [12]. Anisotropic electrofluidic elastomers are molded under an electric field, forming a particle orientation structure that enhances surface polarization and local electric fields, achieving a high electrofluidic effect of 17,160% in shear mode. In this racket, applying an electric field reduces the ball release angle by 11% and increases speed by 2%, demonstrating its potential as a new tool for changing the trajectory of the ball. Morisaki et al. proposed a ping-pong system that changes the trajectory of a ping-pong ball by applying ultrasonic waves to it [13]. Their HoppingPong system sends ultrasonic waves to the ping-pong balls from the left and right sides of the ping-pong table to bend the balls' trajectories to the left or right. The user decides which direction to change using a game controller attached to the racket.

There is research on entertainment systems that extend our familiar objects [14]. Hirai et al. developed an interface system that allows users to enjoy scratch music like a DJ by utilizing the sound produced when rubbing the bathtub [15]. This project aims to transform everyday actions into creative musical experiences. In this research, we will continue our research as an entertainment sport that extends sports, and show a new direction in the fusion of smartphones and analog tools.

As research on sports training using rhythm, studies have been conducted on the effects of exercising while listening to music on performance, and how movements in time with rhythm can improve performance [16] [17] [18]. These studies show that music is an effective tool for supporting physical activities such as sports and improving rhythm and performance. In this research, we will combine the rhythm game and table tennis in the form of rallies at regular intervals to improve the stability of rallies and technical skills in play, and to make it an entertaining sport that even beginners can enjoy playing. Since it is difficult to visually check the evaluation part while playing table tennis, we will use auditory feedback by using audio during the game.

3 Ping × Phone

"Ping × Phone" is a project in which we are developing an entertainment sport that combines a table tennis racket and a smartphone. Figure 1 shows an overview of "Ping × Phone". The use of smartphones is justified by their built-in sensors, which enable simultaneous sensing and feedback.

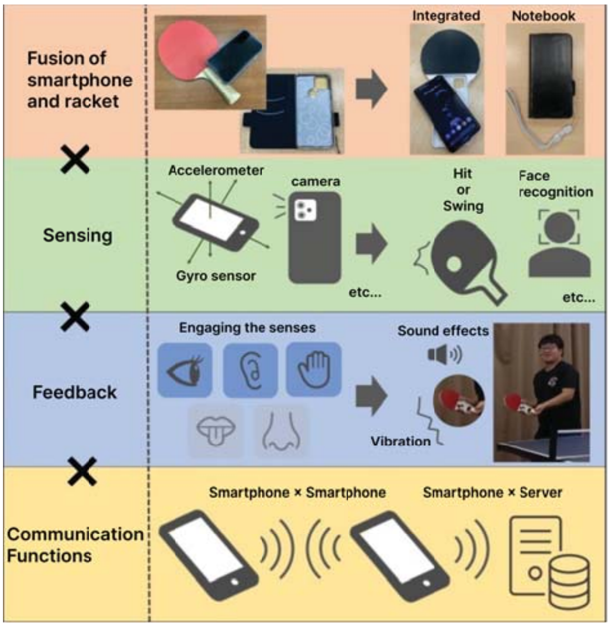


Figure 1: Diagram of of Ping × Phone
P

3.1 Fusion of Smartphone and Racket

First, we will consider whether to use the smartphone itself as a racket, whether to combine the smartphone with a ping-pong racket, and what kind of racket should be created to integrate the smartphone and racket. However, if the smartphone is used as a racket in its original form , there is a risk of screen cracks and other malfunctions, so a protective case is considered necessary. Therefore, as a method of integrating a smartphone and a racket, we considered the method of using the smartphone cover itself as a racket. It is easy to use the smartphone cover itself as a racket, and it can be expected to hit back and forth on both sides by using not only the case type but also the notebook type cover. However, all shapes of rackets have certain limitations and are difficult to be used in regular sports, so a new entertainment sport like table tennis is used as a way to take advantage of the shape. So far, we have created three types of cases: a nylon handbook type, a separate type, and an integrated type, as well as a wooden case. Figure 2 shows the rackets created.

3.2 Fusion of Smartphone and Racket

From the left side, there are a nylon pocketbook type, a separate type, an integrated type, and a wooden integrated case. First, a nylon pocketbook case was created. The organizer-type case is characterized by the fact that rubber is attached to both sides of the organizer-type smartphone cover so that

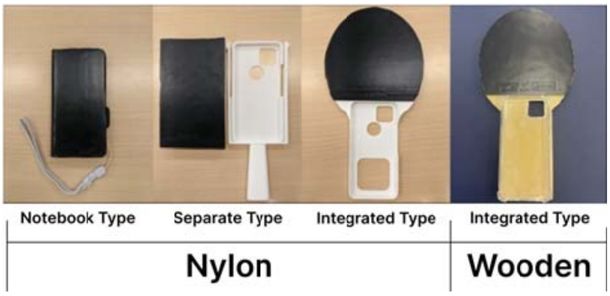


Figure 2: Fusion of smartphone and racket

it can be used as is. To prevent the case from slipping and flying off, a hand strap is used to secure it in place. The advantage of the organizer-type case is that it is easy to use, and while the case-type cover can only be used on one side, the organizer-type cover can be used on both sides. However, the disadvantage is that the shape of the smartphone remains unchanged, making it difficult to hit the screen.

To improve the difficulty of hitting, we next created a separable case made of nylon. The feature of the detachable case is that the smartphone is mounted on the hitting surface, leaving the grip part of the ping-pong racket. The advantage of the detachable case is that it looks like a feather board and has a grip, making it easier to hit and hold than a pocketbook case. However, although it is easier to hit, it is difficult to make fine adjustments when hitting, and because the surface itself is in the shape of a smartphone, the ease of playing ping-pong is not as good as it could be.

To realize the ease of playing ping-pong, we next created an integrated case made of nylon. The integrated case is characterized by the fact that the smartphone is mounted on the grip part, leaving the hitting surface of the ping-pong racket. The advantage of the integrated case is that the surface of the racket itself remains intact, making it easier to play table tennis, such as serving and spinning. We considered other ways to attach the smartphone, such as placing it below the racket or on the back side. However, attaching it below the racket caused the balance to shift, which made it difficult to play properly. Placing it on the back side would have made it impossible to use that surface during play, which could interfere with important techniques. For these reasons, we decided to attach the smartphone to the grip, where it would not get in the way of gameplay and could still be used for sensing. Although the thicker grip may suggest reduced usability, actual use revealed that players could hold the racket comfortably without any major difficulty. However, the racket has the disadvantage that the overall balance of the racket is not good because the resin density of the face of the racket is large and that of the handle is small.

Therefore, a wooden one-piece case was created. The feature of the one-piece wooden case is that it is close to the original table tennis racket. The advantages of the one-piece wooden case are that it is less expensive than the one-piece wooden case, the racket is better balanced as a whole, and the handle is thicker and easier to hold. Considering the above points, the wooden one-piece case will be used in this project, considering the ease of playing ping-pong as the first priority.

3.3 Sensing

In this project, it is necessary to create entertainment using the functions in smartphones. In this research, we use sensors that can be connected to entertainment. To realize an algorithm that classifies table tennis racket swings and hits, an acceleration sensor that can measure changes in the orientation and movement of the smartphone is necessary. In addition, an algorithm that can make real-time decisions on swings and hits during a table tennis rally is needed. The algorithm should be able to discriminate between a hit and a swing in a rally. In order to determine whether a rally is a hit or a swing, it is necessary to extract high-frequency and low-frequency components using a filter. To enable real-time judgment, a butterworth filter, one of the IIR filters, is used as a low-pass filter and a high-pass filter. The cutoff frequency of the low-pass filter is set at 3 Hz and that of the high-pass filter at 13 Hz. The raw data acquired by the accelerometer is divided into data processed by the low-pass filter and the high-pass filter, respectively. The data are then converted into a norm. The low-pass filtered norm is used to determine swing. The high-pass filtered norm is used to determine the impact. When playing ping-pong, it is normal to swing and hit back a ping-pong ball. Therefore, a hit is judged to be a hit when an impact is judged at the time of swinging and timing of judgment. This enables real-time swing sensing with a smartphone.

To make use of this sensing, a smartphone application is implemented to realize real-time processing of the judgment. The acceleration sensor values are used to identify swings and hits in table tennis. For the hit judgment in a rally, the actual hit is determined by the impact of a ping-pong ball on the smartphone during the rally. The actual hit decision also includes the mere impact. For example, there is the impact of a ping-pong ball hitting the racket while the racket is on the table. In this case, there is no actual hitting in the rally, but there is an impact from the ping-pong ball. In this project, the actual hit decision is implemented as an impact decision, using an algorithm in which both the swing and impact decisions are made to determine the hit.

3.4 Feedback

Real-time feedback is needed when taking action during table tennis play. The advantages of using smartphones are that they have major functions such as sound, camera, communication, and display, which can be utilized for entertainment. In addition, smartphones can be used to engage three of the five senses: sight, hearing, and touch. Feedback is expected to be provided to the sense of sight through the use of visual effects, to the sense of hearing through the production of sound, and to the sense of touch through the generation of vibrations. We propose a variety of entertainment by combining these three senses. We examined feedback that can act on the five senses using the functions of a smartphone. Table 1shows examples for consideration.

Table 1: Example of feedback from sensing a table tennis racket using a smartphone

Examples of working with the senses	Working with the senses	Feedback
Speakers	Hearing	Sound effects Rhythm game
Display	Visual	Screen effects Direction during swing
Vibration	Tactile	Racket Vibration

By combining hearing with sound as a function, feedback such as sound effects can be expected. By combining hearing with sound as a sensor, we can expect speech recognition using a microphone. By combining vision and a camera, we can expect face and ball recognition using an out-camera sensor. The combination of vision and display is expected to provide functions such as effects on the screen and targeting. We believe that the communication function can expand the range of these feedbacks. Real-time feedback will be realized by utilizing these features.

3.5 Communication Functions

There are two main types of communication possible with smartphones. The first is communication between the server and the smartphone. In server communication, for example, sensing data can be accumulated and used for analysis, and data such as how much a user has played can be stored. Since this project is being carried out as an entertainment project, we thought that the accumulation of data is unlikely to lead to entertainment. Second, there is communication between smartphones. Table tennis is basically played by two or four players, so if smartphones can communicate with each other, they can share information. This information sharing enables the implementation of entertainment sports that allow players to accomplish something in cooperative table tennis, or competitive entertainment sports that allow players to interfere with their opponent's play. One method of communication is Nearby Connections. It is a peer-to-peer networking API that allows apps to discover, connect, and exchange in real-time with nearby devices, regardless of network connectivity. Nearby Connections enables offline peer-to-peer advertising, discovery, and connection between nearby devices. Table tennis requires low-latency communication speeds due to the high rally speeds, and Nearby Connections enables fast and secure data transfer. The project is also suitable for locations where table tennis is played that do not require network connectivity, such as gymnasiums and Japanese inns.

3.6 Applications created so far

There are three applications created so far in the Ping × Phone project.

3.6.1 Nabeatsu Rally

Nabeatsu Rally is a cooperative entertainment sport based on a story by the comedian "The World Famous Nabeatsu[4]. We thought that the World Famous Nabeatsu's story "I become an idiot when the number is a multiple of 3 and the number with 3 is " was a good match as an approach to make people naturally want to continue the rally. Figure 3 shows an overview of the Nabeatsu rally.

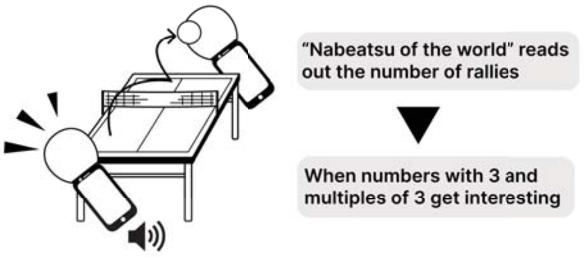

"Nabeatsu of the world" reads out the number of rallies
When numbers with 3 and multiples of 3 get interesting

Figure 3: Diagram of the Nabeatsu Rally

The number of rallies is counted by sensing, and feedback is given in the form of the World Famous Nabeatsu reading out the number of rallies from the speaker of a smartphone. The reading out of the rally is done in a way that is stupid when the number of rallies is a number with 3 or a multiple of 3, so the fun of the simple sounds, the goal of how far to go, and the sense of accomplishment of having made it that far are not found in ordinary rallies, leading to entertainment. The goal is to keep the rally going until the 40th turn, because from the 30th to the 39th turn, all 3s are marked, which is a particularly enjoyable timing in Nabeatsu rally.

3.6.2 Shout Smash

Shout Smash is a cooperative entertainment sport in which players return a ball that is easy to hit after five rallies, and the opponent shouts and hits a smash [4]. Figure 4 shows an overview diagram of Shout Smash.

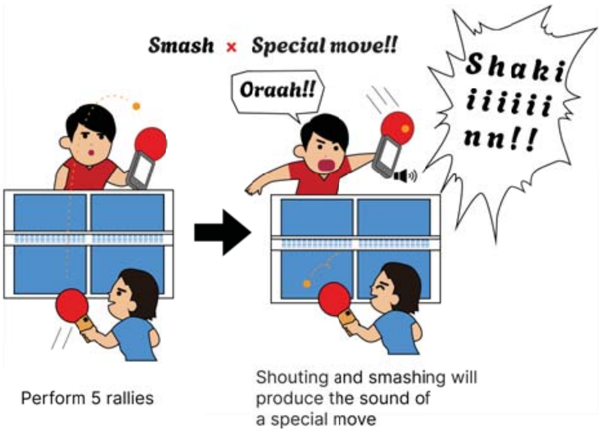

Smash × Special move!!
Oraah!!
Shakiiiiiiinn!!
Perform 5 rallies
Shouting and smashing will produce the sound of a special move

Figure 4: Diagram of the Shout Smash

The volume of the shout is acquired by a microphone, and if the volume of the shout exceeds a threshold value, the sound effect of a special move is played from a smartphone. If the sound effect of the special move is played, it is a success. If the voice is too quiet, the rally is unsuccessful, and a sound reminiscent of a failed special move is played. In addition to the accelerometer used to count the rallies, a microphone is used for voice recognition. The game is a cooperative entertainment sport that can be enjoyed by table tennis beginners because the difficulty level is relatively easy: a smash is made after five rallies. The non-smash player feels a sense of accomplishment by cooperating to return a ball that is easy for the opponent to return, while the smash player feels a sense of exhilaration by making a special move.

3.6.3 Facepong

In addition to conventional table tennis, facepong uses a smartphone attached to the racket and pointed at the opponent's face during table tennis to perform facial recognition [4]. Face recognition refers to the automatic detection and recognition of faces from the camera image. Figure 5 shows an overview of facepong.

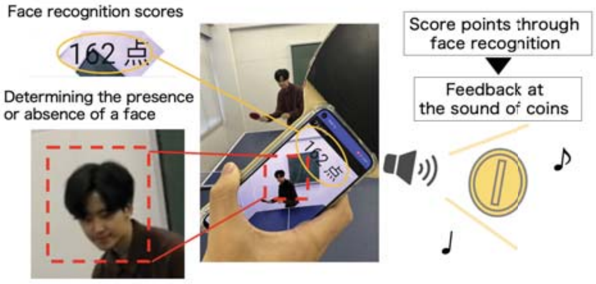

Face recognition scores
162 点
Determining the presence or absence of a face
Score points through face recognition
Feedback at the sound of coins

Figure 5: Diagram of Facepong

The racket fused with a smartphone has a hole for a camera, so the camera recognizes the opponent's face through the hole. Each time an opponent's face is captured in table tennis, one point is added to the face recognition score. At this time, a rally in table tennis is played in parallel. When a face is successfully recognized, feedback like the sound of a coin being dropped is emitted from the speaker of the smartphone in accordance with the score addition. The game is exhilarating when the player scores points by face recognition in succession. Unlike the other projects, facepongpon is a competitive entertainment sport for intermediate players that requires a stable rally. Since rallies are not counted in facepong, the acceleration sensor is not used, but the camera image sensor is used for face recognition.

The calculation method of the score is the multiplication of the table tennis score and the face recognition score. The reason for using multiplication as the calculation method is to create the possibility of a reversal due to score inflation when the player loses in the table tennis score but wins in the face recognition score, leading to an uplifting playing experience. In addition, the incorporation of face recognition creates new tactics that do not exist in original table tennis, such as hiding the face with the racket to prevent the opponent from recognizing the face, or returning the ball loosely to buy time for face recognition.

4 Rhythm Rally

This chapter describes Rhythm Rally, one of the Ping × Phone projects and an application created in this research.

4.1 System Requirements

The purpose of this research is to examine an entertaining sport that can be enjoyed even by table tennis beginners, and to propose a practice support method that aims to improve the technical skill of table tennis players who can keep rallying at regular intervals through this sport. We devised an entertainment sport that not only provides entertainment but also improves the technical skill of rallying at regular intervals. As a method of support, an entertainment sport in which the timing of the return of the ball is specified according to the rhythm was considered. If players can rally at regular intervals in time to the rhythm, they will naturally become better at rallying, and this will lead to improvement of their technical skills. We propose a new approach that combines table tennis and rhythm games as a sport that can be used for both entertainment and exercise support in the form of cooperative rallies timed to music. The feature of our research project is a rhythm game-like ping-pong game that can be played using a racket integrated with a smartphone. Figure 6 shows a schematic diagram of Rhythm Rally.

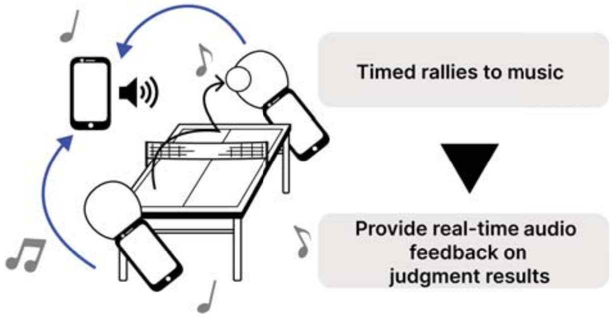


Figure 6: Diagram of of Rhythm Rally

The main devices used are two smartphones (client devices) and two racket cases for playing table tennis used in Ping × Phone, and one smartphone (host device) for outputting audio and displaying results in the rhythm game. Music in the rhythm game is required. We first considered an approach in which the game is played with only two devices. However, when music is played on only one of the devices, there is a noticeable difference in the playing environment. If music is played on both devices, the music will be heard twice, which may affect the play. With these considerations in mind, a separate smartphone for music playback should be provided. The rhythm game was implemented as a simple rhythm game in which two types of evaluations were made: “Timing was good” and ”Timing was not good. The purpose was to make it easy for anyone to play, with an emphasis on ease of use for beginners.

Feedback is indispensable in rhythm games in order to grasp the results of decisions. However, visual feedback, such as displaying the result on the screen, is very difficult in table tennis rallies. Therefore, in this research, auditory feedback using voice is used. The reason for using audio feedback is that the auditory sense is used to listen to music in rhythm games, so we thought that auditory feedback would be appropriate. The addition of auditory information is also appropriate in that it does not significantly interfere with the play of the game.

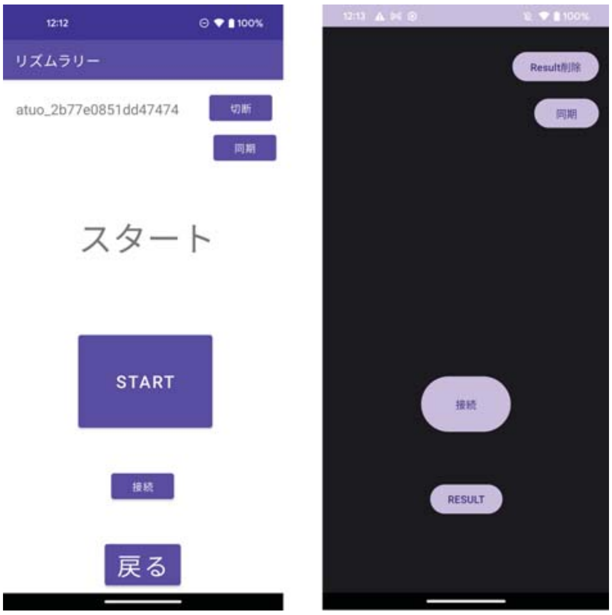


Figure 7: Rhythm Rally application screen

4.2 Implementation of Rhythm Rally

In the implementation of the application, hit judgments are made using the acceleration sensor, and the current time and device ID at the time of the hit obtained from the client device are sent to the host device to judge the rhythm game. The application screen of the client device and the application screen of the host device for sound source playback are shown in Fig. The client device is on the left and the host device is on the right. Figure 8 shows the rhythm rally flow for each device.

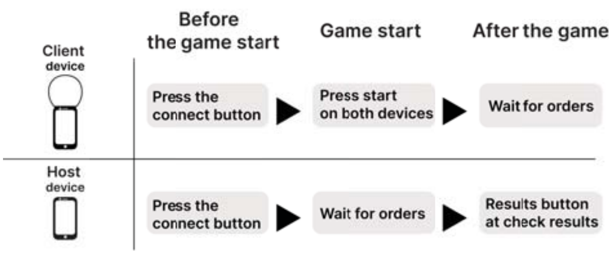
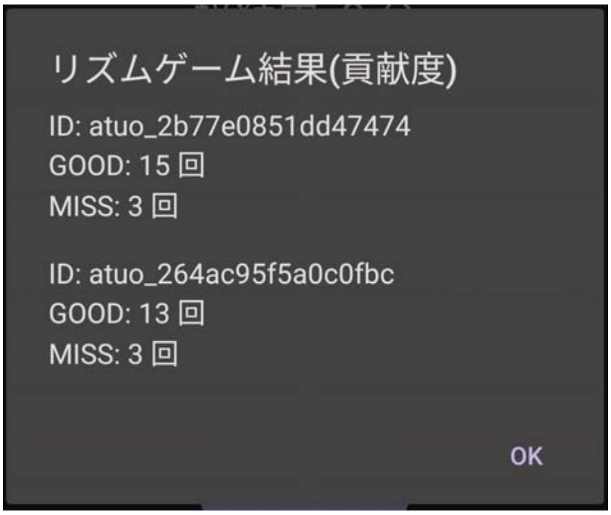


Figure 8: Overall Rhythm Rally per device

The racket case used in the implementation of is a wooden one-piece case for ease of use and cost. It is necessary to add an element of entertainment to the rally by using a smartphone to keep track of the rally during the rally. Therefore, hit detection by the acceleration sensor is performed on the client device used for ping-pong. When a hit is detected on the client device, the current time and device ID are sent to the host device for audio output and result display. In addition to music output, the host device also performs judgment processing for rhythm games. The judgment process mainly acquires the current time at the timing specified for the rhythm game and calculates the difference from the current time received from the client device. The difference between the current time and the received time from the client device is used for judgment. Judgment of the rhythm game is made based on the value of the difference in time. The judgment result is saved for each ID of each device, and the number of judgments is visualized by ID as shown in Figure 9 when the result button on the host device is pressed.



リズムゲーム結果(貢献度)

ID: atuo_2b77e0851dd47474
GOOD: 15 回
MISS: 3 回

ID: atuo_264ac95f5a0c0fbc
GOOD: 13 回
MISS: 3 回

Figure 9: Rhythm Rally results screen

Communication between smartphones in a rhythm rally is also essential to produce judgment results in a rhythm game. Figure 10 shows an overview of the communication between devices in the system.

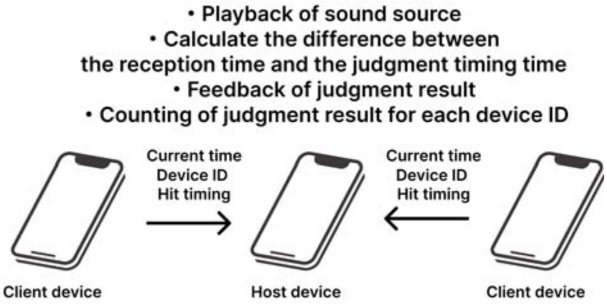

• Playback of sound source
• Calculate the difference between the reception time and the judgment timing time
• Feedback of judgment result
• Counting of judgment result for each device ID
Current time
Device ID
Hit timing
Current time
Device ID
Hit timing
Client device
Host device
Client device

Figure 10: Schematic diagram of communication between devices in a rhythm rally

The connection between smartphones is implemented using Nearby Connections for each of the two client devices and the host device. Although BLE and Wi-Fi Direct are other connection methods, Nearby Connections is used in this research because of its high communication speed and network-independence. Both the client device and the host device have buttons for connection, which are pressed simultaneously to connect the devices. When the connection is completed, the client device can visually confirm whether it is connected and the host device can visually confirm the number of devices that are currently connected. When the start button on each client device is pressed, a start signal is sent, and when the host device receives two start signals, the game begins. When a hit decision is made, the current time and the ID of the device are sent to the host device. While the sound source of the rhythm game is being played, a sound other than the sound source is played at a constant rate in accordance with the timing of the beats. Table 2 shows the information required for rhythm rally and how it can be used.

Table 2: Necessary information and its use in rhythm rallies

Needing information	how to use it
BPM	beat timing
Start of sound source	start of beat timing placement
End of sound source	end of beat timing placement

For the timing of beats, if the timing of the start and end of music playback in the sound source and the BPM of the song are specified on the algorithm side, the timing of a fixed interval is automatically adjusted to match the start of the music, resulting in a sound that matches the timing of the beats. The beat timing is automatically adjusted based on the BPM of the sound source and the start and end information of the sound source. The timing of the beats is automatically adjusted based on the BPM of the audio source and the start and end information of the audio source. The judgment algorithm used in rhythm games is shown in Figure 11.

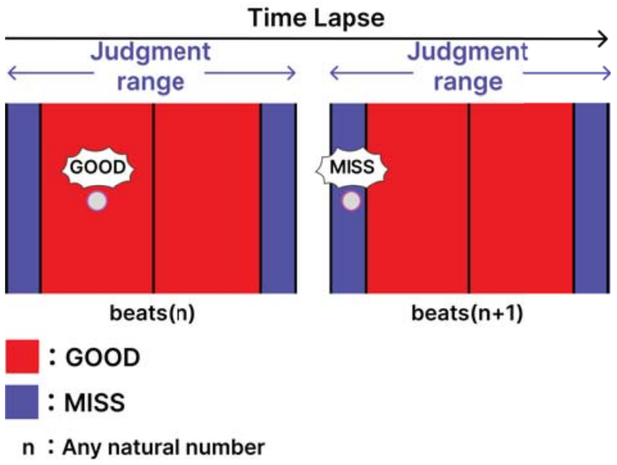


Figure 11: Judgment Algorithm

The circles in Figure 11 represent the difference in current time. If the difference between the current times falls within the red box in Figure 11, the rally in that cycle is a “GOOD” success, and if the difference between the current times falls between the blue boxes, the rally in that cycle is a “MISS” failure. In Figure 11, there is a little space between the judgments, but in reality, this algorithm is executed continuously. If a “MISS” decision is made, the players adjust the speed of the rally to return to the “GOOD” decision cycle, and cooperation unique to Rhythm Rally is born. If multiple data are received within a single judgment cycle, only the first ID received and the current time are saved, and subsequent data are ignored. If the received ID is not in the list, the host device initializes and stores it, and counts the number of “GOOD” or “MISS” responses for that ID. If only the timing of a beat is “GOOD” for a single judgment timing, the rhythm game will become too difficult. Therefore, a judgment range with a certain degree of grace is provided. If a hit judgment signal is not received from the client device for one judgment cycle, the judgment cycle is skipped.

4.3 Feedback

In rhythm games, feedback methods based on judgment results are important. Therefore, we examined two types of feedback methods. The first one is a real-time feedback method using voice according to the judgment result. The feedback method uses a pleasant sound for a “GOOD” judgment, and a sound that evokes the image of failure for a “MISS” judgment. It is thought that this desire to maintain a sense of accomplishment will motivate the player, and furthermore, it is expected to be effective in supporting practice. This feedback is thought to allow users to quickly understand the discrepancy between their own rally timing and the specified timing, and to correct the discrepancy when they cooperate to match the timing of their rallies. Based on this, we asked six men in their 20s to actually experience Rhythm Rally and to respond to a questionnaire. The “Oklahoma Mixer” and “Mime Mime,” which are songs often used at campfires and other events, were used for implementation in this evaluation experiment. The results of the questionnaire are shown in 3.

Table 3: What was it like to play Rhythm Rally?

Evaluation Items	Results
Enjoyment of Rhythm Rally	4.5/5.0 points
Feeling of Improvement in Rally	4.17/5.0 points

As a result, the possibility of entertainment and practice support was demonstrated, but many issues were found in the implementation method. Regarding the combination of ping-pong and rhythm games, there were many positive comments that even beginners can enjoy the game, such as “the rhythm of the rally music is easy for beginners” and “it was interesting” and “it was fun”. There were also comments that indicated the possibility of practice support, such as ”I feel like I am getting better at certain rallies. On the other hand, there were some opinions about the implementation method, such as “it is difficult to know when to start a rally” and “it is difficult to know when to hit the ball” that made it difficult to play the game as a rhythm game. Among them, some players commented that they felt that the visual feedback method of results visualization was connected to the improvement of their technical skills, saying, ”I feel that I am getting better every time I check the results. On the other hand, there was room for improvement in the audio feedback, with comments such as “It would be nice if the client device could play the sound of the judgment result” and “The sound of the judgment result is part of the music and it is hard to hear”. In addition, a fundamental problem was the accumulation of sound discrepancies over time.

In order to solve this problem, a different feedback method was considered. As a feedback method, we considered changing the pitch of the sound source depending on the judgment result. When a “MISS” decision is made, the pitch of the main sound source is lowered by one step. The pitch of the sound source is lowered in five steps in proportion to the number of “MISS” results. If a “GOOD” decision is made, the pitch returns to its initial state. When a “MISS” decision is made and the pitch of the sound source is changed, the player feels discomfort due to the change in the sound source. In order to avoid this discomfort, the player will be aware of the cycle of the “GOOD” decision and rally. If players can rally with the cycle in mind, they will be able to rally at regular intervals, which will support their practice. Since the sound sources are affected by their own play, rallies in which the players try to keep the pitch of the sound sources constant can be entertaining. This method eliminates the need to simultaneously output two sounds, one for the main sound source and another for the judgment result, and thus solves the problem of sound shift. In addition, the change in the pitch of the sound source makes it easy to understand the results of one’s own judgment during a rally in real time.

We conducted an evaluation experiment again to investigate the entertainment value of this rhythm rally and its potential as a practice aid, after modifying the feedback method. We asked eight men and women in their 20s to actually experience the Rhythm Rally and to answer a questionnaire. The results of the questionnaire are shown in Figure 4.

Table 4: What was it like to play Rhythm Rally?

Evaluation Items	Results
Enjoyment of Rhythm Rally	4.5/5.0 points
Feeling of Improvement in Rally	4.0/5.0 points

As a result, the problem of feedback sound misalignment was solved without loss of entertainment value, but the change in pitch caused the sound source to sound out of sync. Although the tempo of the sound source is actually constant, if the player perceives a shift in tempo, it may interfere with play. In addition, the more the player perceives the tempo shift, the more difficult it becomes to use the game as a practice aid, and therefore, it is necessary to improve the game. The possibility of entertainment was also demonstrated by the results of the questionnaire. The feedback method for judging the pitch was divided, with some saying that “changing the pitch is innovative and interesting” and others saying that ”the previous method was more like a rhythm game. Through the initial evaluation and re-evaluation experiments, we were able to demonstrate the entertainment potential of both feedback methods. However, it is difficult to find a method that can be used by beginners to improve their skills and to support their practice, and there are some cases where the feedback method can be used to support practice.

5 Conclusion

5.1 Summary

In the Ping × Phone project, this research examined a rhythm game smartphone application that obtains the current time using hit judgments obtained from a ping-pong racket and a smartphone’s acceleration sensor, and judges the rhythm game based on the difference between the current time obtained at the timing of the judgment. Rallies were performed at regular intervals in time with music, with real-time feedback to the auditory sense through judgments and visualization of the results through results. The timing of the rally is automatically arranged by specifying the BPM of the sound source and the start and end timing of the sound source. In the implementation of feedback, we proposed a form that provides audio feedback for each decision and a form that changes the pitch for each decision. As a result, the possibility of entertainment was demonstrated with both types of feedback, but opinions were divided on the feedback method. In addition, it is necessary for beginners to be able to rally to some extent in order to use the system as a practice support.

5.2 Future Works

There are several challenges for Rhythm Rally going forward, one of which is the need to have more people actually experience it and evaluate it. For that reason, it is necessary to define evaluation criteria and have people with various attributes experience and evaluate the system.

Since the evaluation experiment was conducted using only one song, there is concern that repeated experiences may become less effective due to familiarity or boredom. To address these concerns about user retention, it is necessary to prepare multiple songs, expand the entertainment value, implement features that allow difficulty levels to be adjusted according to the player’s skill level, and create functionality that can be enjoyed by a wide range of people. When using the system for practice support, assuming that there are players who wish to have stricter judging criteria, it may be necessary to make the judging criteria as detailed as “GREAT”, “GOOD”, ‘BAD’, and “MISS”.

There are also issues with the implementation of the system. Since many problems arose with the feedback method, it is necessary to consider a method that can be used without affecting play so that it can actually be used as a practice aid, and to consider new approaches such as that allows players to select a feedback method according to their preferences. Since the accuracy of hits in rallies also lacks precision, it is necessary to improve the accuracy of hit judging by extracting environmental sounds using FFT and using the sound of the hit as the judging criterion.

REFERENCES

[1] Slipper Table Tennis Championship Executive Committee, Slipper Table Tennis. https://www.slipper.yokohama (reference 2025-01-20).(in Japanese)

[2] Cap Baseball Information Bureau, Cap Baseball (Cap Throwing) Information Bureau. https://cap-baseball.com (reference 2025-01-20).(in Japanese)

[3] Tokyo Sports Association for Persons with Disabilities, Balloon Volleyball. `https://parasports-start.tokyo/sports/s35/` (reference 2025-01-20).(in Japanese)

[4] Yusuke Ueji, Katsuhiko Kaji: Ping × Phone: a proposal for entertainment sports by integrating a ping-pong racket and a smartphone. In *Symposium on Multimedia, Distributed, Cooperative and Mobile 2024*, Vol. 2017, pp. 958-967, (2024).(in Japanese)

[5] NOAA TECHNOLOGY PARTNERSHIPS OFFICE, Blue Ocean Gear Smart Buoy. `https://techpartnerships.noaa.gov/blue-ocean-gear-smart-buoy/` (reference 2025-06-08)

[6] Thomas Kosch, Albrecht Schmidt: Enabling Tangible Interaction through Detection and Augmentation of Everyday Objects. In *International Workshop on Interacting with Smart Objects*,pp. 8-13, (2019).

[7] Vera Prohaska, Eduardo Castelló Ferrer: SPICE: Smart Projection Interface for Cooking Enhancement. In *ACM International Conference on Intelligent User Interfaces*, pp. 1-10, (2025).

[8] Peter Blank, Julian Hoßbach, Dominik Schuldhaus and Bjoern M. Eskofier: Sensor-based stroke detection and stroke type classification in table tennis. In *Proceedings of the 2015 ACM International Symposium on Wearable Computers*, pp. 93-100, (2015).

[9] Guo Tian, Weili Deng, Yuyu Gao, Da Xiong, Cheng Yan, Xuebing He, Tao Yang, Long Jin, Xiang Chu, Haitao Zhang, Wei Yan and Weiqing Yang: Rich lamellar crystal baklava-structured PZT/PVDF piezoelectric sensor toward individual table tennis training. *piezoelectric sensor toward individual table tennis training*, Nano Energy, Vol. 59, pp.574-581, (2019).

[10] Craig Wisneski: Pingpongplus. In *ACM SIGGRAPH 98 Conference abstracts and applications*, p. 111, (2011).

[11] Afrooz Arzehgar, Seyedeh Nahid Seyedhasani, Fatemeh Baharvand Ahmadi, Fatemeh Bagheri Baravati, Alireza Sadeghi Hesar, Amir Reza Kachooei and Shokoufeh Aalaei: Sensor-based technologies for motion analysis in sports injuries: a scoping review. In *BMC Sports Science, Medicine and Rehabilitation*, Vol.17, Issue 1, Article No. 15, (2025)

[12] Xuefeng Zhou, Lijuan Wang, Dongyang Huang, Yudai Liang, Quan Shi, Hong Yaying, Mengying Zhang, Huayan Pu, Weijia Wen, Jinbo Wu: Smart Table Tennis Racket with Tunable Stiffness for Diverse Play Styles and Unconventional Technique Training. In *Advanced Materials Technologies*, Vol. 6, Issue 10, Article No. 2100535, (2021).

[13] Tao Morisaki, Ryoma Mori, Ryosuke Mori, Yasutoshi Makino, Yuta Itoh, Yuji Yamakawa, and Hiroyuki Shinoda: Hopping-pong: Changing trajectory of moving object using computational ultrasound force. In *ACM International Conference on Interactive Surfaces and Spaces*, pp. 123–133, (2019).

[14] Fahim Kawsar, Kaori Fujinami and Tatsuo Nakajima: Augmenting Everyday Objects– A Reflection on Reuse as an Interface Design Method. In *MuC'19 Workshops*, pp. 113-117, (2019).

[15] Shigeyuki Hirai, Yoshinobu Sakakibara and Seiho Hayakawa: Bathcratch: Touch and Sound-Based DJ Controller Implemented on a Bathtub. In *International conference on Advances in Computer Entertainment*, pp.44-56, (2012).

[16] N. Illias, R. Adnan, V. Ross, N. Sulaiman, and M. Appukutty: The effects of musical tempo during cycling exercise among undergraduate sport science students. *13th AFSM Congress*, pp.95-106, (2013).

[17] J. Pates, C. Karageorghis, R. Fryer, and I. Maynard: Effects of asynchronous music on flow states and shooting performance among netball players. *Psychology of Sport and Exercise 4*, pp.415-427, (2003).

[18] M. Söğüt, S. Kirazci, and F. Korkusuz: The Effects of Rhythm Training on Tennis Performance. *Journal of Human Kinetics volume 33*, pp.123-132, (2012).

Empowering End-Users with Generative AI: A New Paradigm for Tool Development in Higher Education Portals

Akihiro HAYASHI[†], Michi KOMURA, Midori ISHIHARA, Hajime KANEKO
Department of Information Design, Shizuoka Institute of Science and Technology, Japan
Osaka Metropolitan University, Japan
ATAS Laboratory, Japan
pixysbrain@gmail.com, mitikomm347@gmail.com, Midoripapa@gmail.com, kaneko68@gmail.com

Abstract - End-User Computing (EUC), first introduced in the late 1980s through tools such as Microsoft Excel, has long empowered non-technical users to create customized solutions. Despite its potential, the widespread adoption of Visual Basic for Applications (VBA) has been hindered by its steep learning curve. The emergence of generative artificial intelligence (AI) in 2022 has transformed this landscape, enabling users to generate source code and develop tools without prior programming knowledge. This study explores the extent to which end-users can leverage prompt engineering in conjunction with generative AI to construct functional tools. We applied the EUC paradigm to routine administrative tasks within student portal systems at multiple universities, aiming to simplify complex workflows. By automating these tasks into one-click operations using VBA, we demonstrated that many routine functions can be implemented with minimal code modification. Our findings indicate that integrating EUC with generative AI significantly improves operational efficiency and reduces the technical barriers associated with tool development.

Keywords: EUC, VBA, Generative AI

1 Introduction

End-user computing (EUC) began to be advocated in the late 1980's and 1990's. EUC refers to activities in which end-users of an organization or business improve their daily operations by creating and managing their own tools and solutions, without relying on the IT department. The main targets of EUC are spreadsheets, simple databases, and no-code/low-code application development tools that are used on a daily basis.

Around the time that EUC was proposed in 1990, Microsoft Excel was released. Excel gained many users because it had a user-friendly interface that was easy to use despite its advanced functions and high performance. It replaced spreadsheet software such as Lotus 1-2-3, which was used before 1990, and relational databases, and came to be used as the de facto standard for database applications. Recently, Excel has become compatible with Apple computers running MacOS, so its uses have become even more widespread.

In addition, in enterprise applications used by universities, the final output is exported to Excel and utilized. Recently, manaba of Asahinet, Active Acacemy Advance of Densho, and University Passport RX of Nihon System Technology are being used. These applications are output to Excel for use of end users.

Another important aspect is that many of the daily tasks operated by universities and other institutions are routine work, such as daily, weekly, monthly, quarterly, and annual operations. In the case of a university, these include the registration of new students for the new academic year, registration for the first and second semesters, grading and grading of reports, rubric evaluation at the end of the academic year, and evaluation of students for promotion. These operations are repeated at regular intervals. The routine work is the same even when the person in charge is replaced.

In other words, if most business data is managed in Excel and most work is routine work, once routine Excel works are made more efficient with EUC, the benefits of efficiency can be enjoyed continuously thereafter. A dramatic increase in productivity of white-collar work can be expected. In order of simplicity, the most common uses of Excel are spreadsheets calculation, functions, pivots, and VBA (Visual Basic for Applications). Spreadsheet calculation and Pivot do not require programming. Everyone can use them once they become familiar with them. There are many Excel functions available, but when you can remember even just the 60 most commonly used functions, you will have no trouble with your everyday Excel work.

However, VBA is Microsoft's Basic language, and it was not easy for end users with insufficient programming literacy to develop the desired tools. This is because VBA is difficult for end-users. In other words, the lack of VBA literacy among end users is thought to have been the cause of the difficulty in realizing EUC based on Excel.

In this study, we propose a method to realize EUC using Generative AI proposed in 2022. Generative AI is a field of artificial intelligence that uses algorithms and models to generate new data and content. Generative AI is used for various purposes such as idea generation, text summarization, and translation into foreign languages. As a result, end users with insufficient programming literacy can use VBA source code if they make good use of the Generate AI, eliminating an unachievable bottleneck in EUC.

Therefore, this study evaluates the possibility of EUC realization by assuming daily Excel work in a university institution and improving daily work processes by general end-users using the generated AI.

As for the previous studies in this research field, Yaghoe

[1] stated that the opportunity to discover the strategic use of information technology is not provided by management strategy theorists or analysts, but by the users of information technology who are performing their tasks in various departments of business management, and that EUC is significant as an emergent method for this purpose. The significance and importance of EUC in the strategic use of information technology was pointed out by Ahnaf Chowdhury Niloy et al [2], who studied the factors that influence university students' use of ChatGPT in a triangulation approach combining exploratory, qualitative, and quantitative analyses. Based on existing literature and narratives from qualitative analysis, they quantitatively measured relevance using SEM techniques and found that it strongly influences students' intention to use and actual use.

Tianzhu Liang et al [3] investigated the evolutionary impact of AI technology on foreign language education. They explore the relationship between foreign language education and AI technology using complex systems thinking and analyze its evolutionary patterns in a dynamic model. Finally, they make recommendations to promote the development of foreign language education and AI technology Alejandro Barredo Arrieta et al [4] assessed the impact of artificial intelligence (AI) in education, which has been widely adopted by educational institutions, initially from computer technology to web-based educational systems, to The evolution to the use of humanoid robots and chatbots has allowed faculty to efficiently review and grade student assignments, improving the quality of education; they report that AI personalizes the curriculum and facilitates learning according to student needs, leading to an improved learning experience.

These previous studies point out the effectiveness of using AI for teaching and learning, but do not discuss the use of ChatGPT among end users who lack programming literacy.

In Chapter 3, we explain the procedure for creating Excel VBA macros by using the generative AI such as ChatGPT to improve the efficiency of the target's Excel work. In Chapter 4, we report on how three steps of work, mainly related to folders, are made more efficient by an Excel VBA program using a generative AI, based on the assumption of a white-collar worker at a university. In Chapter 5, we discuss the proposed method and its application cases, and finally, in Chapter 6, we discuss future issues.

2 Capability of Excel Literary and End Users

2.1 Human resources to realize EUC

In Japan, the bubble economy collapsed around 1991, and the lost 30 years have continued to the present. During that time, digitalization and networking progressed around the world, with the United States in particular making remarkable progress toward an ICT society. The U.S. in particular has made remarkable progress toward an ICT society, and has achieved historic prosperity, including the birth of tech giants such as GAFA. In the meantime, Japan has remained stagnant for a long period of time due to the delay in digitalization and the lack of digital literacy. According to the "DX White Paper 2023" published by the Information-technology Promotion Agency, Japan (IPA) in 2023, 10.9 % of respondents in Japan and 73.4 % in the U.S. answered that the "quantity" of human resources to promote DX was sufficient in the FY2022 survey. The percentage of respondents who answered that there is a "significant shortage" decreased from 20.9 % in the FY2021 survey to 3.3 % in the FY2022 survey in the U.S., while the percentage in Japan increased from 30.6 % in the FY2021 survey to 49.6 % in the FY2022 survey.

As for the "quality" of human resources to promote DX, the number of companies in Japan that responded that there is a "major shortage" increased from 30.5 % in the FY2021 survey to 51.7 % in the FY2022 survey, with half of the companies responding that there is a clear shortage. In the U.S., the situation is exactly the opposite, with 50.8 % answering "not inadequate" alone. Therefore, it is clear that digitalization must be promoted in Japan, and at the same time, digital human resources must be developed.

Two occupational categories of workers are used: white-collar and blue-collar. White-collar workers are those who work mainly in offices and business offices. It includes managerial, professional, clerical, and other jobs that are mainly knowledge work or desk work. Blue-collar workers (blue-collar) are those who work mainly in physical labor. It includes occupations that involve manual labor and machine operation, such as manufacturing, construction, and transportation. In this study, we focus on white-collar reskilling, since Microsoft Excel has been used as a standard and the ICT society has developed, it is fair to say that most white-collar work is done with Excel. This is because most of the data in companies are now managed by Excel. Most of the work in an organization is routine work. Therefore, if the productivity of Excel work can be improved, the productivity of white-collar workers in the entire organization can be expected to improve significantly.

In this study, the skills that white-collar workers themselves recommend for improving their work by using Excel VBA are called digital transformer skills. The above-mentioned IPA document also mentions the shortage of digital human resources. However, there is no shortage of digital human resources themselves, since almost all workers use Excel to process digital data on a daily basis. What is lacking is not digital talent but transformers.

2.2 Mismatch of training and human resources in a company

Until now, training provided by company organizations has included group training, on-the-job training, and mentoring. These efforts have achieved a certain level of success. However, even if these training programs enable employees to become proficient in using Excel on a daily basis, they have yet to develop digital transformers who can acquire Excel VBA programming skills and improve their daily work by themselves.

Before Windows 95 was released in 1995, Microsoft users were familiar with MS-DOS. MS-DOS commands to create folders. Everyone knew that the MS-DOS command for creating folders was mkdir. Later, when programming VBA, it was no surprise when mkdir appeared in the source code. However, after Windows 95 was released in November 1995, the general white-collar workers no longer use MS-DOS commands by themselves since Windows is used as an operating system and a graphical user interface (GUI) is used. As a result, the term "mkdir" does not ring a bell anymore. This is one of the reasons why end-users are not able to use VBA freely.

2.3 Issues to be solved

The problem to be solved is to establish a methodology to develop transformers by enabling today's white-collar workers to use Excel VBA and manipulate it without any MS-DOS or programming experience.

3 EUC realization method by intelligent white-collar using generative AI

3.1 3.1. Policy

Since the emergence of ChatGPT in November 2022, several companies and organizations have already released generative AIs. The fact that they are provided by commercial companies does not mean that their outputs are manipulated for profit, so there is basically no significant difference in the results regardless of which AI is used. In this study, we use ChatGPT with the GPT-4o model (short for "omni"), which is the latest version as of 2025, released by OpenAI in the U.S. as a generative AI.

In addition, the EUC targeted in this study covers daily operations by end-users. As mentioned earlier, most data is managed in Excel, and most tasks are routine. The Excel work used in the routine work refers to Excel work using a combination of Excel spreadsheets and functions that can be performed by end-users' literacy. Therefore, when outputting VBA source code by ChatGPT, even if the VBA source code generated by ChatGPT is not completely complete, we judge that the completion criteria for this study have been met if it can be made complete by modifying it with the end user's normal skills. Repeatedly prompting engineers for perfect output is a waste of time and effort.

3.2 Procedure

In this study, we propose a white-collar EUC realization method using ChatGPT according to the following procedure.

3.2.1 Assignment of Tasks

The tasks are those that are feasible by using spreadsheets and functions that are within the end-user's Excel literacy, but that would require a great deal of time and effort if done as-is. It is better to select operations with a large amount of data and a high frequency of handling, so that the end user can enjoy the benefits of one-click processing. It is even more desirable if the operation can be horizontally deployed not only to a single end user but also to other users in the same department.

Most of such Excel work is likely to be done when dealing with multiple files contained in a folder. This is a very powerful tool when repeating similar operations on multiple files, each of which is created in a similar format, and therefore it would be more efficient if this part could be solved by Excel VBA.

3.2.2 Prompt Input

Enter a prompt that producesd VBA source code that solves the problem. Here are some things to keep in mind

- Specify as per the procedure the end user himself is working on
- Paraphrase to make it easier for ChatGPT to understand before specifying
- Specify each and every action to avoid leaps and bounds.

3.2.3 Localization of output source code

ChatGPT seems to have some difficulties when used with Microsoft OneDrive. For example, if you specify "this folder" or "current directory", the output may be different from what you expect. To avoid such rework, ChatGPT will specify the following folder name in the output of the source code if nothing is specified to the contrary.

Then modify it to the real folder name in an editor. This is faster than spending time on prompt engineering. The process of the proposed method is complete when the output is sufficiently correctable within the end user's literacy level.

3.2.4 Execution in Excel VBA

When the VBA source code is completed, open Excel and copy and paste it into "Development" -¿ "Visual Basic". Now that the preparation is complete, the VBA program will be executed when the macro is executed.

3.2.5 Evaluation

Evaluate the results for yourself, and if the problem is solved, you are done. If not, go back to the prompt input and try prompt engineering to enter a more appropriate prompt.

4 Case Verification

In order to illustrate the usefulness of Excel VBA, we evaluate the effectiveness of the proposed method using an example of Excel work done by white-collar workers at a university. This is the case of a large number of files in a folder.

Therefore, as a case study of this research, we evaluate the effectiveness of the proposed method by actually operating files in folders at the university where we work. The efficiency of the proposed method is evaluated based on the improvement of the manual operation and the operation with Excel VBA in three levels of difficulty.

4.1 Case 1: Registering in the attendance book from a submitted file in a folder

The first example supports the creation of an attendance record by reading the student number folder in the working folder and writing it to a specific column in an Excel file. The example of Academ Academy Advance provided by Densho Corporation will be used for the explanation.

Universities use a variety of methods to check student attendance, including roll call, card readers, and registration using Microsoft Forms on smart phones. Another method is to assign a report for each class, and attendance is counted only when the report is submitted by the due date. In other methods, attendance is taken as long as the student responds even if he or she has not listened much to the class. Substitution may also be possible. On the other hand, if the assignment is to submit a report, if the student does not listen to the class and does not write correctly, the report will not be accepted and the student will be considered absent. From the teacher's point of view, this method kills two birds with one stone, as students concentrate on the class.

If the number of attendees is small, the registration will be done manually, and if the number of attendees is large, the registration will be done in Excel and uploaded. Since more than 100 students are enrolled in university classes at times, there is a limit to the number of students who can be registered manually. It is a challenge to reduce this burden.

In Active Academy, a list of reports is displayed when a student submits a report within the deadline provided on the class website. However, the report submission is not linked to the attendance record, so you need to put the student ID number of the submitted report in column A and the name in column B of the Excel file, save the file in CSV format, and upload the file to Active Academy.

Active Academy has a function to download all submitted reports as a Zip file. Once downloaded and extracted in this way, a folder with the student ID number will be created under the file for each report. The file name is written in Excel, and the Excel file is uploaded to the system to enable attendance registration.

4.1.1 Prompt Engineering

In this case, there is one type of student ID number in one folder, and it is a simple case of reading it and putting it in column A of Excel. At the prompt, type "In VBA, please create source code to line up the folder names in the folder named report in column A of Excel and fill in ● in column B". In other words, the prompt is completed with only one statement.

4.1.2 Output VBA source code

The source code is output from ChatGPT as shown in Figure 1. In this source code, the folder path is not specified, so the end user can change it to the actual folder name to complete the source code.

Next, open "Development" in Excel and select "Visual Basic" to open the new "Microsoft Visual Basic for Application Module2" window. Paste this source code in the area displayed. Select "Insert" - "Standard Module" and paste this source code in the displayed area.

In this case, the first time, the system outputs the result as expected at the prompt. Note that column B actually contains the student's name, but the system does not recognize it, so a dummy ● is inserted. Uploading this file to Active Academy completes the registration of students in the attendance register.

4.1.3 Evaluation

In Case 1, the folder names are changed using an editor as described above, but this has nothing to do with VBA literacy as it is simply changed using an ordinary text editor. The end user outputs the completed VBA source code only by prompt engineering, improving routine work.

4.2 Case 2: Grading final grades based on the number of reports submitted in multiple folders

The second case study is an example of how to grade a class by counting the number of attendance by reading the attendance count for each class in a folder under the system. As in Case 1, we will use the example of Academ Academy Advance provided by Densho.

At the end of each semester, a final exam is given to students to determine their grades. However, the prerequisite for taking the final exam is that the student has attended at least 2/3 of the classes. If a class does not take attendance during class and attendance is counted only by submitting a report, the number of times the student attends the class must be counted out of 15 classes.

In this case, attendance is listed for each class, repeated for the number of classes (usually 15), and the cumulative total is calculated. This is an example of a simple Excel-based routine.

4.2.1 Prompt Engineering

Unlike Case 1, the complexity of the prompts cannot be expressed in a single statement. Therefore, as explained in the procedure, the prompts should be written to explain the way the end user actually performs the work. Specifically, the prompt should be as follows.

Open a subfolder under the folder reports in VBA. Then, retrieve the name of the folder that is contained in the subfolder. This is the student number. Prepare a counter for each student number, and if the same student number appears twice, add one to the student number counter. After reading all the folders, sort the student numbers in ascending order and post them to column A of Excel; do not show the same name twice in column A. In column B, show the student number counter corresponding to column A.

4.2.2 Output VBA source code

The source code is output from ChatGPT as shown in Figure 2. The same output does not specify the folder path, so the end user can complete the source code by changing the actual folder name. Then, in the same way, open "Development" in Excel, select "Visual Basic," choose "Insert" - "Standard Module" in "Microsoft Visual Basic for Application Module2," paste this source code and "Run" to obtain the result. The result can be obtained by pasting this source code and "Execute."

4.2.3 Evaluation

As in Case 1, the desired result is obtained simply by changing the folder name in the editor.

4.3 Case 3: Final grades based on the number of reports per class in multiple Excel files

The third case is that the reports submitted in each class are output to an Excel file. If there are 15 classes, 15 Excel files will be output. In this case, instead of a folder, the 15 files are opened to check whether or not a report has been submitted for each student number, and the number of attendance and report submissions are counted.

4.3.1 Prompt Engineering

Open ChatGPT and type the following at the prompt Open all the files in the folder called "Reports". If the file has a student number in column F, and the report has been submitted in column M, add one to the counter you prepared for the student number in column A. If the file has a student number in column M, add one to the counter you prepared for the student number in column A. For all files, tally up the number of times each student number has submitted the report. Open a new file and create the VBA source code to list all student numbers in column A and output the number of submissions in column B

4.3.2 Output VBA source code

The source code is output from ChatGPT as shown in Figure 3. The same output does not specify the folder path, so the end user can complete the source code by changing it to the actual folder name. The end user can complete the source code by changing the folder name to the actual folder name.

4.3.3 Evaluation

As in Cases 1 and 2, the desired results are obtained simply by changing the folder names in the editor.

5 consideration

In this study, we replaced the Excel work required for registering attendance and assigning grades at a university with macros using VBA. In this case, the end-users are university faculty members, but they are still end-users as long as they have no programming experience.

In all three cases evaluated for application in this study, the VBA source code sought by prompt engineering was output simply by modifying the folder names. It should be noted that the folder names are modified only once when the VBA source code is output. Once the folder name is modified once, there is no need to modify it again. It can be used continuously in end-user's routine work.

In this case, whether EUC by VBA is effective or not depends on the number of students in the class. If there are only a few students in the class, it would be faster to open Active Academy from a browser and input data. However, the number of students in a university class may exceed 100. When the number of students is large, the time and effort required to open a browser and input the information is considerable, and the numbers may shift during the course of the class. Once a macro is created using EUC, these tasks can be done with a single click, and it can be said that IT has improved productivity.

The method proposed in this study improved the end-user's Excel work by making macros for the end-users themselves, and made it possible for them to process their routine work with a single click.

+

6 Conclusion

EUC, which was proposed more than 30 years ago, has not faded at all. In the future, when the existence of DX (Digital Transformation) will be the source of competitiveness, the concept of EUC will become more and more important. In particular, end-users can lay the groundwork for the DX era by streamlining their own Excel work, which they use almost every day.

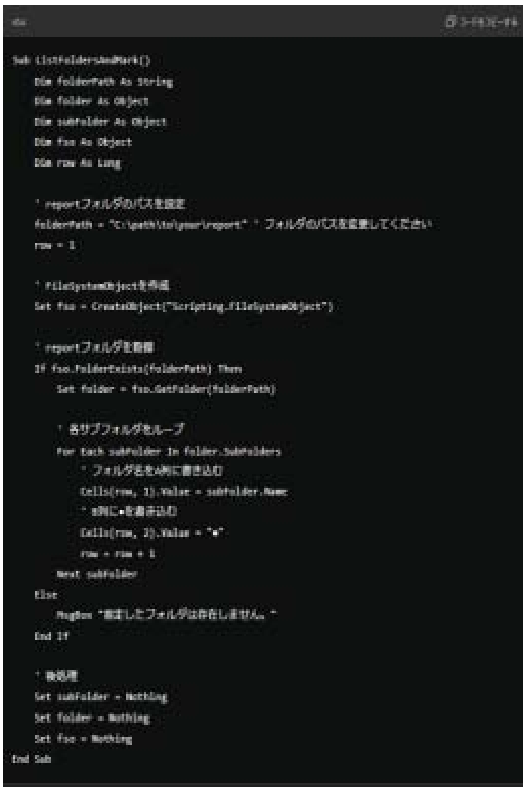
REFERENCES

[1] Yukinobu Yaghoe, EUC as an emergent means to realize strategic information systems, Journal of Economic and Economic Studies, Vol.57, No.4, PP.1-12, 2008

[2] Niloy, Ahnaf et al, Why do students use ChatGPT? An-swering through a triangulation approach, Comput-

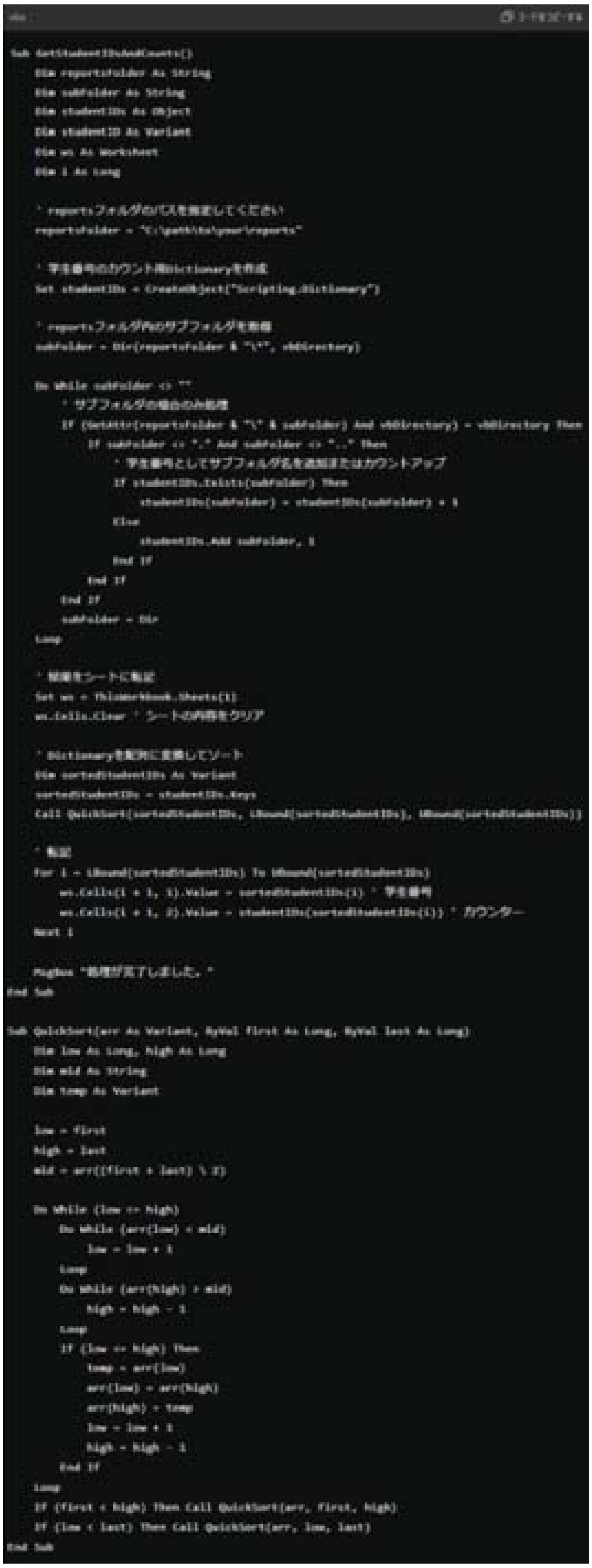
ers andEducation: Artificial Intelligence, Vol . 6, PP. 100208, 2024
[3] Liang, Tianzhu et al, An Evolutionary Study of the Impact of Artificial Intelligence Technology on Foreign Language Education, International Journal of Emerging Technologies in Learning (iJET), Vol. 18, PP.190-204, Oct, 202
[4] Alejandro Barredo Arrieta et al, Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opporunities and Challenges toward Responsible AI, Information Fusion Volume 58,Pages 82-115, June 2020.

A Appendix Section



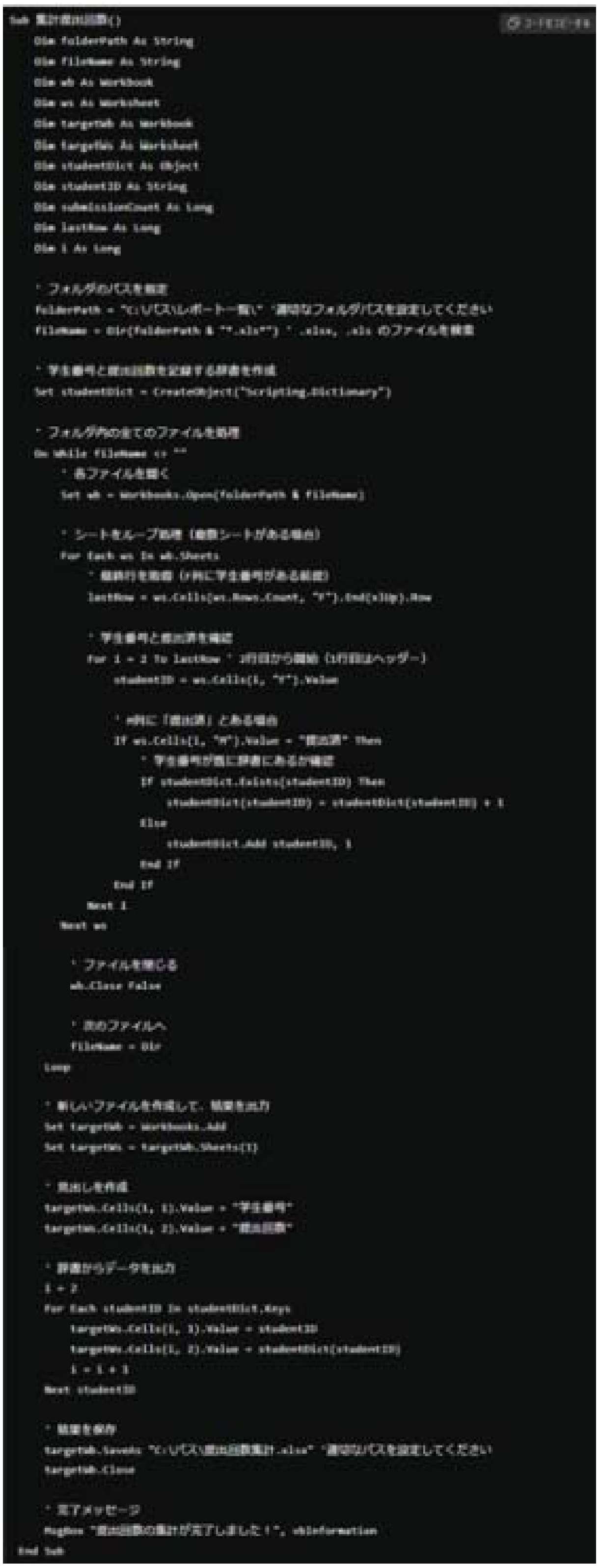
```
Sub ListFoldersAndMark()
    Dim folderPath As String
    Dim folder As Object
    Dim subfolder As Object
    Dim fso As Object
    Dim row As Long

    ' reportフォルダのパスを設定
    folderPath = "C:\path\to\your\report" ' フォルダのパスを変更してください
    row = 1

    ' FileSystemObjectを作成
    Set fso = CreateObject("Scripting.FileSystemObject")

    ' reportフォルダを取得
    If fso.FolderExists(folderPath) Then
        Set folder = fso.GetFolder(folderPath)

        ' 各サブフォルダをループ
        For Each subfolder In folder.SubFolders
            ' フォルダ名をA列に書き込む
            Cells(row, 1).Value = subfolder.Name
            ' B列に●を書き込む
            Cells(row, 2).Value = "●"
            row = row + 1
        Next subfolder
    Else
        MsgBox "指定したフォルダは存在しません。"
    End If

    ' 後処理
    Set subfolder = Nothing
    Set folder = Nothing
    Set fso = Nothing
End Sub
```

Figure 1: Registering in the attendance book from a submitted file in a folder



```
Sub GetStudentIDsAndCounts()
    Dim reportsFolder As String
    Dim subFolder As String
    Dim studentIDs As Object
    Dim studentID As Variant
    Dim ws As Worksheet
    Dim i As Long

    ' reportsフォルダのパスを指定してください
    reportsFolder = "C:\path\to\your\reports"

    ' 学生番号のカウント用Dictionaryを作成
    Set studentIDs = CreateObject("Scripting.Dictionary")

    ' reportsフォルダ内のサブフォルダを取得
    subFolder = Dir(reportsFolder & "\*", vbDirectory)

    Do While subFolder <> ""
        ' サブフォルダの場合のみ処理
        If (GetAttr(reportsFolder & "\" & subFolder) And vbDirectory) = vbDirectory Then
            If subFolder <> "." And subFolder <> ".." Then
                ' 学生番号としてサブフォルダ名を追加またはカウントアップ
                If studentIDs.Exists(subFolder) Then
                    studentIDs(subFolder) = studentIDs(subFolder) + 1
                Else
                    studentIDs.Add subFolder, 1
                End If
            End If
        End If
        subFolder = Dir
    Loop

    ' 結果をシートに転記
    Set ws = ThisWorkbook.Sheets(1)
    ws.Cells.Clear ' シートの内容をクリア

    ' Dictionaryを配列に変換してソート
    Dim sortedStudentIDs As Variant
    sortedStudentIDs = studentIDs.Keys
    Call QuickSort(sortedStudentIDs, LBound(sortedStudentIDs), UBound(sortedStudentIDs))

    ' 転記
    For i = LBound(sortedStudentIDs) To UBound(sortedStudentIDs)
        ws.Cells(i + 1, 1).Value = sortedStudentIDs(i) ' 学生番号
        ws.Cells(i + 1, 2).Value = studentIDs(sortedStudentIDs(i)) ' カウンター
    Next i

    MsgBox "処理が完了しました。"
End Sub

Sub QuickSort(arr As Variant, ByVal first As Long, ByVal last As Long)
    Dim low As Long, high As Long
    Dim mid As String
    Dim temp As Variant

    low = first
    high = last
    mid = arr((first + last) \ 2)

    Do While (low <= high)
        Do While (arr(low) < mid)
            low = low + 1
        Loop
        Do While (arr(high) > mid)
            high = high - 1
        Loop
        If (low <= high) Then
            temp = arr(low)
            arr(low) = arr(high)
            arr(high) = temp
            low = low + 1
            high = high - 1
        End If
    Loop
    If (first < high) Then Call QuickSort(arr, first, high)
    If (low < last) Then Call QuickSort(arr, low, last)
End Sub
```

Figure 2: Grading final grades based on the number of reports submitted in multiple folders



```
Sub 集計提出回数()
    Dim folderPath As String
    Dim fileName As String
    Dim wb As Workbook
    Dim ws As Worksheet
    Dim targetWb As Workbook
    Dim targetWs As Worksheet
    Dim studentDict As Object
    Dim studentID As String
    Dim submissionCount As Long
    Dim lastRow As Long
    Dim i As Long

    ' フォルダのパスを指定
    folderPath = "C:\パス\レポート一覧\" '適切なフォルダパスを設定してください
    fileName = Dir(folderPath & "*.xls*") ' .xlsx, .xls のファイルを検索

    ' 学生番号と提出回数を記録する辞書を作成
    Set studentDict = CreateObject("Scripting.Dictionary")

    ' フォルダ内の全てのファイルを処理
    Do While fileName <> ""
        ' 各ファイルを開く
        Set wb = Workbooks.Open(folderPath & fileName)

        ' シートをループ処理（複数シートがある場合）
        For Each ws In wb.Sheets
            ' 最終行を取得（F列に学生番号がある前提）
            lastRow = ws.Cells(ws.Rows.Count, "F").End(xlUp).Row

            ' 学生番号と提出済を確認
            For i = 2 To lastRow ' 2行目から開始（1行目はヘッダー）
                studentID = ws.Cells(i, "F").Value

                ' H列に「提出済」とある場合
                If ws.Cells(i, "H").Value = "提出済" Then
                    ' 学生番号が既に辞書にあるか確認
                    If studentDict.Exists(studentID) Then
                        studentDict(studentID) = studentDict(studentID) + 1
                    Else
                        studentDict.Add studentID, 1
                    End If
                End If
            Next i
        Next ws

        ' ファイルを閉じる
        wb.Close False

        ' 次のファイルへ
        fileName = Dir
    Loop

    ' 新しいファイルを作成して、結果を出力
    Set targetWb = Workbooks.Add
    Set targetWs = targetWb.Sheets(1)

    ' 見出しを作成
    targetWs.Cells(1, 1).Value = "学生番号"
    targetWs.Cells(1, 2).Value = "提出回数"

    ' 辞書からデータを出力
    i = 2
    For Each studentID In studentDict.Keys
        targetWs.Cells(i, 1).Value = studentID
        targetWs.Cells(i, 2).Value = studentDict(studentID)
        i = i + 1
    Next studentID

    ' 結果を保存
    targetWb.SaveAs "C:\パス\提出回数集計.xlsx" '適切なパスを設定してください
    targetWb.Close

    ' 完了メッセージ
    MsgBox "提出回数の集計が完了しました！", vbInformation
End Sub
```

Figure 3: Final grades based on the number of reports per class in multiple Excel files

Challenges of a Self-Learning Support System Using Augmented Reality

Yuga Ito, Yuta Takenaka, and Shoji Sano[†]

[†]Kanazawa Institute of Technology, Japan
sano@neptune.kanazawa-it.ac.jp

Abstract - This paper proposes a self-learning support system using Augmented Reality (AR). In the proposed system, when learners point a camera at a part they do not understand, contextual advice is overlaid via AR. To realize this AR-based self-learning support system, we have developed prototypes capable of providing guidance tailored to the learner's programming proficiency and recognizing their implementation processes. This paper reports on those developments.

Keywords: Self-Learning Support System, Augmented Reality, Hints, Camera, Metacognition

1 INTRODUCTION

Currently, universities provide a variety of learning opportunities beyond traditional lecture-based instruction, including problem-solving approaches such as active learning [1, 2] and project-based learning (PBL) [3–5]. In addition, the widespread adoption of online classes accelerated during the COVID-19 pandemic due to restrictions on face-to-face teaching. Even today, after the impact of COVID-19 has subsided, online learning continues to be adopted as one of the learning formats that offers their advantages. Furthermore, generative AI systems such as ChatGPT and Gemini are attracting attention and are increasingly being utilized in school education.

Although there are various learning formats, when learners study according to their own interests and concerns, they often study independently. In school classes or study groups, questions can be resolved by asking the instructor. However, when learning alone—including reviewing class material or study group content—one must research independently. When considering internet research for this purpose, it's ideal if questions can be resolved quickly, but sometimes it requires significant time.

Many learning support systems have been developed using Virtual Reality (VR) [6, 7] or Augmented Reality (AR) [8–10]. However, most learning materials present specific content related to the learning material, no systems are available that can answer questions.

We focus on learning through classroom instruction or self-study at home. When encountering difficulties, learners require hints to overcome them. In this paper, we propose a self-learning support system. Our proposed system aims to enhance learning effectiveness by adopting an approach where learners consciously seek guidance on what they do not understand. Specifically, as illustrated in Figure 1, when the smartphone is pointed at an unclear section, hints are displayed via AR. Challenges include presenting hints tailored to the learner's comprehension level and accurately recognizing the unclear sections. We created two prototypes for each approach and conducted evaluation experiments. The evaluation results confirmed that users could complete tasks while viewing hints and that the recognition accuracy for unclear sections was reasonably high.

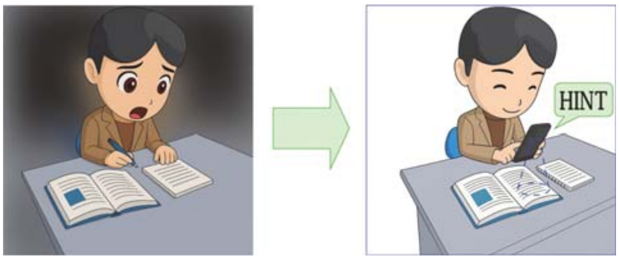


Figure 1: Assumed scenario

2 RELATED WORK

2.1 Metacognition in Learning and Education

Metacognition, which refers to an individual's awareness of their own learning processes, is fundamental to learning and education and has been shown to improve learning outcomes.

Mani et al. investigated the effects of incorporating metacognitive elements into instruction [11]. Students reported confidence levels for each exam question, and the relationship between confidence and performance was analyzed. The findings indicated that high-performing students demonstrated more accurate self-perception, whereas lower-performing students tended to overestimate their abilities.

Caratozzolo et al. implemented an educational program aimed at fostering creative thinking and metacognitive awareness in engineering education [12]. The program included activities designed to encourage learners to shift between associative and analytical thinking modes. Results of evaluation experiments showed that receiving metacognitive instruction significantly improved students' self-report scores for creativity and enhanced their ability to transition between analytical and associative thinking.

Increasing learners' awareness of their own learning processes has been shown to improve learning effectiveness, which also informs our approach.

2.2 VR-Based Learning System

A number of VR-based learning systems have been designed for use in educational settings.

Abulrub et al. reported on the effectiveness of an immersive 3D visualization environment. [13] Using a 4K resolution 3D visualization system, they simulated real industrial challenges and built a learning environment in which students could practically apply theoretical knowledge. The use of VR technology enabled realistic, safe, and cost-efficient experiences in product design and evaluation, contributing to the development of creativity, problem-solving skills, teamwork, and communication abilities.

Sun et al. investigated the educational effectiveness of a VR-based engineering training support system through practical exercises [14]. In a project-based course, students designed the motion of equipment such as drilling machines and lathes based on real-world rules using Unreal Engine 4. The developed VR content was used to support lower-level students during their practice sessions and to provide feedback. When these students progressed to upper-level courses, they participated in the project-based course and created VR contents themselves. Overall, 81 % of the students expressed interest in this instructional method.

2.3 AR-Based Learning System

Numerous AR-based systems have also been developed to support instruction.

Venigalla et al. developed FlowARP [15], which dynamically visualizes program control flow with AR. This system dynamically provides real-time visualizations of algorithms and control flow in C/C++ programming. An evaluation experiment with 44 university students demonstrated that the time required to understand code decreased, although students needed time to become familiar with the system.

Suselo et al. designed an AR tool that enables users to manipulate 3D objects through translation, rotation, and scaling along coordinate axes [16]. In an evaluation experiment involving 22 university students, comparison of performance on 3D transformation problems before and after using the tool revealed an improvement in the average correct response rate from 13% to 71%.

AR technology enables additional information to be overlaid on conventional teaching materials. However, most previous AR learning support systems provide the same content to all learners and do not provide guidance tailored to individual proficiency levels.

3 PROPOSED SYSTEM

3.1 System Overview

We often encounter parts of the learning material that we do not understand while learning. When studying alone, there is frequently no one nearby to ask for assistance. In such situations, learners typically rely on web searches or generative AI tools to look up information. However, when the search approach or the prompt is not effective, the process can become time-consuming.

To solve this problem, this paper proposes a self-learning support system using AR. when a learner points the camera at a part they do not understand, the proposed system provides the corresponding advice through AR, as shown in Figure 2. By pointing the camera, learners implicitly indicate the parts they find difficult, which is intended to enhance their metacognition during the learning process. The usage flow of the proposed system is as follows:

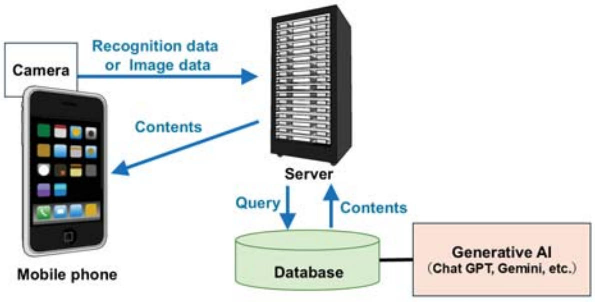

Camera
Recognition data or Image data
Contents
Server
Query
Contents
Mobile phone
Database
Generative AI (Chat GPT, Gemini, etc.)

Figure 2: Proposed system

1. When a learner encounters a part they do not understand, they capture the area using the built-in camera of a smartphone.
2. The captured image is sent to the server, where the system analyzes the situation and queries the database.
3. A relevant hint is retrieved from the database and returned to the smartphone.
4. The hint is displayed in AR, overlaid on the corresponding part of the image on the smartphone screen.

3.2 Issues

To realize the proposed system, we believe the following three issues must be addressed:

Issue 1 Confirming the usefulness and benefits of hints provided to learners

Issue 2 Recognizing the learner's unclear parts situation from a camera image

Issue 3 Building a database that associates the learner's unclear parts with the appropriate hints

To resolve Issue 1 and Issue 2, we developed the two prototypes and conducted the evaluation experiments, which are reported in this paper. Issue 3 has not yet been addressed, as it will be tackled after verifying Issue 1 and Issue 2. To approach Issue 3, we are considering the use of generative AI for this purpose, although we will need to customize the system so that it can generate hints appropriate for the learning content.

4 EVALUATION OF THE HINTS PROVIDED TO LEARNERS

4.1 Implementation of the Prototype 1

We developed Prototype 1 to confirm whether the hints provided to learners are useful. In Prototype 1, we use Scratch as the programming learning material, and hints are provided according to the learner's programming proficiency. In this paper, we classify programming proficiency into the following three levels:

- Beginner: Learners who have no programming experience or or lack confidence in programming.
- Novice: Learners who have programming experience but have never used Scratch.
- Intermediate: Learners who have used Scratch and can use it with some proficiency.

Prototype 1 uses the Japanese language, since the evaluation experiment was conducted with Japanese university students in Japanese. Learners aim to create a Scratch game in which a sprite catches falling apples. Six tasks are prepared for this purpose, as follows.

1. Place the sprite at the bottom and move it left and right using the keyboard.
2. Let the apple fall.
3. Make apples fall randomly.
4. Detect collisions between the basket and falling apples.
5. Increase the score whenever the sprite catches an apple.
6. Set an end condition that finishes the game when the score reaches a certain threshold.

Hints can be provided when the learner gets stuck on a task. As shown in Figure 3, learners with lower programming proficiency receive more detailed explanations, whereas learners with higher programming proficiency receive shorter explanations designed to help them understand quickly. In detail, for beginner-level learners, code composed of Scratch blocks is displayed for beginner-level programmers. This allows them to execute the program by writing it almost verbatim. For novice-level learners, the blocks to be used are displayed; that is, algorithmic concepts are presented, and the learners are instructed to construct the corresponding code. For intermediate-level learners, broken-down tasks are presented, and learners create the corresponding code.

We implemented Prototype 1 using Scratch and ARKit. The learner's programming proficiency level is set at the beginning, and six AR markers corresponding to the tasks are prepared. Learners can display hints by pointing their smartphone at the appropriate AR marker.

4.2 Evaluation of Prototype 1

We conducted an evaluation experiment of Prototype 1 using nine university students aged 21 to 23 as participants, referred to as A through I. The participants were divided into three groups according to programming proficiency—beginner, novice, and intermediate—with three participants in each group. Participants A, B, and C were classified as beginners; D, E, and F as novices; and G, H, and I as intermediates. Before conducting the experiment, we explained Prototype 1 to all participants.

The evaluation tasks consisted of six questions embedded in the process of creating the game. A time limit of 40 minutes was set for completing the entire game.

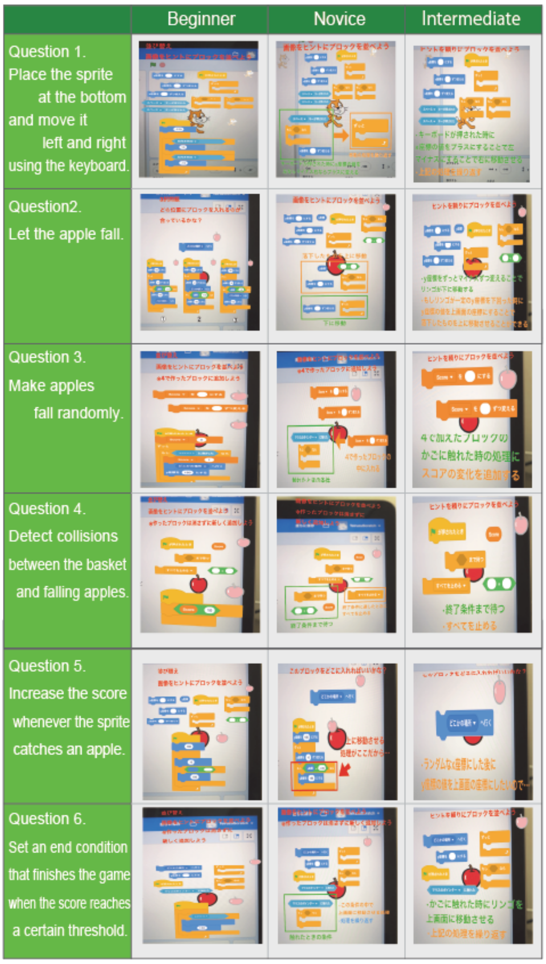

Beginner
Novice
Intermediate
Question 1. Place the sprite at the bottom and move it left and right using the keyboard.
Question2. Let the apple fall.
Question 3. Make apples fall randomly.
Question 4. Detect collisions between the basket and falling apples.
Question 5. Increase the score whenever the sprite catches an apple.
Question 6. Set an end condition that finishes the game when the score reaches a certain threshold.

Figure 3: Hints provided according to proficiency levels

The results of the evaluation experiment are shown in Table 1.

To begin with, out of the nine participants, seven correctly completed all six tasks.

Participant B made an error in the processing of the falling apple: the program lacked the process that makes the apple fall again from a random position after it has fallen. As a result, if the apple was not caught, it continued to fall repeatedly from the same position. Since this mistake is difficult to notice unless a missed catch occurs, it is considered that the error went unnoticed during this task, which focuses on catching apples. Participant C made an error regarding the coordinates of the ring: instead of using a block that specifies the coordinates directly, the participant used a block that changes the coordinates, causing the falling position of the apple to vary. Since this error can sometimes behave similarly to the correct behavior, it is considered that the mistake was not noticed during the task. Based on these results, the errors made by Participants B and C were difficult to detect.

Moreover, regarding the number of times hints were used, there was variability among participants; however, when viewed by programming proficiency, beginner and novice participants used hints on average more than five times as often as intermediate participants. Moreover, since all participants used hints at least once, the necessity of providing hints was demonstrated.

Table 1: Result of the evaluation experiment

Participant		A	B	C	D	E	F	G	H	I
Programming level		Beginner	Beginner	Beginner	Novice	Novice	Novice	Intermediate	Intermediate	Intermediate
Rates of correct answers		6 / 6	5 / 6	6 / 6	5 / 6	6 / 6	6 / 6	6 / 6	5 / 6	6 / 6
Number of hints provided		23	10	9	21	2	16	1	9	12
Details	Question 1	2	0	2	3	0	0	0	3	3
	Question 2	6	3	2	6	0	6	0	3	2
	Question 3	3	0	1	2	1	2	1	1	3
	Question 4	5	4	1	7	0	5	0	1	1
	Question 5	4	2	2	2	0	1	0	1	2
	Question 6	3	1	1	1	1	2	0	0	1
Time		36min 29s	26min 56s	21min 40s	31min 53s	18min 0s	32min 57s	20min 27s	20min 55s	22min 8s
Details	Question 1	4min 17s	3min 17s	3min 7s	7min 32s	3min 38s	6min 6s	3min 38s	5min 23s	4min 23s
	Question 2	8min 42s	5min 29s	1min 18s	7min 8s	2min 57s	9min 50s	2min 57s	5min 1s	3min 9s
	Question 3	2min 20s	1min 53s	0min 12s	0min 46s	2min 51s	2min 32s	2min 21s	1min 30s	1min 25s
	Question 4	6min 6s	5min 20s	1min 44s	3min 56s	3min 3s	4min 77s	3min 3s	2min 32s	1min 45s
	Question 5	4min 17s	0min 17s	1min 0s	0min 22s	1min 19s	1min 59s	1min 19s	1min 13s	1min 25s
	Question 6	6min 3s	1min 21s	0min 37s	0min 29s	4min 12s	3min 16s	6min 39s	1min 29s	1min 20s
	Hint 1	0min 32s	0min 0s	2min 57s	0min 46s	0min 0s	0min 0s	0min 0s	0min 37s	0min 46s
	Hint 2	0min 58s	2min 42s	4min 24s	3min 37s	0min 0s	1min 46s	0min 0s	2min 12s	3min 9s
	Hint 3	0min 44s	0min 0s	2min 10s	0min 54s	0min 36s	0min 28s	0min 30s	0min 6s	3min 26s
	Hint 4	0min 59s	2min 7s	1min 41s	3min 4s	0min 0s	1min 20s	0min 0s	0min 42s	0min 8s
	Hint 5	1min 6s	2min 34s	1min 47s	1min 49s	0min 0s	0min 37s	0min 0s	0min 10s	0min 27s
	Hint 6	0min 25s	1min 56s	1min 32s	1min 30s	0min 0s	0min 16s	0min 12s	0min 0s	0min 45s

Participants A, F, and H created the program while understanding each step of the process. Participants B, D, and I initially proceeded by thinking on their own, but when they became stuck, they compared the hints with the blocks they had created. Participant C used hints from the very beginning and proceeded while referring closely to them. Participants E and G did not require process-related hints and focused only on the blocks to be used.

Furthermore, the time spent using Scratch was longest for novice-level participants. When the average time spent viewing hints is considered by proficiency level, there is a tendency for this time to decrease as proficiency increases. Participants A and F, who spent relatively more time using Scratch, showed a strong desire to create based on their own ideas and tended to glance only briefly at the hints. As a result, the time spent using Scratch was longer, and the time spent viewing hints was shorter. In contrast, the extremely short Scratch usage time for Participant C was due to frequently referring to the hints while creating the program.

After the evaluation experiment, we asked participants for their impressions of the proposed system. Participants who had never used a visual programming language appreciated being able to create the game in a short time. Those who had used such languages liked the fact that hints were displayed in AR. However, some participants reported that text displayed in AR was sometimes difficult to read; therefore, alternative methods of providing information other than text will be considered in future work.

5 EVALUATION OF PROCESS RECOGNITION

5.1 Implementation of Prototype 2

In Prototype 1, learners can understand the process and selected hints accordingly; however, to provide advice based on the process, it is necessary to recognize the process itself. Therefore, we implemented Prototype 2, which aims to recognize processes using image processing. Prototype 2 targets electronic work, where the smartphone camera captures images of electronic circuits, and process recognition is performed by detecting electronic components using image processing and machine learning. Each electronic component is identified and counted, and the detected circuit configuration is compared with the predefined component configurations for the six processes. The process whose component counts satisfy the required configuration is output as the recognition result.

The target electronic circuit is shown in Figure 4. The target microcontroller is an Arduino UNO. The components used include an operational amplifier (JRC H002B), temperature sensors (LM35DZ and TC622EPA), an acceleration sensor (KX9R4 2050), red and green LEDs, 1 k Ω and 100 k Ω resistors, and four types of wires (red, blue, yellow, and green). Among these, the temperature sensor LM35DZ does not have markings on its top surface, unlike the operational amplifier JRC H002B or the temperature sensor with automatic output switching, TC622EPA, and therefore appears simple and entirely black. This makes them difficult to detect using image processing. Therefore, the temperature sensor LM35DZ is excluded from recognition in the process recognition results.

We designed the exercise in which learners constructed the following circuits on two breadboards:

Circuit 1 A circuit that measures ambient temperature using a temperature sensor and controls green LEDs as follows: all green LEDs turn off if the temperature is between 0°C and 10°C; one green LED turns on between 10°C and 20°C; and two green LEDs turn on between 20°C and 30°C.

Circuit 2 A circuit that turns on a red LED when the ambient

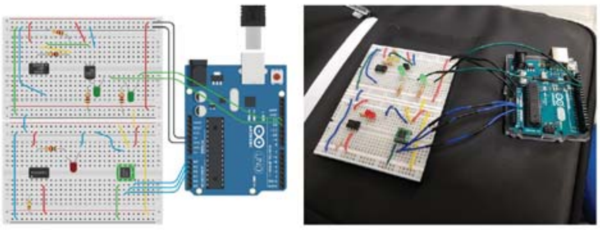

Target circuit Example implementation of the circuit

Figure 4: Target circuit and implementation example

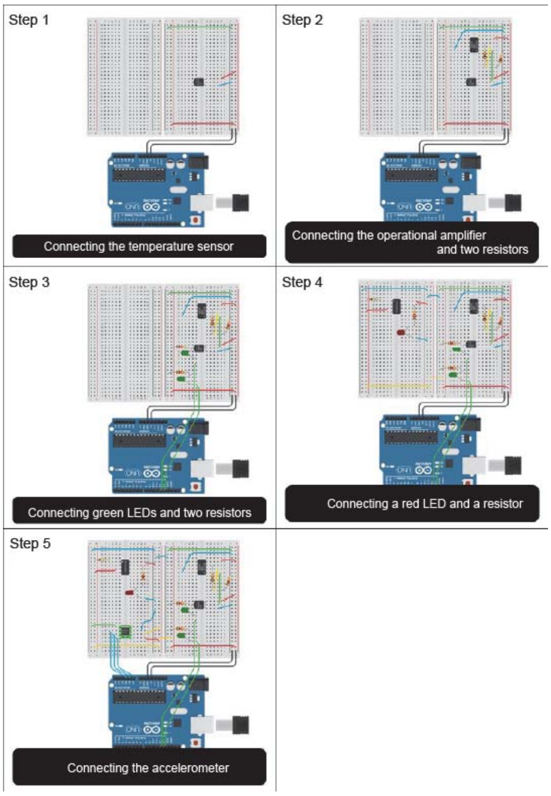

Step 1
Connecting the temperature sensor
Step 2
Connecting the operational amplifier and two resistors
Step 3
Connecting green LEDs and two resistors
Step 4
Connecting a red LED and a resistor
Step 5
Connecting the accelerometer

Figure 5: Construction steps

temperature reaches 10°C, using a temperature sensor (TC622EPA).

Circuit 3 A circuit that acquires acceleration data using an acceleration sensor.

The details of the construction steps are shown in Figure 5. We defined five construction steps for the target circuits as follows:

Step 1 Connect the temperature sensor.

Step 2 Connect the operational amplifier and two resistors.

Step 3 Connect two green LEDs and two resistors.

Step 4 Connect a red LED and a resistor.

Step 5 Connect the accelerometer.

In the process of constructing the target circuits, each step corresponds to adding one function. Circuit 1 is divided into three sub-steps(Step 1–Step 3): placing wires related to the temperature sensor, placing wires related to the operational amplifier that amplifies the temperature sensor signal, and installing resistors. Moreover, Circuit 2 was assigned to Step 4, and Circuit 3 was assigned to Step 5.

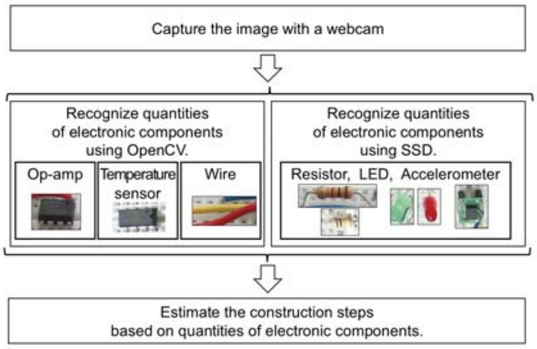

Capture the image with a webcam
Recognize quantities of electronic components using OpenCV.
Op-amp
Temperature sensor
Wire
Recognize quantities of electronic components using SSD.
Resistor, LED, Accelerometer
Estimate the construction steps based on quantities of electronic components.

Figure 6: Process recognition flow

We implemented Prototype 2 for recognizing the process of electronic exercise. Although the final goal is to operate the system on a smartphone, Prototype 2 currently processes images captured by a webcam connected to a laptop. For process recognition, OpenCV and the Single Shot MultiBox Detector (SSD) were used. Image acquisition from the webcam, component identification using OpenCVSharp, and process recognition were implemented using Unity and C#. Component identification using SSD was performed with Python and PyTorch. The object recognition results obtained via SSD in Python were imported into Unity, and combined with the component recognition results performed in Unity to determine the construction steps.

As shown in Figure 6, the types and quantities of electronic components were recognizedfrom images captured by the webcam using the following procedures:

- Recognition of IC types and quantities using template matching focused on the printed markings on the ICs.
- Recognition of wire types and quantities based on hue information in the HSV color space.
- Estimation of component types and quantities using an SSD model trained in advance with electronic component image datasets.

During processing in Prototype 2, the captured images undergo preprocessing using OpenCVSharp. The four corner points of the breadboard are detected, and an affine transformation is applied to the image so that the transformed image forms a rectangle with the same aspect ratio as the breadboard. The method for obtaining the four corner points of the breadboard is performed as follows:

1. Apply threshold-based binarization to the captured image of the electronic circuit.
2. Extract contours (coordinate arrays) from the binarized image.
3. Compute the convex hull of the contours.

Table 2: HSV range settings for detecting the wires

Wire color	The range of H	The range of S	The range of V
Red	165 - 180	30 - 255	50 - 255
Blue	100 - 115	80 - 255	50 - 255
Yellow	23 - 39	100 - 255	50 - 255
Green	70 - 100	50 - 255	0 - 200

4. Obtain a polygonal approximation of the convex hull.
5. Regard as the four corner points of the breadboard, when the polygonal approximation has four vertices,

The method for obtaining the four corner points of the breadboard is performed according to the following steps:

1. Perform thresholding binarization on the captured image of the electronic circuit.
2. Extract contours (arrays of coordinates) from the binarized image.
3. Calculate the convex hull of the contours.
4. Obtain the polygonal approximation of the convex hull contours.
5. If the polygonal approximation has four vertices, these points are regarded as the four corner points of the breadboard.

To make the top surface of the breadboard easier to recognize as a rectangle and to simplify circuit detection, the sides of the breadboard were covered with black masking tape. Additionally, to ensure a black background in the captured images, the circuit construction was performed on a black cloth. To avoid capturing objects other than electronic components and to improve detection accuracy, the transformed images were used as input for wire recognition via image processing, IC recognition via template matching, component counting with SSD, and for generating SSD training data.

Wire recognition was performed by converting the input image into the HSV color space. The HSV value ranges are shown in Table 2. For each wire color to be detected, HSV value ranges were defined, and using these values, a mask image was generated by extracting the regions corresponding to each wire color. Furthermore, the extracted contour mask for each color was multiplied by the corresponding masked image for that wire color. The area of the resulting contour was then calculated, and if it exceeded a predefined threshold, the presence of one wire of that color was recognized. This procedure was repeated to determine the total number of wires.

Template matching was used to recognize the operational amplifier and the temperature sensor (TC622EPA). As a preparation step, images of the IC components to be detected were captured by placing them on a breadboard, and the portions containing the printed markings were cropped to create template images for template matching using OpenCVSharp. From each cropped image, three additional images were generated: one flipped along the X-axis, one flipped along the Y-axis, and one flipped along both axes. Together with the original image, these four images were used as template images for recognition. The procedure for detecting IC components is outlined below.

1. Generate grayscale images from the input images, including the four template images.
2. Execute threshold-based binarization on both the template images and the input images.
3. Perform template matching between the binarized template images and the binarized input images; if the similarity exceeds a predefined threshold, the corresponding IC component is deemed to be installed.

During system operation, grayscale images of the four template images and the input image generated beforehand were produced, followed by thresholding. Template matching was then performed between the four thresholded template images and the thresholded input image. If the similarity exceeded a certain threshold, the corresponding IC component was judged to be installed. Furthermore, since the threshold value used during thresholding affected the output depending on the brightness of the system's processing environment, the threshold was set in advance to match the brightness of the input image.

Resistors, acceleration sensors, and other components that were difficult to recognize using template matching or feature point detection were recognized using SSD. Images of resistors, acceleration sensors, LEDs, and other electronic components, together with pre-trained VGG16 weights for transfer learning, were used to train the model. The size of detectable objects was constrained to between 20% and 70% of the image frame.

5.2 Evaluation of the Prototype 2

We conducted an evaluation experiment of Prototype 2 using seven university students between 21 and 22 years old as participants, referred to as J through P.

During the circuit construction task, Prototype 2 intermittently performed process recognition to achieve real-time identification of the construction step. The training data used for the SSD model implemented in Prototype 2 were prepared by the authors as original labeled datasets and used for transfer learning. Cross-validation was conducted to examine the effectiveness of Prototype 2 under different training data conditions,

The participants were instructed to assemble electronic circuits using components prepared by the authors. Table 3 shows the types and quantities of electronic components used in the evaluation. The participants were provided with instructional materials that included written descriptions of the construction steps, the types and quantities of components required, and sample illustrations of the circuits to be built at each step.

While the participants worked on the task, Prototype 2 continuously recognized the current construction process in real

Table 3: Number of electronic components used in the evaluation experiment

	Operational amplifier	Temperature sensor TC622EPA	Acceleration sensor	1k Ω resistor	100k Ω resistor	Red LED	Green LED	Red wire	Blue wire	Green wire	Yellow wire
Step 1	0	0	0	0	0	0	0	2	1	1	0
Step 2	1	0	0	2	0	0	0	3	2	2	1
Step 3	1	0	0	4	0	0	2	3	2	4	1
Step 4	1	1	0	5	1	1	2	5	5	4	3
Step 5	1	1	1	5	1	1	2	6	9	5	4

Table 4: Result of process recognition

	Step 1	Step 2	Step 3	Step 4	Step 5
Participant J	1.000	0.500	0.375	0	0
Participant K	0.750	1.000	0.375	0.120	0.875
Participant L	0.875	0.625	0	0.375	0
Participant M	1.000	0.500	0.370	0.250	0
Participant N	0.875	0.875	0	0.125	0.125
Participant O	1.000	1.000	0.375	0.025	0.625
Participant P	1.000	0.750	0	0.125	0

Table 5: Recognition rate of the components

Operational amplifier	Temperature sensor TC622EPA	Acceleration sensor	1k Ω resistor	100k Ω resistor	Red LED	Green LED	Red wire	Blue wire	Green wire	Yellow wire
0.664	0.504	0.882	0.579	0.864	0.896	0.568	0.282	0.529	0.454	0.664

time and recorded the recognized step, the detected component types and quantities, and the elapsed time. Additionally, a webcam recorded video of the circuit throughout the task. Still images were extracted from the recorded footage for evaluating step recognition and component detection using static images, as well as for constructing training and validation datasets for cross-validation. The footage was also used to verify the accuracy of the proposed system by comparing the actual construction steps and component placement.

For evaluation, eight images per process were randomly selected at one-second intervals from the recorded footage of each participant's circuit construction. To ensure accurate assessment of process recognition, images in which the circuit was obscured by the participant's hands or materials—making detection impossible—were excluded.

We compared the processes recognized by Prototype 2 with the actual construction steps and evaluated both the process recognition accuracy and the component detection accuracy. Table 4 demonstrates the process recognition accuracy for each participant, the average accuracy for each step. Table 5 presents the average detection accuracy by component type across all participants.

According to Table 4, the average accuracy across all participants was high for Step 1 and Step 2 at 0.929 and 0.880, respectively. In contrast, the average accuracy for Step 3, 4, and 5 was low, at 0.214, 0.489, and 0.489, respectively. The overall average accuracy across all participants and all steps was 0.461.

Participant O achieved the highest process recognition accuracy, with an overall rate of 0.650. In contrast, Participants J, L, and P had the lowest overall accuracy, each with a recognition rate of 0.375. Although Participant J had the same overall accuracy as Participants L and P. Moreover, Participant J showed 0 accuracy for Steps 4 and 5, and had the lowest accuracy for Step 2, which made J the weakest in overall process recognition. Additionally, Participants N and P each forgot to install a required component during Step 2. Both later noticed the omission during Step 4 and subsequently installed the missing component. As a result, their step recognition accuracy for Step 3 is presumed to have been 0.

For all participants, the lower accuracy in recognizing Step 3, 4, and 5 compared to Step 1 and 2 is attributed to failures in detecting the green LEDs, which caused Prototype 2 to be unable to advance beyond Step 2.

Furthermore, Prototype 2 recognizes processes based on the number of detected components. If certain components are not detected, Prototype 2 does not judge the process as complete until those missing components are recognized. This likely affected the recognition accuracy for Step 3 and beyond. As more components that are difficult to detect are added, correctly identifying the construction step becomes increasingly challenging.

6 CONCLUSION

In this paper, we proposed a self-learning support system using Augmented Reality (AR). In the proposed system, when a learner points a smartphone camera at a section they do not understand, the system provides hints through AR. To encourage learners to recognize what they do not understand—that is, to promote metacognition—the system requires them to perform the action of directing the camera toward the unclear section. To realize the proposed system, we developed two prototypes and addressed two associated challenges. Prototype 1 presents hints in Scratch that are tailored to the learner's proficiency level. The results of the evaluation experiment confirmed that this approach contributed to facilitating learners' understanding to a certain extent. Prototype 2 recognizes construction processes in electronic circuit building. The evaluation results showed that Prototype 2 was able to recognize the processes with a reasonable level of accuracy.

As future work, we plan to construct a database that associates the results of situation recognition with appropriate hints generated using generative AI APIs, and to develop a complete AR-based self-learning support system based on the findings obtained to date.

REFERENCES

[1] N. Dehbozorgi, M. L. Maher, C. Latulipe, H. Ramaprasad: "Active Learning Design Patterns for CS Education", In the proceedings of the ACM Conference on International Computing Education Research (ICER 2017), pp. 291-292 (2017).

[2] Q. Hao, B. Barnes, E. Wright, E. Kim: "Effects of Active Learning Environments and Instructional Methods in Computer Science Education", In the proceedings of the 49th ACM Technical Symposium on Computer Science Education (SIGCSE 2018), pp.934–939 (2018).

[3] S. Yeom, N. Herbert, R. Ryu: "Project-Based Collaborative Learning Enhances Students' Programming Performance", In the proceedings of the 27th ACM Conference on Innovation and Technology in Computer Science Education (ITiCSE 2022), Vol. 1, pp.248-254 (2022).

[4] A. Lara, L. Quesada: "A Project-based Learning Experience in a Compilers Course", In the proceedings of the ACM Conference on Global Computing Education (CompEd 2019), pp.136-142 (2019).

[5] M. Moalagh, S. S. Hussain, B. A. Farshchian, S. G. Selassie: "Which Teamwork Challenges Do Computing Students Face in a Project-Based Learning Course in Research Methods?", In the proceedings of the 24th ACM Koli Calling International Conference on Computing Education Research (Koli Calling 2024), No.46, pp.1–12 (2024).

[6] J. Nie, B. Wu: "Investigating the Effect of Immersive Virtual Reality and Planning on the Outcomes of Simulation-Based Learning: A Media and Method Experiment", In the proceedings of the IEEE 20th International Conference on Advanced Learning Technologies (ICALT 2020), pp.329-332 (2020).

[7] A. Arntz, D. Kessler, S. C. Eimler: "EnLighten: A Photovoltaics Learning Environment in Virtual Reality", In the proceedings of the IEEE 21st International Conference on Advanced Learning Technologies (ICALT 2021), pp.221-223 (2021).

[8] R. Reuter, M. Knietzsch, F. Hauser, J. Mottok: "Supporting Abstraction Skills Using Augmented Reality?", In the proceedings of the ACM the 24th ACM Conference on Innovation and Technology in Computer Science Education (ITiCSE 2019), p.320 (2019).

[9] A. Sarkar, K. Arya: "Teaching Marker-based Augmented Reality in a PBL-Based Online Robotics Competition", In the proceedings of the IEEE 20th International Conference on Advanced Learning Technologies (ICALT 2020), pp.338–340 (2020).

[10] C. Arce, A. Hernandez, R.T. Castillo, T. Valdez, R. Cota, E. Pontelli: "Learning with the Rashomon Augmented Reality Cube (RARc)", In the proceedings of the 55th ACM Technical Symposium on Computer Science Education (SIGCSE 2024), pp. 1877-1882 (2024).

[11] M. Mani, Q. Mazumder: "Incorporating Metacognition into Learning", In the proceedings of the 44th ACM Technical Symposium on Computer Science Education (SIGCSE 2013), pp.53–58 (2013).

[12] P. Caratozzolo, A. Alvarez-Delgado, S. Hosseini: "Metacognitive Awareness and Creative Thinking: The Capacity to Cope with Uncertainty in Engineering", In the proceedings of the IEEE Global Engineering Education Conference (EDUCON 2020), pp.638-643 (2020).

[13] A. G. Abulrub, A. Attridge, M. A. Williams: "Virtual Reality in Engineering Education: The Future of Creative Learning", In the proceedings of the IEEE Global Engineering Education Conference (EDUCON 2011), pp.751-757 (2011).

[14] Z. Sun, D. Zhang, X. Luo, Q. Cao, Z. Li: "An Open-Source Engineering Practice Assistant Training System Based on Virtual Reality", In the proceedings of the IEEE Frontiers in Education Conference (FIE 2020), pp.1-4 (2020).

[15] A. S. M. Venigalla, S. Chimalakonda, "FlowARP-Using Augmented Reality for Visualizing Control Flows in Programs", In the proceedings of the 2023 ACM Conference on Computing Education Practice (CompEd 2023), Vol. 1, pp. 161–167 (2023).

[16] T. Suselo, B. C. Wünsche, A. Luxton-Reilly, "Using Mobile Augmented Reality for Teaching 3D Transformations", In the proceedings of the 52nd ACM Technical Symposium on Computer Science Education (SIGCSE 2021), pp. 872-878 (2021).

Educational Question Insertion into Lecture Videos Using Large Language Models: An Initial Exploration of Chat GPT Prompt Design for Question Content and Insertion Points

Shuo Liu*, Tomoo Inoue**

*Graduate School of Comprehensive Human Sciences, University of Tsukuba, Japan
**Institute of Library, Information and Media Science, University of Tsukuba, Japan
*s2421724@u.tsukuba.ac.jp, **inoue@slis.tsukuba.ac.jp

Abstract - Determining when to ask questions in lecture videos is as important as determining what to ask. Manual authoring of both the question content and the insertion points is labor-intensive and scales poorly. We present an initial approach that uses Large Language Models (LLMs) to propose pedagogically coherent question insertion points together with exemplar questions for monologue-style lecture videos. Grounded in established instructional strategies, we operationalize three insertion types: pre-questioning (before technical terms), post-questioning (immediately after complex explanations), and interpolated questions (prior to extended segments to sustain attention). We then design a structured prompting workflow with (1) transcript segmentation, (2) insertion-point judgment, (3) question generation, and (4) self-critique.

Keywords: Educational technology, Question insertion, Question generation, Large language model, Prompt design

1 INTRODUCTION

On-demand, monologue-style lecture videos support self-paced study but often lead to passive viewing, making attention and comprehension difficult to sustain [1]. To address this, researchers insert in-video questions or brief dialogues to prompt retrieval, direct attention, and improve learning effectiveness [2-7].

Dialogue-style videos let learners vicariously follow a teacher-learner exchange, which can deepen understanding and promote metacognition; however, the content of inserted questions and, crucially, their insertion points have traditionally depended on manual authoring, creating a practical bottleneck for large-scale use [2,3,4]. Recent advances in Large Language Models (LLMs) present an opportunity to automate parts of this workflow.

Work on LLM-based automatic question generation (AQG) shows that models can produce high-quality, content-aligned questions, yet most prior systems still hand-specify the timing of insertion. Because learning impact depends not only on what is asked but also on when it is asked, we investigate whether an LLM can propose pedagogically appropriate insertion points from transcript context while generating example questions for those points [2-5, 8-21].

This paper explores a prompting approach that maps instructional intent to three insertion types of 1) pre-questioning (before key terms), 2) post-questioning (after complex explanations), and 3) interpolated questions (before extended segments to sustain attention), as a basis for scalable, LLM-assisted authoring of more interactive lecture videos.

2 RELATED WORK

2.1 Role of Questions in Education

In education, questions function not only as assessments but as instructional interventions that activate thinking, stimulate interest, deepen understanding, and support long-term retention [13]. Well-designed prompts help learners reorganize information and shift from passive reception to active knowledge construction, particularly when tightly coupled to instructional content [14,15,]. Prior studies show benefits of learner-generated questions for conceptual understanding and inquiry orientations and highlight the value of environments that foster divergent questioning and metacognitive engagement—especially for highly curious learners who tend to self-generate more questions [14,15]. These results underscore that the effectiveness of questions depends on their form and on when and how they are posed within instruction.

2.2 Question Insertion Points in Education

Placement substantively shapes comprehension and downstream learning outcomes. Three placements recur across the literature:

1) Pre-questioning before instruction directs attention to key elements and primes selective processing, improving comprehension and learning outcomes [16-18].
2) Interpolated questions presented before extended segments sustain engagement and orient attention toward essential information [19].
3) Post-questioning immediately after dense explanations consolidates understanding and enhances durable retention via retrieval practice [20,21].

Designing insertion points that explicitly match these temporal functions is therefore central to instructional effectiveness.

2.3 Question Insertion in Video-Based Learning

Monologue lecture videos can induce passive viewing and increased cognitive load when technical terms or dense logic appear abruptly [22]. Embedding questions into the video stream has been shown to improve attention and learning. Empirical work supports insertions before, during, and after viewing; within-video prompts are particularly effective in lecture formats, where they counteract attention decay and help maintain narrative flow [1-7, 23-28]. In practice, representative moments mirror the pedagogical functions above: before definitions (attentional guidance), immediately after complex explanations (comprehension check), and prior to long uninterrupted spans (re-engagement) [2-6]. Collectively, these findings motivate an approach that couples timing with pedagogical intent rather than treating it as an afterthought.

2.4 LLMs for Educational Authoring: Prompting, Segmentation, and Self-Critique

LLMs have advanced automatic question generation (AQG), producing fluent, content-aligned questions tailored to lecture material [8-11]. However, most systems still hand-specify insertion timing [2-5]. Two practical mitigations from recent work are pertinent for timing-aware authoring.

1) Segmentation of long inputs, concept-coherent units to counter the lost-in-the-middle effect and improve coverage of mid-lecture content [29.30].
2) Self-critique pass that checks placement naturalness, contextual fit, and wording, then proposes revise/move/delete edits to reduce instruction mismatch and lower manual correction burden [31-33].

In the present study, these mitigations are combined with explicit intent labels (Pre / Interpolated / Post) and anchor excerpts so that every insertion is auditable against its local context and pedagogical purpose [12,28,34,]

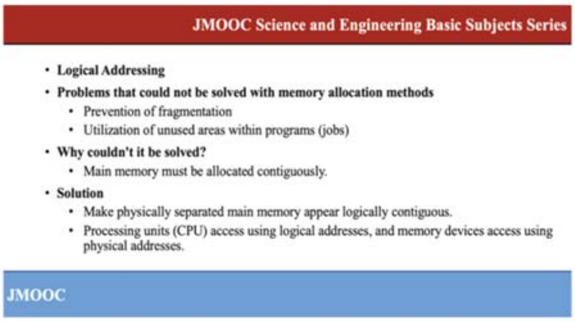
2.5 Human-AI Collaboration on Insertion Point Design

Human-AI workflows illustrate complementary roles for instructors and models. VIVID enables instructors to mark potentially difficult script segments and then uses GPT-4 to generate interactive questions at those human-specified points [5]. Other approaches segment at regular intervals or along structural boundaries to standardize placement, and some detect visual cues (e.g., slide transitions, diagrams, dense subtitle regions) to trigger insertions, enabling context-sensitive strategies for diverse learners (e.g., learners with hearing impairments) [35-37]

These studies collectively argue for flexible, context-aware placement that considers instructional structure, learner characteristics, multimodal information density, and cognitive load. At the same time, they reveal that autonomous timing by LLMs, which infer optimal insertion points directly from transcript context, remains underdeveloped and is precisely the gap this work targets.

3 PROMPT DESIGN

3.1 Problems Observed from the Pilot Study



JMOOC Science and Engineering Basic Subjects Series

- **Logical Addressing**
- **Problems that could not be solved with memory allocation methods**
  - Prevention of fragmentation
  - Utilization of unused areas within programs (jobs)
- **Why couldn't it be solved?**
  - Main memory must be allocated contiguously.
- **Solution**
  - Make physically separated main memory appear logically contiguous.
  - Processing units (CPU) access using logical addresses, and memory devices access using physical addresses.

JMOOC

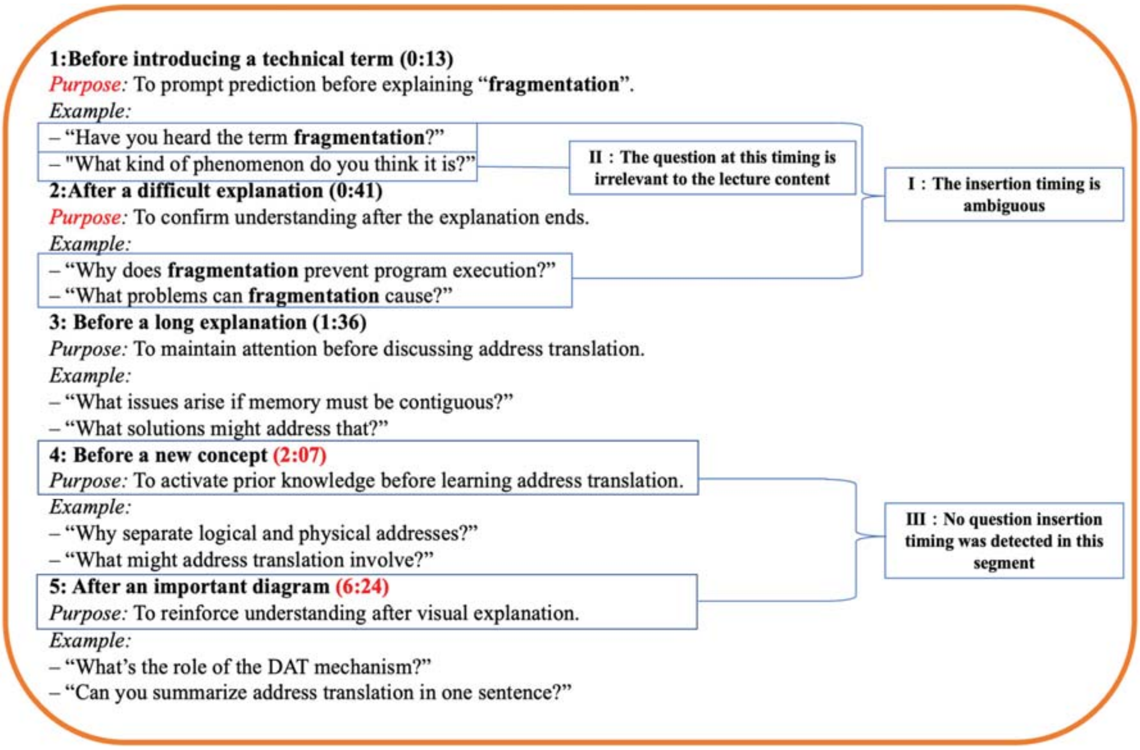
Figure 1: A slide (scene) used in a sample educational video. (Source: JMOOC [38])

As a pilot study, we used the transcript (total 6m24s) of an example educational video "Operating Systems" provided by JMOOC [38] (Figure 1). Based on a zero-shot configuration, the specific instruction given was: *" Please read the following lecture transcript and identify appropriate points for inserting questions. For each point, provide a reason for the insertion and an example question that enhances learning effectiveness."* This configuration yielded the output shown in Figure 2, revealing the following problems.

3.1.1. Accuracy of Insertion points

As indicated by I in Figure 2, Some questions were ambiguous as to whether they should be asked before a term was introduced or after it had been explained. For example, two outputs related to “fragmentation” made it unclear whether the question should appear before or after the explanation. Such ambiguities align with Keluskar et al., who note that when given prompts like “*Where is the Cardinals' home base?"*, LLMs tend to make unfounded assumptions and produce confident yet incorrect outputs (whether it's the baseball's St. Louis Cardinals or American football's Arizona Cardinals) and generate incorrect information with high confidence [39]. Similarly, even with vague insertion instructions, LLMs might infer incorrect premises from incomplete context and generate questions at inappropriate positions. This issue stems from the lack of formality and specificity in prompt design, making it difficult to pinpoint insertion positions and consequently hindering the flow of learning. Therefore, ambiguity in specifying question insertion points is a significant challenge in prompt design.

Furthermore, as shown by III in Figure 2, there was an absence of question insertion in the middle sections of the lecture (e.g., 2’07’’-6’24’’). This pattern reflects the "Lost-in-the-Middle" effect identified by Liu et al., where LLMs are less sensitive to content located in the middle of long inputs. Their study demonstrated that response accuracy varies significantly depending on the position of information, with information placed in the middle being particularly prone to being overlooked [29]. Thus, the phenomenon of questions not being appropriately inserted in the middle of video lectures suggests that LLMs are not adequately recognizing and processing important learning content, revealing a structural limitation in their use for educational support.



1:Before introducing a technical term (0:13)
Purpose: To prompt prediction before explaining "**fragmentation**".
Example:
– "Have you heard the term **fragmentation**?"
– "What kind of phenomenon do you think it is?"
2:After a difficult explanation (0:41)
Purpose: To confirm understanding after the explanation ends.
Example:
– "Why does **fragmentation** prevent program execution?"
– "What problems can **fragmentation** cause?"
3: Before a long explanation (1:36)
Purpose: To maintain attention before discussing address translation.
Example:
– "What issues arise if memory must be contiguous?"
– "What solutions might address that?"
4: Before a new concept (2:07)
Purpose: To activate prior knowledge before learning address translation.
Example:
– "Why separate logical and physical addresses?"
– "What might address translation involve?"
5: After an important diagram (6:24)
Purpose: To reinforce understanding after visual explanation.
Example:
– "What's the role of the DAT mechanism?"
– "Can you summarize address translation in one sentence?"

II：The question at this timing is irrelevant to the lecture content
I：The insertion timing is ambiguous
III：No question insertion timing was detected in this segment

Figure 2：Example output of automatic question insertion using Zero-shot prompting

3.1.2. Appropriateness of Question

As illustrated by II in Figure 2, Although the insertion point was appropriate, the content of the generated question did not correspond to the lecture material. For example, the question *"What kind of phenomenon do you think it is?"* was output before the explanatory text, but it did not align with the preceding or following context. Such outputs represent a type of faithfulness hallucination, where the generated content appears fluent and confident but is misaligned with the underlying source context. Among the types of hallucinations classified by Huang et al., this corresponds to the instruction mismatch pattern, where the model misunderstood the user's intent and output a question based on an inappropriate context [40].

3.2 Prompt Design Strategy

This section introduces two prompt strategies developed to address the limitations identified in the zero-shot configuration. Specifically, they are ambiguity in insertion points and inconsistencies in the pedagogical relevance of generated questions. The goal is to improve both the accuracy of question generation and its educational effectiveness.

3.2.1. Explicit Insertion Points Specification

To reduce ambiguity in LLM outputs, the prompt explicitly specifies when questions should be inserted, according to pedagogical intent [41]. Accordingly, in this study, we incorporated a structure that clearly communicates the pedagogical purpose of question insertion to the LLM. Based on insights from prior research (Section 2.2), we defined three categories—pre-questioning, interpolated questions, and post-questioning—according to their timing within the lecture and their pedagogical objectives. This design enhances the explicitness of prompt formulation [16,18,20,21, 26,28,34,].

Pre-questioning: Questions inserted before the introduction of technical terms or new concepts, aiming to elicit learners' prior knowledge and imagination [16,18].

Post-questioning: Questions inserted immediately after complex explanations or difficult logical developments, to confirm learner comprehension [20,21,34].

Interpolated questions: Questions inserted during lengthy explanations or at topic boundaries to prompt re-engagement and attention [26,28].

By embedding concrete instructions corresponding to these categories into the prompt, the LLM is guided to understand *why* a question should be inserted at a particular point. This design is intended to align the insertion point of the generated questions with their pedagogical objectives.

Additionally, to mitigate the Lost-in-the-Middle problem [29], where LLMs underrepresent mid-sequence content, we segmented the lecture transcript into 30-second intervals [30]. Each segment was then processed independently. This segmentation helps the model attend to important instructional moments throughout the lecture while maintaining coherence within each segment.

3.2.2. Ensuring Educational Appropriateness of Output

To ensure that the generated questions are logically aligned with the lecture content, contextually coherent with the specified insertion point, and whether they fulfill their intended educational purpose, we adopted a Self-Critique approach. This approach prompts the LLM to evaluate its own output based on predefined criteria and revise it if necessary [31-33].

The LLM is guided to assess whether each insertion point aligns with the lecture flow and whether both the placement and content of each question serve pedagogical objectives. By performing this self-critique after generation, the model helps reduce the need for manual correction or evaluation.

3.3 Proposed Prompt Design

Based on the above design principles, question insertion was performed according to the following procedure.

1) Transcript Segmentation

The lecture audio was transcribed using Automatic Speech Recognition (ASR) to obtain a full transcript (Figure 3). The LLM was then instructed to divide the transcript into semantically coherent segments, with an approximate target length of 30 seconds per segment [30].

Prompt example*: "Please divide the following lecture transcript into segments based on conceptual or explanatory units. Use approximately 30 seconds as a guideline for segment length. Then, for any segment that meets the conditions below, insert an appropriate question."*

2) Identifying Insertion Points

For each segment, the LLM identifies appropriate insertion points for questions, based on three pedagogically grounded categories—pre-questioning, post-questioning, or interpolated questions—as defined in prior educational literature [16,18,20,21, 26,28,34].

Prompt example*: "For each segment, determine whether a question should be inserted based on the following conditions:*

- Before important terms or concepts (Pre-questioning)

- After difficult explanations (Post-questioning)

*- Before long or uninterrupted segments (*Interpolated questions*)"*

3) Generating Questions

If the insertion point is judged to be appropriate, a question is generated at that position.

Prompt example *: "If the conditions are met, please generate a question at that position."*

4) Self-Critique

The purpose of this stage is to have the LLM itself evaluate the generated questions and their insertion points [31-33].

This enables the LLM to autonomously assess whether each question is appropriately placed and whether its insertion disrupts the lecture's flow or comprehension.

Prompt example*: "Please check the coherence of the lecture transcript after inserting questions throughout. Evaluate whether the questions are inserted at natural positions and if their insertion point does not hinder the lecture's flow or comprehension."*

4 CONCLUSION

This study tackles the problems of passive engagement often associated with lecture-style videos by using large language models (LLMs), specifically ChatGPT-4o, to automatically insert pedagogically motivated questions at three insertion types: Pre, Post, and Interpolated.

By emphasizing not only question quality but also when questions are posed, this work addresses a gap in prior research. The study aims to investigate if LLMs can meaningfully enhance the interactivity of lecture videos, laying groundwork for AI-assisted, personalized learning support that reduces instructor burden and promotes deeper understanding. Future work will include the evaluation study of the proposed prompt design.

ACKNOWLEDGMENT

This research was partially supported by the JSPS KAKENHI Grant Number 25K15367.

REFERENCES

[1] Chi, M. T., Kang, S., and Yaghmourian, D. L. (2017). Why students learn more from dialogue-than monologue-videos: Analyses of peer interactions. Journal of the Learning Sciences, 26(1), 10-50.

[2] Nugraha, A., Wahono, I. A., et al. (2020). Creating dialogue between a tutee agent and a tutor in a lecture video improves students' attention. In International Conference on Collaboration Technologies and Social Computing (pp. 96-111). Cham: Springer International Publishing.

[3] Nugraha, A., Inoue, T., et al. (2023). A Dialogue-Like Video Created From a Monologue Lecture Video Provides Better Learning Experience. International Journal of Distance Education Technologies (IJDET), 21(1), 1-21.

[4] Tanprasert, T., Fels, S. S., et al. (2023). Scripted vicarious dialogues: Educational video augmentation method for increasing isolated students' engagement. In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (pp. 1-25).

[5] Choi, S., Lee, H., et al. (2024). VIVID: Human-AI collaborative authoring of vicarious dialogues from lecture videos. In Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (pp. 1-26).

[6] Haagsman, M. E., Scager, K., et al. (2020). Pop-up questions within educational videos: Effects on students' learning. Journal of Science Education and Technology, 29(6), 713-724.

[7] Wachtler, J., Khalil, M., et al. (2016). On using learning analytics to track the activity of interactive MOOC videos. In LAK 2016 Workshop on Smart Environments and Analytics in Video-Based Learning.

[8] Lee, U., Jung, H., et al. (2024). Few-shot is enough: exploring ChatGPT prompt engineering method for automatic question generation in english

education. Education and Information Technologies, 29(9), 11483-11515.
[9] Al Faraby, S., and Romadhony, A. (2024). Analysis of llms for educational question classification and generation. Computers and Education: Artificial Intelligence, 7, 100298.
[10] Mulla, N., and Gharpure, P. (2023). Automatic question generation: a review of methodologies, datasets, evaluation metrics, and applications. Progress in Artificial Intelligence, 12(1), 1-32.
[11] Scaria, N., Dharani Chenna, S., and Subramani, D. (2024). Automated Educational Question Generation at Different Bloom's Skill Levels Using Large Language Models: Strategies and Evaluation. In International Conference on Artificial Intelligence in Education (pp. 165-179). Cham: Springer Nature Switzerland.
[12] Little, J., and Bjork, E. (2011). Pretesting with multiple-choice questions facilitates learning. In Proceedings of the annual meeting of the cognitive science society (Vol. 33, No. 33).
[13] Tofade, T., Elsner, J., and Haines, S. T. (2013). Best practice strategies for effective use of questions as a teaching tool. American journal of pharmaceutical education, 77(7), 155.
[14] Chin, C., and Osborne, J. (2008). Students' questions: a potential resource for teaching and learning science. Studies in science education, 44(1), 1-39.
[15] Alaimi, M., Law, E., et al. (2020). Pedagogical agents for fostering question-asking skills in children. In Proceedings of the 2020 CHI conference on human factors in computing systems (pp. 1-13).
[16] Carpenter, S. K., Rahman, S., and Perkins, K. (2018). The effects of prequestions on classroom learning. Journal of Experimental Psychology: Applied, 24(1), 34.
[17] Carpenter, S. K., and Toftness, A. R. (2017). The effect of prequestions on learning from video presentations. *Journal of Applied Research in Memory and Cognition*, *6*(1), 104-109.
[18] Richland, L. E., Kornell, N., and Kao, L. S. (2009). The pretesting effect: Do unsuccessful retrieval attempts enhance learning?. Journal of Experimental Psychology: Applied, 15(3), 243.
[19] Andre, T. (1979). Does answering higher-level questions while reading facilitate productive learning?. Review of educational research, 49(2), 280-318.
[20] Adesope, O. O., Trevisan, D. A., and Sundararajan, N. (2017). Rethinking the use of tests: A meta-analysis of practice testing. Review of educational research, 87(3), 659-701.
[21] Dunlosky, J., Rawson, K. A., et al. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. Psychological Science in the Public interest, 14(1), 4-58.
[22] Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. Cognitive science, 12(2), 257-285.
[23] Schacter, D. L., and Szpunar, K. K. (2015). Enhancing attention and memory during video-recorded lectures. Scholarship of Teaching and Learning in Psychology, 1(1), 60.
[24] Szpunar, K. K., Jing, H. G., and Schacter, D. L. (2014). Overcoming overconfidence in learning from video-recorded lectures: Implications of interpolated testing for online education. Journal of Applied Research in Memory and Cognition, 3(3), 161-164.
[25] Kovacs, G. (2016). Effects of in-video quizzes on MOOC lecture viewing. In Proceedings of the third (2016) ACM conference on Learning@ Scale (pp. 31-40).
[26] Szpunar, K. K., Khan, N. Y., and Schacter, D. L. (2013). Interpolated memory tests reduce mind wandering and improve learning of online lectures. Proceedings of the National Academy of Sciences, 110(16), 6313-6317.
[27] Van der Zee, T., Davis, D., et al. (2018). Evaluating retrieval practice in a MOOC: how writing and reading summaries of videos affects student learning. In Proceedings of the 8th international conference on learning analytics and knowledge (pp. 216-225).
[28] Weinstock, M., Pallaci, M., et al. (2020). Effect of interpolated questions on podcast knowledge acquisition and retention: a double-blind, multicenter, randomized controlled trial. Annals of Emergency Medicine, 76(3), 353-361.
[29] Liu, N. F., Lin, K., et al. (2024). Lost in the middle: How language models use long contexts. Transactions of the Association for Computational Linguistics, 12, 157-173.
[30] Shi, W., Li, S., et al. (2024). Segment+: Long text processing with short-context language models. In Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing (pp. 16605-16617).
[31] Saunders, W., Yeh, C., et al. (2022). Self-critiquing models for assisting human evaluators. arXiv preprint arXiv:2206.05802.
[32] Gou, Z., Shao, Z., et al. (2023). Critic: Large language models can self-correct with tool-interactive critiquing. arXiv preprint arXiv:2305.11738.
[33] Yao, Z., Parashar, A., et al. (2025). Mcqg-srefine: Multiple choice question generation and evaluation with iterative self-critique, correction, and comparison feedback. In Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers) (pp. 10728-10777).
[34] Rickards, J. P. (1979). Adjunct postquestions in text: A critical review of methods and processes. Review of Educational Research, 49(2), 181-196.
[35] Kang, Y. B., Forkan, A. R. M., et al. (2021). An AI-based Solution for Enhancing Delivery of Digital Learning for Future Teachers. arXiv preprint arXiv:2112.01229.
[36] Gala, R., Vijayaraghavan, R., et al. (2021). Real-time cognitive evaluation of online learners through automatically generated questions. In 2021 IEEE 20th

International Conference on Cognitive Informatics & Cognitive Computing (ICCI* CC) (pp. 53-58).

[37] Cheng, S., Huffman, S., et al. (2024). "Real Learner Data Matters" Exploring the Design of LLM-Powered Question Generation for Deaf and Hard of Hearing Learners. arXiv preprint arXiv:2410.00194.

[38] JMOOC. (n.d.). Japan Massive Open Online Education Promotion Council. https://www.jmooc.jp/

[39] Keluskar, A., Bhattacharjee, A., and Liu, H. (2024). Do LLMs Understand Ambiguity in Text? A Case Study in Open-world Question Answering. In 2024 IEEE International Conference on Big Data (BigData) (pp. 7485-7490).

[40] Huang, L., Yu, W., et al. (2025). A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions. ACM Transactions on Information Systems, 43(2), 1-55.

[41] Rawte, V., Priya, P., et al. (2023). Exploring the relationship between llm hallucinations and prompt linguistic nuances: Readability, formality, and concreteness. arXiv preprint arXiv:2309.11064.

Keynote Speech 2:
Dr. Van Cu Pham

(Research Assistant Professor,
Center for Digitalization Endeavors,
Japan Advanced Institute of Science and Technology)
(Chair: Shoji Sano)

Advancing Smart Home Interoperability: Our Current Efforts with ECHONET Consortium

Van Cu PHAM

Center for Digitalization Endeavors

Research Assistant Professor

cupham@jaist.ac.jp

September 2, 2025

Today's Agenda

1. Smart Homes Market Overview
2. ECHONET Lite and ECHONET Lite Web API
3. ECHONET Lite Web API x HL7 FHIR
4. ECHONET Lite (Web API) x W3C Web of Things
5. ECHONET Lite (Web API) x Matter

1. Smart Homes Market Overview

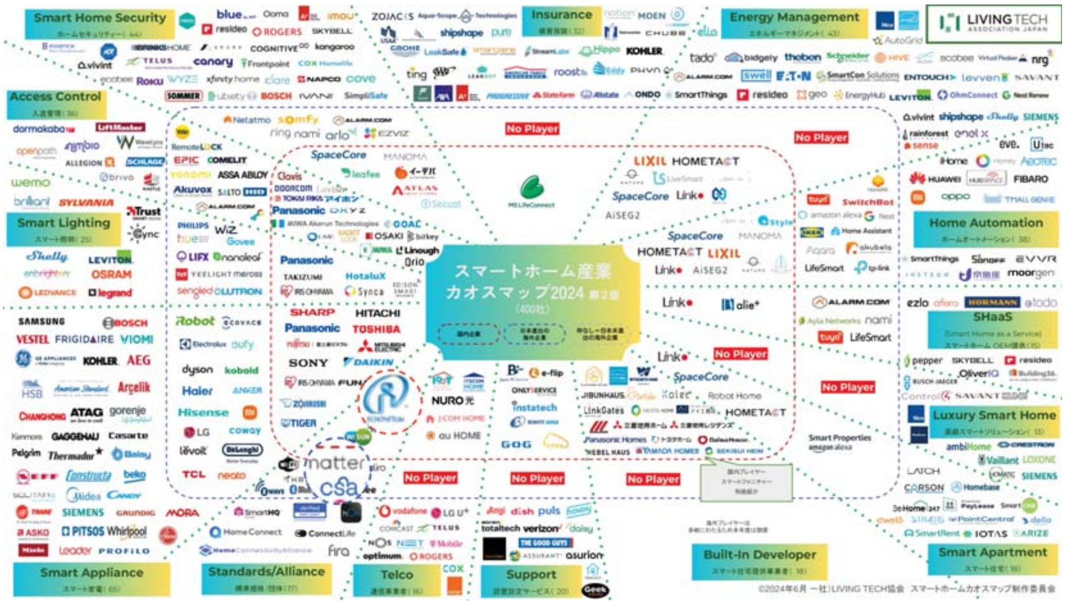
Smart Homes: The Place You Spend the Most of Your Time

- **We spend most of our lives at home** → safety, comfort, well-being, energy saving...
- Smart home = smart appliances + HEMS/smart meters + personal healthcare devices + entertainment systems + sensors/actuators + security cameras + EVs ...

  → Collaboration of devices from different domains can enhance the quality of life for us!

iHouse, our advanced experimental environment for Japan's future smart homes, with commercial and experimental smart home appliances (more than 200) utilizing ECHONET Lite version 1.1. (Photo credits `https://article.murata.com/en-global/article/smart-home-1`)

Smart Home Market is Fragmented



ECHONET

Japanese manufacturers

Release: ECHONET Lite (Device Layer) and ECHONET Lite Web API (Cloud Layer)

Matter

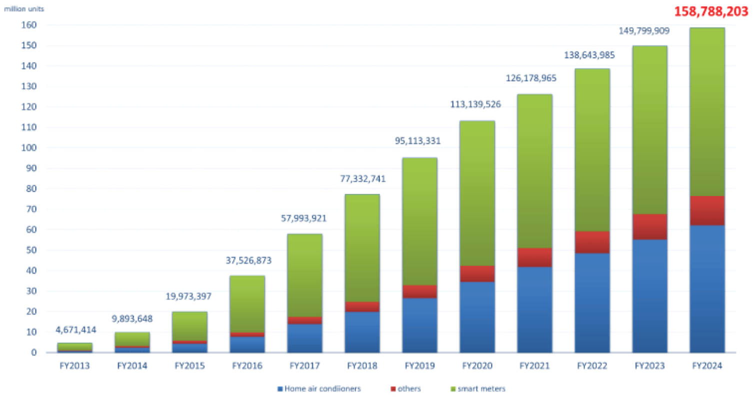
Introduced by GAFA

Emerging smart home protocol worldwide

Interoperability of Standards/Alliance (ECHONET x HL7 FHIR, Matter, W3C Web of Things)

2. ECHONET Lite and ECHONET Lite Web API

ECHONET Lite and ECHONET Lite Web API



Source: `https://echonet.jp/`

Shipment Report

Total: Approximately 159 million units

Smart meter, Home Airconditioner, etc

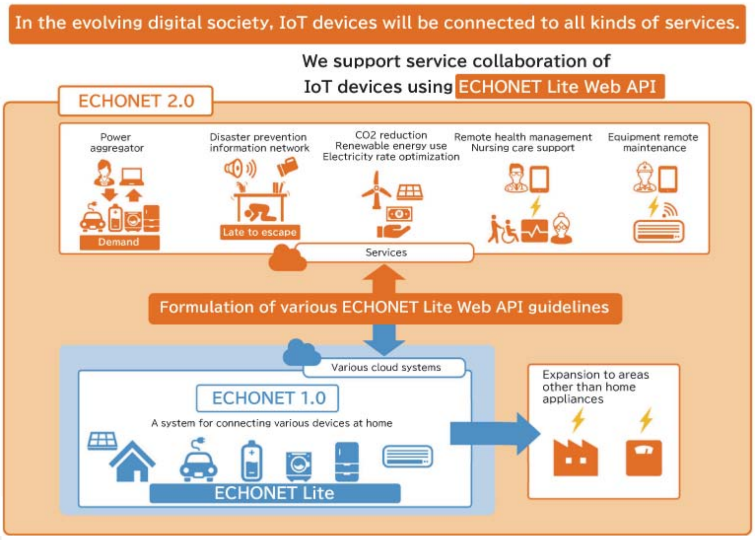
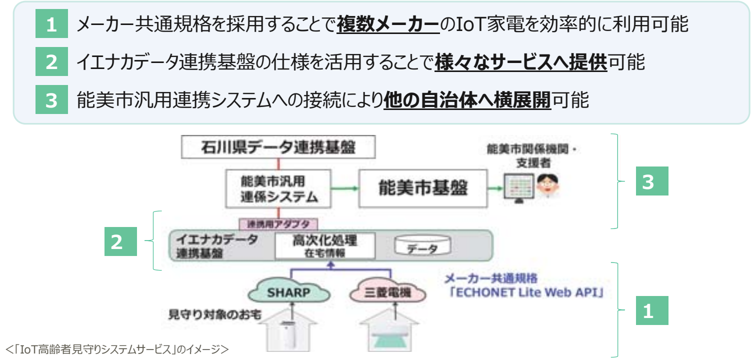

In the evolving digital society, IoT devices will be connected to all kinds of services.
We support service collaboration of IoT devices using ECHONET Lite Web API
ECHONET 2.0
Power aggregator
Demand
Disaster prevention information network
Late to escape
CO2 reduction
Renewable energy use
Electricity rate optimization
Remote health management
Nursing care support
Equipment remote maintenance
Services
Formulation of various ECHONET Lite Web API guidelines
Various cloud systems
ECHONET 1.0
A system for connecting various devices at home
ECHONET Lite
Expansion to areas other than home appliances

Image Credits: `https://echonet.jp/`

ECHONET Lite: ISO/IEC international standard

ECHONET Lite Web API In Use (In Japan)

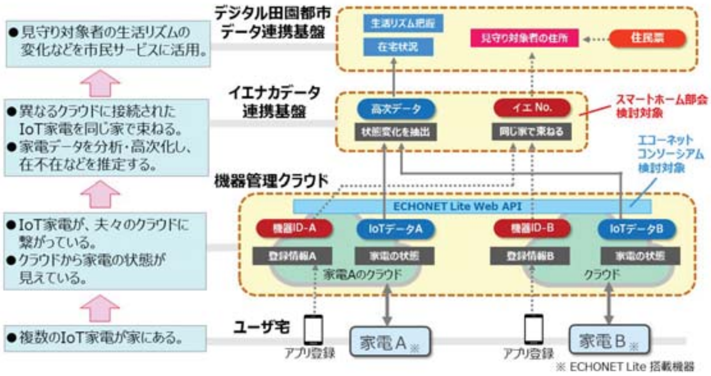


1 メーカー共通規格を採用することで**複数メーカー**のIoT家電を効率的に利用可能

2 イエナカデータ連携基盤の仕様を活用することで**様々なサービスへ提供**可能

3 能美市汎用連携システムへの接続により**他の自治体へ横展開**可能

<「IoT高齢者見守りシステムサービス」のイメージ>


●見守り対象者の生活リズムの変化などを市民サービスに活用。
●異なるクラウドに接続されたIoT家電を同じ家で束ねる。
●家電データを分析・高次化し、在不在などを推定する。
●IoT家電が、夫々のクラウドに繋がっている。
●クラウドから家電の状態が見えている。
●複数のIoT家電が家にある。
デジタル田園都市データ連携基盤
生活リズム把握
在宅状況
見守り対象者の住所
住民票
イエナカデータ連携基盤
高次データ
状態変化を抽出
イエ No.
同じ家で束ねる
スマートホーム部会検討対象
エコーネットコンソーシアム検討対象
機器管理クラウド
ECHONET Lite Web API
機器ID-A
IoTデータA
登録情報A
家電の状態
家電Aのクラウド
機器ID-B
IoTデータB
登録情報B
家電の状態
クラウド
ユーザ宅
アプリ登録
家電A※
アプリ登録
家電B※
※ ECHONET Lite 搭載機器

- ELWA as an interface to develop elderly care services in Nomi City, Ishikawa Prefecture
- ELWA as an interface to connect smart city assets (in the Smart City Reference Architecture White Paper 3rd Edition by Japan Cabinet Office (in Japanese))

3. ECHONET Lite Web API x HL7 FHIR

ECHONET Lite Web API x HL7 FHIR

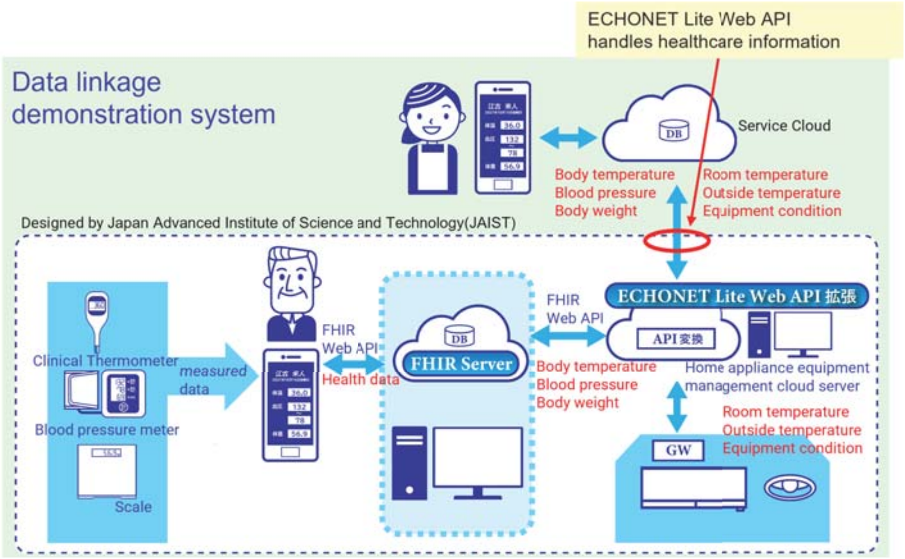


Photo credits: `https://echonet.jp/activities_with_pcha_en/`
Video → `https://www.youtube.com/watch?v=i0QiWTU0zAE`

ECHONET Lite Web API

Common API at the cloud level = Common Data Model, Common API used by different vendors

HL7 FHIR: Fast Healthcare Interoperability Resources

Set of rules and specifications for the exchange of electronic health care data (cloud level)

Health-related data and living condition data can be remotely accessed

ECHONET Lite Web API x HL7 FHIR (Cont.)

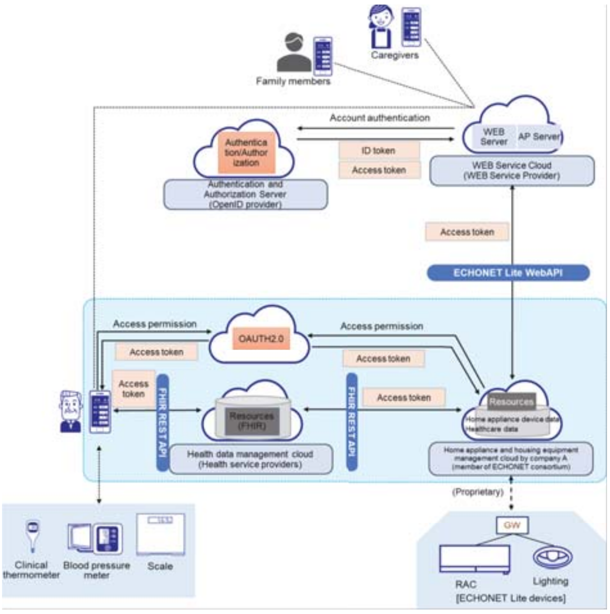


Photo credits: `https://echonet.jp/`

- Access control: User-Managed Access 2.0
  - Resource owners control and grant access permission

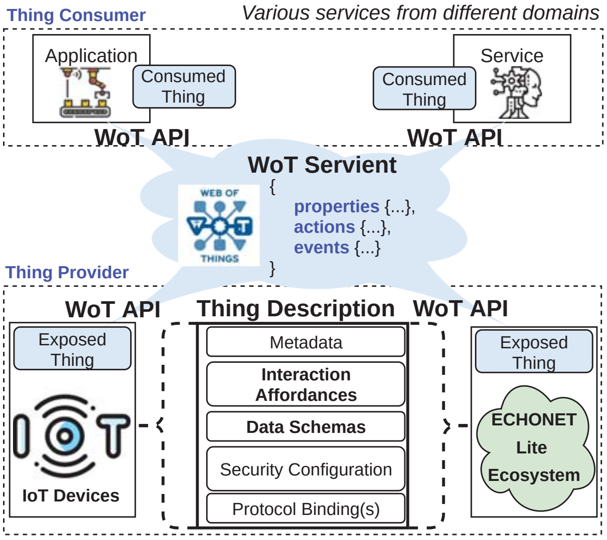
Possible use-cases

1. Caregivers can know the living condition and health condition → Reduce workload and time
2. Early signs of dementia (e.g, users tend to forget), early signs of abnormal ADL (e.g, users do not go to the toilet...)
3. Save lives in critical conditions (such as natural disaster, earthquake...)

---

4. ECHONET Lite (Web API) x W3C Web of Things

Web Access for Smart Appliances via W3C Web of Things



- W3C WoT = *Lingua Franca* for the IoT via Web technologies

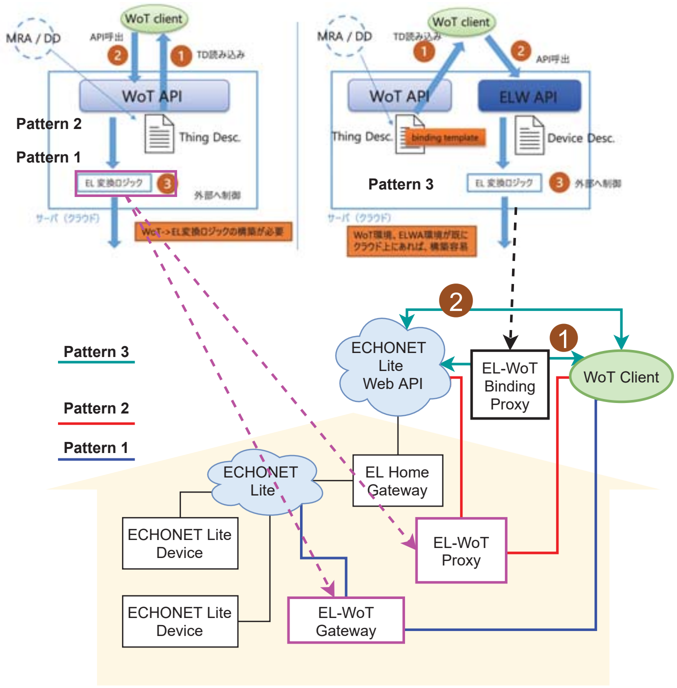
  → This integration improves compatibility and brings more services for ECHONET based smart homes in Japan

X Current solutions: **outdated**, **fragmented** data models for the ECHONET Lite (ENL) -WoT integration

Contributions

- A unique WoT **Thing Description** for the ECHONET Lite
- New approaches for ENL-WoT integration with the **ECHONET Lite WebAPI**

Web Access for Smart Appliances via W3C Web of Things (cont.)



Pattern 1 and **Pattern 2** are similar

- WoT Clients talk to a **Proxy/Gateway** → No need to adjust ELW API and WoT API

Pattern 3

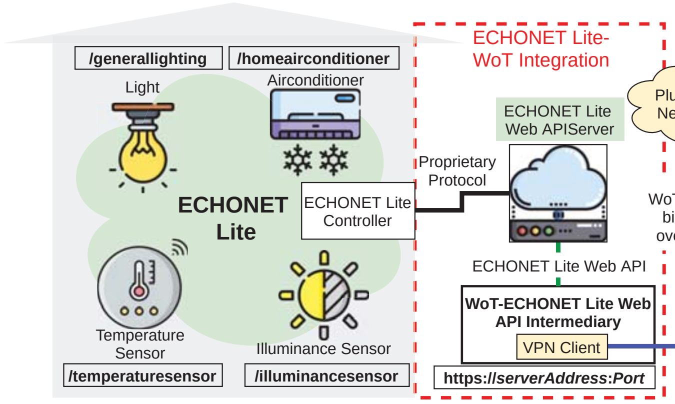
- WoT Clients talk **directly** with **ELW API** using **information** from **binding templates**

Conclusion

Pattern 2 is the suitable solution (Support not only ECHONET Lite but other protocols)

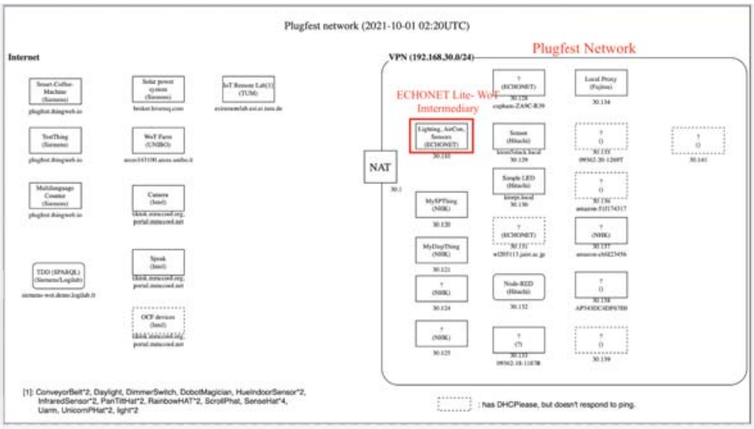
Pattern 3 is desirable (No need translator/adaptor), but not yet achievable

At W3C WoT Plugfest/Testfest 2021



Japan: NHK, Fujitsu, Hitachi, ECHONET Consortium

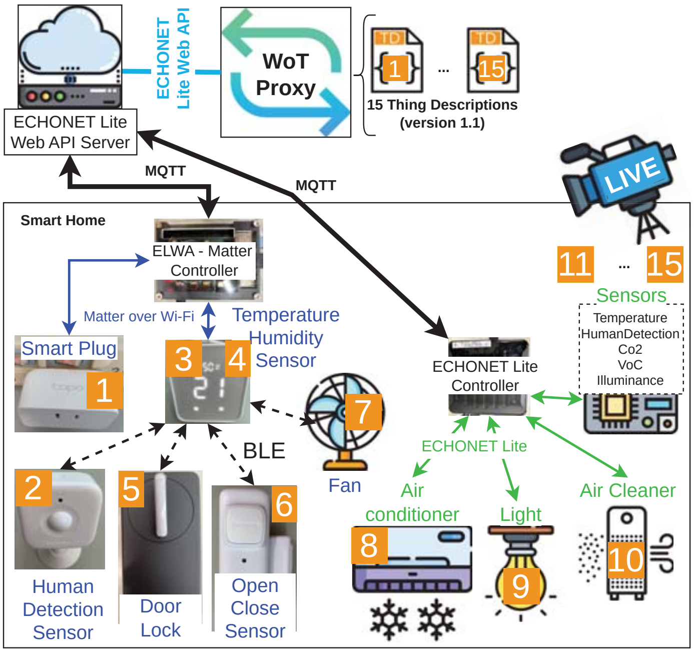
Europe: Intel, Siemens, University of Bologna, Technical University of Munich


Plugfest network (2021-10-01 02:20UTC)
Internet
VPN (192.168.30.0/24)
Plugfest Network
ECHONET Lite- WoT Intermediary
NAT

Results

- We are the only team that got 100% **Pass** on all tests

At W3C WoT Plugfest/Testfest 2024



Added Matter devices to the test

- But use same source of the ELWA-WoT Proxy

Matter and ECHONET Lite devices are similar by WoT Clients

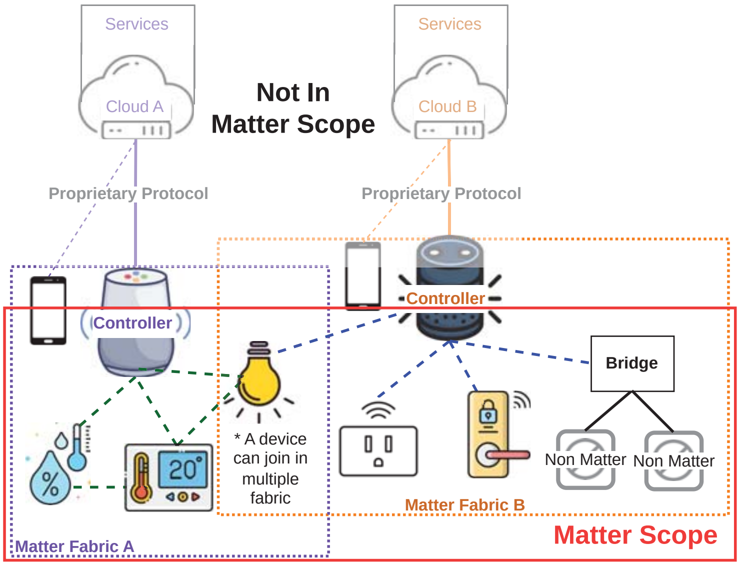
Results

- No evaluator

Has been invited for the Plugfest/Testfest 2025 (Kobe, Japan)

5. ECHONET Lite (Web API) x Matter

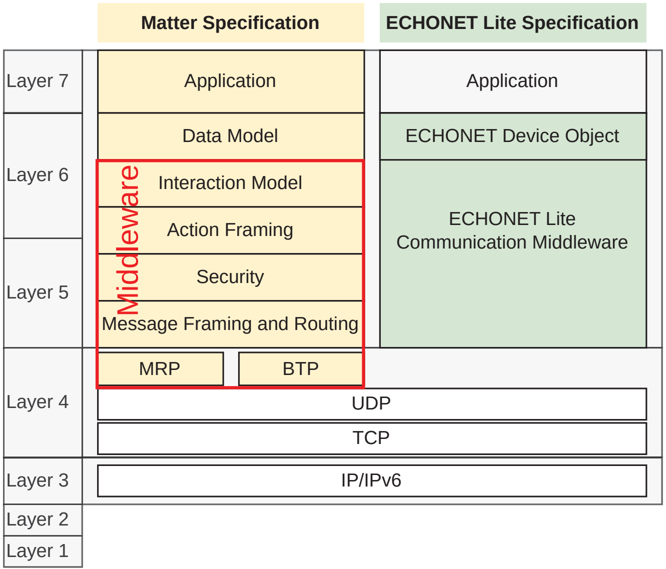
Matter Overall



A Simplified Matter Topology: Controller A supports Thread, Controller B supports Wi-Fi

- Matter: open protocol supports multiple vendors' interoperability for smart homes/IoT
- Scope:
  Device ↔ Device,
  Device↔Controller, Non Matter↔Matter
- Fabric: A logical group of devices. A device can join multiple fabrics (connect to multiple controllers)

Basic Concept: Matter Protocol Stack

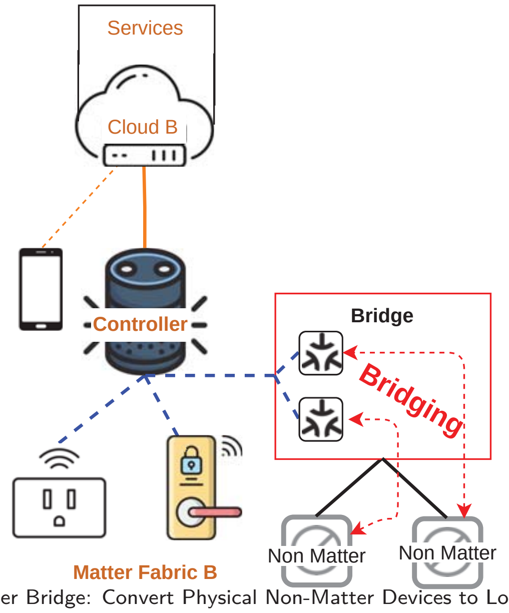


MRP: Matter Reliable Protocol **BTP**: Bluetooth Transport Protocol

- Matter *Middleware* open sources:
  - ▸ C++
  - ▸ TypeScript
  - ▸ Rust
- Data Model: Device Definition
  - v1.3 : Added enegery related cluster → For energy related devices
  - v1.4 (Release November 7th, 2024): More energy-related devices, better support for battery-operated devices.

Matter (and ECHONET Lite) Protocol Stack

Basic Concept: Matter Bridge

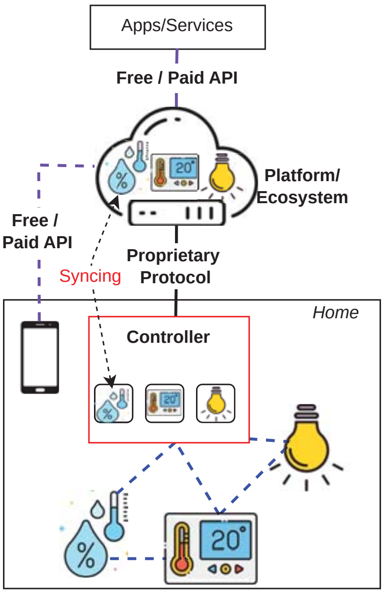


- Matter Bridge is defined in the Matter Core Specification
  - Device Type: → **Aggregator** (`0x000E`)
  - How to expose functionality and metadata
  - How to construct endpoints
- Commercial Bridges:
  - Phillip Hue Bridge: Convert Hue light bubs (Zigbee) into Matter lights
  - Switchbot Hub 2: Convert sensors/actuators (Bluetooth Low Energy) into Matter sensors, actuators
  - ...

→ ECHONET Lite - Matter Bridge

Matter Bridge: Convert Physical Non-Matter Devices to Logical Matter Endpoints

Basic Concept: Matter Controller



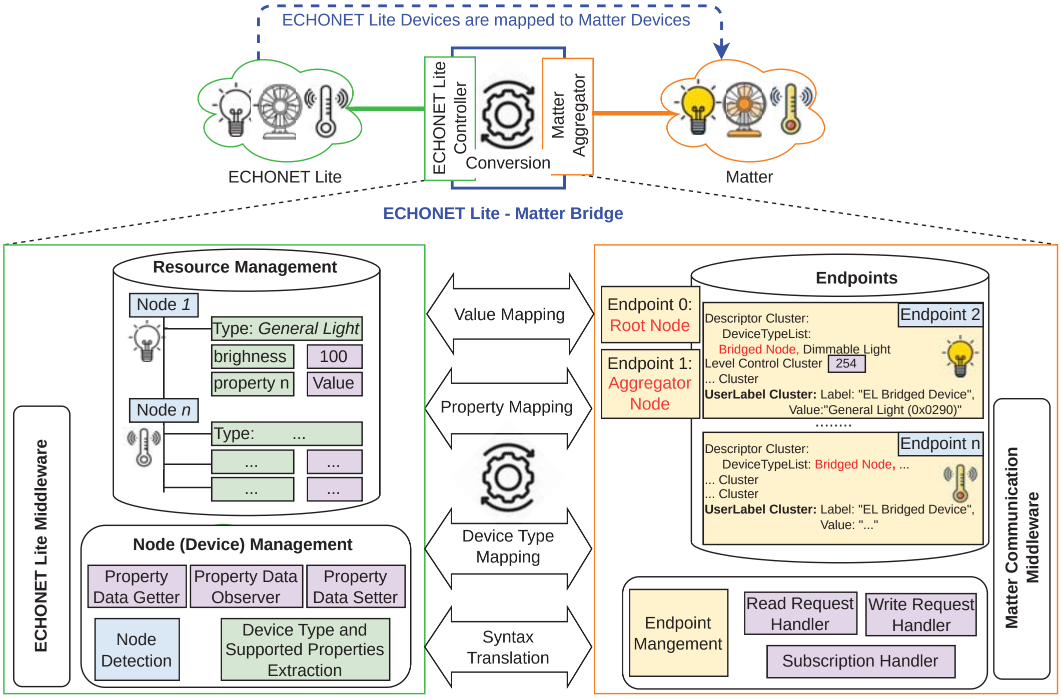
Matter Controller + Platform: Transforming Physical Devices

- Matter Controller:
  - Form a fabric, commissioning Matter devices into a fabric
  - Discover, Read/Write/Invoke/Subscribe and Interact with Matter device
  - Stick with a Platform/ Ecosystem.
- Commercial Controllers/Hubs:
  - Amazon Echo: Device ↔ Alexa Skill
  - Apple Homepod: Device ↔ Home Kit
  - Google Nesthub: Device ↔ Google Home
  - ...
- Experimental/ Open-source Controller
  - `chip-tool`: general purpose controller

→ (2) Matter-ECHONET Lite Web API Controller

5.1. ECHONET Lite → Matter Bridge

ECHONET Lite → Matter Bridge Concept



---

(1) ECHONET Lite Matter Bridge's Logic

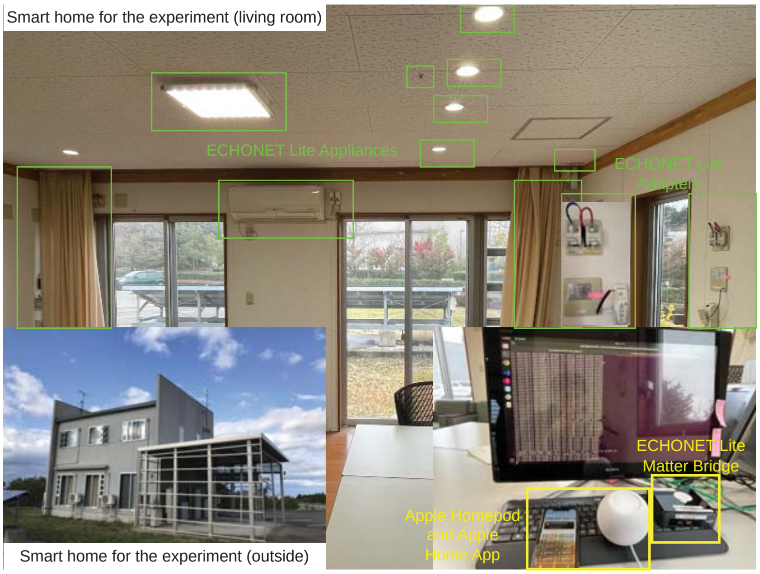
1. Instantiate `endpoint 0`: type `Root Node` and `endpoint 1`: type `Aggregator`
2. Discover ECHONET Lite devices (multicast)

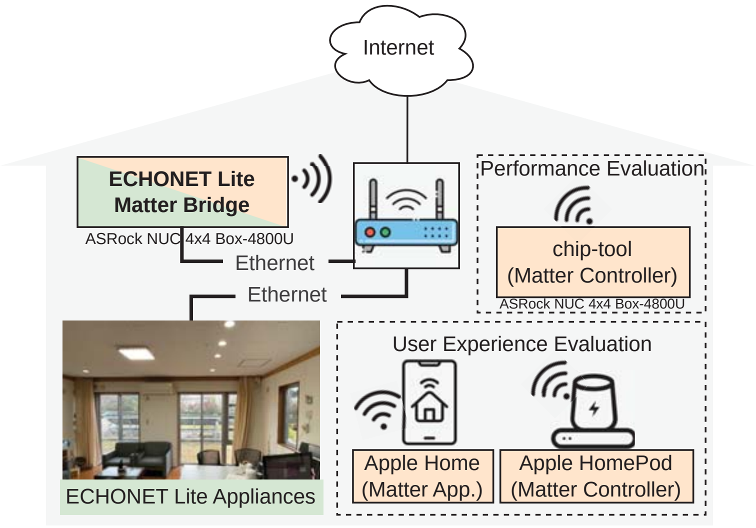
Foreach new ECHONET Lite node:
- 2.1 Extract device type, supported properties and data
- 2.2 Apply rule to convert to Matter endpoint
- 2.3 Add endpoint to the bridge and register interaction handlers

3. Periodically check the network of ECHONET Lite devices

(1) new ECHONET Lite node is detected → 2.1, 2.2, 2.3
(2) ECHONET Lite node left the network
- 3.1 Find the endpoint ID
- 3.2 Unregister the endpoint

(1) Implementation and Deployment


Smart home for the experiment (living room)
ECHONET Lite Appliances
ECHONET Lite Adapters
ECHONET Lite Matter Bridge
Apple Homepod and Apple Home App
Smart home for the experiment (outside)



Result: 116 endpoints are successfully mapped in less than 7 seconds (conducted 50 times)

---

(1) ECHONET Lite - Matter Bridge: Remark

Results

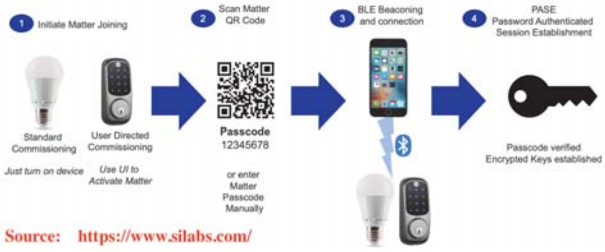
- ECHONET Lite devices (in the iHouse) are exposed as Matter devices effortlessly. Source code + video → `https://github.com/Tan-Lab/connectedhomeip`
- The bridge is tested with experimental controller (`chip-tool`) and commercial controllers (**Apple Homepod**, **Google Nesthub**)

Limitations

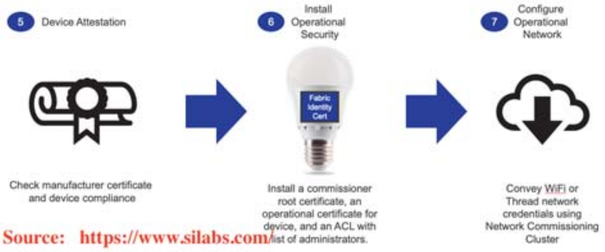
- ECHONET Lite → Matter mapping is implemented, but ECHONET Lite ← Matter mapping is not supported. Theoretically, it is impossible to have a two-direction mapping bridge on a single Matter bridge
- Properties (Functionalities) are lost when mapping to Matter is unavoidable

5.2. Matter - ECHONET Lite Web API Controller

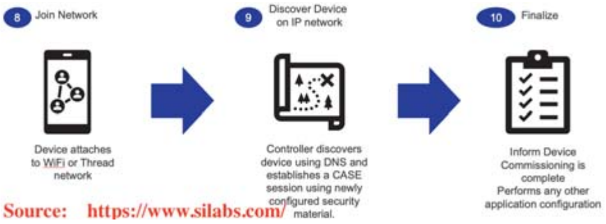
Basic Concept: Commissioning A Node


1 Initiate Matter Joining
Standard Commissioning
Just turn on device
User Directed Commissioning
Use UI to Activate Matter
2 Scan Matter QR Code
Passcode
12345678
or enter Matter Passcode Manually
3 BLE Beaconing and connection
4 PASE Password Authenticated Session Establishment
Passcode verified Encrypted Keys established

Source: https://www.silabs.com/


5 Device Attestation
Check manufacturer certificate and device compliance
6 Install Operational Security
Fabric Identity Cert
Install a commissioner root certificate, an operational certificate for device, and an ACL with list of administrators.
7 Configure Operational Network
Convey WiFi or Thread network credentials using Network Commissioning Cluster

Source: https://www.silabs.com/


8 Join Network
Device attaches to WiFi or Thread network
9 Discover Device on IP network
Controller discovers device using DNS and establishes a CASE session using newly configured security material.
10 Finalize
Inform Device Commissioning is complete
Performs any other application configuration

Source: https://www.silabs.com/

A Matter Controller will

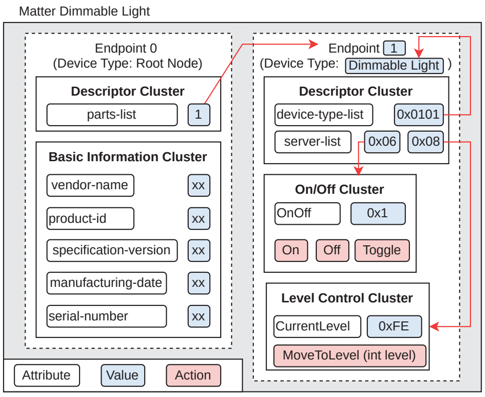
Check certificates

- **Authorized Cert.**: Add to fabric
- **Experimental Cert.**: Notify and ask for the user's approval, then add to fabric
- **Non Authorized Cert.**: Reject

On Success: Return `node ID` (for the specific fabric)

Then browse the node to get its resources

Basic Concept: Browsing a Node

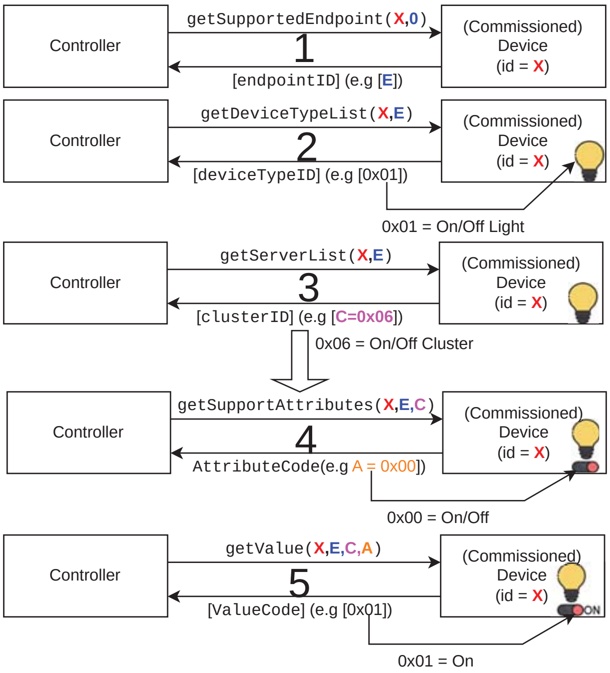


- Node ← Physical Device:
  - Endpoint 0 ← README
    - Cluster `x`
      -Properties
      -Actions
      -Events
    - Cluster `y`
      -Properties
      -Actions
      -Events
  - Endpoint `n`

How to browse a matter node

1. Endpoint 0 → `Descriptor cluster`→ `parts-list` → **supported endpoint id(s)**
2. For each `endpoint id`:
   `Descriptor cluster` → `device-type-list` → **device type**
   `Descriptor cluster` → `server-list` → **support cluster id(s)**
3. For each `cluster id`: Refer to the specification to get properties (and values), actions, events

Basic Concept: Browsing a Node (Cont.)



Observation

- Not so many properties can report status changed event
- Need to poll/subscribe the device to get the latest value
- Can use `Wildcard (0xFF)` to get all properties

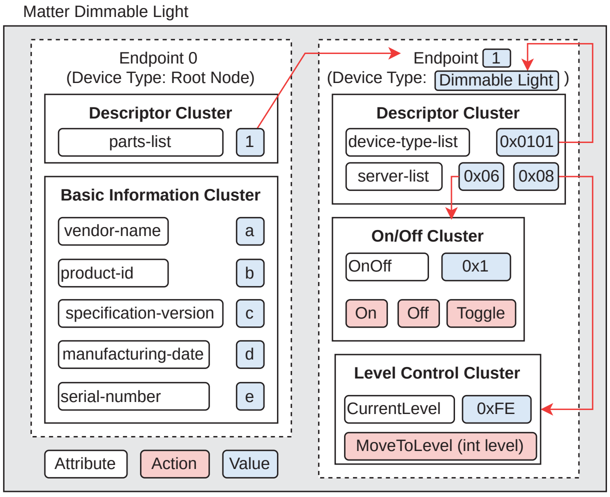
→ Trade-off: real time data vs battery life

→ Controller's Implementations are different on how to sync data from devices

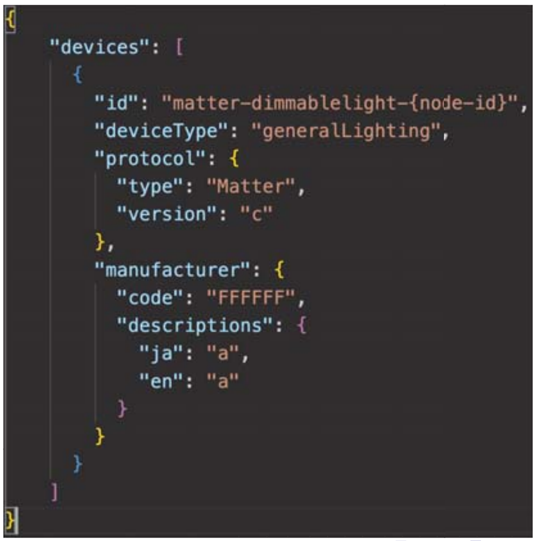
Sample Mapping (Device Registration)

When? An app sends request **HTTP Get** : `x/elapi/v1/devices`

→ JSON response is as in the figure.



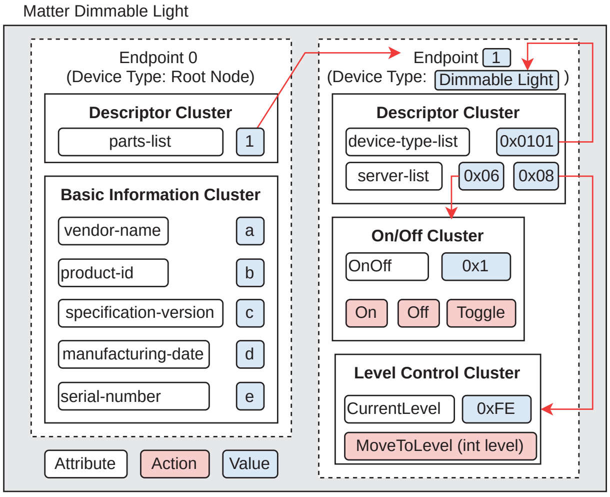
A dimmable light has been commissioned to the *ELWA* fabric



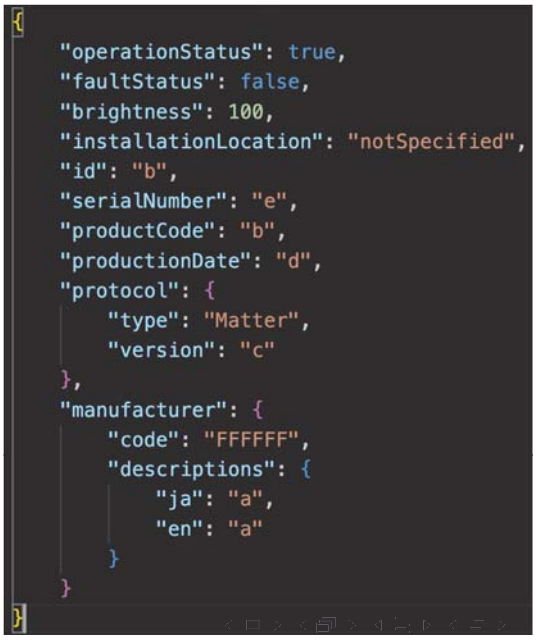
```
{
  "devices": [
    {
      "id": "matter-dimmablelight-{node-id}",
      "deviceType": "generalLighting",
      "protocol": {
        "type": "Matter",
        "version": "c"
      },
      "manufacturer": {
        "code": "FFFFFF",
        "descriptions": {
          "ja": "a",
          "en": "a"
        }
      }
    }
  ]
}
```

Sample Mapping (Device Property Retrieval)

Request **HTTP Get** : `x/elapi/v1/devices/{node-id}/properties`



A dimmable light has been commissioned to the *ELWA* fabric

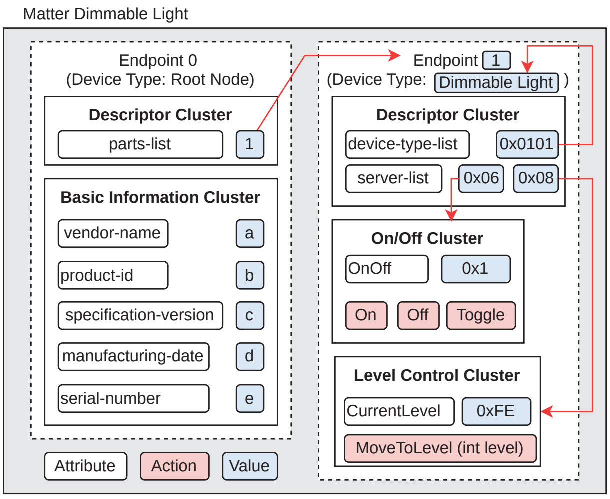


```
{
  "operationStatus": true,
  "faultStatus": false,
  "brightness": 100,
  "installationLocation": "notSpecified",
  "id": "b",
  "serialNumber": "e",
  "productCode": "b",
  "productionDate": "d",
  "protocol": {
      "type": "Matter",
      "version": "c"
  },
  "manufacturer": {
      "code": "FFFFFF",
      "descriptions": {
          "ja": "a",
          "en": "a"
      }
  }
}
```

Sample Mapping (Device Property Update)

Request: **HTTP Put** : `x/elapi/v1/devices/{node-id}/properties/operationStatus`

Request Body : `operationStatus: false`



Matter Dimmable Light

Endpoint 0
(Device Type: Root Node)

Descriptor Cluster
parts-list | 1

Basic Information Cluster
vendor-name | a
product-id | b
specification-version | c
manufacturing-date | d
serial-number | e

Attribute | Action | Value

Endpoint 1
(Device Type: Dimmable Light)

Descriptor Cluster
device-type-list | 0x0101
server-list | 0x06 | 0x08

On/Off Cluster
OnOff | 0x1
On | Off | Toggle

Level Control Cluster
CurrentLevel | 0xFE
MoveToLevel (int level)

Response

`200 OK`

Physical Light

Turned Off

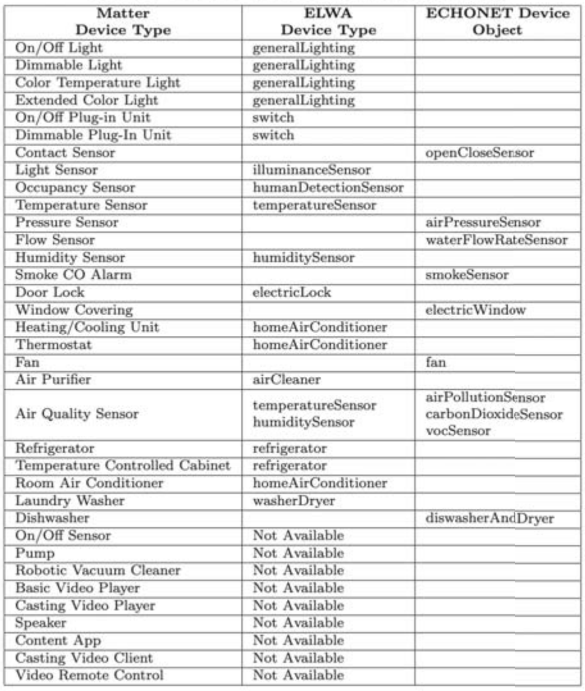
Matter Endpoint

On/Off cluster's `OnOff` attribute value = `0x0`

A dimmable light has been commissioned to the *ELWA* fabric

Matter - ELWA Data Model Mapping

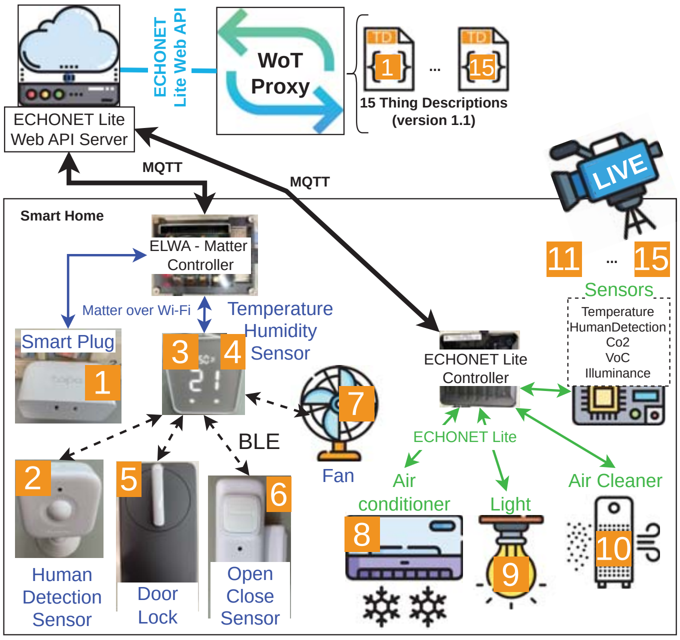
Table 2 Matter -ELWA Data Model Mapping



Matter Device Type	ELWA Device Type	ECHONET Device Object
On/Off Light	generalLighting	
Dimmable Light	generalLighting	
Color Temperature Light	generalLighting	
Extended Color Light	generalLighting	
On/Off Plug-in Unit	switch	
Dimmable Plug-In Unit	switch	
Contact Sensor		openCloseSensor
Light Sensor	illuminanceSensor	
Occupancy Sensor	humanDetectionSensor	
Temperature Sensor	temperatureSensor	
Pressure Sensor		airPressureSensor
Flow Sensor		waterFlowRateSensor
Humidity Sensor	humiditySensor	
Smoke CO Alarm		smokeSensor
Door Lock	electricLock	
Window Covering		electricWindow
Heating/Cooling Unit	homeAirConditioner	
Thermostat	homeAirConditioner	
Fan		fan
Air Purifier	airCleaner	
Air Quality Sensor	temperatureSensor humiditySensor	airPollutionSensor carbonDioxideSensor vocSensor
Refrigerator	refrigerator	
Temperature Controlled Cabinet	refrigerator	
Room Air Conditioner	homeAirConditioner	
Laundry Washer	washerDryer	
Dishwasher		diswasherAndDryer
On/Off Sensor	Not Available	
Pump	Not Available	
Robotic Vacuum Cleaner	Not Available	
Basic Video Player	Not Available	
Casting Video Player	Not Available	
Speaker	Not Available	
Content App	Not Available	
Casting Video Client	Not Available	
Video Remote Control	Not Available	

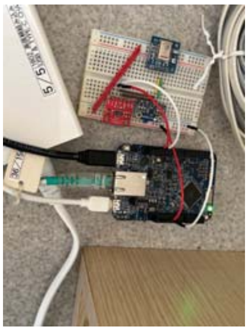
- 1 Matter device can be mapped to 0..* ELWA device type (e.g Air Quality Sensor → temperature, humidity, ... sensors)
- Multiple Matter device types can be mapped to 1 ELWA device type (e.g On/Off, Dimmable, Color Temperature, Extended Color Lights → generalLighting)
- In Matter Device Library (v1.2) 26/35 (74%) devices are mapped to ELWA Device Description (as native ELWA device)

Tested at Plugfest/Testfest 2024



Smart Devices SwitchBot and Tapo (Matter devices)

Testbed (view from outside)


ECHONET Lite Sensors

ECHONET Lite Air Cleaner

ECHONET Lite
Ceiling light and home airconditioner

(1) Matter - ECHONET Lite Web API Controller: Remark

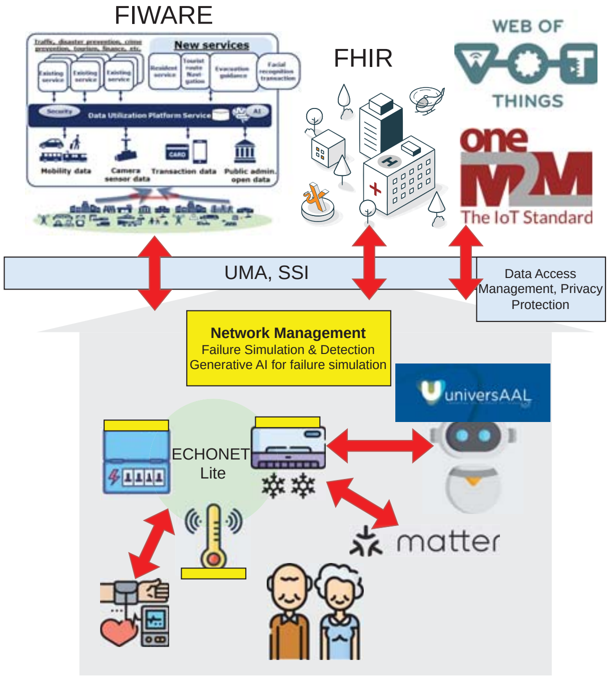
Results

- This work (controller) is the **first** to support ELWA for Matter
- The controller is tested with an experimental Matter device and Commercial Matter devices and has been running stably for months
- This solution outperforms the original `chip-tool` → more stable

Limitations

- There is no ELWA APIs for commissioning → need customized API for such tasks
- Properties (Functionalities) are lost when mapping to ELWA is unavoidable
- Success mapping rate is fair (**74%**) but will be lower when new Matter device types are introduced → new data model for Matter?

Others



Compatibility

Integrated Approach: ECHONET Lite (EL) ↔ universAAL, EL ↔ Matter

Unified Approach: EL ↔ oneM2M , EL↔FIWARE, EL↔WoT

Federated Approach: EL↔ *FHIR*

Smart Home Simulator

Device Failure Simulation

Closing: From Fragmentation to a Harmonized Home

Recap

Health integration, Matter compatibility, Web access enables richer services

Industry collaboration is required to unlock the full potential of smart homes

Vision: a world where homes are safe, healthy, efficient, and truly interoperable.

Call to Action

More use cases for smart homes with AI/ML

I personally like the idea of `Disaster Preparedness in Everyday Life - BOSAI: Science that Can Save Your Life` → `https://youtu.be/RAHzPXhVBHI?si=xBGMko5r4yVlpYED`

Q&A

Thank You!

cupham@jaist.ac.jp

Session 7:
Network and Security Technologies

(Chair: Yoshitaka Nakamura)

An Evaluation and Analysis of a Zero Trust Based IoT Security Framework

Hibiki Oizumi† Nobuhiro Kobayashi‡ Ryozo Kiyohara†

†Kanagawa Institute of Technplogy, Japan
‡ University of Nagasaki, Japan

Abstract - In the era of Society 5.0, where cyberspace and physical space are deeply integrated, ensuring the security of IoT actuators—particularly in mobile robotic systems—has become a critical challenge. As malware grows increasingly stealthy and AI-driven, conventional detection-based security approaches are proving insufficient. This study proposes and evaluates a Zero Trust IoT Security Framework (ZeTiots FW) specifically designed for mobile robot platforms, with implementation on the TurtleBot3. The framework combines Zero Trust principles with physical-layer safety mechanisms utilizing infrared sensors. The objective is to verify the feasibility and effectiveness of this framework in preventing real-world damages before they occur.

Keywords: Zero Trust Architecture, IoT Security, TurtleBot3, AI driven Malware, Cyber Physical Systems

1 INTRODUCTION

Japan's vision of Society 5.0 aims to establish a human-centered society in which cyberspace and physical space are seamlessly integrated [1]. In this context, IoT actuators—such as autonomous vehicles, service robots, and smart home systems—play a vital role in connecting the physical and digital domains. However, their continuous network connectivity exposes them to an expanding range of security threats.

The number of IoT devices continues to grow rapidly, reaching 16.6 billion in 2023 and projected to exceed 18.8 billion by 2024 [2]. According to NICTER observations, many of the most frequently targeted network ports are associated with IoT devices [3]. Moreover, new IoT botnet variants have emerged that exhibit no typical scanning behavior, unlike Mirai [4] or Satori [5]. These trends indicate that traditional detection-based security technologies are becoming increasingly ineffective.

To address these challenges, the Zero Trust IoT Security Framework (ZeTiots FW) has been proposed [6], which applies the core principles of Zero Trust architecture to IoT actuator systems. In this study, ZeTiots FW is implemented on the TurtleBot3 mobile robot [7] to evaluate its effects on operational reliability and security performance in real-world environments.

2 FRAMEWORK OVERVIEW

The Zero Trust IoT Security Framework (ZeTiots FW) applies the principles of Zero Trust architecture to the actuator control layer of IoT devices. As illustrated in Fig. 1, the architecture consists of two logical components:

- Policy Decision Point (PDP) located in the cloud
- Policy Enforcement Point (PEP) implemented on the actuator side [6]

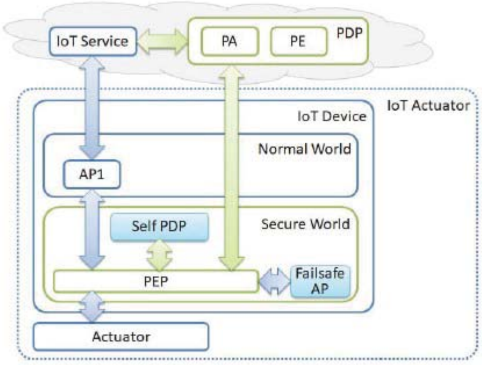


Figure 1: Zero Trust IoT Security Framework
(Source: Kobayashi, IWIN 2022 [6])

To maintain functionality under unstable network conditions or during distributed denial-of-service (DDoS) attacks, a portion of the PDP logic is embedded locally within the actuator as a Self-PDP. This component enables autonomous decision-making when the connection to the central controller is disrupted.

This layered architecture enforces strict verification of all commands prior to execution, thereby minimizing the risk of physical or operational damage caused by compromised nodes. Furthermore, it incorporates a fail-safe mechanism at the physical layer, which prevents hazardous movements even if upper software layers—such as ROS or the operating system—are compromised.

3 IMPLEMENTATION ON TURTLEBOT3

The TurtleBot3 Burger employs a two-layer control architecture: the Raspberry Pi 4 handles high-level processing, while the OpenCR 1.0 board manages motor and sensor control. Key ROS topics include /cmd_vel (velocity commands), /navigate_to_pose (navigation goals), /scan (sensor data), and /odom (odometry). Ultimately, /cmd_vel commands are transmitted to the OpenCR via USB serial communication. Fig. 2 illustrates the autonomous navigation workflow. Navigation goals sent through /navigate_to_pose are processed by the ROS navigation stack (Nav2), which publishes velocity commands to /cmd_vel. The turtlebot3_node then relays these commands to the OpenCR, which executes them as actual motor movements.

In this study, an autonomous proximity-based stop mechanism was implemented at the firmware level of the OpenCR board. The system employs a SHARP GP2Y0A21YK0F infrared distance sensor to continuously

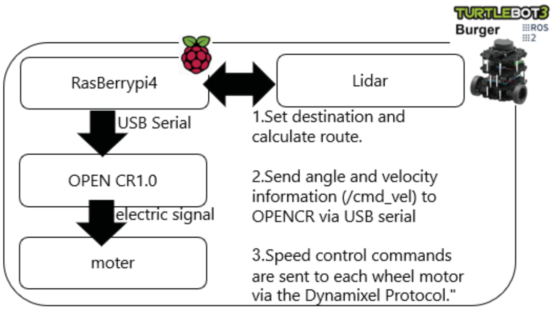

TURTLEBOT3
Burger
ROS 2
RasBerrypi4
Lidar
USB Serial
1.Set destination and calculate route.
OPEN CR1.0
2.Send angle and velocity information (/cmd_vel) to OPENCR via USB serial
electric signal
moter
3.Speed control commands are sent to each wheel motor via the Dynamixel Protocol."

Figure2 System Flow

measure the frontal distance [8]. When an object is detected within a 20 cm safety threshold, the firmware immediately halts the motors by sending stop signals directly to the motor driver—completely bypassing the ROS layer.

Although the OpenCR does not natively include a Trusted Execution Environment (TEE), the firmware was designed to conceptually emulate TEE-like isolation. Specifically:

- Execution is confined to a small, independent memory area (<32 KB).
- Only essential peripherals (motor driver, sensor) are accessible.
- The control logic follows secure-boot and non-bypassable control principles.

This approach ensures that even if the ROS node or Linux system is compromised, the emergency stop logic remains trusted and active.

4 PROPOSED SYSTEM ARCHITECTURE REAL TIME PERFORMANCE AND FUTURE EXTENTIONS

The proposed system also emphasizes real-time operability and control stability. By minimizing processing overhead and isolating the emergency control logic within the firmware, the architecture achieves low latency and continuous monitoring without interfering with normal driving commands (see Fig. 3). Potential future enhancements include:

- Integration of AI-based anomaly detection
- Fusion with additional sensors (e.g., ultrasonic or vision-based sensors) to reduce false positives.
- Migration to hardware platforms equipped with built-in Trusted Execution Environments (TEEs) to strengthen security guarantees.

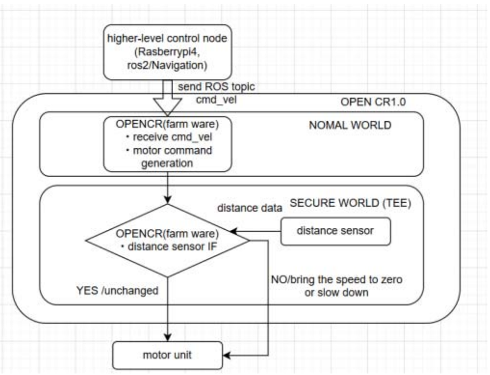

higher-level control node (Rasberrypi4, ros2/Navigation)
send ROS topic cmd_vel
OPEN CR1.0
OPENCR(farm ware) • receive cmd_vel • motor command generation
NOMAL WORLD
distance data
SECURE WORLD (TEE)
distance sensor
OPENCR(farm ware) • distance sensor IF
YES /unchanged
NO/bring the speed to zero or slow down
motor unit

Figure3 The control flow of the proposed approach

Finally, a quantitative comparison of system behavior, response latency, and control performance before and after integrating the proposed mechanism will be essential to fully evaluate its effectiveness and practical feasibility..

5 CONCLUSION AND FUTURE WORK

In this study, we proposed a multi-layered IoT security approach for robotic control systems by integrating the Zero Trust security model with physical-layer sensors (infrared distance sensors). By embedding a mechanism for directly detecting physical risks and autonomously stopping the robot at the firmware layer, we demonstrated the feasibility of safer control strategies for IoT and robotic platforms operating in real-world environments.

Future work includes the following:

- Implementation on the actual TurtleBot3 platform and scenario-based evaluation.
- Improvement of detection accuracy and reduction of false positives or malfunctions through the integration of additional sensor types and multiple safety thresholds.
- Enhancement of operational design and robustness through integration with secure execution environments such as Trusted Execution Environments (TEEs).

Ongoing research into accident-preventive security implementations remains essential for advancing the safety and practical deployment of IoT and cyber-physical systems.

ACKNOWLEDGMENT

This work was partially supported by JSPS KAKENHI Grant Number JP23K16872.

REFERENCES

[1] Cabinet Office, Government of Japan, Society 5.0. https://www8.cao.go.jp/cstp/society5_0/

[2] IoT Analytics, State of IoT – Summer 2024, IoT Analytics, 2024.

[3] National Institute of Information and Communications Technology (NICT), NICTER Observation Report, NICT, Japan, 2024.

[4] J. Gamblin, Mirai Source Code, GitHub Repository, 2016. https://github.com/jgamblin/Mirai-Source-Code

[5] S. Yamaguchi and B. B. Gupta, “Malware threat in Internet of Things and its mitigation analysis,” in Research Anthology on Combating Denial of Service Attacks, IGI Global, pp. 371–387, 2021.

[6] N. Kobayashi, “A study on application of zero trust architecture to IoT actuator,” in Proc. Int. Workshop on Informatics (IWIN 2022), pp. 191–197, 2022.

[7] ROBOTIS Inc., TurtleBot3 Manual Overview, ROBOTIS Inc., 2023. https://emanual.robotis.com/docs/en/platform/turtlebot3/overview/

[8] Sharp Corporation, GP2Y0A21YK0F Analog Distance Sensor Datasheet, Sharp Corporation, Japan, 2020. [https://global.sharp/products/device/lineup/selection/opto/haca/

A Proposal for LDoS Attack Method Using MPTCP Signal Manipulation

Hiromichi Hagiwara[†], Hiroshi Inamura[†], Shigemi Ishida[†]

[†]Graduate School / School of Systems Information Science, Future University Hakodate, Japan
{g2123048, inamura, ish}@fun.ac.jp

Abstract - This study proposes a Low-rate Denial of Service (LDoS) attack method using Multipath TCP (MPTCP). LDoS attacks exploit TCP's retransmission timeout characteristics to degrade throughput with minimal bandwidth. Our method leverages MPTCP's multi-path capabilities, combining Optimistic ACKing with DSS manipulation. We developed two strategies: a "Distributed Attack Strategy" that distributes attack rates across multiple subflows, and a "Cycle Attack Strategy" that modifies attack timing patterns. These strategies enable effective attacks from a single node while avoiding the complexity of traditional distributed attacks. Future work will focus on evaluation in real-world environments and development of defense mechanisms.

Keywords: Low-rate DoS Attack, Multipath TCP, Network Security

1 Introduction

Low-rate Denial of Service (LDoS) attacks that degrade communication Quality of Service (QoS) using intermittent burst traffic are being discussed as threats on the Internet [1]. These attacks exploit the characteristics of TCP's Retransmission Time Out (RTO) to maintain TCP flows in a continuous congestion avoidance state, thereby reducing throughput. While conventional DoS attacks aim to saturate 100% of available bandwidth, LDoS attacks typically use only 10-30% of the attack period, making them difficult to detect due to their low average attack traffic volume [2].

On the Internet, Low-rate Distributed DoS (LDDoS) attacks that execute LDoS attacks distributed across multiple nodes have also been observed [1]. In this method, traffic sent from multiple attack nodes is aggregated at the target's bottleneck link to form an LDoS attack waveform. LDDoS attacks have the characteristic that detection becomes difficult because the communication volume of individual nodes is suppressed, and the attack effect is enhanced by aggregating traffic from each node at the target.

However, executing LDDoS attacks requires securing multiple nodes and coordinated control, presenting implementation complexity as a challenge. To address this challenge, this research focuses on Multipath TCP (MPTCP).

MPTCP is an extension of TCP that enables communication over multiple paths, and recent research has revealed that DoS attacks using MPTCP are feasible [3]. Since MPTCP has the characteristic that a single client can communicate simultaneously over multiple links, it is conceivable that effects similar to LDDoS attacks could be generated from a single node. However, the feasibility of LDoS attack methods using MPTCP has not yet been clarified.

Although MPTCP adoption is currently limited, this research holds the following academic significance: (1) novel insights into multipath protocol vulnerabilities, (2) implications for future protocol design, (3) foundational research for defense mechanism development.

This research proposes an MPTCP-based LDoS attack method. Furthermore, we propose two attack strategies called Distributed Strategy and Cycle Strategy that utilize multi-path characteristics. We demonstrate that the proposed method can potentially achieve effects equivalent to conventional LDDoS attacks from a single node.

2 Constructing LDoS Attack

This chapter explains the technologies that form the foundation of the proposed method. First, we describe TCP's retransmission timeout mechanism, then explain the mechanism and characteristics of LDoS attacks using this, and finally outline LDDoS attacks executed by multiple nodes.

2.1 TCP Retransmission Timeout

TCP's Retransmission Time Out (RTO) is an important mechanism that determines the waiting time before starting data retransmission when packet loss occurs. RTO is calculated using the following formula (1) [4].

$$\begin{aligned}\mathrm{RTO} = \max\{&\mathrm{minRTO}, (\mathrm{Smoothed\ RTT}) \\ &+ \max[(\mathrm{Clock\ Granularity}), 4(\mathrm{RTT\ variation})]\}\end{aligned} \tag{1}$$

Smoothed RTT represents the smoothed round-trip time calculated from RTT measurements, Clock Granularity represents the operating system-dependent clock granularity, and RTT variation represents the mean deviation of RTT. It is recommended that minRTO be set to 1 second [4]. Generally, since RTT is small compared to RTO, minRTO (=1 second) is selected as the initial RTO.

When retransmission is not acknowledged, the new RTO value increases by exponential backoff according to the following formula (2).

$$\mathrm{RTO}_i = \mathrm{RTO}_{i-1} \times 2 \tag{2}$$

When the same packet is determined to be lost, it is specified that RTO should be doubled and retransmitted according to this formula [4]. However, it is shown that the maximum RTO value should be 60 seconds or more. When a packet is transmitted normally, RTO is reset to the initial value minRTO. This is called Karn's algorithm and is implemented as the retransmission control algorithm in most TCP implementations [5]. This predictability of RTO makes LDoS attacks easier to execute.

2.2 Low-rate DoS Attack

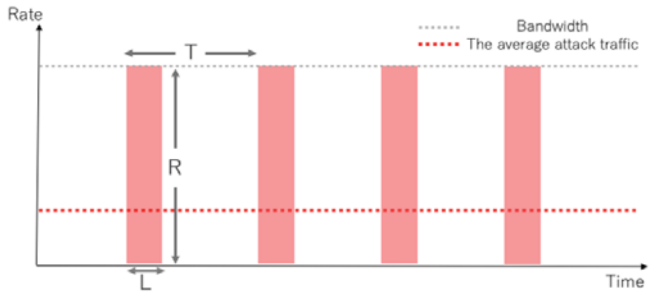


Figure 1: Model of LDoS attack

LDoS attacks are efficient DoS attack methods that exploit vulnerabilities in TCP's congestion control mechanisms [6]. These attacks use burst traffic at specific intervals matched to TCP's RTO to cause congestion and keep TCP connections in a continuous congestion avoidance state [7].

As shown in Figure 1, the LDoS attack model is characterized by three main parameters:

- R (attack rate): Traffic transmission speed during bursts
- T (attack period): Time interval between attack bursts (RTO)
- L (burst length): Duration of each attack burst

LDoS attacks are most effective when the following conditions are met [6]:

- R >bottleneck link bandwidth
- T = length of RTO
- L <time sufficient to fill buffer

The average attack traffic volume (R_{ave}) of LDoS attacks is expressed as $R_{ave} = R \times L/T$. Since effective implementation is possible when the attack burst length (L) is about 10-30% of the attack period (T), effective attacks are possible with low average attack traffic, making detection difficult [7].

2.3 Low-rate DDoS Attack

LDDoS attacks are methods that distribute LDoS attacks across multiple nodes. Multiple botnet nodes are coordinated by a Command & Control (C&C) server, and timing-adjusted instructions are sent to each node so that burst traffic from each node is aggregated at the bottleneck link to form LDoS attack waveforms.

LDDoS attacks have two main advantages. The first is enhancement of attack effects, where attack traffic generated from multiple nodes accumulates at the target's bottleneck, enabling total attack rates that exceed the resource constraints (bandwidth, processing power) of a single node. This cumulative effect enables attack intensities that are difficult to achieve with a single node. The second is improvement in attack detection avoidance, where distributing attack traffic across multiple nodes reduces the communication volume of individual nodes, enabling effective attacks overall while maintaining low attack communication volume from each node.

The control methods for LDDoS attacks to achieve such cumulative effects are broadly classified into two types [2]. One is synchronous attacks, where each node cooperatively sends periodic burst traffic. The other is asynchronous attacks, where different instructions are sent to each node and they operate in non-periodic patterns, but are designed so that attack pulses converge periodically at the target side.

In recent years, diversification of such attack methods has progressed, with proposals for composite attack models combining increased attack frequency, extended attack pulse duration, enhanced attack rates, and combinations of these [2]. Due to such characteristics and diversification, detection and defense of LDDoS attacks have become more difficult, making them an important issue in network security.

However, executing LDDoS attacks to achieve these attack effects requires coordinated control of multiple nodes, involving the following attack costs: (1) costs for securing and managing multiple bot nodes, (2) costs for constructing and operating control infrastructure for time synchronization and command distribution between nodes, (3) complexity of coordinated control of multiple geographically distributed nodes. These costs make the method a high execution barrier for attackers.

As described above, LDDoS attacks require securing and synchronizing multiple nodes, making implementation with a single node difficult. This point provides the motivation for new attack methods using MPTCP.

3 Multipath TCP

3.1 Architecture

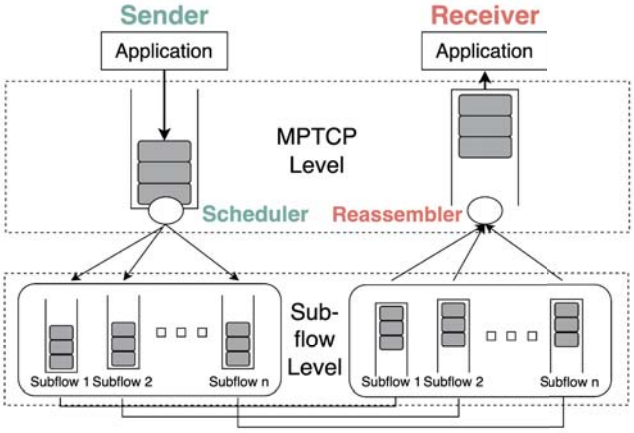


Figure 2: Overview of MPTCP architecture

Multipath TCP (MPTCP) is an extension protocol of TCP that enables the use of multiple physical paths in a single logical connection [8]. MPTCP provides terminals with the ability to communicate simultaneously using multiple network interfaces (e.g., Wi-Fi and LTE), achieving improved throughput and connection reliability [9]. The MPTCP architecture provides a standard TCP socket interface to the application layer and sends TCP-compatible segments to the lower network layer. Each subflow functions as an independent TCP connection and includes MPTCP options in the TCP option field.

3.2 Sequence Space Management

MPTCP operates in two different sequence spaces [10]:

- Connection level: Maintains global sequence numbers (Data Sequence Numbers, DSN) across all subflows
- Subflow level: Each subflow uses conventional TCP sequence numbers and acknowledgment mechanisms

As shown in Figure 2, MPTCP manages communication consistency across multiple paths using two sequence spaces: **MPTCP level** and **subflow level**. The MPTCP level maintains global Data Sequence Numbers (DSN) for all application data, while the subflow level uses independent TCP sequence numbers for each path.

On the sender side, the Scheduler performs connection-level scheduling of application data and assigns data to multiple subflows. Each data is assigned a DSN, and Data Sequence Signal (DSS) options provide mapping with subflow sequence numbers.

On the receiver side, the Reassembler reassembles data arriving from each subflow using DSS mapping information. Data arriving from different subflows is sorted into the correct order based on DSN and provides a consistent data stream to the application.

This mechanism is realized through dual acknowledgment of conventional TCP acknowledgments at the subflow level and DATA acknowledgments at the connection level. Data is considered properly received only when acknowledged at both levels, ensuring data consistency across all communication paths.

3.3 Data Consistency Maintenance

MPTCP implements special mechanisms to ensure data consistency between subflows. The receiving side uses DSS mapping information to reconstruct data arriving from different subflows in the correct order. Data is considered properly received only when acknowledged at both levels (connection and subflow):

- Subflow level acknowledgment:
  Uses conventional TCP ACK fields
- Connection level acknowledgment:
  Uses DSS options with DATA ACK fields

This dual-level acknowledgment mechanism ensures data consistency across all communication paths. However, using DSS manipulation, malicious receivers can bypass this consistency management mechanism and illegally manipulate sequence spaces.

4 Related Work

4.1 TCP Optimistic ACKing

Optimistic ACKing is a technique where receivers send fake ACKs for segments they have not yet received [11]. This technique allows receivers to artificially increase the sender's data transmission volume by returning ACKs for unconfirmed data, bypassing congestion control to achieve high-speed transmission.

The operating principle of Optimistic ACKing is as follows. Upon receiving the first segment, the receiver returns both the normal ACK for the first segment and ACKs for subsequent unconfirmed segments. The number of ACKs for unconfirmed segments can be increased as needed. Since TCP congestion control is determined by ACK arrival speed, when ACKs arrive quickly, the congestion window size increases and the transmission rate increases.

TCP congestion control is receiver-driven, and data transmission is controlled by ACKs generated by the receiver, so malicious receivers can dominate senders by manipulating ACKs. Optimistic ACKing invalidates the sender's congestion window size, making it possible to extract unlimited traffic from TCP.

Furthermore, methods for implementing LDoS attacks using TCP have been established [12]. The LDoS attack applying TCP Optimistic ACKing called Induced-shrew is a method where TCP receivers (attackers) remotely control the transmission rate and pattern of TCP senders, exploiting them as attack traffic generation sources for LDoS attacks [12]. This attack is realized by malicious receivers sending preemptive ACKs for data that senders are expected to send, in contrast to normal receivers that send ACKs only after receiving data.

Unlike conventional UDP-based LDoS attacks, TCP is not a unidirectional communication protocol, so direct attacks from the sender side are impossible. In TCP-based LDoS attacks, TCP receivers become attackers and induce attack traffic from legitimate TCP senders.

Implementation of LDoS attacks using Optimistic ACKing is performed in the following steps:

1. The attacker establishes a legitimate connection with the TCP sender
2. After receiving the first data segment, the attacker predicts sequence numbers of subsequent data not yet received based on the sender's maximum segment size (SMSS)
3. The attacker generates forged ACKs with predicted sequence numbers and sends them to the sender
4. The sender increases the congestion window in response to each ACK and sends new data
5. The attacker can induce LDoS attack pattern traffic from the sender by sending ACKs at specific timing patterns

LDoS attacks are expected to be composed of UDP, and composition with UDP is adopted as one of the detection indicators [13]. Therefore, composing LDoS attacks with TCP may avoid protocol-based detection and potentially make detection more difficult.

4.2 DSS Manipulation Vulnerability

In the MPTCP protocol, it has been shown that receivers can manipulate acknowledgments at both subflow level and connection level through a vulnerability in the connection establishment process [3]. To exploit this vulnerability, attackers (malicious receivers) need to obtain Initial Data Sequence Numbers (IDSN) and Initial TCP Sequence Numbers (ISN).

The acquisition of IDSN and ISN by attackers is realized through the following procedure. First, attackers participate in legitimate MPTCP connection establishment processes with senders. During MPTCP connection establishment, connection-level IDSNs are exchanged via MP_CAPABLE options in SYN packets during the 3-way handshake. Similarly, during each subflow establishment, subflow-level ISNs are exchanged via MP_JOIN options in SYN packets. Since this information is communicated as part of normal connection establishment procedures, attackers can obtain IDSN and ISN simply by acting as legitimate receivers without requiring special privileges or complex techniques.

Additionally, senders notify receivers of the Maximum Segment Size (SMSS) for each subflow. By combining this information (IDSN, ISN, SMSS), attackers can accurately predict sequence numbers of subsequent data segments that senders are expected to transmit and generate valid acknowledgment segments.

To execute attacks using DSS manipulation, attackers need to generate valid MPTCP acknowledgment segments. To accomplish this, the following condition (3) must be satisfied:

$$\begin{aligned}&\text{MPTCP.DATA\_SEQ\_SND\_UNA}\\&\quad < \text{MPTCP.DATA\_ACK} \leq \text{MPTCP.DATA\_SEQ\_SND\_NXT}\end{aligned} \tag{3}$$

When an MPTCP connection is established, the receiver (attacker) grasps the connection-level IDSN and ISNs of all subflows. Also, the sender notifies the receiver of the Sender Maximum Segment Size (SMSS) for each subflow. From this information, it is possible to predict the subflow-level TCP sequence numbers and connection-level DSNs of subsequent data segments that the sender is expected to send.

Furthermore, MPTCP receivers can control the sender's transmission rate. The condition necessary for the sender to perform flow control is expressed by the following formula (4):

$$\begin{aligned}&\text{MPTCP.DATA\_SEQ\_SND\_NXT} - \text{MPTCP.DATA\_ACK}\\&\qquad < 2^{wso} \times \text{RCV\_WND}\end{aligned} \tag{4}$$

Here, $2^{wso} \times$ RCV_WND represents the receiver's available receive buffer, where wso is the window scale option notified during subflow establishment, and RCV_WND is the value of the window field in the TCP fixed header. According to RFC specifications, wso can be set up to a maximum of 14, and RCV_WND can be set up to a maximum of 65535 ($2^{16} - 1$) [14]. Attackers can use this functionality to disguise very large receive buffers (about 1GB) and induce senders to send data without restrictions.

By combining this vulnerability with Optimistic ACKing, attackers can potentially confuse TCP congestion control mechanisms and induce DoS attack traffic. Particularly when combined with MPTCP's multiple subflow functionality, this vulnerability may enable efficient attacks across multiple paths from a single node.

While prior research [3] has shown the DSS manipulation vulnerability mechanism itself, no research has developed this into LDoS attack methods. This research addresses this challenge and demonstrates the feasibility of LDoS attacks in MPTCP combining DSS manipulation and Optimistic ACKing, while proposing two attack strategies that utilize multi-path characteristics.

5 Proposed Attack Method

5.1 Attack Overview

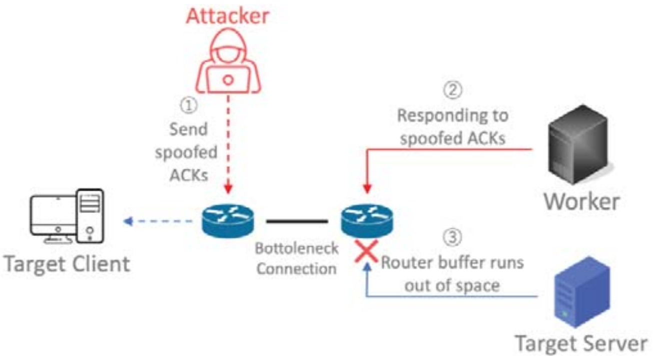


Figure 3: Overview of MPTCP-based LDoS attack

The proposed method realizes LDoS attacks using MPTCP traffic instead of commonly used UDP to enhance stealth characteristics. Furthermore, it establishes multiple subflows from a single MPTCP client to achieve effects equivalent to conventional LDDoS attacks. Figure 3 shows an overview of the proposed method.

In this attack scenario, a Worker Node refers to a legitimate MPTCP server or service that the attacker exploits. Specifically, it is a node capable of transmitting large amounts of data using the MPTCP protocol, such as file transfer services, web servers, or streaming services. The attacker establishes connections to this Worker Node as a legitimate client and sends forged ACKs using DSS manipulation and Optimistic ACKing to intentionally induce large amounts of attack traffic from the Worker Node.

An important characteristic of this attack method is its low implementation cost compared to conventional distributed attack methods. Attackers do not need to directly control Worker Nodes and can execute attacks by accessing them as legitimate users. The prerequisites for attack execution are as follows: (1) attackers can access Worker Nodes as general users, (2) Worker Nodes provide transfer functions using the MPTCP protocol, (3) attackers can send traffic generation requests such as data downloads. This allows effective attacks to be executed while behaving as normal users, avoiding the complex infrastructure construction required by conventional distributed attacks, such as securing and managing multiple bot nodes.

This method constructs attacks using forged ACKs with TCP-level Optimistic ACKing and MPTCP-level DSS manipulation. LDoS attacks are performed by congesting bottleneck queues shared on paths with target traffic using attack traffic.

The transmission timing of these forged ACKs is controlled based on the LDoS attack model, creating 1-second periodic On/Off traffic patterns.

The detailed attack mechanism is explained below. The red lines in Figure 3 represent attack traffic, while the blue lines represent target traffic. An MPTCP connection is established between the Attacker and Worker, and the attacker sends forged ACKs using DSS manipulation and Optimistic ACKing in this connection to induce large amounts of pulsed attack traffic from the Worker. Meanwhile, normal TCP communication is conducted between the Target Client and Target Server, which becomes the target connection of the attack. The pulsed attack traffic controlled by the attacker periodically fills the buffer of the bottleneck link (Router part in Figure 3), causing packets of the target TCP connection to be dropped. As a result, the target connection repeatedly experiences continuous timeouts and retransmissions, significantly reducing throughput.

This attack method represents a new threat that extends conventional TCP LDoS attacks by utilizing MPTCP's multiple subflow characteristics, enabling efficient attacks with low traffic volume. Furthermore, the ability to use multiple paths from a single node reduces attack costs compared to conventional LDDoS attacks.

5.2 Technical Challenges

To realize the proposed method, the following technical challenges must be addressed:

5.2.1 Transmission Timing Control Based on LDoS Attack Model

Precise transmission timing control based on LDoS attack parameters (R, T, L) is necessary. To create 1-second periodic On/Off traffic patterns, transmission intervals must be calculated.

Attack cycles consist of alternating repetitions of attack traffic On (forged ACK transmission, sequence number tracking) and attack traffic Off (sleep state). It is necessary to precisely control 1-second periodic On/Off traffic patterns and generate effective burst traffic timed to RTO.

This timing control enables efficient attack effects with low average attack communication volume.

5.2.2 Consistency Maintenance Between Subflow and MPTCP Connection Levels

For successful attacks, consistency maintenance based on accurate knowledge of Initial Data Sequence Numbers (IDSN) and Initial TCP Sequence Numbers (ISN) becomes a challenge. This challenge has already been solved in prior research on DSS manipulation described in Section 4.2, and this research applies this existing technology.

5.3 Attack Strategies

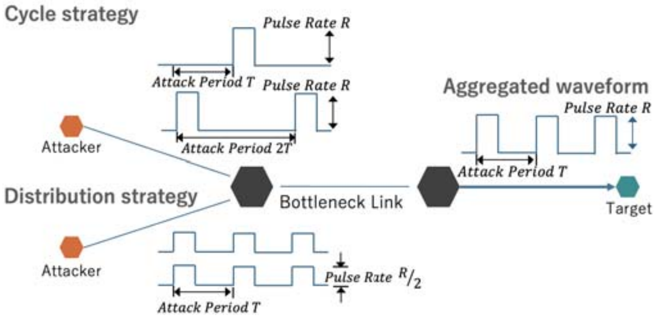


Figure 4: Overview of the Attack Strategy

In addition to basic MPTCP-based LDoS attacks, we propose two attack strategies (Distributed Strategy and Cycle Strategy) that utilize MPTCP's multiple subflow characteristics. As shown in Figure 4, these strategies aim to enhance attack efficiency while improving detection avoidance capabilities. Details of each strategy are explained below:

5.3.1 Distribution Strategy

Distributed Strategy has the following two advantages:

- **Detection avoidance effect (R/n distribution)**:
  By distributing attack traffic across n subflows, the attack rate of each flow is reduced to R/n. This causes traffic from each subflow to fall below detection thresholds, making detection difficult for flow-based anomaly detection systems.
- **Enhanced attack intensity (n×r accumulation)**:
  Attack traffic from multiple subflows is aggregated at the bottleneck link, achieving high overall attack rates. Since attack rates that can be generated by a single node are limited, using n subflows can theoretically achieve attack intensity equivalent to n×r.

This method enables attackers to execute effective distributed attacks from a single node without physically securing and coordinating multiple nodes. Also, since each subflow communicates at low rates, distinguishing from normal traffic becomes potentially difficult when monitoring individual flows.

5.3.2 Cycle Strategy

Cycle Strategy is a method that makes temporal pattern recognition difficult by shifting traffic patterns of each subflow. Phase differences between subflows are set to T/n (where T is the attack period) to maintain overall continuous attack effects. For example, with n=2 and T=1 second, one subflow sends burst traffic from 0.0 to 0.3 seconds, while the other sends alternating burst traffic from 1.0 to 1.3 seconds, enabling temporal detection pattern avoidance while maintaining attack effects when viewing attack periods of each node.

Furthermore, as described in Section 2.3 regarding LDDoS attack control methods, asynchronous control of traffic patterns for each node is also conceivable. It is possible to send different instructions to each node and design them so that individual nodes operate in non-periodic patterns while attack effects converge systematically at the target side. Such dynamic changes in traffic patterns can hide periodic characteristics of attack traffic, further improving detection system avoidance capabilities while maintaining attack efficiency.

These strategies can realize attacks with low average attack traffic volume similar to conventional LDoS attacks, and by aggregating attack traffic from multiple links to form LDoS attack waveforms, they make detection more difficult similar to LDDoS attacks.

5.4 Attack Design

This section explains implementation details of the proposed attack method. We focus particularly on forged ACK generation mechanisms and timing control mechanisms.

5.4.1 Forged ACK Generation

Forged ACK generation is a method that manipulates MPTCP's dual sequence spaces. To realize this dual manipulation (subflow level and connection level), implementation follows these steps:

1. **Connection establishment and initial information collection**:
   During establishment of each subflow, collect connection-level Initial Data Sequence Numbers (IDSN) and subflow-level Initial TCP Sequence Numbers (ISN). Also record Maximum Segment Size (SMSS) notified by the sender.
2. **Initialization of legitimate acknowledgments**:
   During initial legitimate data reception, initialize sequence numbers at two levels:
   - Subflow level: ACK_i values for each subflow (TCP ACK field)
   - Connection level: $DACK$ values for entire MPTCP (DSS DATA_ACK field)
3. **Dual sequence space manipulation**:
   During forged ACK packet generation, manipulate sequence numbers at both levels:
   - Subflow level (Optimistic ACKing):
     Increase ACK_i values to forge acknowledgments for TCP segments not yet received
   - Connection level (DSS Manipulation):
     Increase $DACK$ values to forge acknowledgments for MPTCP data not yet processed

   Increase values at both levels by $skip\_ack\_count \times SMSS_i$.
4. **DSS manipulation packet construction**:
   Construct and send ACK packets containing manipulated sequence numbers at both levels. This includes both normal TCP ACK fields and MPTCP DSS DATA_ACK fields.

This dual-level manipulation causes subflow-level and connection-level manipulations to bypass MPTCP's data flow control. By generating forged ACKs with consistency at both levels, MPTCP protocol stacks can be prevented from detecting manipulations.

5.4.2 Timing Control Mechanism

To maximize LDoS attack effects, precise control of forged ACK transmission timing is necessary. Timing control is implemented as follows:

1. **Period parameter setting**:
   Set attack period T (1.0 seconds) and burst length L (0.3 seconds). Burst length L has been adopted in related research, and 0.3 seconds is considered appropriate as a practical range that is sufficiently longer than RTT to enable timeout induction while avoiding detection, so it was adopted [7].
2. **Attack rate calculation based on strategy**:
   Based on the selected strategy, determine attack rates and phases for each subflow:

   Distribution Strategy:
   Two approaches are conceivable depending on implementation methods. When emphasizing detection avoidance, distribute the total attack rate R across n subflows, reducing each flow's rate to R/n. When emphasizing attack intensity, each subflow can maintain target rate R, achieving overall attack intensity of $n \times R$.

   Cycle Strategy:
   Apply phase differences of $i \times (T/n)$ to each subflow i to achieve continuous effects by shifting attack cycles. This research uses periodic phase differences, which can further improve detection avoidance capabilities. For example, randomizing attack timing of each subflow can make detection by temporal pattern analysis more difficult.
3. **On/Off pattern control**:
   Divide attack cycles into On periods (burst transmission) and Off periods (rest):
   - On period $[0, L)$ seconds: Continuously send forged ACKs at calculated transmission intervals
   - Off period $[L, T)$ seconds: Stop transmission and wait until the next On period

This timing control mechanism enables effective LDoS attacks with low average traffic volume.

6 Conclusion

This paper proposed an LDoS attack method using MPTCP.

The proposed method utilizes MPTCP's multiple subflow functionality to establish multiple logical subflows on the same path and constructs attacks using forged ACKs that combine DSS manipulation and Optimistic ACKing. This method can potentially achieve effective attacks with low average attack traffic volume.

An important characteristic of the proposed method is that it can realize the advantages of conventional LDDoS attacks from a single node. LDDoS attacks have the advantages of enhanced attack effects through accumulation of attack traffic from multiple nodes at bottlenecks and improved detection avoidance through distribution of attack traffic. The proposed method is expected to achieve equivalent advantages to LDDoS attacks from a single node by aggregating attack traffic from multiple subflows at bottlenecks through MPTCP's multiple subflow functionality and realizing detection avoidance effects through attack rate distribution across individual subflows.

Furthermore, this paper proposed two attack strategies, Distributed Strategy and Cycle Strategy, that utilize MPTCP's multiple subflow characteristics. These strategies are expected to improve detection avoidance capabilities through attack method diversification.

This work suggests the possibility that distributed attack effects similar to LDDoS attacks can be realized while avoiding implementation challenges required by conventional LDDoS attacks, such as securing and managing multiple bot nodes, constructing control infrastructure for time synchronization and command distribution between nodes, and coordinated control of geographically distributed nodes.

A limitation of this research is that MPTCP adoption is currently limited. However, with the expansion of adoption in mobile devices and cloud environments, it may become an important security threat in the future.

Future challenges include verification in real environments and development of defense methods.

REFERENCES

[1] M. Delio, “New breed of attack zombies lurk,” *Wired*, 2001, accessed: 2025-05-15.

[2] Zhijun *et al.*, “Low-rate dos attacks, detection, defense, and challenges: A survey,” *IEEE Access*, vol. 8, pp. 43 920–43 943, 2020.

[3] Kumar *et al.*, “Data sequence signal manipulation in multipath tcp: The vulnerability, attack and its detection,” *Computers & Security*, vol. 103, p. 102180, 01 2021.

[4] M. Sargent, J. Chu, D. V. Paxson *et al.*, “Computing TCP's Retransmission Timer,” RFC 6298, Jun. 2011. [Online]. Available: https://www.rfc-editor.org/info/rfc6298

[5] P. Karn and C. Partridge, “Improving round-trip time estimates in reliable transport protocols,” *ACM Trans. Comput. Syst.*, vol. 9, no. 4, p. 364–373, Nov. 1991. [Online]. Available: https://doi.org/10.1145/118544.118549

[6] A. Kuzmanovic and E. W. Knightly, “Low-rate tcp-targeted denial of service attacks: the shrew vs. the mice and elephants,” in *Proceedings of the 2003 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications*, ser. SIGCOMM '03. New York, NY, USA: Association for Computing Machinery, 2003, p. 75–86. [Online]. Available: https://doi.org/10.1145/863955.863966

[7] Y. Takahashi, H. Inamura, and Y. Nakamura, “A low-rate ddos strategy for unknown bottleneck link characteristics,” in *19th IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events, PerCom Workshops 2021, Kassel, Germany, March 22-26, 2021*. IEEE, 2021, pp. 508–513. [Online]. Available: https://doi.org/10.1109/PerComWorkshops51409.2021.9430992

[8] A. Ford, C. Raiciu, M. J. Handley *et al.*, “TCP Extensions for Multipath Operation with Multiple Addresses,” RFC 6824, Jan. 2013. [Online]. Available: https://www.rfc-editor.org/info/rfc6824

[9] Y.-C. Chen, Y.-s. Lim, R. J. Gibbens *et al.*, “A measurement-based study of multipath tcp performance over wireless networks,” in *Proceedings of the 2013 Conference on Internet Measurement Conference*, ser. IMC '13. New York, NY, USA: Association for Computing Machinery, 2013, p. 455–468. [Online]. Available: https://doi.org/10.1145/2504730.2504751

[10] A. Ford, C. Raiciu, M. Handley *et al.*, “Tcp extensions for multipath operation with multiple addresses,” in *RFC 6824, Internet Engineering Task Force (IETF)*, Jan. 2013. [Online]. Available: https://tools.ietf.org/html/rfc6824

[11] S. Savage, N. Cardwell, D. Wetherall *et al.*, “Tcp congestion control with a misbehaving receiver,” *ACM SIGCOMM Computer Communication Review*, vol. 29, no. 5, pp. 71–78, 1999.

[12] Kumar *et al.*, “On remote exploitation of tcp sender for low-rate flooding denial-of-service attack,” *IEEE Communications Letters*, vol. 13, no. 1, pp. 46–48, 2009.

[13] Tang *et al.*, “Low-rate DoS attack detection based on two-step cluster analysis and UTR analysis,” *Human-centric Computing and Information Sciences*, vol. 10, no. 1, p. 6, Feb. 2020. [Online]. Available: https://doi.org/10.1186/s13673-020-0210-9

[14] D. A. Borman, R. T. Braden, and V. Jacobson, “TCP Extensions for High Performance,” RFC 1323, May 1992. [Online]. Available: https://www.rfc-editor.org/info/rfc1323

Survey and Challenge on ICN for Industrial IoT Services

Atsuko Yokotani*/**/***, Yuuki Hatanaka**, Koichi Ishibashi**, Tetsuya Yokotani**, Satoshi Ohzahata***

* The Telecommunication Technology Committee, Japan
yokotani@s.ttc.or.jp

**Kanazawa Institute of Technology
{"c7400034@st, k_ishibashi@neptune, yokotani@neptune}.kanazawa-it.ac.jp

***University of Electro-Communications
ohzahata@is.uec.ac.jp

Abstract -
Information Centric Network (ICN) transfers information with independence of IP addresses. It provides name-base transfer on information. Therefore, it invokes the simplified mechanisms without address translation by Domain Name System. It also contributes to reduction of traffic volume by the networked temporary storage for mitigation of duplicate information transfer. Recently, ICN seems to be applied to IoT services which require information with tiny size at a number of devices, because of these simplified mechanisms. Particularly, ICN seems to be a promising method in control communication for industrial fields, e.g., factory, industry plant. Some researches including authors' activities for this purpose have been activated. This paper indicates features of ICN and discusses possibility that it can be applied to next generation network infrastructure. Then, it surveys researches for control communication using ICN and identifies future issues.

Keywords: ICN, IoT services, CCN, IoT DEP

1 INTRODUCTION

Information-centric networks are generally referred to as Information Centric Networks (ICN). ICNs can be classified into several categories [1]. Common to all of them is that they use the information itself as the address rather than network-dependent addresses when transferring information. In this paper, we focus on the representative Content Centric Network (CCN) method [2] among the classifications in [1] and define it as ICN.

ICN was originally designed for hop-by-hop transmission of large-capacity content, but due to its simple communication method, its application to Internet of Things (IoT) services that transmit small amounts of information from multiple locations is being explored. This paper summarizes research and technical trends related to the application of ICN to IoT services. Particular focus is placed on its application to control communication in industrial settings such as manufacturing and industrial plants.

2 OVERVIEW OF ICN

According to [2], Figure 1 shows the structure and communication mechanism of ICN, which operates with two messages: "Interest" for information request and "Data" for information transfer. This mechanism is performed on a hop-by-hop basis at each Interworking Point (IWP). Information stored in Cache in the IWP is transferred as Data when an Interest is received. In this mechanism, the duplicate transfer is mitigated. This corresponds to the case where User #2 obtains the same information as User #1 in Figure 1. This eliminates the need for information transfer from the Server.

Next, the processing of information in the IWP is shown in Figure 2.

The IWP consists of three facilities as follows; Pending Information Table (PIT), Forwarding Information Base (FIB), and Content Store (CS). Figure 2(a) shows the operation when an Interest message is received, and (b) shows the operation when a Data message is received.

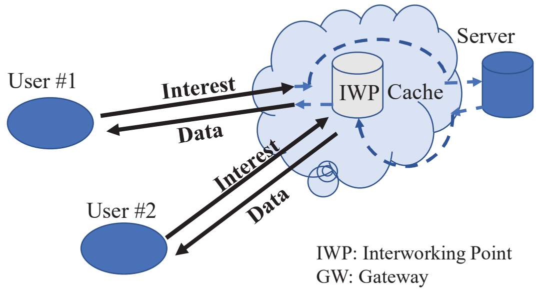
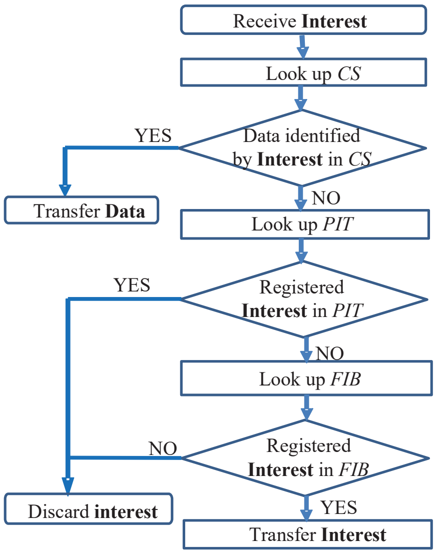
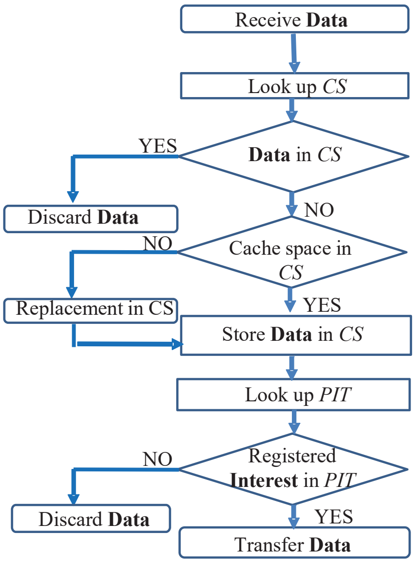


Figure 1: System configuration in ICN



(a) Processing sequences of **Interest** message



(b) Processing sequences of **Data** message

Figure 2: Processing flows in an IWP

3

3 REQUIREMENTS OF INDUSTRY IOT SERVICES

In this section, requirements on communication for IoT industry services are surveyed, because possibility of these services applying ICN are discussed. A number of literatures has been published. For example, the performance requirements for factories, industrial plants, smart grids, and automated operations are described in [3], and a summary of essential metrics is given in Table 1. The performance requirements for each specific operation in a factory are described in [4] and summarized in Table 2.

Tables 1 and 2 show that control communication requires low latency and reliable information transfer. In addition, they also show that communication is generally cyclic.

Table 1 Requirements for communication quality in industrial feilds

	Latency (ms)	Packet loss ratio	Cycle (ms)	Size (B)
Factory	0.25～10	1.0E-9	0.5～50	10～500
Industrial plant	50～100	1.0E-4 - 1.0E-3	100～5000	40～100
Smart grid	3～20	1.0E-6	10～100	80～1000
Transportation (Safety drive)	10～100	1.0E-5 - 1.0E-3	100～1000	～1000

Table 2: Requirements by operation at the factory

Conventional applications				
Appl.	Reliability (Failure rate)	Latency /Cycle time (ms)	Data rate (Mbps)	Density (nodes /area)
Monitor	1.0E-3	50 - 100	0.1 - 0.5	100 - 1000
Safety Control	1.0E-5	5 - 10	0.5 - 1.0	10 - 20
Closed-loop control	1.0E-5	2-10	1.0 - 5.0	100 - 150
Motion Control	1.0E-6	0.5 - 2	1.0 - 5.0	10 - 50
Emerging applications				
Mobile workforce	1.0E-5	5 - 10	10.0 - 50.0	50 - 100
Augmented Reality	1.0E-4	5 - 10	500.0 - 1000.0	10 - 20
Remote Maintenance	1.0E-4	20 - 50	1.0 - 2.0	500 - 1000
Remote Operation	1.0E-5	2 - 10	100 - 200	1 – 5

4 PROBLEMS AND SOLUTIONS OF ICN FOR INDUSTRIAL IOT SERVICES

As described in Section 2, ICN is expected to simplify communication control and reduce processing load of communication sequences. Therefore, there is a high possibility that ICN is applied to control communication to reduce processing delay and achieve low latency. In this section, authors extract the items to realize communication for industrial control, and to provide solutions for items.

4.1 Low latency communication

To achieve the low latency required for control communication, Cache control specified in ICN is required in addition to bandwidth reservation and priority transfer control that have been used in conventional networks, i.e., the Internet.

Regarding Cache control, Least Frequency Used (LFU) and Least Recent Used (LRU) [5] have been used in distributed processing in Cache control, and their performance improvement as a network has been analyzed in [6]. In addition, it has been considered to create priority areas in the cache to improve the cache hit ratio of information that should be prioritized for forwarding [7]. Furthermore, a method to dynamically change the priority area by monitoring traffic [8] has also been proposed by authors. A study described in [9] summarizing ICN-aware control schemes for Cache has also been published.

On the other hand, control communication requires information to be updated in absolute time. When information is transferred by cyclic communication, as shown in Figure 3, old information continues to be stored in the Cache if information is discarded in the process. This means that when there is a request for information acquisition in an Interest message, the old information is transferred, resulting in inconsistency of information between servers. To avoid this problem, it is effective to forcibly discard information stored in the Cache when there is no information update for a certain period of time, and to acquire information from the adjacent IWP when information acquisition by an Interest message occurs. The basic concept and a method for predicting the forced discard time are proposed by authors described in [10].

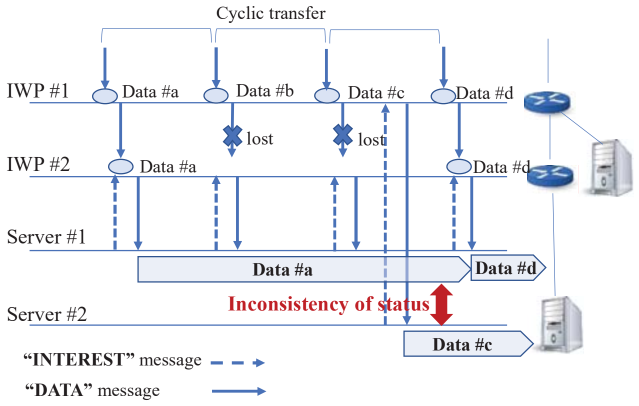


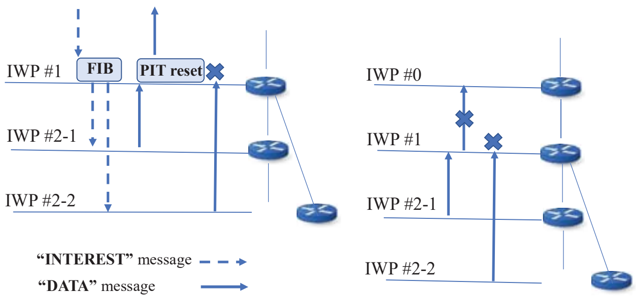
Figure 3 Inconsistency in Server by lost information

4.2 High reliability communication

In control communication, high reliability is important as well as low latency. To achieve high reliability, it is necessary to set up redundant paths and control switching with hitless or near hitless.

Authors have proposed the concept for traffic reduction by providing a protection function at the physical layer of ICN [11]. Detailed scheme is described as follows.

Figure 4 shows the behavior of Cache in ICN when it is not considered. The Interest message to obtain information is forwarded from the IWP concerned to the IWPs connected by the FIB. The information itself is transferred from another IWP to the IWP concerned by a Data message, which is then transferred to the IWP that issued the Interest message. Since the PIT is reset at this time, duplicate transmission of the same Data message does not occur. By using these operations, redundancy is basically possible. However, as shown in Figure 4(b), when Data messages are aggregated at the intermediate IWPs, countermeasures against the loss of Data messages in the downstream direction are required.



(a) Transfer control using FIB/PIT (b) Discarding Data message by aggregation

Figure 4 High reliability and issues using FIB/PIT

4.3 Reduction of processing load

ICN does not necessarily require the protocol stack to be in a layer configuration unlike the conventional Internet. In other words, if ICN is positioned as the application layer of the conventional Internet, processing delays increase due to buffer copying that occurs when each layer is processed in the IWP. Therefore, it is not possible to take advantage of the lightweight communication features of ICN.

To solve this issue, there is a method of processing by Smart NICs on the data link layer. Another method is to connect directly to the data link layer using raw sockets to reduce buffer copying in the OS. For the former, there is a study that describes general-purpose packet processing for each Smart NIC architecture [12], although it is not intended for ICN, and a study that evaluates the performance, when packet processing is programmed in the P4 programming language, see [13] and [14].

On the other hand, there are also studies, e.g., [15] and [16], and others that focus on reducing the number of buffer copies for methods using raw sockets, although they do not assume ICN. The authors have conducted experiments to compare the performance of raw sockets and existing TCP sockets [17].

4.4 Support of push type communication

In ICN, the sequence is command-response type with Interest and Data messages, as shown in Figure 1. On the other hand, in control communication, in most cases, it is a push type communication from the end system. For this reason, as shown in Figure 5, a method is being considered in which information is aggregated at the IWP and information is acquired from the server to that IWP with Interest messages. For example, the realization of ICN-based IoT is described in papers, e.g., [18] - [21]. These papers propose or survey trends in the realization of various IoT services, including control communication using ICN.

However, when end systems, e.g., sensors, are distributed over a wide area, the overhead for aggregation becomes one of critical problems. To solve this problem, a special Interest message. i.e., referred to as Dummy Interest message, is created at the end system to elicit an Interest message as shown in Figure 6.

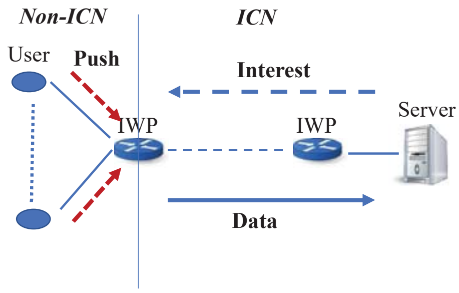


Figure 5 Information aggregation and access by Interest messages

In Figure 6, the control loop for information transfer will become long. For this reason, authors have proposed the method in which Interest messages are issued by IWP, focusing on the fact that ICN is a hop-by-hop communication. This scheme is called CCN with Network initiative And Traffic control (C-NAT). The basic operation is described in [22]. However, the method described in [22] only specifies for cyclic communications. Non-cyclic communication should be supported. Figure 7 shows an overview of C-NAT operation in cyclic and a proposal on non-cyclic communications focusing on features of C-NAT.

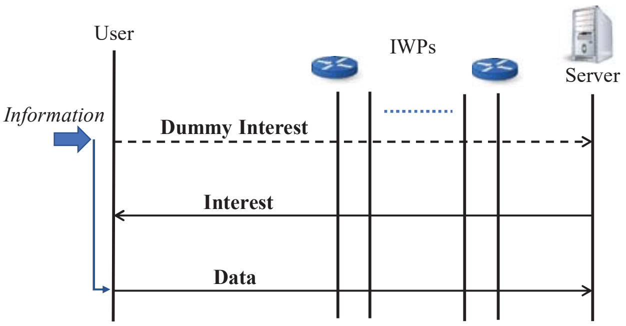
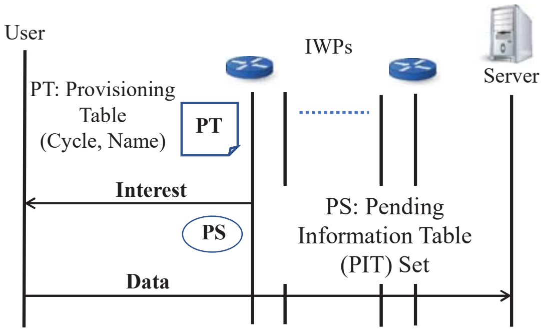
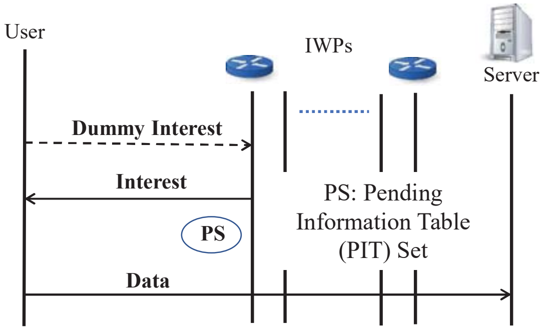


Figure 6 Information Transfer by Dummy Interest

In Figure 7 (b), the IWP accommodating end systems creates Interest message according to basic operations in C-NAT. Therefore, Dummy Interest messages can be terminated at this IWP. Data messages triggered by Interest messages are transferred from an end point in the user to an opposite end point according to operations of C-NAT described in [22]. This proposed method is referred to as C-NAT for Acyclic (C-NAT-A). When messages are lost by bit errors, buffer overflows, etc. and are retransferred, C-NAT-A contributes to reduce traffic volume. A sample of performance evaluation by simulations is shown in Figure 8. This figure shows the case where the ratio of messages lost is 1.0E-5 per physical link. The number of IWPs is indicated by N.


User
IWPs
Server
PT: Provisioning Table (Cycle, Name)
PT
Interest
PS
Data
PS: Pending Information Table (PIT) Set

(a) Cyclic communication sequence in C-NAT



(b) Non-Cyclic communication sequence in C-NAT

Figure 7: Operations of information transfer by C-NAT

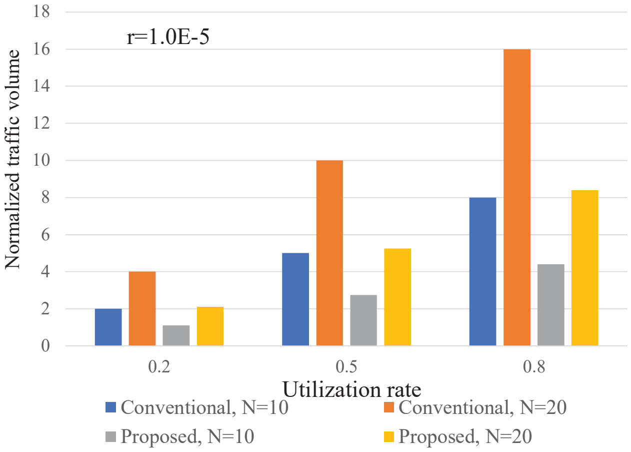


Figure 8: Performance evaluation in C-NAT-A

5 CONCLUSIONS

This paper has discussed the applicability of ICN, which realizes lightweight communication, to control communication based on related trends. It has identified problems for this purpose and has described challenges based on C-NAT by authors. Although there are various issues to be considered for ICN in the future, e.g., operational rules and security, it is possible that ICN will be realized as a new framework that will replace the Internet.

REFRENCES

[1] B.Ahlgren, C. Dannewitz, C. Imbrenda, D. Kutscher, B. Ohlman, “A Survey of Information-Centric Networking,” IEEE Communications Magazine, Vol. 50, No.7, pp. 26-36, 2012

[2] V. Jacobson, D. K. Smetters, J. D. Thornton, M. F. Plass, H. Briggs, R. L. Braynard, “Networking Named Content,” Proceedings of the 5th international conference on Emerging networking experiments and technologies (CoNEXT 2009), pp.1-12, 2009

[3] P. Schulz, M. Matthe, H. Klessig, M. Simsek, G. Fettweis, J. Ansari, S. A. Ashraf, B. Almeroth, J. Voigt, I. Riedel, A. Puschmann, A. Mitschele-Thiel, M. Muller, T. Elste, M. Windisch, “Latency critical IoT applications in 5G: Perspective on the design of radio interface and network architecture,” IEEE Communication Magazine, February, 2017, pp. 70-78, 2017

[4] A. Aijaz, “Private 5G: The Future of Industrial Wireless,” IEEE Industrial Electronics Magazine, Vol. 14, Issue 4, pp. 136-145, 2020

[5] H. Li, H. Nakazato, A. Detti, N. B. Melazzi, "Popularity Proportional Cache Size Allocation Policy for Video Delivery on CCN", Proceedings of EuCNC 2015, pp. 434-438, 2015

[6] S. Podlipnig, L. Böszörmenyi, “A survey of Web cache replacement strategies”, ACM Computing Surveys, Volume 35, Issue 4, pp. 374-398, 2003

[7] H. Qian, W. Muqing, H. Hailong, W. Ning, Z. Chaoyi, "In-Network Cache Management Based on

Differentiated Service for Information-Centric Networking", IEICE Transactions on Communications, Vol.E97-B, No.12, pp. 2616-2626, 2014

[8] A. Yokotani, S. Ohzahata, R. Yamamoto, T. Kato, “A Dynamic Cache Size Assignment Method with Bandwidth Reservation for CCN,” Proceedings of the 33th International Conference on Information Networking (ICOIN 2019), P2-15, 2019

[9] C. Barrios, M. Kumar, “Service Caching and Computation Reuse Strategies at the Edge: A Survey,” Journal of ACM Computer Survey. Vol.56, Issue 2, Article 43, 2023

[10] A. Yokotani, H. Mineno, T. Yokotani, “A cache control mechanism in CCN for cyclic communication base IoT services,” IEICE Communication Express, Vol.11, No.8, pp. 521 - 526, 2022

[11] A. Yokotani, H. Mineno, T. Yokotani, “High reliable information transfer in the IoT platform: C-NAT,” Proceedings of IEEE International Japan-Africa Conference on Electronics, Communications, and Computations (JAC-ECC 2023), Paper ID:42, 2023

[12] E. F. Kfoury, S. Choueiri, A. Mazloum, A. Alsabeh, J. Gomez, and J. Crichigno, "A Comprehensive Survey on SmartNICs: Architectures, Development Models, Applications, and Research Directions," IEEE Access Vo. 12, pp. 107297- 107336, 2024

[13] H. Harkous, M. Jarschel, M. He, R. Priest, and W. Kellerer, "Towards Understanding the Performance of P4 Programmable Hardware ," Proceeding of the 2019 ACM/IEEE Symposium on Architectures for Networking and Communications Systems (ANCS), 2019, IEEE Xplore DOI: 10. 1109/ANCS.2019.8901881

[14] J. Xing, Y. Qiu, K. Hsu, S. Sui, K. Manaa, O. Shabtai, Y. Piasetzky, M. Kadosh, A. Krishnamurthy, T. S. E. Ng, A. Chen, "Unleashing SmartNIC Packet Processing Performance in P4," Proceedings of the ACM SIGCOMM 2023, pp. 1028 - 1042, 2023

[15] P. Balaji, S. Bhagvat, H. W. Jin, and D. K. Panda, "Asynchronous Zero-copy Communication for Synchronous Sockets in the Sockets Direct Protocol (SDP) over InfiniBand," Proceedings of the 20th international conference on Parallel and distributed processing (IPDPS'06), pp. 271-278, 2006

[16] V. Tikhonov, S. Nesterenko, O. Tykhonova, O. Tsyra, O. Yavorska, V. Hlushchenko, "Management of Digital Streams of an Autonomous System by Management of Digital Streams of an Autonomous System by the Raw Socket Ethernet Channel Virtualization Method in Linux OS," Proceedings of The 11thInternational Conference on Applied Innovation in IT (ICAIIT 2024), pp. 1-6, 2024

[17] M. Fukuda, K. Ishibashi, T. Yokotani, K. Akta, and T. Suzuki, "A proposal on raw socket communication for lightweight transport protocol for robot applications," Asia Pacific Conference on Robot IoT System Development and Platform 2024 (APRIS 2024), A4, 2024

[18] M. Frey, C. Gundogan, P. Kietzmann, M. Lenders, H. Petersen, T. C. Schmidt, F. Juraschek, M. Wahlisch, "Security for the Industrial IoT: The Case for Information-Centric Networking," in Proceedings of the IEEE 5th World Forum on Internet of Things (WF-IoT), IEEE Xplore DOI: 10.1109/ WF-IoT.2019.8767183, 2019

[19] S. S. Adhatarao, M. Arumaithurai, D. Kutscher, and X. Fu, "ISI: Integrate Sensor Networks to the Internet with ICN," IEEE Internet of Things Journal, Vol. 5, Issue 2, pp. 491 - 499, 2017

[20] J. Chen, S. Li, H. Yu, Y. Zhang, D. Raychaudhuri, R. Ravindran, H. Gao, L. Dong, G. Wang, and H. Liu, "Exploiting ICN for Realizing Service - Oriented Communication in IoT. Oriented Communication in IoT," IEEE Communications Magazine - Communications Standards Supplement, pp. 24 - 30, December, 2016

[21] I. U Din, H. Asmat, and M. Guizani, "A review of information centric network-based internet of things: communication architectures, design issues, and research opportunities," Vol. 78, pp. 30241 - 30256, multimedia Tools and Applications, 2019

[22] A Yokotani, H. Mineno, S. Ohzahata, T. Yokotani, "A Proposal on New Control Mechanisms based on ICN for Low Latency IoT services," International Journal of Informatics Society (IJIS), Vol. 14, No. 2, pp. 95-104, 2022

Anomaly Detection of Network Traffic Based on Frequency Analysis with Consideration of Time Scales

Rintaro Okamoto*, Koichi Ishibashi**

* Graduate School of Engineering, Kanazawa Institute of Technology, Japan
c6500735@st.kanazawa-it.ac.jp
** College of Information Science and Engineering, Kanazawa Institute of Technology, Japan
k_ishibashi@neptune.kanazawa-it.ac.jp

Abstract - Recent studies have focused on detecting and controlling anomaly network traffic. In particular, with the widespread adoption of the Internet of Things (IoT), the threat of attacks targeting IoT devices has become more apparent, leading to the increased deployment of Intrusion Detection Systems (IDSs). Signature-type IDSs cannot detect communications and attacks by unknown malware. Anomaly-based IDSs use statistical and machine learning methods to detect deviations, but there are still issues to be solved, such as the need to define various types of traffic behavior. To address the above issues, research on methods for detecting the anomaly traffic and changes in the statistical characteristics of network traffic has been studied in various research communities. This paper proposes and evaluates a method to efficiently detect changes in statistical characteristics of network traffic by frequency analysis based on multiple time scales. The proposed method compares the features of each time-series data by frequency analysis, leading to the detection of statistical characteristic changes.

Keywords: anomaly detection, frequency analysis, IDS

1 INTRODUCTION

In recent years, the rapid proliferation and acceleration of the Internet have led to a dramatic increase in Internet traffic, accompanied by a diversification of usage patterns and applications. In particular, the widespread adoption of mobile devices such as smartphones, tablets, and mobile PCs has resulted in Internet access via mobile devices accounting for a significant share of communication traffic.[1] Consequently, the efficient operation and management of networks has become increasingly important, making the adoption of traffic monitoring technologies essential for detecting and controlling anomaly traffic that can cause resource wastage and degradation of the quality of services. Traditionally, numerous studies have been conducted on methods for monitoring and analyzing anomaly traffic that consumes network resources and decreases the quality of services. For example, [2] has proposed a traffic monitoring method on backbone networks provided by network operators. It has presents the effectiveness of an approach that automatically determines thresholds based on time-series predictions using time-series analysis techniques. From a security perspective, extensive research has also been conducted on the detection of attacks and malware infections through network traffic analysis. Furthermore, with the recent rapid expansion of the Internet of Things (IoT), the threat of attacks targeting IoT devices has become more pronounced, leading to the deployment of intrusion detection systems (IDS).

IDS are generally classified into two types: signature-based and anomaly-based. Signature-based IDS detect attacks by searching for specifically defined patterns, such as byte sequences within network traffic or known malicious instruction sequences used by malware. Anomaly-based IDS, on the other hand, define expected behaviors and detect deviations from these behaviors to identify attacks. Owing in part to the rapid development of malware, anomaly-based IDS are primarily introduced to detect unknown attacks. Deviations are detected using statistical and machine learning techniques; however, there are many challenges, such as the need to define the behaviors of various types of traffic. Moreover, the frequent emergence of diverse variants of typical malware makes it difficult to detect malware based on the same set of characteristics.

In addition, machine learning has been widely utilized in traffic analysis in recent years. However, training data for supervised learning is essential for traffic analysis, and re-training of traffic models due to changes in traffic trends has become a major challenge.

In consideration of the above, we propose an anomaly detection method that efficiently detects statistical trends change in traffic through frequency analysis based on multiple time scales. In this study, to evaluate the feasibility of the proposed method for generating time-series data based on multiple time scales from network traffic data and extracting characteristics of time and frequency within specified time windows.

The remainder of this paper is organized as follows. Section 2 provides an overview of related works and the challenges remained by each work. The proposed method are described in Section 3. Section 4 evaluates the proposed method by way of using a opened datasets . Finally, Section 5 presents the conclusions of this paper.

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 anomaly traffic caused by infected mobile devices connecting to the network has also become a problem. For example, [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 anomaly 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, 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 area[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, which gives distance between data of different lengths while allowing for time axis deviations. There is also a proposal for dissimilarity based on the magnitude of the frequency component of time-series data[7], and a proposal for dissimilarity focusing on the frequency component using the Discrete Fourier Transform (DFT). In addition, analysis of communication traffic using the Wavelet transform has been proposed in [8][9], among others.

In [10][11], studies have also been conducted to detect change in patterns and determine when to reconstruct clustering by calculating the distance between cluster transition distributions.

In contrast to the above, this study discusses an anomaly detection mechanism that uses detected abnormal traffic data as reference material, analyzing network traffic within each time window.

3 PROPOSAL

In this study, we aim to detect statistical trends change of network traffic by frequency analysis based on multiple time scales. Specifically, we propose a method of change point detection in traffic trends that takes into account the Euclidean distance between characteristics of time and frequency derived in adjacent time windows, as well as the Euclidean distance between characteristics of time and frequency derived in time windows belonging to different time scales. The proposed method is analyzed and evaluated according to the following procedures.

1) By way of a network tool such as Wireshark, monitor traffic on a target network and/or network interface.
2) Create time-series data from the monitoring data based on sampling time.
3) Extract characteristics of time and frequency at each time window from the created time-series data.
4) Detect a change point in traffic trends that takes into account the Euclidean distance between characteristics derived in adjacent time windows, as well as the Euclidean distance between characteristics derived in time windows belonging to different time scales.

According to the above procedure, the proposed method detect the change points of traffic trends.

Deviation of traffic trends, as shown on equation (1), is obtained based on the following two factors:

1) The difference (diff_ne) based on the results derived from frequency analysis (feature extraction up to the Nth order by Wavelet transform) for the time window wb(t) and the adjacent time window wb(t-1), as shown on equation (2).
2) The difference (diff_sc) based on the results derived from frequency analysis for the time-series data at different time scales.

$$d\_fr(t) = f(\,\mathrm{diff\_ne(t)}\,,\mathrm{diff\_sc(t)}\,) \tag{1}$$

$$diff\_ne(t) = \sum_{i=1}^{N} k_i \times 1/\left(1 + eu\_len\,(fc_{b(t,i)}, fc_{b(t-1,i)})\right)\,,\ \sum_{i=1}^{N} k_i = 1 \tag{2}$$

$$diff\_sc(t) = \sum_{i=1}^{N} k_i \times 1/\left(1 + eu\_len\,(fc_{b(t,i)}, fc_{s(t,i)})\right)\,,\ \sum_{i=1}^{N} k_i = 1 \tag{3}$$

It is noted that where $fc\_b(t,i)$ is the i-th order characteristic of time and frequency in time window wb(t) and $fc\_s(t,i)$ is the i-th order characteristics in time window ws(t) at different time scales with respect to time window wb(t). eu_len(x, y) is the Euclidean distance between x and y. ki is the weight for the i-th characteristics and is a parameter for more accurately tuning the detection of change points.

Figure 1 shows an overview of the proposed method.

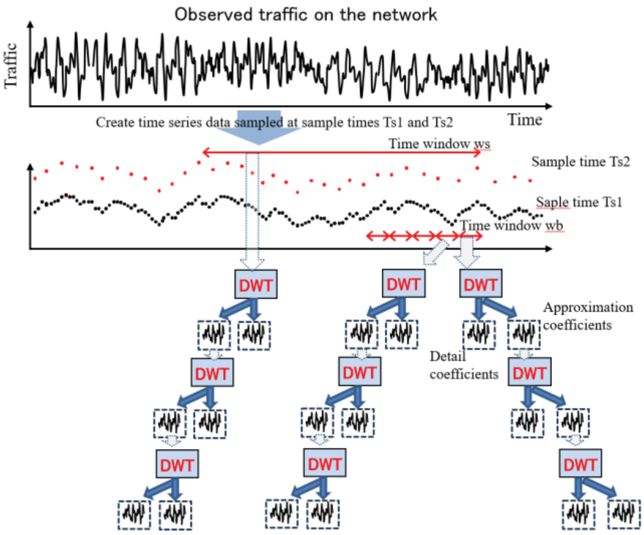

Observed traffic on the network
Traffic
Time
Create time series data sampled at sample times Ts1 and Ts2
Time window ws
Sample time Ts2
Saple time Ts1
Time window wb
DWT
Approximation coefficients
Detail coefficients

Figure 1: Overview of the proposed method.

4 EVALUATION

In this section, we evaluate the proposed method for detecting trends change in network traffic by using Wavelet transform with consideration of multiple time scales.

4.1 Datasets

Various datasets are available for intrusion detection and other research purposes, including KDD99[12], ISCXIDS2012[13], UNSW-NB15[14], and CICIDS2017[15]. To evaluate the effectiveness of Network Intrusion Detection Systems (NIDS) based on their ability to identify attacks, a comprehensive dataset containing both normal and anomalous behaviors is essential. While older benchmark datasets like KDD99 have been widely adopted for assessing NIDS performance, improved datasets have emerged to address limitations such as redundant records, discrepancies from real-world conditions, and outdated attack patterns in legacy datasets.

This paper utilizes the UNSW-NB15 dataset, developed by the Cyber Security Research Group at the University of New South Wales as a modern evaluation resource for NIDS. UNSW-NB15 generates nine categories of network traffic samples encompassing both cyberattacks and normal traffic. The dataset was designed to reflect the diversity and realism of actual network traffic, containing over one million records with 49 feature attributes. It includes attacks such as Denial-of-Service (DoS), fuzzing, analysis, backdoor, exploit, generic, reconnaissance, shellcode, and worms.

4.2 Creation of time-series data

First, a time-series data is extracted from the CSV file provided by the UNSW-NB15 dataset, including transmitted bytes, received bytes, and packet arrival times from source and destination nodes. That is, the time-series data is obtained by accumulating the transmitted and received bytes at specified monitoring time window intervals, called sampling time based on the packet arrival times. The resulting time-series data forms a two-dimensional waveform, which is used for characteristics extraction in each evaluation time window. Figure 2 shows an example of waveform data obtained from the UNSW-NB15 dataset when the sampling time is 100msec.

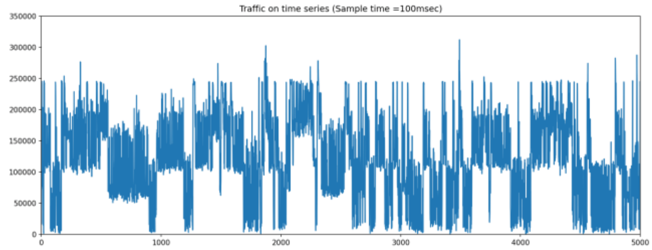


Figure 2: Example of time-series data.

4.3 Estimation by datasets

One of the characteristics of wavelet transform is its ability to capture detailed changes in frequency components at specific points in time. Furthermore, it is well-suited for extracting both global and local characteristics of a signal. By leveraging these properties, it is possible to analyze traffic patterns and perform characteristics extraction and classification. So, in this paper, we analyze time-series data using the wavelet transform.

The deviation between adjacent time windows and those between time-series data with varying time scales are evaluated. Figures 3 and 4 provide the results of the difference evaluations. That is, Figure 3 shows the results based on time-series data created with a sampling time of 1 msec, and the top graph is the created time series-data.

The middle graph in Figure 3 shows the Euclidean distance between an given time window and its adjacent time window for the characteristics up to the third order when the Wavelet transform is applied to a time window consisting of 512 data, and the Euclidean distance between an given time window and a time window with different time scale with a sampling time of 1 msec x 32. And the bottom graph in Figure 3 shows the difference between the features in the time window according to equations (2) and (3).

On the other hand, Figure 4 shows the result based on time-series data created with a sampling time of 10 msec.

In the dataset studied in this evaluation, there are periods of time when traffic increases spike-like, but is not classified as an anomaly traffic.

In response to the above, it seems possible to suppress the impact on the detection of traffic trend changes by adjusting the weights of the high-frequency component characteristics based on the frequency analysis.

In the comparison of features by frequency analysis based on time-series data with different time scales, it appears that there are time periods that show different trends from the comparison of features between adjacent time windows.

As a result, the possibility of improving the detection accuracy for detecting traffic trend changes on a single time scale is expected.

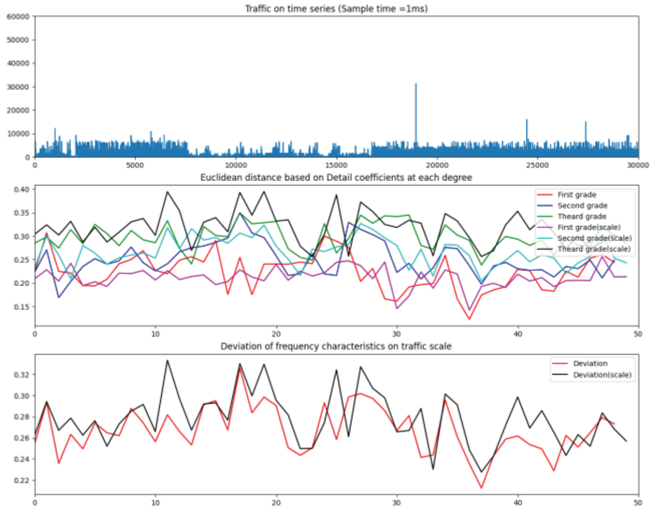


Figure 3: Graph with sampling time of 1ms.

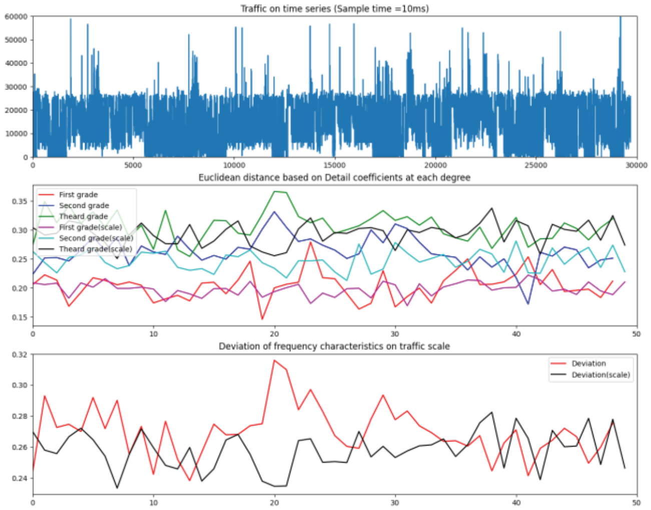


Figure 4: Graph with sampling time of 10ms.

5 CONCLUSION

In this paper, we propose a frequency analysis method that considers multiple time scales for the purpose of trends change detection in network traffic. As an initial step, we evaluated the results of applying wavelet transform in terms of time scales. In future work, we plan to investigate algorithms for anomaly detection by evaluating the proposed approach using time-series data with various time scales.

REFERENCES

[1] Ericson Mobility Report, June 2015, 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] M ehedee 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] Arash Habibi Lashkari, Andi Fitriah A.Kadir, Hugo Gonzalez, Kenneth Fon Mbah, and Ali 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] Takanori Yoshikawa, Masahiro Ishikawa, Hanxiong Chen, Kazutaka Furuse, and Nobuo Ohbo, "A Study on Similarity Search for Very Long Time-Series Data," Data Engineering Workshop (DEWS) 2007, February 2007.

[7] Katsumasa Koyama, Tbruhisa Hochin, Hideya Nakanishi, and Mamoru Kojima, "Similarity Based on Spectrum of Time Series," IPSJ SIG technical reports, November 2006.

[8] Darya Lavrova, Pavel Semyanov, Anna Shtyrkina, and Peter 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-TESC2018), Volume 44, June 2018.

[9] Josef Koumar, Karel Hynek, and Tomáš Čejka, "Network Traffic Classification based on Single Flow Time Series Analysis," The 19th International Conference on Network and Service Management (CNSM), October 2023.

[10] Shoko Takahashi, Kei Takeshita, Kazuhisa Yamagishi, Akihiro Shiozu, "Change Point Detection Based on Cluster Transition Distributions", IEEE Access, September 2024.

[11] Dang-Hoan Tran, "Automated Change Detection and Reactive Clustering in Multivariate Streaming Data", 2019 IEEE-RIVF International Conference on Computing and Communication Technologies (RIVF), May 2019.

[12] Tavallaee Mahbod, Bagheri Ebrahim, Lu Wei, and Ghorbani 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.

[13] 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.

[14] 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.

[15] Iman Sharafaldin, Arash Habibi Lashkari, and Ali A. Ghorbani, "Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization", 4th International Conference on Information Systems Security and Privacy (ICISSP), Purtogal, January 2018.

Session 8:
Computing in Society and Culture

(Chair: Takuya Yoshihiro)

Automatic Translation from PlantUML description to NuSMV using LLM

Kansei Inoue[†], Shinpei Ogata[†], Toshifusa Sekizawa[‡]and Kozo Okano[†]

[†]Faculty of Engineering, Shinshu University, Japan
[‡]College of Engineering, Nihon University, Japan
[†]{25w6012a, ogata, okano}@shinshu-u.ac.jp
[‡]sekizawa.toshifusa@nihon-u.ac.jp

Abstract - In software development, requirement specifications written in natural language often cause ambiguity, leading to unintended system behavior. To address this, developers create state transition diagrams and perform model checking, but these tasks require advanced expertise and considerable effort. This study proposes an approach using ChatGPT, a large language model (LLM), to support this process. We developed two methods: one automatically generates NuSMV code from state transition diagrams described in PlantUML, and the other generates verification properties, including CTL formulas, from the NuSMV code. Experiments using diagrams for the boil, keep-warm, and timer buttons of an electronic hot water pot (7th Edition) showed that the generated code and tests contained minor errors but were still practically usable.

Keywords: Formal Methods, NuSMV, LLM, Prompt Engineering

1 Introduction

A requirements specification document describes the behavior and functions of software in natural language and plays a crucial role in development. However, natural language inherently contains ambiguity—sentences can be interpreted in multiple ways—causing miscommunication, design errors, and system defects [1]. Such ambiguities often lead to rework, redesign, and reimplementation, increasing development time and cost [2]. Especially in large-scale projects, early misunderstandings can have serious downstream effects, making it vital to minimize these risks.

One effective approach to verifying requirements is to create a state transition diagram and perform model checking. A state transition diagram visually represents possible system states and transitions, helping clarify system behavior. Model checking automatically verifies whether the system design satisfies the specified requirements, detecting unintended behavior or unsafe states early [3]. Conducting this verification at the design stage enhances quality assurance in system development.

However, these tasks demand high expertise. Engineers must deeply understand system behavior to create accurate diagrams and express specifications formally for model checking. Transforming ambiguous natural language into formal expressions is labor-intensive and error-prone, posing challenges to efficiency and accuracy [4], [5].

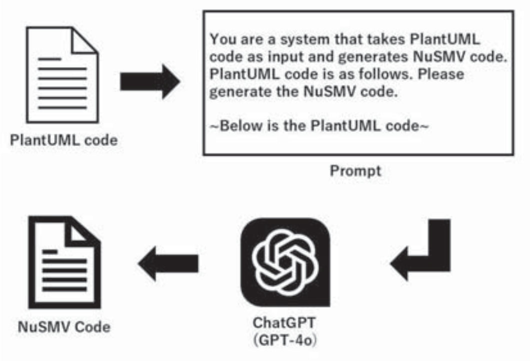
To overcome these issues, recent studies have explored automatic analysis of requirement specifications and automated model generation [6]-[8]. Last year ' s research group proposed an LLM-based method for automatically generating state transition diagrams [9]. The use of large language models (LLMs) has also expanded to code conversion between programming languages [10], [11], but research on conversions from minor languages—such as from PlantUML to model checking code—remains limited. If LLMs can automate both state transition diagram generation and model checking code conversion, it could significantly reduce rework risks and improve development efficiency. Furthermore, even engineers with limited expertise could perform precise verification tasks. Since ChatGPT can already derive state transitions in generating NuSMV code from PlantUML, this approach could eventually enable direct derivation from natural language requirements themselves. Thus, LLM-based support for software development is expected to become increasingly significant [12]-[15]. The rest of the paper is organized as follows. Chapter 2 describes the proposed methods. Chapter 3 presents experimental results. Chapter 4 discusses these results. Chapter 5 concludes the paper and outlines future work.

2 Material and Method

This chapter describes the proposed methods for automatic generation of model checking code and verification properties, as well as the details of the target systems.

2.1 Automatic Generation of Model Checking Code

We propose a method to automatically convert models described using PlantUML, which is widely used for software system modeling, into the specification description language used by the formal verification tool NuSMV. Our research group has previously proposed a method for deriving PlantUML state transitions from requirement specifications written in natural language[17]. By combining this technology, requirement specifications in natural language can be automatically converted into model-checkable descriptions using LLMs. In our method, we input text-based PlantUML code into ChatGPT to efficiently ingest and transform the model. The actual generation process of model checking code is shown in Figure 1. We use PlantUML code as input and perform NuSMV code generation through Zero-Shot Prompting. GPT-4o is used as the version of ChatGPT[16].



You are a system that takes PlantUML code as input and generates NuSMV code. PlantUML code is as follows. Please generate the NuSMV code.

~Below is the PlantUML code~

Figure 1: How to generate model checking codes

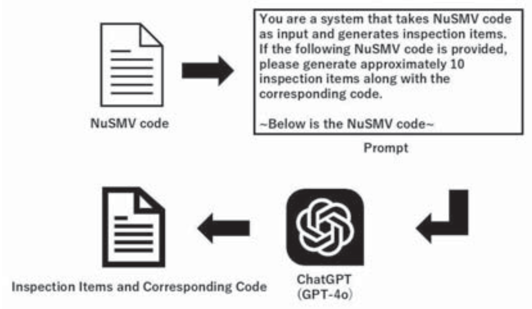
Our proposed method analyzes the input PlantUML code to extract elements such as states, transitions, and condition branches. Since PlantUML code is text-based, it is easy to analyze automatically and avoids data loss or ambiguity compared to image-based inputs.

Next, based on the model elements described in PlantUML, the process of conversion into NuSMV specification language is performed. This conversion accurately maps the state transitions and conditions of PlantUML into the syntax used for formal verification in NuSMV, generating a verifiable specification. The process is fully automated, and developers can use it without needing specialized knowledge in formal verification.

In designing prompts for ChatGPT, we adopt the Persona pattern proposed by White[18], which catalogs prompts according to use cases. We instructed ChatGPT to act as an ”expert in formal verification tools.” This allows ChatGPT to accurately interpret PlantUML code and generate appropriate NuSMV code. By setting such a persona, we expect improved consistency in the generated code and deeper understanding of the specifications of PlantUML and NuSMV.

2.2 Automatic generation of model verification properties

This section describes the method for automatically generating model verification properties. The actual generation process is shown in Figure 2. Similar to the method in Figure 1, we use Zero-Shot Prompting with NuSMV code as input to generate verification properties. Using NuSMV code as input allows us to consistently treat both the generation of verification properties and the corresponding model checking code. In order to evaluate the validity and effectiveness of the automatic generation technology in stages, the prompt is set to generate a small number of verification properties, specifically 10.



You are a system that takes NuSMV code as input and generates inspection items. If the following NuSMV code is provided, please generate approximately 10 inspection items along with the corresponding code.

~Below is the NuSMV code~

Figure 2: How to generate verification properties

To establish the process for generating model verification properties from NuSMV code, we use the Persona pattern to design prompts for ChatGPT, as shown in Figure 1.

2.3 Model Checking Target

In this study, we selected the operational specifications of the timer button in the product ”electronic hot water pot (7th Edition)” as the target for model checking[19]. In addition to basic functions like boiling and dispensing hot water, this product features a timer for time measurement and a water level indicator. For the experiments, we selected 3 of the 8 requirement verification properties for the pot’s software: the boiling button, keep warm button, and timer button. As an example, I will explain the state transition diagram for one of them, the timer button.

The timer button allows users to easily set the time in minutes and count down, and is used to measure the time required to make cup ramen. Upon connecting the power, the timer resets to ”0 minutes 0 seconds” and enters a stopped state. Each press of over 100 milliseconds adds one minute, up to a maximum of 60 minutes. Additional time can be set even during countdown, and each press triggers a 50-millisecond buzzer sound. Releasing the button (after 1 second) starts the countdown. When less than 5 minutes remain, the display switches to show seconds. A long press of 3 seconds resets and stops the timer with a 100-millisecond buzzer. When the countdown ends, the buzzer sounds three times at 100-millisecond intervals to indicate time up. The state transition diagram for the timer button is shown in Figure 3.

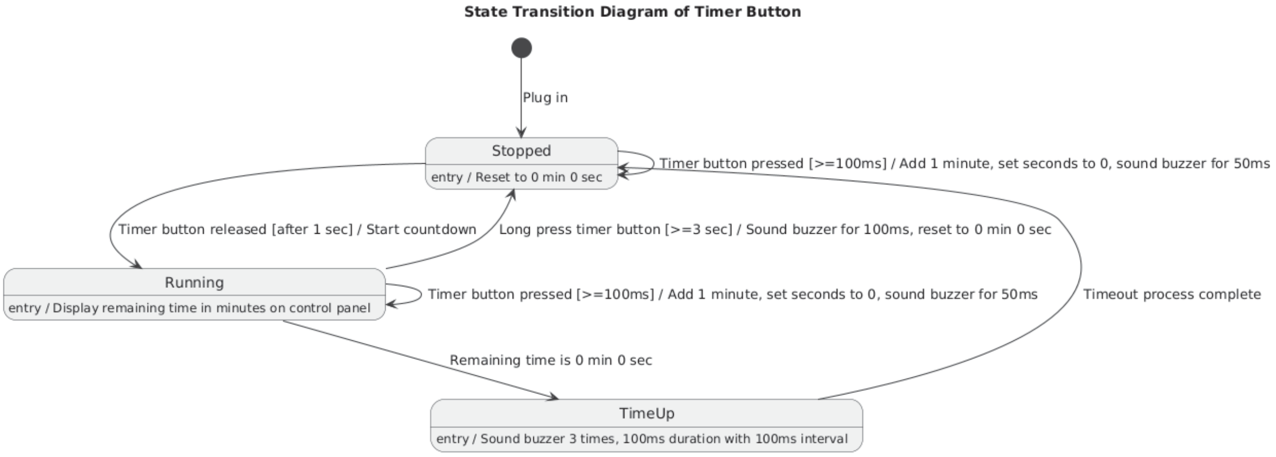

State Transition Diagram of Timer Button
Plug in
Stopped
entry / Reset to 0 min 0 sec
Timer button pressed [>=100ms] / Add 1 minute, set seconds to 0, sound buzzer for 50ms
Timer button released [after 1 sec] / Start countdown
Long press timer button [>=3 sec] / Sound buzzer for 100ms, reset to 0 min 0 sec
Running
entry / Display remaining time in minutes on control panel
Timer button pressed [>=100ms] / Add 1 minute, set seconds to 0, sound buzzer for 50ms
Timeout process complete
Remaining time is 0 min 0 sec
TimeUp
entry / Sound buzzer 3 times, 100ms duration with 100ms interval

Figure 3: State transition diagram for the timer button

3 Results

This section presents the verification results of the automatically generated NuSMV code and verification properties for the boil button, keep warm button, and timer button. First, manually created verification properties are applied to automatically generated NuSMV code to verify the validity of the code's behavior and transition specifications. Then, automatically generated verification properties are applied to the same model to verify that the test expressions correctly capture the properties of the model.

3.1 Boil Button

For the boil button, we prepared 5 manually written verification properties and generated 10 additional ones. When testing the 5 prepared items with the generated NuSMV code, all produced the expected results. The actual verification properties are as follows:

1. When the button is pressed during warm-keeping and not during dispensing, the state transitions to boiling.
   $AG(state = hoon\_action \& button \& !is\_hotwater \rightarrow AF state = hutto\_action)$

2. After the initial state, the system always transitions to warm-keeping.
   $AG(state = initial \rightarrow AX state = hoon\_action)$

3. The system eventually transitions to warm-keeping.
   $AF state = hoon\_action$

4. While in boiling state and the boiling is not completed, the system does not return to warm-keeping.
   $AG(state = hutto\_action \& !boil\_done \rightarrow AX state != hoon\_action)$

5. The buzzer does not sound in the initial state.
   $AG(state = initial \rightarrow buzzer = FALSE)$

The $is\_hotwater$ variable in the inspection items is used to manage whether hot water is being supplied. Due to the specifications of the boiling pod, boiling does not occur while hot water is being supplied, so this variable is used as a condition.

Next, we evaluated the ten automatically generated verification properties using NuSMV code. While eight of them produced the expected results, Items 1 and 5 did not meet expectations. The reason why the expected results were not obtained is thought to be that, although the test content and code were correct, there was a defect in the generated NuSMV code.

1. If the button is not pressed, the buzzer does not sound.
   $AG(!button \rightarrow buzzer = FALSE)$

2. During keep-warm mode, if the button is pressed and hot water is not being dispensed, transition to boiling mode.
   $AG(state = hoon\_action \& !is\_hotwater \& button \rightarrow AX\ state = hutto\_action)$

3. During keep-warm mode, if hot water is being dispensed and the button is pressed, remain in keep-warm mode.
   $AG(state = hoon\_action \& is\_hotwater \& button \rightarrow AX state = hoon\_action)$

4. During boiling mode, when boiling completes (boil_done), transition back to keep-warm mode.
   $AG(state = hutto\_action \& boil\_done \rightarrow AX\ state = hoon\_action)$

5. The buzzer does not sound twice in succession.
   $AG(buzzer = TRUE \rightarrow AX(buzzer = FALSE \&$
   $AX\ buzzer = FALSE))$

6. During keep-warm mode, if hot water is not being dispensed and the button is pressed, the buzzer should sound.
   $AG(state = hoon\_action \& !is\_hotwater \& button \rightarrow$
   $AX\ buzzer = TRUE)$

7. If the buzzer sounds, it should stop in the next step.
   $AG(buzzer = TRUE \rightarrow AXbuzzer = FALSE)$

8. There is no direct transition from the initial state to boiling mode.
   $AG(state = initial \rightarrow AXstate != hutto\_action)$

9. During keep-warm mode, if the button is not pressed, the state does not change.
   $AG(state = hoon\_action \& !button \rightarrow$
   $AXstate = hoon\_action)$

10. During boiling mode, if boiling does not complete, transition to keep-warm mode does not occur.
    $AG(state = hutto\_action \& !boil\_done \rightarrow AXstate = hutto\_action)$

3.2 Keep Warm Button

For the Keep Warm button, five items were manually prepared and ten items were automatically generated for testing. All five manually prepared items produced the expected results, similar to the Boil button. Of the ten generated items, only one failed to meet expectations:

- Transition from High mode to Eco mode is possible.
  $AG(state = HIGH \rightarrow AX(button\_long\_press$
  $\&buzzer \rightarrow EXstate = ECO))$

This is an error caused by inconsistency with the generated NuSMV code, and the expected result was obtained by modifying the code to $SPECAG(state = HIGH \&$ $button\_long\_press \& buzzer- > AX(state = ECO))$.

3.3 Timer Button

For the Timer button, nine items were manually prepared and ten items were automatically generated. Among the manually prepared items, six produced expected results while three did not, as shown below.

1. When you press and hold the timer button during startup, will it transition while stopped?
   $AG(state = running \& timer\_button\_press\_time >= 3000 \rightarrow AX(state = stopped))$

2. When the timer button is pressed and held for 0 seconds or longer during startup, will it transition to stop mode?
   $AG(state = running \& timer\_button\_press\_time > 0 \rightarrow AX(state = stopped))$

3. When the minute timer is set to 60 minutes, will pressing the button reset it to 0 minutes?
   $AG(state = running \& timer\_button \&$
   $timer\_button\_press\_time = 100 \& timer\_count = 60 \rightarrow$
   $AX(timer\_count = 0))$

Next, for the ten items automatically generated, the expected results were obtained for seven items, but the expected results were not obtained for three items. The three items are as follows:

1. Does the timer always decrease while the device is running?
   $AG(state = running \rightarrow A(timer\_count > 0 \rightarrow$
   $AX(timer\_count <= timer\_count)))$

2. Does it not start up unless you press and hold the button?
   $AG(state = stopped \& timer\_button \rightarrow$
   $AX(timer\_button\_press\_time >= 100 \rightarrow state = running))$

3. As long as the timer does not reach 0, there will be no timeout.
   $AG((timer\_count > 0 \mid sec\_count > 0) \rightarrow AG!(state = timeout))$

Regarding the timer button, most of the errors were related to counter variables in the generated NuSMV code.

3.4 Overall results

The overall results are summarized in Table 1. The boil button showed the expected results in 13 out of 15 items, with an accuracy rate of 86.7%. The keep warm button worked correctly in 14 out of 15 items, showing the highest accuracy rate of 93.3%. On the other hand, the timer button showed the expected results in 13 out of 19 items, with an accuracy rate of 68.4%. From the above results, it is clear that the keep warm button operates most stably and accurately.

4 Discussion

This section discusses the results of verifying the automatically generated NuSMV code and verification properties

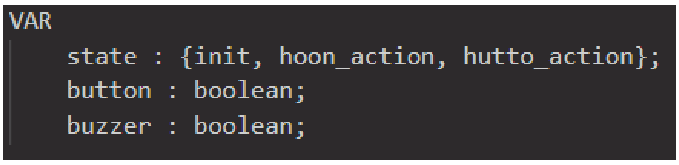
4.1 Automatically generated NuSMV code

The automatically generated NuSMV code generally reflected the states and transition conditions of the original state transition diagram. However, some verification properties did not

Table 1: Total number of verification properties and success rate

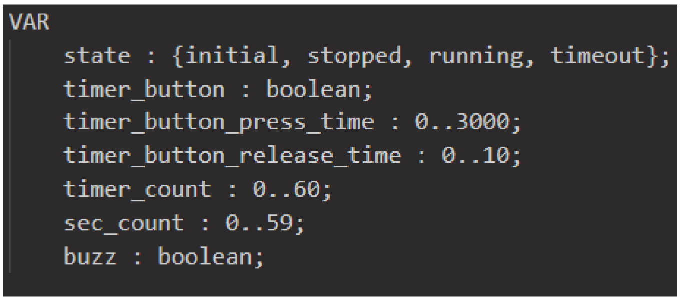
	Number of Properties	Number of Successes	Success Rate(%)
Boil Button	15	13	86.7
Keep Warm Button	15	14	93.3
Timer Button	19	13	68.4

yield the expected results. As shown in Figure 6, the causes of these errors fall into three categories: inconsistencies between verification properties and NuSMV code, missing information in the original diagram, and time handling in the NuSMV code. The most frequent cause was improper handling of time, mainly observed in model checking of timer buttons. While the boil and keep-warm buttons use simple Boolean control without time processing, the timer button involves counter variables to record both press and release times, resulting in more complex time-dependent behavior.



```
VAR
    state : {init, hoon_action, hutto_action};
    button : boolean;
    buzzer : boolean;
```

Figure 4: VAR section for the boil button



```
VAR
    state : {initial, stopped, running, timeout};
    timer_button : boolean;
    timer_button_press_time : 0..3000;
    timer_button_release_time : 0..10;
    timer_count : 0..60;
    sec_count : 0..59;
    buzz : boolean;
```

Figure 5: VAR section for the timer button

When handling time in NuSMV, it is necessary to use counter variables to represent time pseudo-logically, which complicates the program. For the two items, the boil button and the keep warm button, success rates of 86.7% and 93.3% were achieved, respectively. Therefore, when time is not involved, it is thought that useful NuSMV code can be generated even with Zero-Shot-Prompting. The automatic generation of NuSMV code using LLM is expected to be a useful approach for efficiently conducting formal verification of dynamic time management specifications, but there is room for improvement in the current automatic generation process. Regarding the reflection of specifications related to dynamic time management, further refinement and improvement are required.

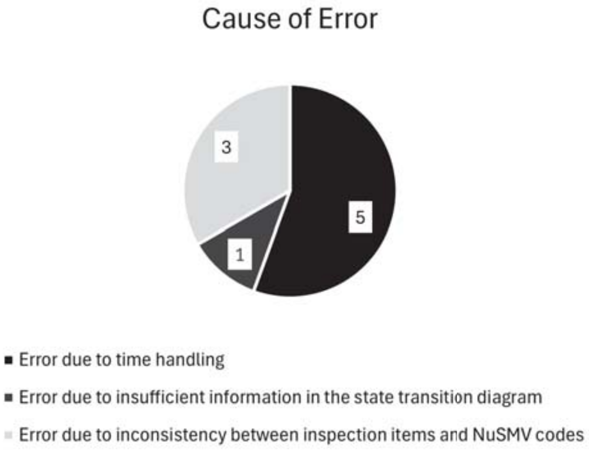


Figure 6: Cause of error

4.2 Automatically generated verification properties

Consider whether the automatically generated verification properties are appropriate. The verification properties automatically generated by ChatGPT can be classified into the following four categories.

1. Validity of state transitions
   Verification that transitions between states are performed correctly as designed.
   (Example:$AG(state = hoon\_action \& !is\_hotwater \& button \rightarrow AX state = hutto\_action)$)

2. Safety and immutability
   Verification that the state does not change under inappropriate conditions.
   (Example:$AG(state = hoon\_action \& is\_hotwater \& button \rightarrow AX state = hoon\_action)$)

3. Output operation control
   Verification of buzzer sounding conditions, timing, and sounding restrictions.
   (Example:$AG(buzzer = TRUE \rightarrow AXbuzzer = FALSE))$

4. Timer and count behavior
   Verification that timers and count processing are working properly.
   (Example:$AG(state = timeout \rightarrow$
   $AX(timercount = 0\&seccount = 0)))$

By combining the above four categories, it is possible to not only comprehensively check the contents of the requirements specifications, but also to broadly check specifications that are not directly written in the requirements specifications. For example, the specification that the buzzer should not sound twice was not written in the requirements specifications. The number of verification properties for each category is shown in Figure 7. The generated categories showed the highest number of items related to the validity of state transitions. The verification properties automatically generated by ChatGPT are considered to be sufficiently valid from the perspective of basic system specifications, operability, and safety, and to reflect the design intent. The three patterns verified this time were verified for all transitions within the model, and it was also possible to verify whether the state is maintained under inappropriate conditions for each state, so it can be said that the verification is comprehensive. In the future, we would like to compare the state transition table and conduct a detailed verification of the comprehensiveness. On the other hand, there are no tests to check that the system does not fall into infinite loops or deadlocks, and it is difficult to check the overall progress and stability of the system based on the generated content.

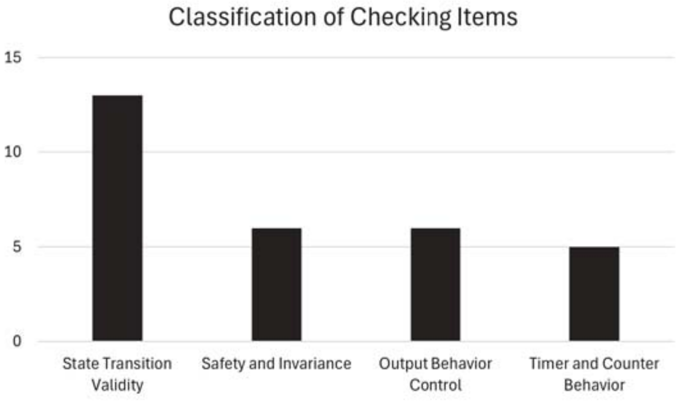


Figure 7: Classification of verification properties

4.3 Improvements to the Proposed Method

The current method involves creating a model as a state transition diagram from the requirements specification and using that state transition diagram as input. As a future plan, using the requirements specification itself as input is expected to prevent information leaks and interpretation discrepancies that often occur during the creation of state transition diagrams. When manually creating state transition diagrams from requirements specifications, there is a risk that the original requirements will not be accurately reflected in the diagram due to differences in developers' understanding and expression. In fact, Figure 6 shows one error caused by information leakage in the state transition diagram. By directly extracting state transitions from the requirements specification and automatically performing model checking, we believe it is possible to reduce manual work in the design process, contributing to improved development efficiency and reduced working hours. We consider this to be a significant advantage in software development, as it enables both quality assurance and optimization of development costs to be achieved simultaneously.

When performing model checking using the generated model checking code and verification properties, we judge whether the manually generated items are correct, but we believe that LLM can also be used for this task. Research is also being conducted on self-evaluation of LLM responses, which we believe will be an approach to ensure the reliability of LLM responses[20].

5 Conclusions

In this paper, we propose two methods: automatic generation of NuSMV code from PlantUML code and automatic generation of verification properties from the NuSMV code. These methods were applied in practice to generate NuSMV code and corresponding verification properties. Verification based on transition and timer operation conditions revealed several minor defects in the generated code, but it remained practically usable with slight adjustments. The generated verification properties were evaluated in terms of system specifications, operability, and safety, and were found to contain sufficient content for comprehensive verification. Their high quality is expected to improve the efficiency of the formal verification process. Thus, combining code and verification properties generation can reduce manual workload during the design phase.

One concern is the need for knowledge of CTL to evaluate the correctness of logical expressions produced by the method. To address this, future work aims to use LLMs for self-verification of verification properties, allowing the entire formal method process to be supported even by non-experts.

Furthermore, since ChatGPT has demonstrated the ability to derive state transitions from PlantUML and generate practical NuSMV code, future research will focus on deriving transitions directly from requirements specifications. We also plan to refine prompt design using Chain-of-Thought techniques, apply the method to other formal verification tools, and optimize the overall efficiency of formal methods.

acknowledgement

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

REFERENCES

[1] 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)

[2] Maiko Onishi, Shinpei Ogata, Kozo Okano, Daisuke Bekki: "Reducing Syntactic Complexity for Information Extraction from Japanese Requirement Specifications," Proceedings of 29th Asia-Pacific Software Engineering Conference (APSEC 2022) pp.387-396 (2022)

[3] Georges Morbé, Florian Pigorsch, Christoph Scholl: "Fully Symbolic Model Checking for Timed Automata," Lecture Notes in Computer Science (LNTCS), vol.6806, pp 616–632 (2011)

[4] Mengyan Zhao, Ran Tao, Yanhong Huang, Jianqi Shi, Shengchao Qin, Yang Yang: "NL2CTL: Automatic Generation of Formal Requirements Specifications via Large Language Models," Lecture Notes in Computer Science (LNCS), vol.15394, pp1-17, (2024)

[5] Anmol Nayak, Hari Prasad Timmapathini, Vidhya Murali, Karthikeyan Ponnalagu, Vijendran Gopalan Venkoparao, Amalinda Post: "Req2Spec: Transforming Software Requirements into Formal Specifications Using Natural Language Processing," Lecture Notes in Computer Science (LNCS), vol.13216, pp87-95, (2022)

[6] Maiko Onishi, Shinpei Ogata, Kozo Okano, and Daisuke Bekki: "Temporal relation identification in functional requirements," Proceedings of 27th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems, pp.1161-1170 (September 2023)

[7] Kozo Okano, Maiko Onishi, Jo Otsuka, Shinpei Ogata, Toshifusa Sekizawa, Keishi Okamoto, Daisuke Bekki: "A Bounded Model Checker for Timed Automata and Its Application to LTL Properties," Proceedings of 26th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems, pp.532-541 (2022)

[8] Ahmed R. Sadik, Siddhata Govind: "Model Generation with LLMs: From Requirements to UML Sequence Diagrams," 2024 IEEE 32nd International Requirements Engineering Conference Workshops (REW), vol.24, pp.291-300 (2024)

[9] Takeki Ninomiya, Maiko Onishi, Shinpei Ogata, Kozo Okano: "A Method for Automatic Creation of State Transition Diagrams from Software Requirement Specifications," Proceedings of International Workshop on Informatics 2024 (IWIN2024) pp.17-22 (2024)

[10] Qingxiao Tao, Tingrui Yu, Xiaodong Gu, Beijun Shen: "Unraveling the Potential of Large Language Models in Code Translation: How Far Are We?" arXiv preprint arXiv:2410.09812 (2024)

[11] Weixiang Yan, Yuchen Tian, Yunzhe Li, Qian Chen, Wen Wang: "CodeTransOcean: A Comprehensive Multilingual Benchmark for Code Translation," EMNLP 2023, pp 5067–5089 (2023)

[12] Zhiqiang Yuan, Yiling Lou, Mingwei Liu, Shiji Ding, Kaixin Wang, Yixuan Chen, Xin Peng: "Evaluating and Improving ChatGPT for Unit Test Generation," Proceedings of the ACM on Software Engineering, vol.1, No.76, pp.1703-1726 (2024)

[13] Arifa Islam Champa, Md Fazle Rabbi, Costain Nachuma, Minhaz F. Zibran: "ChatGPT in Action: Analyzing Its Use in Software Development," Proceedings of the 21st International Conference on Mining Software Repositories , pp.182-186 (2024)

[14] Tao Xiao, Christoph Treude, Hideaki Hata, Kenichi Matsumoto: "DevGPT: Studying Developer-ChatGPT Conversations," Proceedings of the 21st International Conference on Mining Software Repositories , pp.227-230 (2024)

[15] Pekka Abrahamsson, Tampere University, Tatu Anttila, Jyri Hakala, Juulia Ketola: "ChatGPT as a Fullstack Web Developer - Early Results," Agile Processes in Software Engineering and Extreme Programming – Workshops : pp.201-209 (2023)

[16] ChatGPT-OpenAI, https://openai.com/chatgpt

[17] Koki Shimokawa, Hiroya Ii, Maiko Onishi, Shinpei Ogata, and Kozo Okano: "Automatic Derivation of a Transition Model from a Japanese Requirement Specification under a Restricted Grammar," Proceedings of International Workshop on Informatics 2022 (IWIN2022) pp.13-20 (2022)

[18] White, Jules, Quchen Fu, Sam Hays, Michael Sandborn, Carlos Olea, Henry Gilbert, Ashraf Elnashar, Jesse Spencer-Smith, Douglas C. Schmidt:"A prompt pattern catalog to enhance prompt engineering with chatgpt," arXiv preprint arXiv:2302.11382 (2023)

[19] SESSAMI, electronic hot water pot, https://www.sessame.jp

[20] Jie Ren, Yao Zhao, Tu Vu, Peter J. Liu, Balaji Lakshminarayanan : "Self-Evaluation Improves Selective Generation in Large Language Models ," NeurIPS 2023 Workshops, PMLR 239 : pp.49-64 (2023)

Emotional Parameters for Evacuation Agents via Live Earthquake Stream Analysis

Kei Hiroi[†], Masami Shinoda[††], Kanae Matsui[††], Akihito Kohiga[‡], Sho Fukaya[‡‡], Yoichi Shinoda[§]
[†]Disaster Prevention Research Institute, Kyoto University, Japan
[††]Tokyo Denki University, Japan
[‡]Doshisha University, Japan
[‡‡]Suwa University of Science, Japan
[§]NTT Communications Co., Japan
`hiroi@dimsis.dpri.kyoto-u.ac.jp`

Abstract - Understanding how disaster-warning media shape viewers’ emotions is essential for designing messages that prompt timely information seeking and protective actions. This study proposes a data-driven framework for quantifying emotions evoked by earthquake-warning media using publicly available YouTube live-stream data. Frame-by-frame detection of alert-color areas, warning sounds, and text updates is synchronized with MiniLM-embedded comments to extract major emotions such as fear, surprise, information sharing, and calmness on a one-second timeline. The obtained parameters are intended to serve as inputs for emotion-aware evacuation-agent simulators and the design of multimodal disaster communications.

Keywords: Multimodal Emotion Analysis, Disaster Cyber-Physical System, Earthquake Live-Stream

1 Introduction

In recent years the frequency and human impact of diverse natural hazards—earthquakes, typhoons, floods, and landslides—have been increasing worldwide because of the compound effects of climate change, urbanization, and aging infrastructure[1]. By 2050 the urban population is projected to exceed six billion, concentrating in areas where multiple hazards such as floodplains and active fault zones overlap. In this multi-hazard risk society, physical phenomena, social structures, and information environments interact, making it indispensable to build a cyber-physical system (CPS) that comprehensively represents the entire disaster process to mitigate damage.

We have previously developed a “structured evacuation-simulation framework” that integrally handles disaster phenomena, urban infrastructure, evacuation behavior, and information services[**?**]. As shown in Fig.,1, the framework comprises four layers: a main loop, a set of modules, a block of psychological and environmental variables, and external coupling functions. Hydrodynamic flood analysis and meteorological warnings can be ingested sequentially from external simulators and reflected in agent decision making. However, the actual timing of evacuation initiation and information-seeking behavior is influenced not only by physical factors. Emotions such as fear and surprise that arise when people receive disaster information act as mediating variables that dynamically shift decision thresholds, so quantifying these emotions as model inputs is essential.

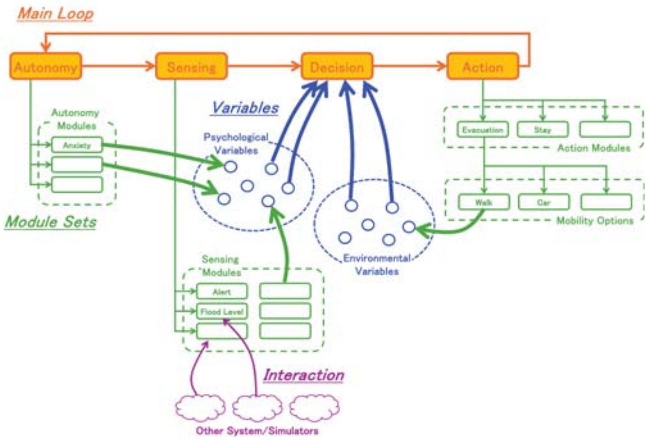


Figure 1: Structured evacuation simulator framework

Empirical knowledge of which emotions are elicited by which informational stimuli during disasters remains limited. In particular, few studies have simultaneously observed multimodal stimuli—video, warning chimes, and tickers—and viewer emotions, and parameterization in a form that can be embedded in existing simulators has not progressed.

This study aims to clarify the correspondence between disaster information presentation and viewer emotions by utilizing public data consisting of Earthquake Stream videos and live-chat comments (hereafter “comments”). Specifically, this paper outlines a method to examine (i) whether major viewer emotions—fear, surprise, information sharing, calmness—can be extracted using only public videos and comments, and (ii) how visual, auditory, and linguistic elements can be quantified as emotional stimuli. The analytical framework is intended to provide parameters for future integration into an urban-disaster cyber-physical system that incorporates emotional dynamics.

2 Related Works

2.1 Emotion Dynamics in Disaster-Oriented Multi-Agent Simulators

Disaster-time multi-agent simulation (MAS) models each evacuee as an agent and reproduces emergent crowd behavior from their interactions, thereby providing an experimental environment” indispensable for designing disaster-prevention plans and evacuation guidance strategies. Discoveries such as the 10-minute threshold after onset” that separates life and death in tsunami evacuation simulations and the counter-intuitive suggestion that temporarily turning back toward the tsunami" constitutes an optimal action demonstrate that insights from bottom-up models can revise real disaster-management protocols[3]. Nevertheless, the timing and selection of these actions are influenced more by victims' emotional states than by geographical conditions. Fear and anxiety consume attentional resources, narrow information processing, strengthen normalcy and confirmation biases, and delay evacuation initiation[4]. Conversely, if heightened panic spreads, even a slight delay can create crowd bottlenecks, leading to blocked exits and fall accidents[5][6]. In other words, emotion is a mediating variable that determines not only when to flee " and where to flee " but also how the crowd behaves, " and without quantifying emotion it is difficult to ensure the realism and policy relevance of simulations. Yet traditional MAS research has treated emotion insufficiently, hindered by the lack of objective behavioral data and by the computational cost of incorporating emotion and cognition into models[7][8].

To bridge this gap, recent studies have advanced in tandem models with richer emotional expressiveness" and mechanisms that reproduce the dynamics by which emotion influences decision making." The continuous valence–arousal space can encompass subtle fluctuations of emotion, but systematic alignment with real disaster data remains undeveloped[9]. Accordingly, methods that formalize psychological traits such as normalcy, conformity, and altruism to modulate evacuation speed and cooperative behavior, and deliberative agents based on the Belief–Desire–Intention (BDI) architecture that sequentially infer information $\rightarrow$ cognition $\rightarrow$ intention $\rightarrow$ action," have emerged, enabling explanatory representation of the emotion-generation process and its reflection in behavior[10]. Furthermore, generative AI agents centered on large language models (LLMs) have been suggested to reproduce complex social and emotional interactions through natural-language dialogue and introspection[11][12]. However, these advances simultaneously expose challenges such as black-box characteristics, controllability, and the reliability of synthetic data. Current research therefore focuses on combining methods for collecting multimodal behavioral and emotional data with techniques that visualize model inference processes in a verifiable form, aiming to balance expressive power" and "verifiability."

2.2 Emotion Modeling

To overcome the dichotomy between quantifying emotion and ensuring verifiability presented in the previous section, the key is how far emotion itself can be reduced to a form a computer can process. Early computational affect studies mapped the psychological "multi-layer theory" and "appraisal theory" directly onto agents and proposed a three-layer framework—sensorimotor $\rightarrow$ schema $\rightarrow$ conceptual—that generates emotions ranging from "reflexive surprise" to "abstract fear"[13]. Subsequently, the Ortony–Clore–Collins (OCC) model derived twenty-two emotion-elicitation conditions from combinations of belief $\times$ goal $\times$ plan and, by numerically expressing intensities and thresholds, enabled emotions to function as heuristics for action selection[14]. Affective Knowledge Representation (AKR) further superimposed stimulus checks such as "novelty" and "certainty" onto physiological axes such as valence, intensity, and duration, extending the probabilistic framework to handle "fluctuating emotions." These rule-/probability-based approaches can explicitly describe behavioral transitions such as plan abandonment and replanning, yet they leave challenges such as cultural differences, mixed emotions, and large-scale computational burden[15].

Attempts to bridge these gaps include a two-dimensional model that modulates evacuation behavior by fear intensity $\times$ duration and a set of norm-oriented agents that formalize a norm $\longleftrightarrow$ emotion loop mediated by shame and pride[16][17]. The former reproduced the nonlinear improvement in survival rates produced by fear contagion, and the latter quantified emotional mechanisms that suppress norm violations, demonstrating that emotion is not merely an internal state but a "mediating variable that drives collective emergence." However, data on mixed emotions in real disasters and culturally dependent expression patterns remain scarce, and model validation often relies on indirect comparison with behavioral indicators. Standardizing direct validation through human annotation and role-play experiments is therefore required[18].

Current attention is shifting toward generative modeling centered on large language models (LLMs). AgentSociety, which incorporates an LLM, reproduced the emergence of social norms among 10,000 agents in response to polarization and UBI policy shocks, and Project Riley prototyped five-emotion agents—"joy," "sadness," and others—that generate post-disaster support dialogues through cooperative and critical voting[19][20]. An LLM $\times$ MAS system achieved a strategic-reasoning accuracy of 88% in the Ultimatum Game, surpassing a standalone LLM, and a Stanford study reported 80–85% fidelity in mimicking the personalities and attitudes of 1,052 real individuals[21][22]. Nevertheless, LLM-driven models carry risks of black-box opacity, uncontrollability, and manipulation through fabricated emotions, so ensuring transparency and excluding synthetic artifacts through human-ground-truth cycles will be essential conditions for future work[21][23][24].

2.3 Extraction of Human Emotions and Intentions and Challenges

The issues concerning human emotions in disaster-time simulations discussed so far can be summarized as follows. First, the emotions incorporated into models are approximated by fixed parameters and therefore fail to adequately reproduce the emotions that actually arise in response to the presentation of disaster information. Second, although highly expressive models such as the OCC type and BDI/LLM families have emerged, parameter identification and theoretical grounding are difficult because of computational cost and black-box characteristics.

Regarding the extraction of human emotions and intentions, the richest window of observation is social networking services (SNS). Short-text SNS (Twitter/X) functions as a "digital early warning network" immediately after a disaster, and emotion analysis using SVM and LSTM/BERT has rapidly extracted polarity time series for fear, anger, joy, and other emotions[25][26]. Combining topic modeling allows the structuring of rescue requests and infrastructure damage, and research has evolved toward geospatial information extraction with LLMs and rumor detection that takes diffusion structures into account[27][28].

This study focuses on YouTube comments as a data source in which “stimulus” can be matched to “response.” Because the videos provide a common stimulus in the form of disaster information and the comments accumulate viewers’ emotions and behavioral intentions as a time series, it is highly likely that context-rich emotional parameters can be extracted. Nevertheless, existing disaster studies have limited YouTube to metadata statistics and information-quality evaluations, and no example has been found in which the comment set is systematically analyzed and linked to model inputs[29]. To establish a method that bridges crowd emotions and intentions to agent models by using comments tied to video content, this study extracts viewers’ emotional states from disaster videos and analyzes them in relation to visual and auditory stimuli.

3 Proposal

3.1 Overview

The objective of this study is to analyze Earthquake Stream videos and their accompanying comments in an integrated manner and to verify, in terms of temporal structure, how the presentation of disaster information influences viewers ’ emotional changes. Specifically, two research questions are set: first, whether the major viewer emotions—fear, surprise, information sharing, calmness, and so on—can be reliably extracted using only the public videos and comments (RQ1); second, to what extent each of the three types of informational stimuli—visual, auditory, and linguistic—elicits emotional responses in terms of time lag and intensity (RQ2).

The research flow comprises three stages. In the first stage, YouTube Earthquake Stream videos are decomposed into frames, and comments are simultaneously retrieved through the YouTube Data API. In the second stage, visual, auditory, and linguistic stimuli are extracted from the video, and the comment texts are clustered by morphological analysis and MiniLM embedding; the stimulus series and the emotion series are then aligned on a common timeline with a resolution of 1 s. In the third stage, the four resulting series (visual, auditory, linguistic, and emotion) are integrated to construct an analytical dataset that serves as the basis for testing the two research questions.

3.2 Data Sources

The video data analyzed in this study consist of archive recordings from “Japan Nationwide Earthquake Early Warning Live: The Real-time Earthquake Alert Channel for Japan (Tokyo) since 2012,” distributed by T5 Project Co., Ltd. We selected nine MP4 files that capture characteristic earthquakes occurring in 2025 and classified them, according to their seismic features, into five categories shown in Table 1: “medium scale (seismic intensity 5 lower),” “small scale (seismic intensity 4),” “perceptible shaking,” “distant earthquake,” and “earthquake in a previously affected area.” Each video was processed with OpenCV to obtain its frame rate and then divided into PNG images at intervals equivalent to 1 fps.

Comments were collected from the same live broadcasts by means of the YouTube Data API v3. Using the video IDs as keys, we acquired comments via the API and stored them in a format that allows alignment on a per-second timeline. This process is implemented as a script periodically executed on AWS Lambda, and the collected comment data are stored in JSON format on Amazon S3. For each archive video, the number of comments was aggregated by retrieving liveChatMessages through the YouTube Data API v3.

3.3 Information Elements Influencing Emotional Changes During Disaster-Information Acquisition

Next, to analyze how viewers ’ emotions change while obtaining disaster information, we aim to identify the elements of information expression that affect emotional change. Based on psychological findings accumulated in prior research, we examine how the presentation format of disaster information influences people ’ s emotions and behavior and clarify the categories of information expression to be analyzed in this study.

Disaster information not only conveys facts but is also known to affect the recipient ’ s emotions and psychological state. In particular, modes of expression such as color, sound, and text are positioned as important factors related to hazard perception and emotional responses during disasters. On the visual side, adopting hazard colors such as red and yellow increases perceived danger, and emphasizing those colors through display area or zoom operations dynamically shifts alert levels[30]. On the auditory side, emergency earthquake-warning chimes and sirens induce anxiety and tension and have been reported to affect arousal levels and the redistribution of cognitive resources[31]. Regarding linguistic expression, the amount of technical terminology and the tone of calls for action influence evacuation decisions, and multiple empirical studies have confirmed that using plain language improves recipients’ understanding and response[32][33].

In view of these findings, this study organizes the information elements that influence emotional changes during disaster-information acquisition into three categories: visual expressions (hazard colors and display area), auditory expressions (warning sounds), and linguistic expressions (text including numerical information, tone, and calls for action). Each category can be automatically extracted from the live videos and associated media targeted in this study, and their temporal correspondence with the emotion series will be examined in subsequent analyses.

3.4 Emotion Extraction from Earthquake Stream

To derive emotion indicators from the live chat accompanying the Earthquake Stream, the comment text is analyzed in four stages: time alignment, preprocessing, semantic vectorization, and clustering. First, the total number of comments

Table 1: Summary of archive videos

Category	Epicenter	Occurrence time	Number of comments
Medium scale (JMA Intensity 5 Lower)	Northern Nagano Prefecture	20:19	656
Small scale (JMA Intensity 4)	Hida Region, Gifu Prefecture	15:29	646
Small scale (JMA Intensity 4)	Amakusa and Ashikita Regions, Kumamoto Prefecture	05:00	598
Small scale (JMA Intensity 4)	Off the eastern coast of the Ōsumi Peninsula, Kagoshima Prefecture	23:04	636
Small scale (JMA Intensity 4)	Off Urakawa, Hokkaido	06:28	256
Small scale (JMA Intensity 4)	Off Kushiro, Hokkaido	17:37	714
Small scale (JMA Intensity 4)	Off Tokachi, Hokkaido	03:52	866
Earthquake in previously affected area	Around the Noto Peninsula	04:08	776
Distant earthquake	Central Myanmar	15:39	212

is aggregated on a per-second basis and overlaid on the same 1 fps timeline as the visual and auditory series, treating the overall volume of utterances as a direct response indicator. For text preprocessing, emojis, URLs, and symbols are removed; surface forms are enumerated by Japanese morphological analysis; and the text is normalized into whitespace-separated token sequences. The cleaned text is converted into 384-dimensional embeddings by the 12-layer MiniLM model, L2-normalized, and projected into two dimensions with UMAP (n_neighbors = 15, min_dist = 0). HDBSCAN (min_cluster_size = 8, min_samples = 2) is then applied to the projected coordinates to extract high-density regions as clusters. Micro sets with three or fewer points in the UMAP space are treated as noise and excluded, suppressing the influence of postings with low semantic coherence. For each cluster, the main words are extracted in order of frequency, and representative sentences are inspected to grasp emotional characteristics. The two resulting series—the total comment count and the super-cluster labels—are combined on the same timeline as the visual and auditory series and used as inputs for the analysis of emotional changes.

3.5 Analysis of Correspondence with Disaster-Information Presentation

In this section, we constructed a time-series dataset for evaluating the correspondence between the stimuli—visual, auditory, and linguistic stimuli obtained from the live video—and viewer responses aligned on the same time axis. First, the video was split at 1 fps, and the right half of each frame (map region) was converted to HSV space to extract hues of 0–35° (red, orange, yellow). The pixel ratio was defined as the “warning-color occupancy rate,” and the frame-to-frame difference Δr_t was calculated; frames satisfying $|\Delta r_t| > 0.005$ were regarded as visual change points. On the same 1 fps timeline, posting times of comments acquired via the YouTube Data API liveChatMessages were converted to integer seconds relative to the start of the video, and occurrences were aggregated per second.

The auditory stimulus was obtained by extracting a WAV file (16 kHz) from the video and scanning it with a 0.5 s window at the same hop size. After converting RMS to dB, a three-frame moving average was applied for smoothing. A dynamic threshold of median $+ 1\sigma$ was set, and the first frame exceeding it was marked as the onset of the warning sound (value 1). For linguistic information, two predefined ROIs (the lower-left ticker and the center of the map) in each frame were analyzed with Tesseract OCR to extract seismic intensity, magnitude, and depth. The moment a numerical value changed within consecutive frames was treated as an event: in the case of seismic intensity, a new value of 4 or higher was labeled 1; for magnitude, the label 1 was assigned when the change exceeded a threshold preset in code (2.0). These visual, auditory, and linguistic series, together with the per-second comment count, were merged to obtain an integrated time series for analyzing the temporal correspondence between disaster-information presentation and emotional responses (comments).

4 Conclusion

In this paper, we proposed a framework for quantifying emotion parameters derived from disaster-warning media, with a focus on live earthquake streams and their accompanying comments. The method integrates frame-level visual, auditory, and linguistic features with natural-language embeddings of live-chat data, enabling the extraction of dominant emotional states—such as fear, surprise, calmness, and information sharing—on a fine-grained temporal scale.

Two research questions were addressed conceptually in this study. RQ1 asked whether major emotions—fear, surprise, information sharing, calmness—can be reliably extracted from public footage and comments alone; RQ2 asked whether the time lag and intensity with which visual, auditory, and linguistic stimuli trigger emotional responses can be represented in a unified analytical framework. To examine these questions, we developed a data-alignment pipeline that synchronizes frame-level visual and auditory cues with text embeddings of comments on a one-second timeline. This multimodal framework provides a foundation for modeling the mechanisms by which disaster-information presentation influences emotional responses, offering parameters applicable to both media design and behavioral-model research.

The proposed findings are directly linked to the implementation of an emotion module in the structured evacuation-simulator framework. First, the extracted emotion parameters can be used to assign the initial proportions of fear and surprise according to seismic intensity and epicentral distance. Second, by incorporating warning-band and auditory cues as external stimuli and defining probabilistic transition matrices, the simulator can reproduce dynamic changes from fear to information sharing and calmness. This enables evaluation experiments on information-service design and alert presentations, extending beyond conventional evacuation-time measurement.

The proposed framework also contributes to establishing guidelines for emotion-driven disaster-information presentation. By linking visual, auditory, and linguistic elements of warning media with corresponding emotional parameters, it becomes possible to evaluate how different modalities influence perception and behavior in near real time. These insights can support the design of multi-modal alert systems that foster timely understanding and decision-making during disasters.

The present framework primarily targets online viewers in areas where communication infrastructure remains functional and therefore does not yet capture situations in large-scale disasters marked by power outages or network failures. Even so, quantifying “what survivors want to know at which stage” and “what triggers a shift from fear to information sharing” offers important clues for improving disaster communication. Moreover, the same approach can be extended to progressive hazards such as floods and typhoons; by tracking emotion transitions and information needs across phases of damage escalation, it can contribute to the development of time-series-aware disaster cyber-physical systems.

Acknowledgement

This work was supported by JST FOREST Program, Grant Number JPMJFR226Z, and the commissioned research(No. 0820103) by National Institute of Information and Communications Technology (NICT), Japan.

REFERENCES

[1] Intergovernmental Panel on Climate Change Fifth Assessment Report (AR5), Retrieved from url-https://www.ipcc.ch/report/ar5/ (accessed: 2025.4)

[2] Hiroi, K., Kohiga A., Fukaya, S., Shinoda, Y., A Structured Evacuation Simulator Framework for Federation Strategy during Flood Disasters, In Proceedings of 16th International Conference on Applied Energy (ICAE2024), 9pages, 2024.

[3] Hatayama, M., Nakai, F., Using Computer Simulation for Effective Tsunami Risk Communication. In: Yamori, K. (eds) Disaster Risk Communication. Integrated Disaster Risk Management. Springer, pp.39–50, 2020.

[4] Yashima, T., Yamasaki, J., Sera, A., Fujita, T., Iijima, T., Building an Agent Simulation Model with Personality and Bounded Processing Resource for Human-like Behavior, Proceedings of The 8th Annual Conference, Information Systems Society of Japan, pp.b2-2, 2012. (In Japanese)

[5] Xu, M., Xie, X., Lv, P., Niu, J., Wang, H., Li, C., Zhu, R., Deng, Z., Zhou, B., Crowd Behavior Simulation with Emotional Contagion in Unexpected Multihazard Situations, Proceedings of IEEE Transactions on Systems, Man, and Cybernetics: Systems, Vol.51, No.3, pp.1567–1581, 2019.

[6] Ta, X. H., Gaudou, B., Longin, D., Ho, T. V., Emotional Contagion Model for Group Evacuation Simulation, Informatica: an international journal of computing and informatics, Vol.41, No.2, pp.169–182, 2017.

[7] Tsurushima, A., Modeling Crowd Evacuation Behavior: New Approach without Assuming Internal Model of Human, Proceedings of The 39th Annual Conference of the Japanese Cognitive Science Society, pp.728–732, 2022. (In Japanese)

[8] Luo, L., Zhou, S., Cai, W., Low, M. Y. H., Tian, F., Wang, Y., Xiao, X., Chen, D., Agent-based Human Behavior Modeling for Crowd Simulation, Computer Animation and Virtual Worlds, Vol.19, No.3-4, pp.271–281, 2008.

[9] Van Haeringen, E., Gerritsen, C., Simulating Panic Amplification in Crowds via Density-Emotion Interaction, In Proceedings of The 2023 International Conference on Autonomous Agents and Multiagent Systems, pp.1895–1902, 2023.

[10] Lee, S., Son, Y. J., Jin, J., An Integrated Human Decision Making Model for Evacuation Scenarios under a BDI Framework, ACM Transactions on Modeling and Computer Simulation (TOMACS), Vol.20, No.4, pp.1–24, 2010.

[11] Tanabe, F., Hayashi, K., Yamazaki, S., Sato, R., Murakoshi H., Hayashi, H., Proposal of Disaster Countermeasures using AI and Verification of Effectiveness by MAS, Bulletin of Advanced Institute of Industrial Technology Vol 18, pp.165–172, 2025. (In Japanese)

[12] Haase, J., Pokutta, S., Beyond Static Responses: Multi-Agent LLM Systems as a New Paradigm for Social Science Research, arXiv preprint arXiv:2506.01839, 2025.

[13] Cami, A., Lisetti, C., Sierhuis, M., Accounting for Emotions in Multi-Agent Modeling and Simulation Systems, In Proceedings of9th International Conference on User Model, pp.22–26, 2003.

[14] Steunebrink, B. R., Dastani, M., Meyer, J. J. C., Emotions as Heuristics in Multi-agent Systems, In D. Reichardt, P. Levi, J.-J. Meyer (eds.), In Proceedings of

The 1st Workshop on Emotion and Computing, University of Bremen, pp.15–18, 2006.

[15] Pei, G., Li, H., Lu, Y., Wang, Y., Hua, S., Li, T., Affective Computing: Recent Advances, Challenges, and Future Trends, Intelligent Computing, Vol.3, pp.1–24, 2024.

[16] Van Minh, L., Adam, C., Canal, R., Gaudou, B., Tuong Vinh, H., Taillandier, P., Simulation of the Emotion Dynamics in a Group of Agents in an Evacuation Situation, In Proceedings of International Conference on Principles and Practice of Multi-Agent Systems, Springer, pp.604–619, 2010.

[17] Argente, E., Del Val, E., Perez-Garcia, D., Botti, V., Normative Emotional Agents: a Viewpoint Paper, IEEE Transactions on Affective Computing, Vol.13, No.3, pp.1254–1273, 2020.

[18] Van Haeringen, E. S., Veltmeijer, E. A., Gerritsen, C., Empirical Validation of an Agent-based Model of Emotion Contagion, IEEE Transactions on Affective Computing, Vol.15, No.1, pp.273–284, 2023.

[19] Piao, J., Yan, Y., Zhang, J., Li, N., Yan, J., Lan, X., Lu, Z., Zheng, Z., Wang, Y. J., Zhou, D., Gao, C., Xu, F., Zhang, F., Rong, K., Su, J., Li, Y., Agentsociety: Large-scale Simulation of LLM-driven Generative Agents Advances Understanding of Human Behaviors and Society, arXiv preprint arXiv:2502.08691, 2025.

[20] Ortigoso, A. R., Vieira, G., Fuentes, D., Frazao, L., Costa, N., Pereira, A., Project Riley: Multimodal Multi-Agent LLM Collaboration with Emotional Reasoning and Voting, arXiv preprint arXiv:2505.20521, 2025.

[21] Sreedhar, K., Chilton, L., Simulating Human Strategic Behavior: Comparing Single and Multi-agent LLMs, arXiv preprint arXiv:2402.08189, 2024.

[22] Stanford HAI, AI Agents Simulate 1,052 Individuals' Personalities with Impressive Accuracy, urlhttps://hai.stanford.edu/news/ai-agents-simulate-1052-individuals-personalities-impressive-accuracy (accessed: 2025.4)

[23] Vinay, R., Spitale, G., Biller-Andorno, N., Germani, F., Emotional Prompting Amplifies Disinformation Generation in AI Large Language Models, Frontiers in Artificial Intelligence, Vol.8, 2025.

[24] Su, Z., Zhou, X., Rangreji, S., Kabra, A., Mendelsohn, J., Brahman, F., Sap, M., Ai-liedar: Examine the Trade-off between Utility and Truthfulness in LLM Agents, arXiv preprint arXiv:2409.09013, 2024.

[25] Vayansky, I., Kumar, S. A., Li, Z., An Evaluation of Geotagged Twitter Data during Hurricane Irma using Sentiment Analysis and Topic Modeling for Disaster Resilience, In Proceedings of 2019 IEEE International Symposium on Technology and Society (ISTAS), pp.1–6, 2019.

[26] Chaudhary, V., Goel, A., Yusuf, M.Z., Tiwari, S., Disaster Tweets Classification Using Natural Language Processing, In: Simic, M., Bhateja, V., Azar, A.T., Lydia, E.L. (eds) Smart Computing Paradigms: Advanced Data Mining and Analytics, Lecture Notes in Networks and Systems, vol.1262, Springer, 2025.

[27] Han, Y., Liu, J., Luo, A., Wang, Y., Bao, S., Fine-Tuning LLM-Assisted Chinese Disaster Geospatial Intelligence Extraction and Case Studies, ISPRS International Journal of Geo-Information, Vol.14, No.2, 2025.

[28] Pattanaik, B., Mandal, S., Tripathy, R. M., A Survey on Rumor Detection and Prevention in Social Media using Deep Learning, Knowledge and Information Systems, Vol.65, No.10, pp.3839–3880, 2023.

[29] Adimoolam, M., Sugumaran, M., Rajesh, R. S., Efficient Encryption Algorithm for Video Data Storage, International Journal of Information and Computing Science, Vol.5, pp.41–49, 2018.

[30] Security and resilience – Emergency management – Guidelines for colour- coded alert, INTERNATIONAL STANDARD ISO 22324, Second edition, 2022.

[31] Nishimura, R., Asaoka, S., The Effects of Alert Sound Reappraisal on Selective Attention, Departmental Bulletin Paper, Edogawa University, Vol.30, pp.349-356,2020.

[32] Iroquois Healthcare Association, Hospital Emergency Codes: Standardization and Plain Language, 2020.

[33] Taylor, R., J., Wogalter, S., M., Specific evacuation instructions enhance spoken fire warnings. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Vol.59, No.1, pp.1486-1490, 2015.