June 3–7, 2024
Minato-ku, Tokyo, Japan

Association for
Computing Machinery

Advancing Computing as a Science & Profession

BTIW2024

Proceedings of the 2024

Behavior Transformation by IoT International Workshop

Sponsored by:

ACM SIGMOBILE in cooperation with ACM SIGOPS

Part of:

ACM MobiSys 2024

Contents

Basic Study on Generation of Dopamine-Secretion-Promoting Intrinsic Motivational Messages Using ChatGPT

Reiko Aruga
Nippon Telegraph and Telephone Corporation
Yokosuka, Japan
reiko.aruga@ntt.com

Yurika Katagiri
Nippon Telegraph and Telephone Corporation
Yokosuka, Japan
yurika.katagiri@ntt.com

Junji Watanabe
Nippon Telegraph and Telephone Corporation
Yokosuka, Japan
junji.watanabe@ntt.com

ABSTRACT

People sometimes put off an action because they do not perceive it as worth doing immediately, despite recognizing they should do it. Procrastination, especially in medical and healthcare fields, can lead to serious consequences, and it is an important social issue to promote cognitive and behavioral change for those reluctant to practice health behaviors. One important factor that can positively influence the psychology of procrastination is intrinsic motivation. Dopamine is known as the primary neurotransmitter associated with motivation and has been shown to relate to positive emotions and cognition. In this study, we defined personal experiences related to emotions and cognitions occurring during dopamine secretion as the person's original favorite activities (FAs). We then attempted to generate intrinsic motivational messages expected to produce dopamine secretion by partially replacing or modifying the activities the subject should do but were disinclined to do (reluctant activities (RAs)) with FAs. Preliminary experiments using questionnaire evaluations showed that dopamine-secretion-promoting intrinsic motivational messages may induce more excited feelings than general motivational messages.

CCS CONCEPTS

• **Human-centered computing → Natural language interfaces**; • **Applied computing → Consumer health**.

KEYWORDS

Intrinsic Motivation, Behavior Change, Dopamine, Prompt Engineering, Large Language Models

ACM Reference Format:
Reiko Aruga, Yurika Katagiri, and Junji Watanabe. 2024. Basic Study on Generation of Dopamine-Secretion-Promoting Intrinsic Motivational Messages Using ChatGPT. In *Behavior Transformation by IoT International Workshop (BTIW '24), June 3–7, 2024, Minato-ku, Tokyo, Japan.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3662008.3662011

BTIW '24, June 3–7, 2024, Minato-ku, Tokyo, Japan

ACM ISBN 979-8-4007-0665-3/24/06…$15.00
https://doi.org/10.1145/3662008.3662011

1 INTRODUCTION

Even when people have knowledge of actions they should take to achieve a desired outcome, they sometimes procrastinate taking those actions if they cannot visualize the actions in their own lives or if they lack proper motivation. Such procrastination is observed in various aspects of human life, and its factors have been studied for many years [22]. In particular, procrastination in medical and healthcare fields can lead to serious consequences. Delaying actions related to disease treatment and prevention leads to disease onset and severity, affecting individuals' well-being, but also increasing healthcare costs. In Japan, lifestyle-related diseases like hypertension and dyslipidemia accounted for approximately 50% of deaths and 30% of general medical care costs in 2021 [13], making health education that motivates individuals to lead a healthy lifestyle a social issue.

To solve this problem, Japan mandates specific health guidance, aimed at eliminating visceral fatty obesity, for insured persons and their dependents aged 40 and older[12]. However, a study evaluating the impact of this guidance reported that the group receiving it showed significantly lower BMI (both men and women) and abdominal circumference (men only) than the control group, but no beneficial effects on cardiovascular disease risk factors[21]. The reductions in BMI and abdominal circumference were minimal, about -0.1 kg/m^2 and -0.36 cm, respectively, for both men and women, suggesting that lifestyle improvements were insufficient to affect risk factors. In situations without physical pain or social inconvenience, individuals may not perceive the immediate value of changing their lifestyle for health reasons, leading them to postpone action. In this study, we focused on the function of dopamine (DA) , a neurotransmitter, in designing an intrinsic motivational message generation system to change these psychological states related to health behavior and promote behavior change.

Dopamine is a neurotransmitter that plays a central role in motivation [1], influencing decision-making processes by dynamically assessing the value of investing limited internal resources like energy, attention, and time [2]. Moreover, dopamine is linked with positive emotions and cognitions such as enjoyment [8], affection [26], and reward expectations [25]. These neuroscience discoveries, indicating that motivational drive and pleasant emotions stem from the same neurotransmitter, align with various psychological findings emphasizing the significance of pleasant factors in motivating and reinforcing behavior [23][5][6].

The first idea of this study is to generate dopamine-secretion-promoting intrinsic motivation messages by capturing personal experiences of positive emotions and cognitions related to dopamine secretion as the person's original favorite activities (FAs) and using

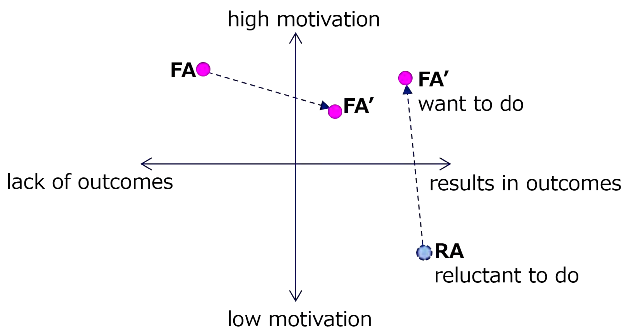

Figure 1: Motivation Outcome Map

these as input. Specifically, FAs and "activities that the user understands he/she should do but is unwilling to do" (reluctant activities (RAs)) are obtained through user input. These are incorporated into the prompts of ChatGPT [16], and new activities are generated by partially replacing or modifying RAs with FAs. Although this study does not make a biological measurement of actual dopamine secretion in the brain, we call the messages generated by this method "dopamine-secretion-promoting intrinsic motivational messages" with the expectation that these messages will promote dopamine secretion in this paper. Since an FA is a favorite activity that users already do in their own lives, messages incorporating it can generate more concrete images to practice in their current lives compared with general health information applicable to everyone (e.g., use the stairs instead of the escalator). In addition, it is expected that positive emotions and perceptions will be induced, and motivation will increase.

To generate messages that evoke more physical sensations during dopamine secretion, eliciting pleasant activities from users in a format likely to evoke physical sensations is necessary. The Japanese language has many mimetic words and interjections, known as embodied emotional expressions (EEEs) [15], that evoke emotional experiences as physical sensations. The second idea of this study was to design an FA acquisition questionnaire that would easily evoke bodily sensations, using EEEs to express pleasant emotions and cognitions corresponding to dopamine secretion-related behaviors.

In this paper, we first propose the concept of a motivation outcome map in Section 2 and present its target in this study. In Section 3, we organize the relationship between motivation and pleasant factors on the basis of previous psychology and neuroscience research. We also present previous research on behavior change support by generative AI. In Section 4, we propose a dopamine secretion-promoting message generation method and describe the details of the message generation process and questionnaire design. Section 5 describes preliminary experiments and results to evaluate the feasibility. The results of the preliminary experiments indicate that the dopamine-secretion-promoting motivational messages may evoke more feelings of excitement than general motivational messages.

2 MOTIVATION OUTCOME MAP

In this paper, we propose a motivation outcome map that shows the relationship between motivation and outcome (Figure 1). In this section, the motivation outcome map is used to explain the target of this study.

In the motivation outcome map, the objective outcome and user's motivation for the target behavior is plotted on the horizontal and vertical axes, respectively. The first quadrant means "I want to do it, and the outcome is obtainable." The second quadrant means "I want to do it, but the outcome is not obtainable or a negative outcome is obtainable." The third quadrant means "not wanting to do it and not obtaining an outcome or obtaining a negative outcome." The fourth quadrant means "not wanting to do it but obtaining an outcome." The second and fourth quadrants have the largest intervention effects, while the first and third quadrants are outside the scope of this study.

An example of the first quadrant is a situation where a person likes to jog and it is good for their health. In this domain, the person voluntarily engages in the behavior, and the behavior has a positive impact on the person, so there is little need for the system to intervene. An example of the third quadrant is that a person refrains from smoking because he/she does not like cigarettes. In this domain, the person voluntarily avoids the behavior, leading to a positive impact, so there is no need for the system to intervene.

An example of the second quadrant is a case where a person likes to read, but reading has a low outcome in terms of physical health activity. When building a system that aims to increase the outcome of a FA, we would propose an FA' with a larger outcome by partially replacing the FA. Behaviors such as reducing alcohol consumption or quitting smoking fall under this domain.

Finally, an example of the fourth quadrant is a situation where a person does not like to exercise but understands he/she should exercise three times a week for his/her health. The target of the proposed method considered in this paper is in this domain, and the goal of this research is to move to the first quadrant by proposing a new activity FA' that incorporates the person's exciting elements into the target behavior RA in the fourth quadrant.

3 RELATED STUDIES

3.1 Psychological findings: Procrastination and intrinsic motivation

A large meta-analytic study of procrastination [22] has shown that strong and consistent predictors of procrastination include self-efficacy and impulsivity, and has proven to be consistent with temporal motivation theory (TMT) [23], an integrated representation of various factors related to motivation. TMT is a simple formulation of an individual's motivation for a task, as shown in the following equation, and is considered applicable to explain procrastination.

$$Motivation = \frac{Expectancy \times Value}{1 + Impulsiveness \times Delay}$$

According to TMT, "Motivation" (the desire for a particular outcome) is enhanced by "Expectancy" (the probability of success or self-efficacy) and "Value" (the reward associated with the outcome). "Motivation" is also reduced by "Impulsiveness" and "Delay" (the period of time until a reward or result is obtained). Impulsivity refers to the tendency to be uncomfortable with the time it takes to obtain results and to give priority to rewards that can be obtained in a short period of time. When a person does not feel any immediate physical pain or social inconvenience, as in the case of lifestyle-related disease prevention, "Motivation" tends to decrease because the numerator of TMT, "Value," is small and the denominator, "Delay," is very long or the time of achievement is unclear (such as the time until the numerical values of health checkup results reach normal values).

In behavior change, both temporarily increasing and maintaining motivation until the outcome is obtained is important. Recent studies on motivation suggest that intrinsic motivation based on internal factors such as curiosity is more effective in sustaining long-term behavior than extrinsic motivation based on external factors such as rewards [9]. The Intrinsic Motivation Inventory (IMI) [5] is a measure of intrinsic motivation. The scale has been evaluated in several experiments, and the results have validated its six subscales: "interest/enjoyment," "value/usefulness," "perceived competence," "perceived choice," "effort," and "felt pressure/tension."

3.2 Findings of Cognitive-Behavioral Therapy: Behavioral Activation Method

Various cognitive-behavioral therapy methods for depression have been proposed, among which behavioral activation therapy stands out for its effectiveness, comparable with comprehensive cognitive therapy [6]. This method gradually introduces activities that are pleasurable to the patient on the basis of the Pleasant Events Schedule, a list of activities that induces pleasant feelings in daily life. It is described as a component of cognitive-behavioral therapy for depression in the Mental Health Gap Action Program (mhGAP) launched by the World Health Organization (WHO) and is recommended as a treatment option for medically unexplained complaints [17]. Although this study does not target mental health care, the "somewhat unmotivated" situation that is the target of this study can be considered a mental condition similar to a medically unexplainable complaint (i.e., feeling somewhat unwell), and focusing on pleasant activities can be expected to be effective. Including pleasant activities increases "Value" in TMT and also enhances "interest/enjoyment " in IMI, strengthening intrinsic motivation.

3.3 Neuroscience findings: dopamine

Dopamine is a neurotransmitter that plays a central role in reward processing, attention, and motivation [1]. Multiple neural pathways through which dopamine is secreted have been identified, and the effects of dopamine on behavior vary depending on the neural pathway activated. In all cases, dopamine dynamically estimates whether it is worth spending limited internal resources such as energy, attention, and time [2]. In recent years, studies on dopamine measurements in the human brain have also been conducted [20]. Mainly using positron emission tomography (PET), these studies have shown, for example, that playing video games increases dopamine levels in the striatum, correlating with game performance [8]. Another study has shown that activation of dopamine nerves in the medial orbitofrontal cortex correlates with the intensity of heightened feelings when viewing pictures of loved ones [26]. Individual differences in dopamine receptor density have also been observed, linked to variances in human behavioral tendencies. For example, higher density of dopamine D1 receptors in the striatum correlates with a propensity to subjectively perceive low probabilities, such as winning the lottery, as high [25]. Thus, dopamine is related to emotions/cognitions such as pleasure, curiosity, infatuation, motivation, and achievement in humans.

Although a review article conceptually analyzes motivational applications for depressed patients, citing the effects of dopamine [14], and laboratory experiments suggest it is possible to spontaneously activate brain regions involved in motivation (VTA) using neurofeedback training [11][24], no studies have yet integrated neuroscience findings into the design of systems intended for intrinsic motivation.

3.4 Prior Research on Automatic Generation Technology to Promote Behavior Change

With the remarkable acceleration of large-scale language model (LLM) research in recent years, numerous studies have reported on using LLMs to support cognitive change [7][10] and behavioral change [3][19][4]. These studies have shown the potential for LLMs to assist cognitive therapy professionals [7], suggesting that LLMs may generate more effective messages [10] that promote cognitive and behavioral change than humans. Several AI chatbots targeting health behaviors have also been proposed and reported to be effective in supporting weight loss [3], improving sleep [19], and increasing physical activity [4]. Although personalization features have been proposed in these studies of behavior change support and motivation, they are based on individual personality traits and behavior patterns, and none, to our knowledge, consider personalization on the basis of neuroscience findings.

4 PROPOSED METHOD

4.1 Outline of the proposed method

The proposed method generates an intrinsic motivation message to stimulate DA secretion by partially replacing or modifying an RA that the user understands he/she should do but feels unwilling to do with an FA that the user originally likes. The process flow is as follows: First, the user inputs the FA and RA through a GUI. Then, by substituting the obtained FA and RA into the pre-designed ChatGPT prompt, a DA secretion-promoting motivational message is output. In the following sections, we describe the design of the questionnaire and the prompts in detail.

4.2 Questionnaire design

In designing the questionnaire to obtain FAs likely to induce DA release, we first surveyed neuroscience literature to identify experimental tasks (specific activities) known to induce dopamine release in humans. In studies of brain dopamine measurements, subjects are given specific tasks in a laboratory. Experimental tasks in studies with healthy subjects, which are strongly related to motivation, dealt with video games [8], financial rewards [27], and looking at pictures of loved ones [26]. We then mapped these dopamine-secreting activities to the corresponding emotions, as shown in Table 1. Here, experiences eliciting these emotions in users would likely trigger dopamine secretion. The idea of this study was that by

Table 1: Behavioral-Emotional-EEE Correspondence Chart

Activities related to dopamine secretion	Emotion/Cognition	EEEs
Playing video game	curiosity, vitality	waku-waku , uki-uki, iki-iki
	enjoyment, pleasure	waa, waRi
	motivation, sense of achievement	yoshi, yoQshaa, shakiiN
See a photo of lover	affection, ecstatic	kyuN, uQtoLi
	peace of mind, relaxed	hoQkoLi
financial rewards	exectations, thrills	waku-waku

Table 2: FA-RA Acquisition Questionnaire

Acquired items	Questionnaire
FA	Please tell us about your favorite activities that make you feel the following emotions. 1. waku-waku, uki-uki, iki-iki 2. waa, waRi 3. kyuN, uQtoLi, hoQkoLi 4. yoshi, yoQshaa, shakiiN
RA	What activities, exercise habits, or eating habits do you think you should do but are unwilling to do?

Table 3: Message Evaluation Questionnaire

	The following message is an idea for an RA (Note: RA with user input). Please see the message and answer questions 1 and 2.
Q.1	Please select the ones you are interested in or would like to try.
Q.2	Please rank the top 5 most exciting messages.

incorporating elements of such FAs into the RA, dopamine secretion would be more likely, thus enhancing motivation.

While direct questions (e.g., "Tell me about an activity you enjoy") could be used to elicit FAs from users, the goal of this study was to evoke the physical sensation of "trying." Japanese speakers often use mimetic words (e.g., "doki-doki" for excitement) or interjections (e.g., "waRi" for joy) rather than adjectives (e.g., English "excited") to convey detailed emotional nuances in daily conversation. Such expressions, called embodied emotional expressions (EEE) [15], are prevalent in Japanese. In addition, several studies have shown that EEE have can evoke emotional experiences as physical sensations. For example, listening to mimetic words for pain sounds (e.g., "zuki-zuki"), activates the anterior cingulate cortex, a brain region related to pain, in Japanese subjects [18]. We considered that it would be effective to use EEE to express emotions to create a questionnaire to obtain FAs that evoke physical sensations. We referenced a detailed and large-scale study on EEE and corresponding emotions [15] to express the emotions we had mapped to dopamine secretion-related behaviors in the previous section. The FA-RA acquisition questionnaire, shown in Table 2, was designed accordingly.

4.3 Prompt design

In designing the prompts for ChatGPT, we iteratively adjusted them through trial and error to ensure that the intended DA secretion-promoting motivational messages were output by assigning individually obtained FA and RA to the parameters within the same prompt. While general advice on RA (general motivational messages) may initially be output as intermediate products, these messages also contain content that increases "value/usefulness" and "perceived competence" in the IMI. Therefore, it is highly likely that users will also prefer these general motivational messages. The primary objective of the proposed method is to enhance intrinsic motivation by strengthening DA-related pleasant factors, including "interest/enjoyment" in accordance with the IMI.

5 PRELIMINARY EXPERIMENTS

The process of behavior change can be broadly divided into two steps: cognitive change as step 1, followed by behavioral change as step 2. Our goal is, of course, behavioral change, but the purpose of this study is to first confirm whether the proposed method changes impressions of RA in a positive direction. In other words, this study is to confirm whether cognition is transformed, and whether behavior is transformed is outside the scope of this study. In the preliminary experiment presented in this section, we will test the hypothesis that DA-secretion-promoting motivational messages promote the emotion "excitement" more than general motivational messages. Since the subject of the evaluation was the change in cognition, the evaluation method was self-report.

5.1 Procedures for generating evaluation messages and questionnaire

To test the hypotheses, a message evaluation questionnaire was developed for each participant using the following procedure. First, an FA-RA acquisition questionnaire was distributed to each participant to obtain RA and FA. The experimenter then generated 10 general motivational messages for each participant using RA as the control condition. Next, 10 DA secretion-promoting motivational messages were generated for each participant using FA and RA. Thus, a total of 20 evaluation messages per participant were generated. Since the RA and FA are different for each participant, the evaluation messages are also all different as well. The method of generating each message is described in Section 4. Following visual inspection by the experimenter, messages that were evidently inappropriate in content were excluded, and additional messages were generated and adopted to ensure at total of 10 messages for each category.

Table 4: Sample of Generated Messages

ID	category	Example of generated message
1	Gen.	There are also possible exercises that can be done in the home while working from home. For example, you can easily incorporate aerobic exercise by taking the stairs.
...	...	...
10	Gen.	Before working late into the night, introduce a short yoga session to prepare your mind and body for work.
11	DA	Take advantage of the stairs at your favorite nature spot. Training in a new location will refresh your mind and body.
...	...	...
20	DA	Practice stress-relieving yoga while watching yoga videos of beautiful actresses. You can pursue beauty and health at the same time.

Table 5: Results of Preliminary Experiment

	1st	2nd	3rd	4th	5th
s1	DA*	Gen.*	Gen.*	DA	Gen.
s2	DA*	DA*	DA*	DA*	Gen.*
s3	DA*	DA*	Gen.	Gen.	DA
s4	DA	DA*	Gen.*	Gen.*	Gen.*
s5	DA*	DA*	DA*	DA*	DA*
s6	Gen.*	Gen.*	DA*	DA	DA*

Several generated messages were adjusted in length by the experimenter while preserving the content. A list of 20 randomly sorted messages was generated, and a message evaluation questionnaire consisting of the questions shown in Table 3 was created.

Examples of general motivational messages (denoted as "Gen.") and DA secretion-promoting motivational messages (denoted as "DA") generated by the system are shown in Table 4. These messages were generated on the basis of the user input of "seeing beautiful actresses and models and taking walks in nature" for FAs and "increasing daily activity" for RAs.

5.2 Participants and experimental procedure

In order to evaluate a prototype of message generation using the proposed method and to obtain future design guidelines, the participants were experts in HCI research. Therefore, the participant size of the experiment was small, and the participants were two women and four men (mean age: 36.5 years). First, the participants received the FA-RA questionnaire electronically from the experimenter, filled in their answers, and returned it to the experimenter. On the basis of the responses obtained from the FA-RA questionnaire, the experimenter generated evaluation messages and created a message evaluation questionnaire in accordance with the procedure described in Section 4. Participants received the message evaluation questionnaire electronically from the experimenter, filled in their answers, and returned it to the experimenter. The message evaluation questionnaire included a comment section for each message, which provided in-depth comments on the prototype from an expert's perspective. Due to the manual effort involved in generating the evaluation messages and creating the message evaluation questionnaires, the series of questionnaire responses were not immediate.

5.3 Results and Discussion

The results of the message evaluation questionnaire are shown in Table 5. It summarizes the order in which participants s1 to s6 responded to Question 2 as "exciting" and categorizes the corresponding message as "DA" for dopamine secretion-promoting motivational messages or "General" for general motivational messages. Those with an asterisk indicate that the respondent answered "would like to try" in Question 1, and those without an asterisk indicate that the respondent answered "interested" only.

Five out of six participants chose the DA secretion-promoting motivational message as the message that evoked the emotion of "excitement" in the first place, and four out of six participants chose it in the second place as well. Five out of six participants responded "I would like to try it" for the message selected in the first place, and all participants responded "I would like to try it" for the message selected in the second place. These results indicate that partially replacing or modifying an FA on the basis of personal experiences with an activity that is "exciting" and that the individual "wants to do" can effectively replace an RA, which is initially met with reluctance. The reason that many subjects selected general motivational messages from the third place onward is that their content may have been specific enough to recognize "value/usefulness" and "perceived competence" in the IMI. This is believed to have led to a sense of "this is going to be useful for me" (value/usefulness) or "I can do this well" (perceived competence), consequently promoting the emotion of "excitement."

According to the results of Question 1, an average of 3.16 "I want to try it" responses were received by all subjects for the 10 DA secretion-promoting motivational messages, and an average of 3.5 for the 10 general motivational messages. While the DA-secretion-promoting motivational messages elicited the most excitement in five out of six participants, which was in line with the design intent, the number of "I want to try it" responses was low, comprising about 30% of the total number of messages generated.

Positive comments from participants regarding the generated messages included "I thought it was intuitively good" (s1) (s2) and "I felt certain that I would naturally increase my physical activity without changing anything in my conventional behavior because it is something I like" (s5), suggesting that the conversion of RA using FA was able to generate acceptable content. There were also comments such as "I never thought of that idea" (s6) and "This idea is new to me" (s1), indicating the possibility of providing new ideas to users. In response to some messages, participants commented "I have done this in the past" (s3) and "I am already practicing this" (s4), suggesting that highly feasible action suggestions were made.

On the other hand, negative comments regarding the message were "I am interested in it, but I wonder how effective it will be" (s6), indicating the need to provide evidence. As comments related to disincentive to the activity, some of the respondents indicated their difficulty with the activity, such as "It is hard for me to get up early in the morning, so I don't think I can do this" (s4) (s5), and weather-related factors, such as "I would like to try it, but it is cold right now, so I don't want to do it" (s2). Participants who suggested activities to do with someone else responded, "I would like to do it, but it is difficult because we live far away from each other" (s4) and "I have tried to do it before, but it is difficult to make a schedule" (s5). These results suggest that the difficulty level of practicing a behavior that is dependent on the partner's situation is high. These factors need to be addressed to increase the probability of generating more "I want to try" and "exciting" messages and promote behavior change.

6 CONCLUSION

In this paper, we proposed a motivational outcome map as a guiding tool for designing systems to control motivation for actions leading to desired outcomes. To implement message generation that replaces an RA, an activity that one understands one should do but feels disinclined to do, with an activity that generates a physical sensation of "let's do it," we organized individuals' pleasant activities, FAs, on the basis of neuroscience findings, and designed a questionnaire to evoke a physical sensation. The results showed that messages generated by ChatGPT with prompts that included FAs and RAs as parameters tended to evoke the emotion of "excitement." In the future, we will further examine the questionnaire and prompts to output messages that evoke excitement but also a desire to try with high accuracy by reflecting the expert opinions obtained in this study. We will also conduct a long-term evaluation to determine whether the messages generated by the proposed method will not only transform cognitions about RA but also change behavior.

REFERENCES

[1] Getinet Ayano. 2016. Dopamine: Receptors, Functions, Synthesis, Pathways, Locations and Mental Disorders: Review of Literatures. *Journal of Mental Disorders and Treatment* 2 (01 2016). https://doi.org/10.4172/2471-271X.1000120

[2] Joshua Berke. 2018. What does dopamine mean? *Nature Neuroscience* 21 (05 2018). https://doi.org/10.1038/s41593-018-0152-y

[3] Han Shi Jocelyn Chew. 2022. The Use of Artificial Intelligence–Based Conversational Agents (Chatbots) for Weight Loss: Scoping Review and Practical Recommendations. *JMIR Med Inform* 10, 4 (13 Apr 2022), e32578. https://doi.org/10.2196/32578

[4] Christopher Cushing, Carolina Bejarano, Adrian Ortega, Nick Sayre, David Fedele, and Joshua Smyth. 2021. Adaptive mHealth Intervention for Adolescent Physical Activity Promotion. *Journal of Pediatric Psychology* 46 (01 2021), 536–546. https://doi.org/10.1093/jpepsy/jsaa125

[5] Center for Self-Determination Theory. [n. d.]. *Intrinsic Motivation Inventory (IMI).* Retrieved April 4, 2024 from https://selfdeterminationtheory.org/intrinsic-motivation-inventory/

[6] Eric Gortner, Jackie Gollan, Keith Dobson, and Neil Jacobson. 1998. Cognitive–Behavioral Treatment for Depression: Relapse Prevention. *Journal of consulting and clinical psychology* 66 (04 1998), 377–84. https://doi.org/10.1037/0022-006X.66.2.377

[7] Kenta Izumi, Hiroki Tanaka, Kazuhiro Shidara, Hiroyoshi Adachi, Daisuke Kanayama, Takashi Kudo, and Satoshi Nakamura. 2024. Response Generation for Cognitive Behavioral Therapy with Large Language Models: Comparative Study with Socratic Questioning. *ArXiv* abs/2401.15966 (2024), 1–17. https://api.semanticscholar.org/CorpusID:267311710

[8] Matthias Koepp, Roger Gunn, Andrew Lawrence, Vincent Cunningham, A Dagher, Terry Jones, David Brooks, C.J. Bench, and P.M. Grasby. 1998. Evidence for striatal dopamine release during a video game. *Nature* 393 (06 1998), 266–8. https://doi.org/10.1038/30498

[9] Lisa Legault. 2016. *Intrinsic and Extrinsic Motivation.* Springer International Publishing, Cham, 1–4. https://doi.org/10.1007/978-3-319-28099-8_1139-1

[10] Sue Lim and Ralf Schmälzle. 2023. Artificial intelligence for health message generation: an empirical study using a large language model (LLM) and prompt engineering. *Frontiers in Communication* 8 (2023), 1–15. https://doi.org/10.3389/fcomm.2023.1129082

[11] Jeff J. MacInnes, Kathryn C. Dickerson, Nan-kuei Chen, and R. Alison Adcock. 2016. Cognitive Neurostimulation: Learning to Volitionally Sustain Ventral Tegmental Area Activation. *Neuron* 89, 6 (3 2016), 1331–1342.

[12] Welfare Ministry of Health and Japan Labour. 2009. *Annual Health, Labour and Welfare Report 2008-2009.* Retrieved April 3, 2024 from https://www.mhlw.go.jp/english/wp/wp-hw3/02.html

[13] Welfare Ministry of Health and Japan Labour. 2021. *Annual Health, Labour and Welfare Report 2021.* Retrieved April 3, 2024 from https://www.mhlw.go.jp/english/wp/wp-hw14/dl/01e.pdf

[14] Stephane Mouchabac, Redwan Maatoug, Ismael Conejero, Vladimir Adrien, Olivier Bonnot, Bruno Millet, Florian Ferreri, and Alexis Bourla. 2021. In Search of Digital Dopamine: How Apps Can Motivate Depressed Patients, a Review and Conceptual Analysis. *Brain Sciences* 11, 11 (nov 1 2021), 1454.

[15] Aiko Murata, Yizhen Zhou, and Junji Watanabe. 2024. Embodied emotional expressions for intuitive experience sampling methods: A demographic investigation with Japanese speakers. *The International Journal of Wellbeing* 14, 1 (2024). https://doi.org/10.5502/ijw.v14i1.2413

[16] OpenAI. 2022. *ChatGPT.* Retrieved April 4, 2024 from https://openai.com/chatgpt

[17] World Health Organization. 2019. *mhGAP Intervention Guide - Version 2.0.* Retrieved April 4, 2024 from https://www.who.int/publications/i/item/9789241549790

[18] Naoyuki Osaka, Mariko Osaka, Masanao Morishita, Hirohito Kondo, and Hidenao Fukuyama. 2004. A word expressing affective pain activates the anterior cingulate cortex in the human brain: an fMRI study. *Behavioural brain research* 153, 1 (August 2004), 123–127. https://doi.org/10.1016/j.bbr.2003.11.013

[19] Pierre Philip, Lucile Dupuy, Patricia Sagaspe, Etienne de Sevin, Marc Auriacombe, Jacques Taillard, Jean-Arthur Micoulaud Franchi, and Charles Morin. 2022. Efficacy of a Smartphone-Based Virtual Companion to Treat Insomniac Complaints in the General Population: Sleep Diary Monitoring Versus an Internet Autonomous Intervention. *Journal of Clinical Medicine* 11 (07 2022), 4387. https://doi.org/10.3390/jcm11154387

[20] Michael Post and David Sulzer. 2021. The chemical tools for imaging dopamine release. *Cell chemical biology* 28 (04 2021). https://doi.org/10.1016/j.chembiol.2021.04.005

[21] Yoichi Sekizawa. 2023. Effects of being eligible for specific health guidance on health outcomes: A regression discontinuity analysis using Japan's data on specific health checkups. *Preventive Medicine* 172 (2023), 107520. https://doi.org/10.1016/j.ypmed.2023.107520

[22] Piers Steel. 2007. The nature of procrastination: a meta-analytic and theoretical review of quintessential self-regulatory failure. *Psychological Bulletin* 133, 1 (2007), 65–94.

[23] Piers Steel and Cornelius König. 2006. Integrating Theories of Motivation. *Academy of Management Review* 31 (10 2006), 889–913. https://doi.org/10.2307/20159257

[24] James Sulzer, Ranganatha Sitaram, Maria Laura Blefari, Spyros Kollias, Niels Birbaumer, Klaas Enno Stephan, Andreas Luft, and Roger Gassert. 2013. WITHDRAWN: Neurofeedback-mediated self-regulation of the dopaminergic midbrain. *NeuroImage* 75 (7 2013), 176.

[25] Hidehiko Takahashi, Hiroshi Matsui, Colin Camerer, Harumasa Takano, Fumitoshi Kodaka, Takashi Ideno, Shigetaka Okubo, Kazuhisa Takemura, Ryosuke Arakawa, Yoko Eguchi, Toshiya Murai, Yoshiro Okubo, Motoichiro Kato, Hiroshi Ito, and Tetsuya Suhara. 2010. Dopamine D-1 Receptors and Nonlinear Probability Weighting in Risky Choice. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 30 (12 2010), 16567–72. https://doi.org/10.1523/JNEUROSCI.3933-10.2010

[26] Kazuhiro Takahashi, Kei Mizuno, Akihiro Sasaki, Yasuhiro Wada, Masaaki Tanaka, Akira Ishii, Kanako Tajima, Naohiro Tsuyuguchi, Kyosuke Watanabe, Semir Zeki, and Yasuyoshi Watanabe. 2015. Imaging the Passionate Stage of Romantic Love by Dopamine Dynamics. *Frontiers in human neuroscience* 9 (04 2015), 191. https://doi.org/10.3389/fnhum.2015.00191

[27] Barbara J. Weiland, Robert A. Zucker, Jon-Kar Zubieta, and Mary M. Heitzeg. 2017. Striatal dopaminergic reward response relates to age of first drunkenness and feedback response in at-risk youth. *Addiction Biology* 22, 2 (2017), 502–512. https://doi.org/10.1111/adb.12341 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/adb.12341

Exploring the Impact of Wearable Continuous Glucose Monitoring on Glucose Regulation and Eating Behavior in Healthy Individuals: A Pilot Study

Zilu Liang
liang.zilu@kuas.ac.jp
Kyoto University of Advanced Science (KUAS)
Kyoto, Japan

ABSTRACT

While continuous glucose monitoring (CGM) technology has been instrumental in managing diabetes, its potential impact on healthy individuals remains underexplored. This pilot study investigates the influence of wearing a CGM device on glucose regulation and eating behavior among healthy participants. Over approximately one month, participants utilized a wearable CGM device, and semi-structured interviews were conducted at the study's conclusion to gather insights. Analysis revealed a significant decrease in mean glucose levels and fewer glucose spikes during the study period for half of the participants. Interview findings supported the hypothesis that CGM usage positively affected behavior, with participants adjusting meal timing but not food choices based on real-time glucose readings. Some participants ignored glucose spikes due to their asymptomatic nature, indicating behavioral inertia. Nonetheless, they expressed intent to implement changes based on what they learnt in the experiment. This study offers initial evidence of CGM's potential as a behavioral moderation tool for improving glucose regulation among healthy individuals.

CCS CONCEPTS

• **Applied computing** → **Consumer health**; • **Human-centered computing** → **Empirical studies in ubiquitous and mobile computing**; **Field studies**.

KEYWORDS

Continuous glucose monitoring, behavior change, wearable computing, personal informatics, glucose regulation

ACM Reference Format:
Zilu Liang. 2024. Exploring the Impact of Wearable Continuous Glucose Monitoring on Glucose Regulation and Eating Behavior in Healthy Individuals: A Pilot Study. In *Behavior Transformation by IoT International Workshop (BTIW '24), June 3–7, 2024, Minato-ku, Tokyo, Japan.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3662008.3662012

BTIW '24, June 3–7, 2024, Minato-ku, Tokyo, Japan

ACM ISBN 979-8-4007-0665-3/24/06
https://doi.org/10.1145/3662008.3662012

1 INTRODUCTION

Controlling the blood glucose level within a narrow range is crucial for maintaining overall health in the human body. Meals exert an immediate and profound impact on glycemic control [1]. The concept of glycemic index (GI) was introduced to measure how quickly foods elevate the blood glucose levels. Prior studies found small but clinically significant effect of low-GI foods on medium-term glycemic control in patients with diabetes [4, 23], and GI has been widely used as a common guideline for food decisions among diabetic patients. However, individual glycemic responses to foods can vary widely [21], and GI does not account for factors such as carbohydrate quality [15], meal composition [9, 12], or the insulinogenic effect of foods [29]. Consequently, opinions on the use of GI as a nutritional therapy intervention remain contentious [10].

Continuous glucose monitoring (CGM) technologies offer a solution to the limitations of traditional diabetes management methods by providing patients with direct feedback on real-time glucose levels and fluctuations. Numerous studies have demonstrated the effectiveness of CGM in improving self-care behaviors among diabetic patients [8, 20, 27, 28, 30], as well as in enhancing glycemic control and reducing hypoglycemia in adults with type 1 and type 2 diabetes [14, 25]. These benefits have been observed even with intermittent CGM use or among type 2 diabetes patients not undergoing intensive insulin therapy [2, 16, 24]. In recent years, advancements in CGM technologies, particularly in terms of accuracy and usability, have made them increasingly appealing for general use. However, research on the use of CGM in healthy populations remains scarce. For the majority of healthy individuals, CGM remains an underutilized tool for self-care and disease prevention.

This study aims to investigate whether the simple act of wearing a CGM sensor and accessing real-time glucose readings could influence users' behavior and glucose levels. While one prior study explored the impact of CGM on behavior change among healthy, non-diabetic individuals [30], it provided users with personalized recommendations alongside glucose readings, making it difficult to isolate the effect of CGM alone. To our knowledge, this study is the first to examine the potential behavior-modulating effect of CGM alone on healthy, non-diabetic individuals.

2 RELATED WORK

Keeping blood glucose levels within a narrow range is crucial for metabolic and cardiovascular health. Diabetes, a condition characterized by high blood glucose levels, typically requires frequent monitoring through methods like self-monitoring of blood glucose (SMBG), which involves painful finger pricking multiple times a day. SMBG has the disadvantage of high cost and causing pain, leading to low adherence [8].However, continuous glucose monitoring (CGM) technology has revolutionized glucose monitoring by automatically measuring interstitial glucose levels at regular intervals and providing real-time readings, day and night. CGM offers a comprehensive view of glucose profiles over extended periods without the need for frequent blood samples. Unlike SMBG, CGM enables monitoring of glucose trends post-meals and during sleep, significantly enhancing convenience and effectiveness. Moreover, CGM devices are user-friendly and do not require constant technical assistance [2]. Studies have validated the accuracy and reliability of CGM technology [22], and its high-resolution data has allowed researchers to identify new patterns of glucose dysregulation [11]. Additionally, by integrating CGM data with other biosignals, researchers have uncovered valuable insights into the relationship between glucose levels and various aspects of human health, in a non-invasive manner [3, 17].

While CGM is not meant to be a behavior change technology, numerous studies have highlighted its potential as a behavioral intervention tool for patients with diabetes. When CGM sensors are used in unblinded mode, allowing patients instantaneous access to their glucose data, they are more likely to adapt their behavior to achieve better glycemic management. This behavioral impact of CGM has been consistently observed in studies involving diabetic patients [8, 20, 27, 28, 30]. For instance, in a small pilot study with type 2 diabetes patients, researchers observed a significant decrease in the consumption of high GI foods and carbohydrates, along with an increase in dietary fiber intake over a three-month period [6]. Additionally, CGM may have more pronounced effect on physical activity than on food selection. In one study, 47.5% of the participants increased their activity levels or exercised more, while only 22.5% reduced rice consumption and 15% reduced cereal intake [7]. On the other hand, research studies applying CGM on normoglycemic populations are scarce. Only one study has investigated the behavioral impact of CGM in conjunction with personalized lifestyle recommendations [30] and found positive outcomes. There has been no study investigating the behavioral impact of CGM alone on healthy individuals. This study aims to fill this knowledge gap.

3 METHOD

3.1 Participants

We recruited 6 participants using convenience sampling, all of whom were graduate students enrolled in a private university in Japan. Inclusion criteria comprised individuals who (1) had not been diagnosed with any chronic diseases and (2) were not pregnant. Exclusion criteria encompassed participants who (1) had allergies to adhesives, (2) exhibited sensitivity to pain, (3) had a history of soft tissue skin infections, or (4) had a blood phobia disorder. As appreciation for their participation, participants received an Amazon Gift Card. This study received approval from the Ethics Review Board of the Kyoto University of Advanced Science.

3.2 Study Procedure

We initiated the experiment with individual meetings with each participant in the laboratory. During the meetings, we explained that the objective of this study was to collect glucose data of healthy individuals in the wild and that they needed to use a FreeStyle Libre flash CGM system (Figure 1) for approximately one month. Participants signed a written consent form prior to commencing the experiment. Following this, we provided instructions on how to install the FreeStyle Libre sensor. Adhering to infection prevention guidelines during the COVID-19 pandemic, participants were advised to apply the sensor themselves on an upper arm, as illustrated in Figure 2.

For those participants whose smartphones were equipped with NFC, we facilitated the download and installation of the FreeStyle LibreLink app, which allowed them to activate their sensors for scanning glucose readings on their smartphones. Participants lacking an NFC sensor on their smartphones were provided with a stand-alone FreeStyle Libre Reader, and we assisted them in activating their sensors using the reader.

Figure 1: The FreeStyle Libre flash CGM system used in this study.

Figure 2: The FreeStyle Libre flash CGM sensor was placed on the back of the upper arm.

The subsequent data collection phase unfolded in participants' natural living environments for 4 weeks. Participants had access to various data on the FreeStyle LibreLink app or the standalone scanner, including daily glucose plots, average glucose patterns, and time spent in different ranges (Figure 3). As each FreeStyle Libre sensor lasted for up to 14 days, participants returned to the laboratory for sensor replacement when the initial sensor expired. Upon completion of the experiment, we conducted semi-structured interviews with each participant to delve deeper into their experiences with the CGM and to ascertain if there were any observable changes in their lifestyle. The duration of these interviews ranged from 1 to 2 hours. All interviews were audio-recorded.

3.3 Data Analysis

We collected both quantitative data on interstitial glucose levels and qualitative data through interviews. The glucose levels were automatically measured at 15-minute intervals throughout the experiment. Utilizing the web-based LibreView platform, we exported the glucose readings of each participant into a CSV file. We then used Python 3.10.5 and several Python packages, including Pandas, NumPy, SciPy, and sklearn to perform statistical analysis. To facilitate analysis, we divided the glucose readings into two halves representing the first and second parts of the experiment. The data from the first and last days of each half were excluded from the analysis due to potential inaccuracies, following a procedure similar to that employed in [30]. For each participant, we calculated the mean and standard deviation of their glucose readings for both halves of the experiment. To explore any potential differences in glucose readings between the first and second halves for each participant, we conducted unpaired t-test because the distribution of glucose readings were approximately normal. The probability densities of the interstitial glucose for the two halves were estimated using the Gaussian kernel.

The interview recordings were transcribed verbatim, and the transcribed scripts were subsequently imported into the software DoveTail for analysis. We conducted the qualitative data analysis following a standard thematic analysis procedure as outlined in Braun and Clarke's work [5].

4 RESULTS

4.1 Glucose Profiles

During the pilot phase we collected glucose data from 6 participants. The demographic information is summarized in Table 1. The participants ranged in age from 25 to 34 years old and were evenly distributed in terms of gender. We were able to collect 27 days of glucose data from five participants, while one participant lost his first sensor after approximately one week, resulting in only 20 days of data. P1, P3, and P4 did not have a habit of regular exercise, whereas P2, P5, P6 engaged in light to intense exercises regularly. Among them, only P4 and P5 followed a consistent meal pattern of three meals per day; the others typically had irregular meal patterns and consumed one to two meals per day. For most participants, diet had never been a concern or a priority, with the exception of P1. In fact, P1 had actively experimented with various types of diets, including vegan and Keto diet, over the years in an effort to lose weight.

Figure 4 depicts participants' estimated glucose probability density for the first (red lines) and the second (blue lines) halves of the experiment. The grey dash-dot lines represent the recommended target range for healthy individuals, which is [70, 140] mg/dL, while the grey dot line represents a wider upper range at 180 mg/dL. The plot shows that the interstitial glucose levels occasionally exceeded 180 mg/dL and frequently rose above 140 mg/dL for most participants. Specifically, P1, P3 and P5 experienced more glucose spikes above 140 mg/dL during the first half of the experiment, while P6 had slightly more spikes during the second half.

On average, the glucose levels hovered around 100 mg/dL for all participants. Upon comparing the first and second halves of the experiment, we observed a notable decrease in mean glucose levels during the latter period for three of the six participants. Conversely, P2 and P6 experienced a slight but statistically significant increase in mean glucose levels, while P4 did not exhibit a significant change. Additionally, the standard deviation of glucose levels decreased during the second half of the experiment for four participants.

4.2 Interview Findings

Overall, participants reported positive experiences with the FreeStyle Libre flash CGM system. None of them had adverse effects such as skin irritation or pain, and the sensor did not disrupt with their normal lifestyle.

4.2.1 Raising awareness of glucose fluctuation. It is well-known that self-tracking can increase people's awareness of their physical and mental health [18, 19]. The use of the CGM sensor had a direct effect on participants, making them more aware of how their glucose levels fluctuated in response to their diet and physical activity. P3 noted, "*I noticed that chicken rice, rice ball, and soba all triggered a spike. Ramen was OK. Taco rice was OK, probably because it was not the sticky Japanese rice. And the peach water bottles I bought from the vending machine. They could trigger a spike as if I just had a big meal. I'm really surprised towards the drinks and how drinks actually spike up glucose really quickly*". P4 noted, "*my glucose at night was higher than I expected. Maybe the condensed milk I added to my tea was the cause. Eating speed does matter*". P5 observed, "*I noticed spike after dinner, but no spike after breakfast or lunch. Perhaps because I had a walk from home to the university after breakfast and lunch*".

The real-time glucose readings also helped reinforce or assure prior knowledge at the individual level, as P6 stated: "*I knew that it is not a good idea to eat a lot of sugar at night. But when I saw the ongoing feedback, not just speculation, like the sensor showed me my glucose went very, very high up. It helped align my emotion with my logic.*" Additionally, two participants mentioned their particular interest in checking and identifying relationship between glucose level and stress.

4.2.2 Inertia in behavior change. Despite the initial hypothesis that immediate feedback of glucose spikes would motivate action, this was not necessarily the case for some participants. P3 mentioned that "*when I saw the glucose spike, I just try to ignore it. I didn't really feel bad*". Similarly, P6 confessed that she did not attempt to mitigate the glucose spikes.

4.2.3 Behavior change during experiment. On the other hand, several participants acknowledged making some changes in the hope of achieving a more ideal glucose curve. Particularly, P1 attempted to reduce her consumption of sweets at night as they "*always trigger a big spike. It was like I ate one big meal. It made me more selective about the kind of food I ate. I didn't want to see glucose spikes that late.*" In addition, P1 adjusted her eating patterns, especially regarding meal timing: "*I'm more careful of eating late night. I tried to eat*

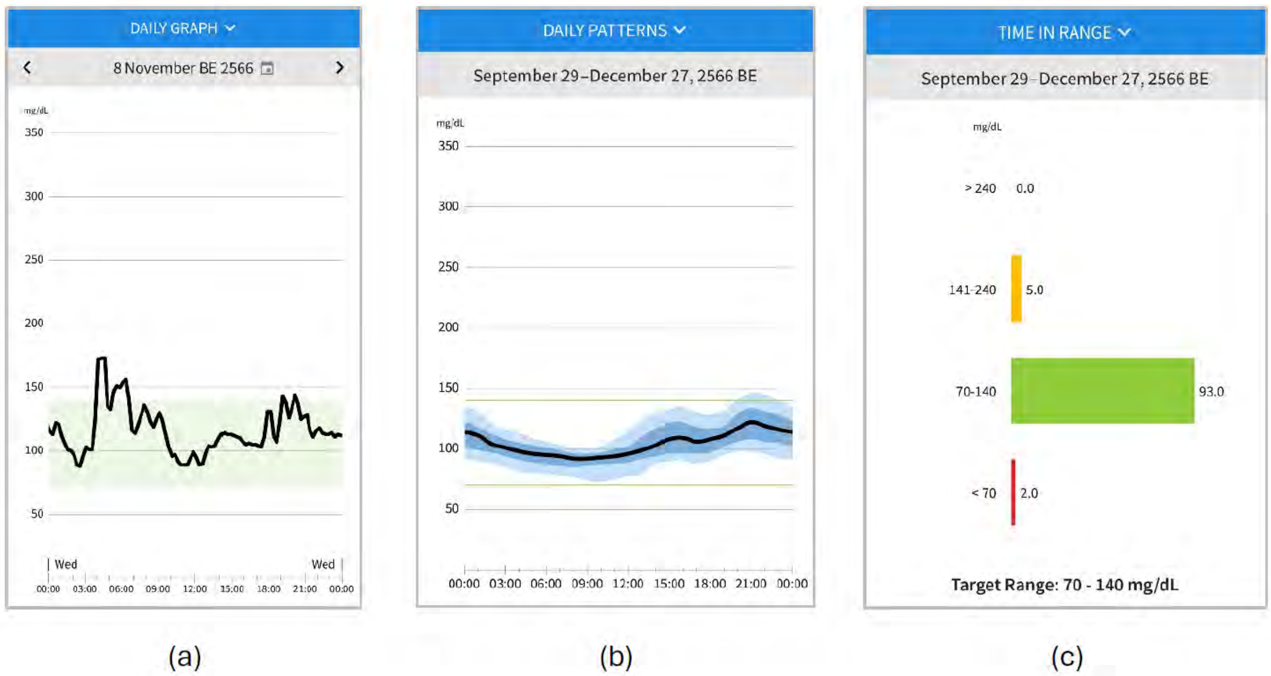

Figure 3: Data accessible to participants during the experiment included: (a) Daily glucose plot. (b) Average daily patterns. (c) Time spent in various ranges.

Table 1: Comparison of Glucose Profiles during the First and Second Half of the Experiment.

Participant ID	Age	Sex	Days of Data	First Half (mg/dL)	Second Half (mg/dL)	*t*-statitic	*p*-value
P1	25	F	27	109.7 ± 17.8	99.7 ± 19.5	12.79	<0.001
P2	34	M	27	97.7 ± 14.6	102.5 ± 12.8	-7.91	<0.001
P3	26	M	20	109.5 ± 22.5	103.5 ± 21.1	5.10	<0.001
P4	25	F	27	106.0 ± 14.2	106.0 ± 13.0	0.13	0.894
P5	34	M	27	99.4 ± 19.7	90.3 ± 18.6	11.3	<0.001
P6	36	F	27	88.8 ± 17.3	91.3 ± 19.2	-3.15	0.002

earlier. And I tried to avoid snacking because I noticed that even just one banana is super aggressive. That was surprising. I used to believe that having more small meals was better. But it was actually not good because the glucose never dropped. So I kind of merged snacks with the meals." P3 also mentioned trying to eat more regularly after observing her glucose curves for a couple of days. Conversely, P2 experienced hypoglycemia events from time to time, prompting him to "*eat something immediately*".

4.2.4 Intention for behavior change after experiment. Some participants, like P2 and P6, expressed intentions to make changes in their future behaviors based on the insights gained from monitoring their glucose dynamics during the experiment. P2 mentioned plans to adjust his eating patterns, aiming to consume more during breakfast and to lower nighttime glucose levels by reducing dinner intake. He noted, "*I have to change a lot. I'm still adapting to the new living environment. Ideally I should eat more for breakfast, so the glucose should be higher in the morning. I would like to reduce my glucose at night because right now I'm eating a lot for dinner*". P6 acknowledged the informative value of the collected data and indicated a willingness to avoid certain types of food in the future based on this newfound knowledge. She stated, "*I don't think I did anything special. I just kind of collecting this information.So now I, I know that this type of food have this type of effect and maybe next time I will not order this type of food*". Similarly, P4 outlined a concrete plan for change, stating "*now I know that meat doesn't raise glucose that fast, I'll try to add some meat when I eat sweet things*".

5 DISCUSSION

In summary, our preliminary findings suggest that CGM technology has the potential to influence the behavior of healthy individuals, albeit in varied ways. We observed a notable decrease in both the mean and standard deviation of interstitial glucose levels in over half of our participants. This suggests that CGM technology may

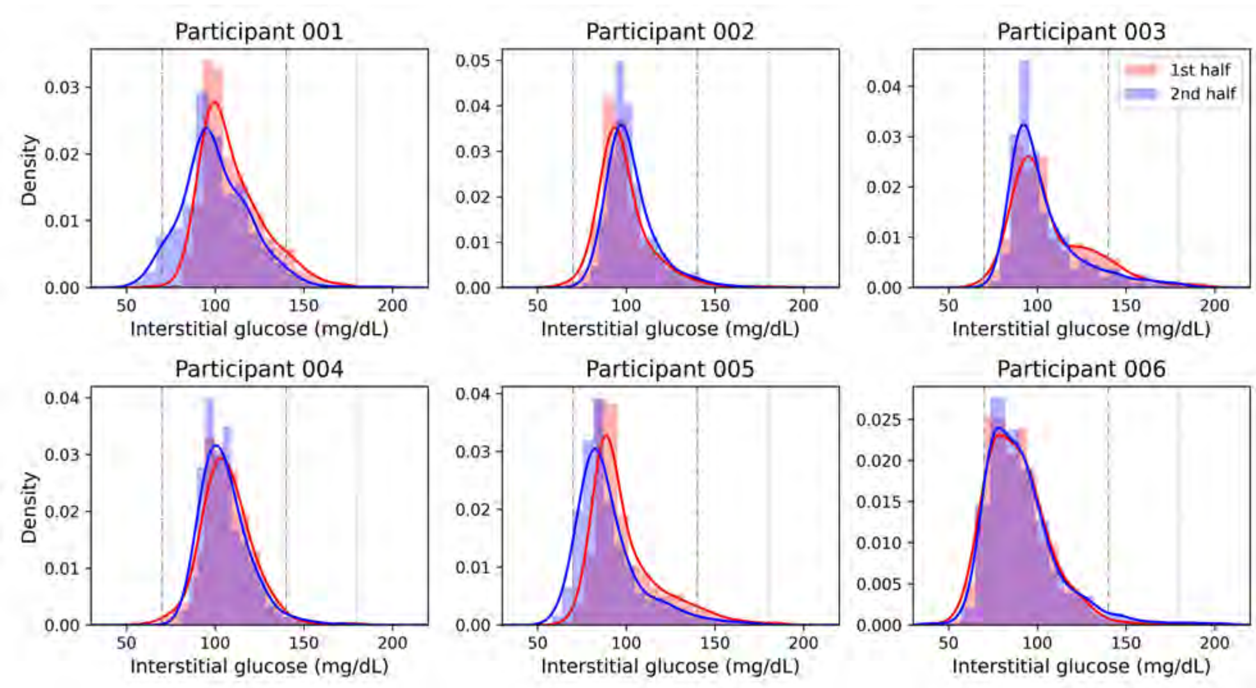

Figure 4: Comparing participants' estimated glucose probability density between the first and second halves of the experiment.

lead to improvements in glucose regulation among healthy individuals. Our qualitative analysis further supports the notion that these changes in glucose levels are likely attributable to participants' behavior modifications. Specifically, some participants reported making conscious adjustments to their eating patterns and habits in response to real-time glucose data.

Our findings align well with the transtheoretical model (TTM) of behavior change [13]. The TTM outlines six stages of health behavior change: precontemplation, contemplation, preparation, action, maintenance, and termination. This model has been widely utilized to inform the design of dietary interventions [26]. At the start of our experiment, all participants were in the precontemplation stage, indicating a lack of concern about their diet. By the end of the study, half of the participants progressed to the action stage, where they actively modified their eating patterns. Two participants were in the contemplation stage, intending to refine their eating schedules based on insights gained during the study. Only one participant remained in the precontemplation stage, expressing on intention to change his dietary habits in the foreseeable future. It is remarkable to observe such significant behavioral impact considering that participants were simply wearing a CGM passively.

Interestingly, behavior changes primarily centered on adjusting meal timing rather than diet composition. Specifically, participants made efforts to regulate their glucose levels by altering when they ate, such as having meals earlier in the day and avoiding snacks. Unlike diabetic patients, whose behavior changes typically focus on the types of foods they consume, our participants prioritized the timing of their meals to manage glucose fluctuations. This suggests a unique approach to glucose regulation among healthy individuals using CGM technology.

However, some participants exhibited behavior inertia when confronted with glucose spikes detected by the CGM. Despite acknowledging these spikes, they chose to disregard them, attributing this indifference to the lack of physical discomfort associated with hypoglycemic events, which are typically asymptomatic in healthy individuals. Consequently, they expressed no urgency or intention to modify their lifestyle habits in response to these glucose fluctuations. It is important to contextualize these findings within the participants' life circumstances, as individuals like P2 and P6, who were undergoing a transitional period, may face challenges in implementing behavioral changes during such times.

Moreover, the potential behavioral moderation effect of CGM may extend beyond the sensor use period. Participants had the opportunity to explore the impact of specific foods, dietary choices, physical activity, and stress on their glucose levels, providing valuable insights that could inform future behavior. For instance, although P6 did not actively modify her habits during the experiment, she found the observations to be informative and anticipates incorporating them into her future diet plans. P2 and P4 expressed similar sentiments, highlighting the lasting impact of CGM.

6 CONCLUSION

In our study involving 6 healthy participants, we observed decreases in the mean and standard deviation of glucose levels for half of the participants when comparing the first and second halves of the study. Qualitative insights from semi-structured interviews further supported the behavioral impact of CGM, particularly regarding eating schedules. However, we also noted instances of behavior inertia, with some participants opting to disregard out-of-range glucose readings due to their asymptomatic nature. Despite of the small sample size of this pilot study, its duration positions it as one of the longest glucose monitoring investigations involving healthy individuals. Moving forward, our next steps will involve expanding the sample size and conducting analysis on additional glucose measures that hold clinical significance. This will allow for a more comprehensive understanding of the potential impact of CGM on behavior and glucose regulation among healthy individuals.

ACKNOWLEDGMENTS

The author would like to thank the participants of this study.

REFERENCES

[1] Fiona Atkinson, Jennie Brand-Miller, Kaye Foster-Powell, Anette Buyken, and et al. 2021. International tables of glycemic index and glycemic load values 2021: a systematic review. *The American Journal of Clinical Nutrition* 114 (2021). https://doi.org/10.1093/ajcn/nqab233

[2] Richard Bergenstal, Jennifer Layne, Howard Zisser, and et al. 2021. Remote Application and Use of Real-Time Continuous Glucose Monitoring by Adults with Type 2 Diabetes in a Virtual Diabetes Clinic. *Diabetes Technol Ther* 23, 2 (2021), 128–132. https://doi.org/10.1089/dia.2020.0396

[3] Lauriane Bertrand, Nathan Cleyet-Marrel, and Zilu Liang. 2021. Recognizing Eating Activities in Free-Living Environment Using Consumer Wearable Devices. *Eng Proc* 6, 1 (2021). https://doi.org/10.3390/I3S2021Dresden-10141

[4] Jennie Brand-Miller, Susan Hayne, Peter Petocz, and Stephen Colagiuri. 2003. Low–Glycemic Index Diets in the Management of Diabetes: A meta-analysis of randomized controlled trials. *Diabetes Care* 26, 8 (2003), 2261–2267. https://doi.org/10.2337/diacare.26.8.2261

[5] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3 (2006), 77–101. https://doi.org/10.1191/1478088706qp063oa

[6] Daniel J Cox, Ann G Taylor, Matthew Moncrief, and et al. 2016. Continuous Glucose Monitoring in the Self-management of Type 2 Diabetes: A Paradigm Shift. *Diabetes Care* 39, 5 (2016), e71–73. https://doi.org/10.2337/dc15-2836

[7] Nicole Ehrhardt and Enas Al Zaghal. 2020. Continuous Glucose Monitoring As a Behavior Modification Tool. *Clinical Diabetes* 38, 2 (2020), 126–131. https://doi.org/10.2337/cd19-0037

[8] Solangia Engler, Sherecce Fields, Walker Leach, and et al. 2022. Real-Time Continuous Glucose Monitoring as a Behavioral Intervention Tool for T2D: A Systematic Review. *Journal of Technology in Behavioral Science* 7 (2022), 1–12. https://doi.org/10.1007/s41347-022-00247-5

[9] Anne Flint, Bente Møller, Anne Raben, and et al. 2004. The use of glycaemic index tables to predict glycaemic index of composite breakfast meals. *The British journal of nutrition* 91 (2004), 979–89. https://doi.org/10.1079/BJN20041124

[10] Marion J. Franz. 2003. The Glycemic Index: Not the most effective nutrition therapy intervention. *Diabetes Care* 26, 8 (2003), 2466–2468. https://doi.org/10.2337/diacare.26.8.2466

[11] Heather Hall, Dalia Perelman, Alessandra Breschi, and et al. 2018. Glucotypes reveal new patterns of glucose dysregulation. *PLoS Biology* 16, 7 (2018), e2005143. https://doi.org/10.1371/journal.pbio.2005143

[12] C J K Henry, H J Lightowler, F L Kendall, and et al. 2006. The impact of the addition of toppings/fillings on the glycaemic response to commonly consumed carbohydrate foods. *European Journal of Clinical Nutrition* 60 (2006), 763–769. https://doi.org/10.1038/sj.ejcn.1602380

[13] Caroline C. Horwath. 1999. Applying the transtheoretical model to eating behaviour change: challenges and opportunities. *Nutrition Research Reviews* 12, 2 (1999), 281–317. https://doi.org/10.1079/095442299108728965

[14] Melanie Jackson, Andrew Ahmann, and Viral Shah. 2021. Type 2 Diabetes and the Use of Real-Time Continuous Glucose Monitoring. *Diabetes Technol Ther* 23 (02 2021). https://doi.org/10.1089/dia.2021.0007

[15] Mitch Kanter, Siddhartha Angadi, Julie Jones, and Katherine Beals. 2021. Limitations of the Glycemic Index and the Need for Nuance When Determining Carbohydrate Quality. *Cardiovascular research* 118 (2021), e38–e39. https://doi.org/10.1093/cvr/cvab312

[16] Wendy H. Updike Kevin Cowart and Rachel Franks. 2021. Continuous glucose monitoring in persons with type 2 diabetes not using insulin. *Expert Rev Med Devices* 18, 11 (2021), 1049–1055. https://doi.org/10.1080/17434440.2021.1992274

[17] Zilu Liang. 2022. Mining associations between glycemic variability in awake-time and in-sleep among non-diabetic adults. *Front Med Technol* 4 (2022), 1026830. https://doi.org/10.3389/fmedt.2022.1026830

[18] Zilu Liang and Bernd Ploderer. 2020. How Does Fitbit Measure Brainwaves: A Qualitative Study into the Credibility of Sleep-Tracking Technologies. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 1, Article 17 (2020), 29 pages. https://doi.org/10.1145/3380994

[19] Zilu Liang, Bernd Ploderer, Wanyu Liu, and et al. 2016. SleepExplorer: a visualization tool to make sense of correlations between personal sleep data and contextual factors. *Pers Ubiquit Comput* 20 (2016), 985–1000. https://doi.org/10.1007/s00779-016-0960-6

[20] Jacquelyn Manfredo, Tyger Lin, Radhika Gupta, and et al. 2023. Short-term use of CGM in youth onset type 2 diabetes is associated with behavioral modifications. *Front Endocrinol* 14 (2023), 1182260. https://doi.org/10.3389/fendo.2023.1182260

[21] Nirupa Matthan, Lynne Ausman, Huicui Meng, and et al. 2016. Estimating the reliability of glycemic index values and potential sources of methodological and biological variability. *Am J Clin Nutr* 104 (2016). https://doi.org/10.3945/ajcn.116.137208

[22] Jordi Merino, Inbar Linenberg, Kate M Bermingham, and et al. 2022. Validity of continuous glucose monitoring for categorizing glycemic responses to diet: implications for use in personalized nutrition. *The American Journal of Clinical Nutrition* 115, 6 (2022), 1569–1576. https://doi.org/10.1093/ajcn/nqac026

[23] A. Maretha Opperman, Christina S. Venter, Welma Oosthuizen, and et al. 2004. Meta-analysis of the health effects of using the glycaemic index in meal-planning. *British Journal of Nutrition* 92, 3 (2004), 367–381. https://doi.org/10.1079/BJN20041203

[24] Banshi Saboo, Ranjit Unnikrishnan, Jothydev Kesavadev, and et al. 2023. Intermittent Use of Continuous Glucose Monitoring: A New Paradigm in Treatment of Type 2 Diabetes. *J Assoc Physicians India* 71 (2023), 11–12. https://doi.org/10.5005/japi-11001-0274

[25] Jan Soupal, Lenka Petruzelkova, George Grunberger, and et al. 2019. Glycemic Outcomes in Adults With T1D Are Impacted More by Continuous Glucose Monitoring Than by Insulin Delivery Method: 3 Years of Follow-Up From the COMISAIR Study. *Diabetes Care* 43 (2019), dc190888. https://doi.org/10.2337/dc19-0888

[26] Deirdre Timlin, Jacqueline McCormack, Maeve Kerr, and et al. 2020. Are dietary interventions with a behaviour change theoretical framework effective in changing dietary patterns? A systematic review. *BMC Public Health* 20 (12 2020), 1–18. https://doi.org/10.1186/s12889-020-09985-8

[27] Michael Vallis, Hana Ryan, Lori Berard, and et al. 2023. How Continuous Glucose Monitoring Can Motivate Self-management: Can Motivation Follow Behaviour? *Can J Diabetes* 47, 5 (2023), 435–444. https://doi.org/10.1016/j.jcjd.2023.04.001

[28] Eugene Wright and Savitha Subramanian. 2021. Evolving Use of Continuous Glucose Monitoring Beyond Intensive Insulin Treatment. *Diabetes Technol Ther* 23 (2021), S12–S18. https://doi.org/10.1089/dia.2021.0191

[29] Zahra Yari, Vahideh Behrouz, Hamid Zand, and et al. 2020. New Insight into Diabetes Management: From Glycemic Index to Dietary Insulin Index. *Curr Diabetes Rev* 16, 4 (2020), 293–300. https://doi.org/10.2174/1573399815666190614122626

[30] Ashkan Dehghani Zahedani, Tracey McLaughlin, Arvind Veluvali, and et al. 2023. Digital health application integrating wearable data and behavioral patterns improves metabolic health. *NPJ Digital Medicine* 6 (2023), Article 216. https://doi.org/10.1038/s41746-023-00956-y

PTAC: Personal Utterance Counter For Well-being

Mina Matoba
Department of Integrated Information
Aoyama Gakuin University
Sagamihara, Japan
mina@rcl-aoyama.jp

Masaki Ishizaka
Graduate School of Science and Engineering
Aoyama Gakuin University
Sagamihara, Japan
zacker@rcl-aoyama.jp

Yoshito Tobe
Department of Integrated Information
Aoyama Gakuin University
Sagamihara, Japan
y.tobe@rcl-aoyama.jp

ABSTRACT

With the recent increase in health consciousness, the use of pedometers and wearable devices that measure the amount of exercise and biometrics in daily life has been increasing. We propose Personal Utterance Counter For Well-being (PTAC), a system that measures a user's utterance amount using a smartphone. The system calculates the user's daily utterance amount based on the daily utterance data recorded using a smartphone. Furthermore, by visualizing the user's daily utterance amount using a smartphone, the user is made to feel a sense of urgency about his or her low utterance amount. This enables the user to notice that his/her daily utterance amount is lower than the average and that he/she tends not to speak on holidays.

CCS CONCEPTS

• Human-centerd computing → Ubiquitous and mobile computing
• Applied computing → Law, social and behavioral sciences

KEYWORDS

Behavioral change, CNN, Speaker recognition

ACM Reference format:

Mina Matoba, Masaki Ishizaka and Yoshito Tobe. 2024. PTAC: Personal Utterance Counter For Well-being. In Behavior Transformation by IoT International Workshop (BTIW '24), June 3–7, 2024, Minato-ku, Tokyo, Japan. ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3662008.3662013.

1 Introduction

In recent years, as a result of the COVID-19 pandemic and the restriction of outings in society as a whole, opportunities for social interaction have been significantly reduced, and the amount of utterance per day by individuals has tended to decrease significantly. For example, in the past, people had opportunities to go to work and communicate with others, but in recent years, remote work has increased due to the Corona Disaster, and communication has decreased. It was found that those who subjectively answered "not healthy" conversed much less than those who answered "healthy" [1]. There have been instances where the spread of COVID-19 and resulting school closures led to an increase in depression rates [2]. However, few people are concerned about the amount of conversation, even though a decrease in the amount of conversation can affect their own physical condition. In this study, we propose Personal Utterance Counter For Well-being (PTAC), which visualizes the amount of utterance recorded by a wearable device and makes users feel a sense of urgency about the low amount of conversation. The PTAC sends the voice file recorded by the wearable device to the server and extracts the part of utterance contained in the voice file. The system uses a Convolutional Neural Network (CNN) to extract only the user's utterance. The system measures the user's daily utterance amount and feeds back the amount to the user.

PTAC reduces the cost of labeling by clustering the user's daily utterance data and roughly classifying the data in advance.

This paper is organized as follows: In this section, we describe the background of the need for this research and the actual proposal. In Section 2, we introduce related works and show the position of this research. Section 3 describes the actual algorithm in detail, and Section 4 evaluates the accuracy and usefulness of the algorithm through subject experiments. Finally, Section 5 summarizes this study and discusses future prospects.

BTIW'24, June 3-7, 2024, Minato-ku, Tokyo, Japan

ACM ISBN 978-8-4007-0665-3/24/06...$15.00

2 Related Work

This study measures and visualizes the amount of utterance produced per day to prevent health hazards caused by reduced utterance production. Therefore, it is necessary to identify which utterance production the user is using.

Tsujita et al. [3] proposed a system to support the improvement of emotional states based on the possibility that smile formation improves emotional states. In this experiment, users were given feedback on their smiles for 10 days. The experimental results showed that the system was effective in promoting smiles by changing the user's state of mind and improving his/her emotional state. From this paper, we

Figure 1: Smartphone during recording

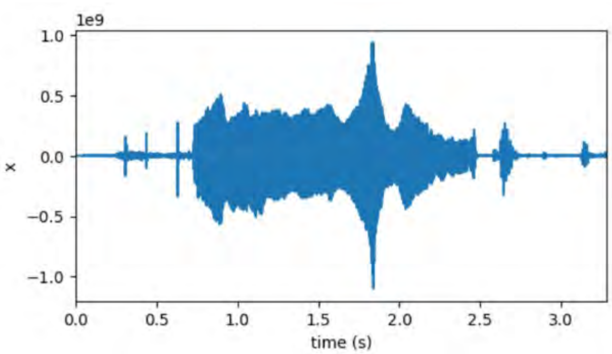

Figure 2a: Graph of utterance in the time domain

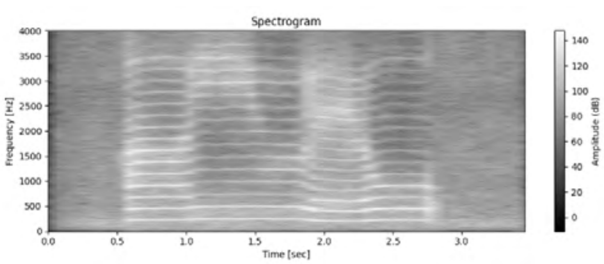

Figure 2b: Utterance graph after short-time Fourier transform

considered that providing feedback could potentially improve behavior.

Yanick et al. [4] proposed a method for automatically identifying and grouping speakers when clustering utterance data containing multiple speakers. In this experiment, features were automatically learned from utterance data using a CNN. The experimental results showed that high clustering accuracy was achieved with only 17% of the training data. From this paper, we gained the inspiration for clustering speakers with minimal data.

Rohan et al. [5] proposed a text-independent speaker recognition method for personalizing users on smart devices. They compared two feature learning methods: one that does not include the usual convolution created by 3000 nodes and four layers, and one that consists of three convolutional layers and one layer using 1000 nodes. Experimental results showed that convolution is effective. This is expected because the convolutional layer can be used to extract temporal and frequency features of the utterance signal. From this paper, we utilized the learning method consisting of three convolutional layers in our experiments.

3 Configuration of PTAC

In this section, we describe a method for calculating and visualizing a user's daily utterance amount by recording a day's utterance with a smartphone, sending the audio file to the smartphone, and extracting only the user's utterance. In this study, a CNN is used to identify the user's utterance, and although a large amount of data is required for CNN training, it is impossible to prepare a large amount of user utterance data in advance. Therefore, we use a day's recorded utterance data and manually label the utterance data according to whether it is the user's utterance or other sounds. However, it is impractical to manually label all of a day's utterance data, so we attempted to reduce the cost of labeling by using clustering.

3.1 Creating a file of one day's worth of recorded utterance

First, record a day's worth of conversation using a smartphone. When the PTAC proposed in this study is started, the buttons "Recording Start," "Recording Stop," and "Send File" are displayed. When "Recording Start" is pressed, recording starts. Figure 1 shows how PTAC looks when the recording button is actually pressed. Then, when "Recording Stop" is pressed, the recording is stopped and the file is saved.

3.2 Sending a recording file to the server

Next, press the "Send File" button on PTAC to send the recorded data of a day's conversation to the server. At this time, the FastAPI of Python is used.

3.3 Clustering

After the recorded audio files are sent to the server, the files are divided into one-second segments, and all the recorded data are sorted according to whether they are the user's voice or other environmental sounds. However, it is impractical to manually label all of the audio data for one day. Therefore, this method reduces the cost of manual labeling by clustering the segmented audio data.

3.4 Feature Extraction

Features are extracted frame by frame from the clustered audio data. First, a window function is applied to the utterance waveform on

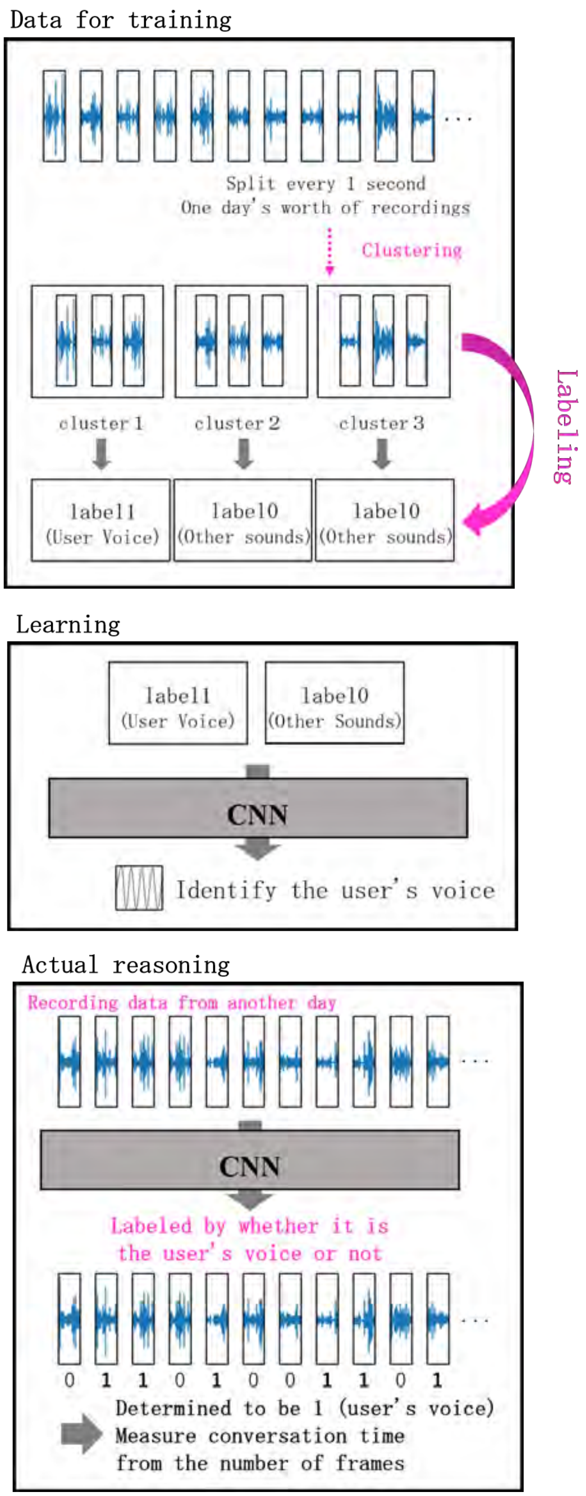

Figure 3: The flow of clustering and actual inference.The data are clustered and manually labeled. After that, the user's audio data is processed using a CNN.Then, the user's voice is inferred using a CNN.

the time axis and a short-time Fourier transform is performed to convert it to a frequency-domain waveform. Figure 2a shows a graph of the utterance in the time domain, and Figure 2b shows the graph after the short-time Fourier transform.

3.5 Identifying the user's voice using CNN

In this study, a CNN is used to identify the user's voice using the labeled data. An overview of the method is shown in Figure 3. Features can be learned by extracting features from the labeled voice data and writing them on a three-dimensional graph. This makes it possible to distinguish whether the voice is the user's voice or some other sound when inputting an audio file segmented into one-second segments.

3.6 Visualization of daily utterance amount on a smartphone

Finally, after calculating the user's daily utterance amount based on the above steps, the results are displayed on the user's smartphone. The system then uses a pre-trained model to determine whether the voice is the user's own and returns a graph showing the total amount of time the user has spoken and at what time. If the amount of utterance is low, the user is expected to feel a sense of urgency about the low amount of utterance through this graph.

4 Model Overview

In this section, an overview of the model is provide. The trained model was created using Google Colaboratory. We then built a server on Raspberry Pi. After that, we set up a server on Raspberry Pi. Then, we sent audio files recorded on Android to Raspberry Pi, and identified the recorded files on Raspberry Pi using the trained model. Finally, we returned the identified results to Android. The procedure for creating the trained model is described below.

4.1 Clustering

We used k-means++ for clustering. k=15 was chosen to reduce the cost of manual clustering considerably. Labeling is done by finding the class that contains the user's voice, and manually binary classifying the voice files sorted into that class. The user's voice is labeled as "Label 1" and other environmental sounds are labeled as "Label 2.

4.2 Feature extraction

Features are extracted from each of the clustered files. Spectral subtraction is used for noise reduction. After noise reduction, the spectra are converted into spectrograms that are aligned in the time direction.

4.3 Creating Training Data

After feature extraction, training data is created. By simultaneously learning spectrograms with white noise, which is noise with the same intensity reproduced at various frequencies, the system is able to identify the user's voice even in situations where the user's voice

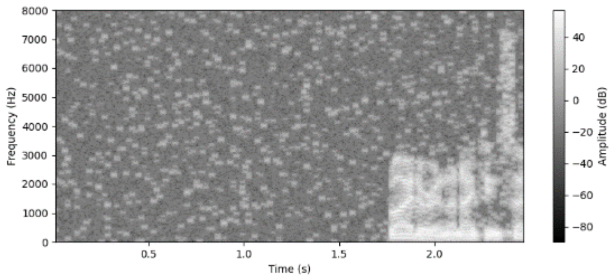

Figure 4a: White noise is removed and spectral subtraction is used to remove noise.Spectrogram with noise reduction using spectral subtraction

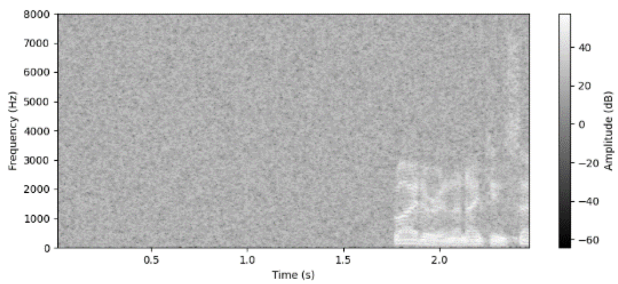

Figure 4b: Spectrogram with white noise

Table 1. Pre-questionnaire items

	Question	Evaluation
Q1	Often to utterance	1・2・3・4・5
Q2	Not stressed about utterance	1・2・3・4・5

Table 2. Items on the post-questionnaire

	Question	Evaluation
Q1	Crisis awareness about utterance amount	1・2・3・4・5
Q2	Increased utterance amount	1・2・3・4・5
Q3	Took action to increase utterance amount after being graphed	1・2・3・4・5
Q4	Stressed about being recorded	1・2・3・4・5
Q5	Did not feel stressed about utterance amount	1・2・3・4・5

is not recorded directly, such as when the smartphone is placed in the pocket of a pair of pants. Figure 4a shows the spectrogram without white noise trained on the CNN, and Figure 4b shows the spectrogram with white noise trained on the CNN. Since each of the utterance files identified as the user's voice is in units of one second, it is possible to calculate the amount of time the user speaks in a day. For example, if there are 7200 voice files identified as the user's voice, the time the user speaks in a day is 7200 seconds, or 2 hours.

4.4 Model

The model created in this study is built using the Sequential API of Keras. It consists mainly of a convolutional layer (Conv2D) and a pooling layer (MaxPooling2D). The first layer is the convolutional layer, which has 10 filters, a kernel size of 3*3, the padding is the same for the input and output space size, and the activation function is ReLU. The geometry of the layers is (image height, image width, number of channels) = (400, 1000, 3). Next, a pooling layer is placed. This is expected to prevent over-training of the model. The convolution and pooling layers are repeated four times in the same way. In each convolution layer, 10 filters are used and the ReLU function is used as the activation function; the Flatten layer needs to be in a form that can be input to all the coupling layers, i.e., it needs to be one-dimensional. Therefore, the feature maps obtained in the convolution and pooling layers are converted to one-dimensional vectors. The full-combining layer has 128 units, and the ReLU function is used as the activation function. Finally, the Dropout layer is set to 0.5 to prevent overlearning of the model. In this way, 50% of the nodes are randomly disabled during training.

5 Experiments

This section describes the evaluation experiments and their results.

5.1 Outline of Experiment

In this experiment, a questionnaire survey was conducted to confirm whether or not the subjects' awareness of the danger of utterance amount had improved. Eleven male and female subjects in their early twenties completed a pre-survey, and a post-survey was conducted after they had used PTAC for four days.

5.2 Experimental details

To demonstrate the effectiveness of the proposed method, an evaluation experiment was conducted with a total of 11 male and female subjects in their early 20s. Before the start of the experiment, we conducted a preliminary questionnaire. Then, the subjects were given a smartphone and recorded their conversations for one day. At the end of the first day of the experiment, the subjects were asked to perform a clustering task to extract their voices from the recorded files. After that, we created a learned model using a CNN model. The PTAC with the learned model was used to record conversations for the remaining three days of the experiment. On the last day of the experiment, a post-recorded questionnaire was administered, and the subjects' responses were used to evaluate the results.

Table3. Preliminary Survey Results

Subjects No.	1	2	3	4	5	6	7	8	9	10	11
Q1-1	4	3	5	2	2	3	5	3	3	3	3
Q1-2	3	4	4	2	3	2	4	3	3	5	5

Table 4. Results of post-survey

Subjects No.	1	2	3	4	5	6	7	8	9	10	11
Q2-1	4	4	4	5	3	4	5	4	3	5	4
Q2-2	2	3	5	1	3	2	4	2	3	3	4
Q2-3	3	2	2	2	1	2	2	2	4	2	3
Q2-4	2	4	4	4	3	4	2	3	2	3	2
Q2-5	1	1	3	1	2	3	4	3	2	2	1

Table 5. T-test for increased awareness of crisis in relation to utterance amount

Null hypothesis	PTAC is not effective in increasing awareness of low utterance amount PTAC is ineffective in increasing awareness of the danger of low utterance amount. The mean of the responses to Q2-1 is equal to 3.5
One-tailed hypothesis	PTAC is effective in increasing the sense of crisis about low utterance amount. PTAC is effective in increasing awareness of the danger of low utterance amount. The mean of the responses to Q2-1 is greater than 3.5.
T-distribution with 10 degrees of freedom (α=0.05)	1.812
P-value	2.797

Table 6. T-test for increased crisis awareness for utterance amount

Null hypothesis	PTAC has no effect on increasing utterance amount itself. The mean of the responses to Q2-2 is equal to 3.5.
One-tailed hypothesis	PTAC is effective in increasing the amount of utterance itself. The mean of the responses to Q2-2 is greater than 3.5
T-distribution with 10 degrees of freedom(α=0.05)	1.812
P-value	0.9424

5.3 Questionnaire items

The contents of the preliminary questionnaire are shown in Table 1. The subjects rated the items on a 5-point scale (1: not applicable, 2: not very applicable, 3: neither applicable nor not applicable, 4: a little applicable, 5: very applicable). Table 2 shows the contents of the post-questionnaire.

5.4 Questionnaire Results

The results of the pre-survey are shown in Table 3. The results of the pre-survey are shown in Table 3. The evaluation was made on a 5-point scale, as described in 5.3. The results of the post-questionnaire are shown in Table 4. In addition, a t-test was conducted to determine whether or not the introduction of the system improved the awareness of low utterance amount in the Q2-1 item. The results are shown in Table 5. In Q2-2, a t-test was conducted to determine whether or not the amount of utterance itself increased. The results are shown in Table 6.

5.5 Discussion

In this section, we discuss our findings based on the experimental results. The above results show that there is a significant difference in awareness of the amount of utterance that rejects the null hypothesis. On the other hand, the t-test was not rejected because the questionnaire survey on whether or not the amount of utterance itself had improved yielded low results. This is because PTAC displays a graph of the user's utterance amount on a smartphone, and if the user's utterance amount is low, the user needs to take some action himself/herself.

6 Conclusion

In this section, we discuss the conclusions and future prospects of this study.

In this study, we proposed a method of recording a user's daily conversation, extracting only the portion of the user's utterance amount from the recorded file, and displaying it on a smartphone. This method is intended to improve the user's awareness of the danger of utterance amount. From the results of evaluation experiments, it can be said that this system is effective in clarifying the user's daily utterance amount and visualizing it as a graph.

This research aimed to measure the improvement of crisis awareness by measuring the user's daily utterance amount and visualizing it as a graph. Therefore, there is a problem that users need to take action by themselves after becoming aware of a crisis. For example, if the amount of utterance per day is 20 minutes, the user needs to take action to increase the amount of utterance, such as contacting other people or talking to chatbots. Therefore, in the future, we would like to develop a system that can be expected to increase the amount of utterance itself by raising awareness of the danger of low utterance amount and then providing applications and systems that respond to that amount.

REFERENCES

[1] Cabinet Office, "Annual Report on the Ageing Society", 2018.

[2] H. L. Weeda, I. Dorsselaer, "Increase of depression among children and adolescents after the onset of the COVID-19 pandemic in Europe: a systematic review and meta-analysis," Child and Adolescent Psychiatry and Mental Health, 2022.

[3] H. Tujita," HappinessCounter: Enhancing Positive Mood and Communication with Smile-Encouraging Digital Appliances", 2012 Japan Society for Software Science and Technology.

[4] T. Levine, "How has COVID-19 affected the way we communicate?", University of Alabama at Birmingham, 2020.

[5] Y. X. Lukic, C. Vogt, O. Dürr and T. Stadelmann, "Learning embeddings for speaker clustering based on voice equality," 2017 IEEE 27th International Workshop on Machine Learning for Signal Processing (MLSP), Tokyo, Japan, 2017, pp. 1-6, doi: 10.1109/MLSP.2017.8168166.

[6] R. Jagiasi, S. Ghosalkar, P. Kulal and A. Bharambe, "CNN based speaker recognition in language and text-independent small scale system," 2019 Third International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Palladam, India, 2019, pp. 176-179, doi: 10.1109/I-SMAC47947.2019.9032667.

[7] B. F. P. Dossou and Y. K. S. Gbenou, "FSER: Deep Convolutional Neural Networks for Speech Emotion Recognition," 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), Montreal, BC, Canada, 2021, pp. 3526-3531, doi: 10.1109/ICCVW54120.2021.00393.

PInT: Person's Interruption Timing Sensor

Kodai Fujita
Department of Integrated Information
Aoyama Gakuin University
Sagamihara, Japan
koh@rcl-aoyama.jp

Masaki Ishizaka
Graduate School of Science and Engineering
Aoyama Gakuin University
Sagamihara, Japan
zacker@rcl-aoyama.jp

Yoshito Tobe
Department of Integrated Information
Aoyama Gakuin University
Sagamihara, Japan
y.tobe@rcl-aoyama.jp

ABSTRACT

When we are working or studying, interruptions occur from time to time. Once an interruption occurs, the thoughts of workers are completely removed from their work, and they have to spend additional time refocusing when they resume their tasks. Returning thoughts takes time and in some cases, it may not be possible to do so. Since multiple people are working in the same space, forcibly preventing interruptions from conversations is not easy, as it is likely to cause friction in human relationships. The main objective of this research is to reduce human-induced interruptions during work.

This research presents Person's Interruption Timing Sensor (PInT). PInT suggests whether it is acceptable to interrupt a worker at present by inferring the nature of the work from the sound of personal computer operation.

CCS CONCEPTS

• Human-centerd computing→Ubiquitous and mobile computing •Applied computing→Law, social and behavioral sciences

KEYWORDS

Support Vector Machine, Acoustic Signal Processing, Concentration Measurement

ACM Reference format:

Kodai Fujita, Masaki Ishizaka, and Yoshito Tobe. 2024 PInT: Person's Interruption Timing Sensor. *In Behavior Transformation by IoT Interrnational Workshop(BTIW'24), June 3-7, 2024, Minato-ku, Tokyo, Japan.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3662008.3662014

ACM ISBN 979-8-4007-0065-3/24/06...$15.00
https://doi.org/10.1145/366208.3662014

1 Introduction

In recent years, due to COVID-19, there has been an increase in the number of situations in which remote work and online tools are used to communicate with colleagues.

On the other hand, this does not mean that the traditional work environment, where people go to a company or laboratory to share space and work with others, has completely disappeared. Although many online communication tools have a 'read message' function, it is unclear whether these tools are truly aware of the content of the communication. In order to avoid this lack of awareness, the value of direct conversation with the person in question remains significant.

Talking to the other person when reporting and consulting among workers is necessary.

If the person spoken to by the other person is performing a task that requires following a complex train of thought, the conversation completely breaks the previous train of thought. The worker is forced to perform a different task when a thought-breaking interruption occurs, which is to return their thinking to the work they had been doing until just before the interruption. Especially when the worker is concentrating on their work, interruptions can seriously impair work efficiency and cause stress.

Several previous studies have proposed methods for estimating worker busyness and estimating the appropriate timing of interruptions. Research has already been conducted to estimate the busyness of workers based on information such as the number of keystrokes and mouse operations of workers using PCs, and to estimate the appropriate timing for interruptions based on breaks in PC work [1]. However, the systems in the preceding studies require the installation of cameras and a system for acquiring PC operations, making their practical use somewhat complicated.

This paper aims to construct an interruptible timing suggestion system that reduces interruptions caused by conversations. This achievement is accomplished solely using audio captured by a commercially available microphone, without directory installing cameras or acquiring PC operations.

2 Related Work

2.1 Uninterruptibility Estimation Method for Office Workers During PC Work

Various methods have been proposed for estimating the condition of workers. Some methods involve keyboard input or mouse operations, while others involve attaching sensors to the body. Additionally, certain approaches employ a combination of sensors, including microphones and cameras, to estimate the state of the workers [7][8]. Externally observable information, such as keyboard and mouse operation, allows for the assessment of the status of workers. Estimating the condition of the worker based on PC operation information mitigates the complexity and privacy concerns inherent in system implementation.

An applied study aims to estimate the states of workers by assessing interrupt rejection levels based on the number of PC operations [1]. PInT incorporates this approach, utilizing the degree of interrupt rejection derived from the study. Notably, interrupt rejection diminishes during task transitions, posing a challenge in comprehensively detecting such transitions. Application-switching (AS) can serve as a viable method for detecting task transitions from PC operations. However, the PInT environment excludes consideration of AS, as its primary focus is determining whether users may interrupt while engaged in a singular task.

In the same study, the period when work continues without application switching, defined as not application-switching (NAS), is identified [1]. The method for calculating interrupt rejection incorporates four characteristics that affect the change in interrupt rejection during NAS, as presented in the previous study. Table 1 presents four features influencing the extent of uninterruptibility during NAS, as documented in prior studies.

	Feature name
T	Manipulation occurred in the last 20 seconds
U	The operation rate of at least 20% within 2 minutes
V	The mouse and keyboard were used together within 2 minutes
W	The trasision from shell occurs within 5 minutes

Table 1: Effects of NAS time features

F_{NAS} expresses the estimation of the degree of uninterruptibility by applying the four features during NAS. F_{NAS} is calculated as follows:

$$F_{\mathrm{NAS}} = d \cdot T + U + V + W \quad (1)$$

where T, U, V, and W are the features listed in Table 1, all of which take only two values, 0 or 1. d is the weight on T, which is set to 2 here [1]. PInT utilizes an application of this formula.

2.2 Method for estimating PC typing sound

L. Zhuang et al. demonstrated that analyzing the sound produced by keyboard keystrokes can be employed to compromise passwords [2]. In their investigation, keyboard input was estimated through a K-means clustering method, with the value of K set to 30 or higher. The authors asserted that, following approximately 10 minutes of training, they achieved a 96% accuracy in inferring the content of typing.

PInT, in contrast, employs support vector machines (SVM) rather than K-means clustering. The method of this research will be used as a method for processing speech information.

Mel-frequency cepstral coefficients (MFCC) are one of the most widely used cepstral features in speech recognition and analysis. Empirically, cepstral features have been verified to be effective for speech signals, and in their study, L. Zhuang et al. used the first 16 MFCCs calculated for a Mel-Scale Filter Bank with 32 channels and a 10 ms window [2].

2.3 Cognitive load due to interruption not of human origin

The consumption of unexpected messages on recently popular social networking sites presents both benefits and disruptions to perceptions [3]. In a survey conducted among Twitter users, participants indicated that the platform's utility aligns more with their long-term goals rather than their short-term objectives [4].

The consensus among users is that the platform proves more advantageous for achieving prolonged objectives. Consequently, it is inferred that user engagement should not be disrupted by real-time notifications. Instead, consideration should be given to methods that prevent notifications as a system of proactive interruption (PInT) mechanism, and efforts should be directed towards presenting notifications at opportune moments.

2.4 A method to treat Uninterruptivility as Content-Awareness

Context awareness is a field of computer science that deals with adapting a computer system to the current context of the user. An entity is a person, place, or object that is considered relevant to the relationship between the user and an application [5]. The information obtained by PInT can be used as context and can be transmitted using Bluetooth and other methods as tag information for sharing content using SNS. As a concrete example, the interrupt rejection level can be used as the context shared in "Hyperlocal Communication," an experiment by A. Joly et al [6].

3 PInT Design

PInT is a system that combines applications that can be run on a Raspberry Pi and peripherals that indicate focus on tasks in the surrounding area, which can be operated by a single peripheral connected to the Raspberry Pi. This section describes the PInT system shown in Figure 1 from front to back.

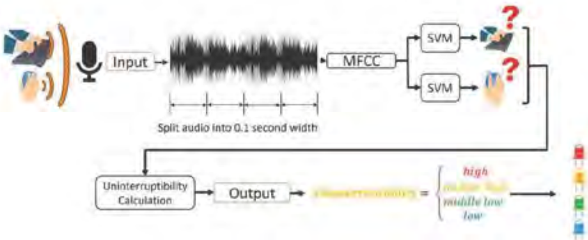

Figure 1: Overview of PInT

3.1 Audio processing

Audio is recorded on the Raspberry Pi through a USB microphone. Subsequently, the recorded audio is segmented into 0.1-second intervals and stored as WAV files. Before the analysis, all audio data will undergo conversion into a format conducive to ease of analysis, specifically through the utilization of a MFCC.

3.2 Develop a classification model using SVM

During training process, each 0.1-second segment of audio data is categorized based on whether it includes keystroke and click sound or not. PinT utilizes an SVM to generate two classifiers, distinguishing between typing sounds and mouse operation sounds.

3.3 Calculation of Uninterruptibility

Acquiring the uninterruptibility of the user from the estimated audio, PInT presents uninterruptibility information to the surroundings.

3.3.1 Definition of Uninterruptibility at NAS. The assumption in this research relies on the estimation of the degree of uninterruptibility during NAS, as well as the calculation of keyboard and mouse operations through guesswork. Consequently, the potential for estimation errors exists. To address this issue, PInT redefined uninterruptibility during NAS.

3.3.2 Obtain feature values as continuous values. As outlined in Section 3.3.1, previous studies defined each feature with only two values, 0 or 1 [1]. However, in the PInT environment, adopting a binary value for expressions would lead to significant ambiguity to uninterruptibility due to false positives. Therefore, PInT takes a continuous quantity between 0 and 1 and handles a transformed formula to reduce the effect of false positives.

In order to reduce the impact of false positives, the PInT defines a false positive incidence rate E calculated as follows:

$$E = 1 - p \quad (2)$$

where p represent the SVM estimation accuracy calculated during training.

Define a sigmoid function so that the features can be output as continuous real values using the false positive incidence rate E, calculated as follows:

$$S(x, E, M, a) = \frac{M}{1 + \exp(-a(x - E))} \quad (3)$$

where x is a variable that contains the number of keystrokes and mouse clicks estimated by the SVM and is subjected to normalization. The number M is an upper limit that must not be exceeded as an output, and x is normalized to a value between 0 and M, with E as the inflection point. a is the slope of the sigmoid function itself.

3.3.3 Calculation of operation rate. The operation rate in U is defined using data obtained during the learning phase. The maximum operation rate is defined as the quickness to perform a task at the time of learning. The maximum operation rate O_{MAX}, calculated as follows:

$$O_{MAX} = \frac{N_k N_c}{N} \quad (4)$$

where N signifies the total number of files acquired during training, N_k is the count of files with recorded keyboard keystrokes, and N_c is the count of files with recorded clicks.

Defining the maximum operation rate as representing 100% operation rate, the operation rate at which the SVM performs the estimation is calculated as follows:

$$O = \frac{D_{nk} + D_{nc}}{N \cdot O_{MAX}} \quad (5)$$

where the number of recorded keyboard keys is D_{nk} and the number of clicks is D_{nc}. However, since this value still has a large impact on the operation rate due to the misestimated files, the estimated operation rate O_e is calculated as follows:

$$O_e = \frac{S\big((D_{nk} + D_{nc}),\ E \cdot N,\ N,\ 1\big)}{N \cdot O_{MAX}} \quad (6)$$

where the variables are those used in equation (2) through (5). This formula is originally used to calculate the feature value V. Since the feature value V is not calculated in PInT, equation (6) is not used in subsequent calculations but is included for reference.

3.3.4 Uninterruptibility formula in PInT. The PInT employs T and U as features to determine uninterruptibility. T signifies whether an operation occurred in the recent tens of seconds, while U indicates whether the operation rate reached 30% or more within a 2-minute timeframe.

T calculated as follows:

$$T = S\left(\frac{N_{ks}}{N_s}, E \cdot N_s, 1.0, 1.0\right) \quad (7)$$

where the number of files acquired in the last few tens of seconds is N_s, the number of keyboard keystrokes estimated in the last few tens of seconds is N_{ks}, and the value obtained by dividing N_{ks} by N_s is substituted into x equation (3) as the operation rate, and the value obtained by multiplying the false positive rate E in equation (2) by N_s is substituted into E in Equation (3) as the number of false positive files to be estimated. The slope a is set to 1.0. The most recent time length obtained by N_{ks} is verified in an evaluation experiment as in previous studies [1].

U calculated as follows:

$$U = S\left(\frac{N_{ol}}{N_l}, 0.3, 1.0, 2.0\right) \quad (8)$$

Where the number of files retrieved in 2 minutes is N_l, and the number of operations within 2 minutes is N_{ol}, which is the sum of the number of keyboard presses and mouse clicks in 2 minutes. The slope, a, is set to 2.0. This choice is informed by the inclusion of data in U about both keystrokes and mouse clicks. Therefore, it does not lead to large deviations in feature values at the boundary of whether the operation rate exceeds 30% or not.

Combining T and U, we derive F_{NAS}, which is calculated as follows:

$$F_{NAS} = \frac{2 \cdot T + U}{3} \quad (9)$$

This value is the vary degree of uninterruptibility presented by PInT. In accordance with previous research, this formula is used to output the value as a probability value with a weight of 2 for T [1].

PInT presents information to the surroundings in four levels based on the results of the F_{NAS} caluculations. For this purpose, the thresholds for the four levels of interrupt rejection are calculated as follows:

$$Rejection\ Level = \begin{cases} high: & F_{NAS} \geq 0.7 \\ middle\ high: & 0.5 \leq F_{NAS} < 0.7 \\ middle\ low: & 0.2 \leq F_{NAS} < 0.5 \\ low: & F_{NAS} < 0.2 \end{cases} \quad (10)$$

This is a further division of the threshold used in the previous study, where the Rejection Level is in the middle [1].

4 Evaluation

This section outlines a validation experiment for the uninterruptibility estimation proposed by PInT. In this experiment, six students, four males, and two females, belonging to the Department of Integrated Information Technology, College of Science and Technology, Aoyama Gakuin University, were tested in a laboratory environment and evaluated in terms of the degree of interruption rejection defined in PInT Design.

4.1 Experiments to evaluate the method of calculating uninterruptibility estimation

4.1.1 Experiment environment. The experiment was conducted in a laboratory with a raspberry Pi 3 Model B and Laptop, as shown in Figure 2. A membrane keyboard with a loud keystroke noise was used, along with a common optical wired mouse. The keyboard and mouse were switched between training and evaluation, and the laptop was used for the evaluation task.

Figure 2: Experiment environment

Equation (10) guided the presentation of colors on a LED connected to the Raspberry Pi, as depicted in Figure 3. Uninterruptibility exhibited an ascending order of improvement: blue, green, yellow, and red.

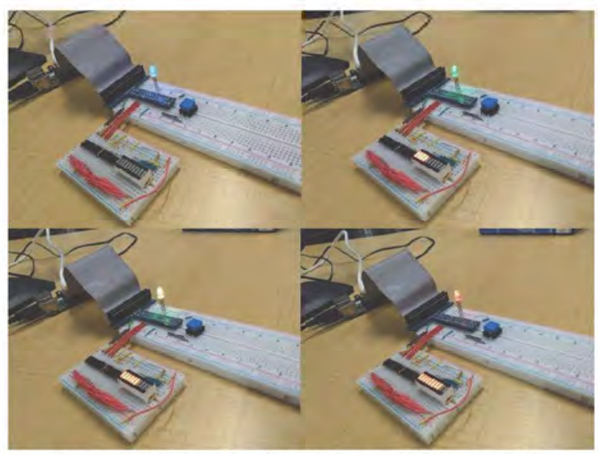

Figure 3: PInT presents uninterruptibility on LED.

4.1.2 Collection of training data. The subjects were asked to perform a task that lasted at least 60 seconds in total and included 100 keyboard keystrokes and 50 mouse clicks, which were used as training data. The task consisted of dictating all

Japanese text of 50 to 100 characters using typing, and then playing Minesweeper, a game that is completed only by clicking, about three times.

4.1.3 Task content at the time of evaluation. Upon completion of the estimator utilizing the training data from the preceding section (4.1.2), the subsequent task involves assessing the degree of uninterruptibility using the Web browser. This evaluation entails performing various tasks such as writing text, playing Minesweeper, and conducting searches on the Web. It is important to note that tasks are to be exclusively carried out within confines of the Web browser.

4.1.4 Evaluation Method. The acquisition range for N_{ks} and N_s employed in deriving equation (7) is initially set at 20 seconds. A total of three experiments were conducted, where in the acquisition time, initially set at 20 seconds, was varied to 10, 20, and 30 seconds. Participants were then requested to subjectively rate, on a 5-point scale (1: not valid, 2: somewhat valid, 3: indifferent, 4: somewhat valid, 5: valid), the appropriateness of uninterruptibility for the given work situation while engaged in the task.

4.2 Evaluation

The results obtained in the evaluation experiment are shown in Figure 4. The horizontal axis is the value answered, and the vertical axis is the number of respondents.

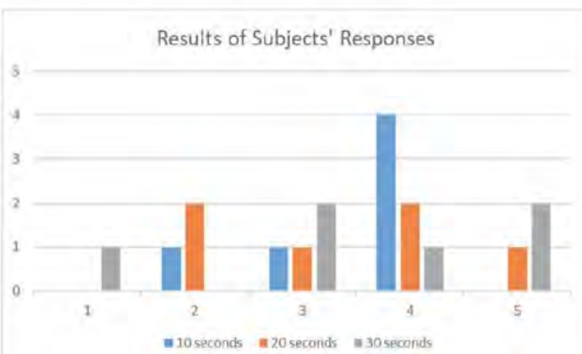

Figure 4: Results of Subjects' Responses.

The validity of different lengths of evaluations is assessed by employing a t-test in equation (7). The comparisons and verifications of results are conducted for acquisition times of 10, 20, and 30 seconds, denoted as V1 for the comparison between 10 and 20 seconds, V2 for the comparison between 20 and 30 seconds, and V3 for the comparison between 20 and 30 seconds, and the results of each t-test are shown in Tables 2 to Table 4.

Null hypothesis	There is no difference in the acquisition time used in equation (7) whether it is 10 or 20 seconds.
Alterative hypothesis	Significant difference between acquisition times of 10 and 20 seconds.
t-value	0.237022732
Critical value	2.015048373
Results	Based on |0.237022732| < |2.015048373|, the null hypothesis cannot be rejected

Table 2: Results for V1

Null hypothesis	There is no difference in the acquisition time used in equation (7) whether it is 20 or 30 seconds.
Alterative hypothesis	Significant difference between acquisition times of 20 and 30 seconds.
t-value	-0.23702
Critical value	2.015048
Results	Based on |-0.23702| < |2.015048|, the null hypothesis cannot be rejected

Table 3: Results for V2

Null hypothesis	There is no difference in the acquisition time used in equation (7) whether it is 10 or 30 seconds.
Alterative hypothesis	Significant difference between acquisition times of 10 and 30 seconds.
t-value	0
Critical value	2.015048
Results	Based on |0| < |2.015048|, the null hypothesis cannot be rejected

Table 4: Results for V3

4.3 Consideration

The results of the t-test showed no significant difference in the acquisition time used in equation (7). Since significant differences appeared in similar experiments conducted in previous studies, the method used by PInT to calculate uninterruptibility is considered to inaccurate. The reasons for this result include the inaccuracy of the calculation of uninterruptibility due to the low estimation accuracy of the SVM itself, and the lack of training and evaluation time in the experiment. The Fnas values measured in this study are based on factors other than keyboard and mouse operation sounds as features to be extracted from the feature audio during the operation of the same application. The Fnas values measured in this study are based on the keyboard and mouse operation sound elements as features to be extracted from the feature audio during the same application operation. The main purpose of the reference paper is to detect conversations in addition to PC operations, and it seems that the discussion of factors that may affect uninterruptibility in NAS has not been fully discussed. Therefore, the accuracy of calculating uninterruptibility may be improved if we can find additional elements in the voice as features to calculate uninterruptibility during NAS.

5 Conclusion and Future Work

This research proposed PInT, a system designed to assess uninterruptibility solely through audio captured by a commercially available microphone, presenting this information to individuals in the vicinity. Although PInT did not achieve results comparable to the system in the prior study, we posit that enhancing the prediction accuracy of typing and clicking sounds, expanding the scope of estimate features, and detecting task breaks can significantly enhance uninterruptibility calculations. In addition, if the presence or absence of human conversation and the factor of body posture due to sound waves could be added as features used during NAS, it would have a significant impact on the calculation of the degree of no-disturbance. Therefore, we would like to improve the estimation accuracy and increase the number of features to be estimated in the future.

REFERENCES

[1] S.Hashimoto, T.Tanaka, K.Aoki, K.Fujita "Estimation of Interruptibility during Office Work Based on PC Activity and Conversation", HIMI Conference paper pp. 297-306 (July 2013).

[2] L. Zhuang, F. Zhou, J. D. Tygar, "Keyboard Acoustic Emanations Revisited", In Proceeding of the 12th ACM Conference on Computer and Communications Security, pp. 373-382 (November 2005).

[3] Y. Miyata, D. A. Norman, "Psychological issues in support of multiple activities" User centered system design, pp.265-284 (January 1986)

[4] A. Joly, P. Maret, J. Daigremont, "Between social awareness and productivity: Results of survey about real-time microblogging" First Monday 15(11) (November 2010)

[5] B. Schilit, N. Adams, R. Want, "Context-Aware Computing Applications" IEEE Workshop on Mobile Computing Systems and Applications (WMCSA) pp.85-90 (1994).

[6] A. Joly, P. Maret, J. Daigremont "Context-Awareness, The Missing Block of Social Networking" International Journal of Computer Science and Applications, Vol. 6, No. 2, pp.50-65 (March 2009)

[7] S. E. Hudson, J. Fogarty, C. G. Atkeson, D. Avrahami, J. Forlizzi, S. Kiesler, J. C. Lee, J. Yang "Predicting Human Interruptibility with Sensors: A Wizard of Oz Feasibility Study", Proc. SIGCHI conf. on Human Interruptibility Study, Proc. SIGCHI conf. on Human factors in computing systems, pp. 257-264 (2003).

[8] T.Tanaka, R.Abe, K.Aoki, F.Fujita, "Interruptibility Estimation Based on Head Motion and PC Operation", Intl Journal of Human-Computer Interaction, 31(3) 167-179 (March 2015)

Laterality of arm movement variability on copying and tracing

Mikimasa Omori
Faculty of Human Sciences
Waseda University
Saitama, Japan
mikkiomori@waseda.jp

INTRODUCTION

Handwriting is a complex behavior requiring the writer's visual, motor, and cognitive integration skills. Previous studies using motion analysis [1,2] have shown that children with handwriting difficulties (HD) showed more vertical variation in the writing arms than typically developing children during figure writing tasks. However, these studies did not evaluate the laterality of both sides of arms during handwriting. Moreover, previous studies [1,2] did not used figure tracing task and this task would help us find the arm movement variabilities without considering the visuo-motor integration difficulties. To investigate the laterality of arm movement variabilities, I asked participants to copy and trace the figures while their body movement was recorded with motion analysis software.

CCS CONCEPTS

• Applied computing → Law, social and behavioral sciences→ Psychology

KEYWORDS

Arm movement variabilities, Motion analysis, Laterality

ACM Reference format:

Mikimasa Omori. 2024. Laterality of arm movement variability on copying and tracing. In *Behavior Transformation by IoT International Workshop (BTIW'24), June 3-7, 2024, Minato-ku, Tokyo, Japan. ACM, New York, NY, USA, 2 pages.* https://doi.org/10.1145/3662008.3662015

Method

Participant. Seven children with chronological ages (CA) ranging from 4 to 6 years participated in the study. All of them haven't diagnosed as any developmental disabilities. However, they have some problem behavior during nursery school activity especially in writing and drawing task.

Stimuli and Apparatus. I prepared eight simple figures including circle, diamond, and cycloids. I also used the Beery-VMI6 to evaluate participants' visuo-motor integration skills. To analyze the body movement, I used VisionPose 3D software and recorded the video camera during handwriting task.

Procedure. Participant started with either copying or tracing task. In copying task, participant wrote the same figure presented on the computer whereas tracing task required them to trace the guided line of the presented figure.

Dependent variables. (1) Response time (RT) of writing figures (2) Coefficient of variation (CoV) on handwriting movement. CoV is a quotient of standard deviation and mean scores and as the CoV increases, the variability increases.

Data Analysis. (1) To compare the RT of two tasks, I used the paired t-test. (2) To analyze the stabilities of handwriting movement, I conducted the analyses of variance (ANOVAs) using a 2 (hand: right vs left) × 2 (tasks: copying vs tracing) × 3 (direction: horizontal vs vertical vs depth) design.

Results & Discussion

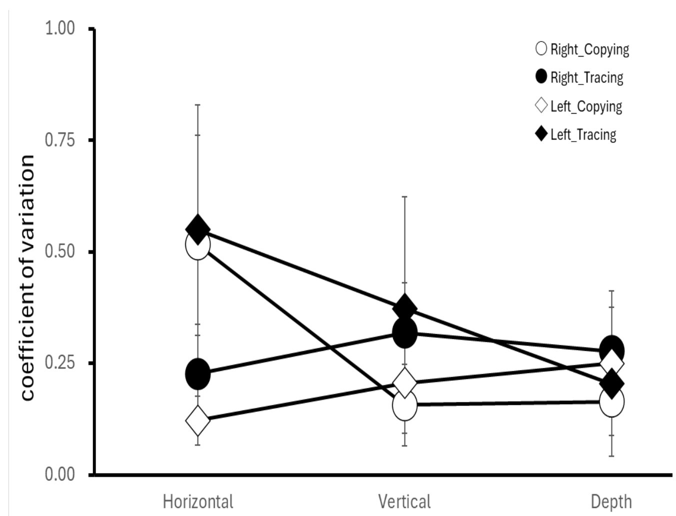

Figure 1: Coefficient of variation on wrist movement during writing figures

BTIW '24, June 3–7, 2024, Minato-ku, Tokyo, Japan

ACM ISBN 979-8-4007-0665-3/24/06…
https://doi.org/10.1145/3662008.3662015

In response time of writing figures, participants took longer to trace figures than copying them [t (6) =3.93, p<.005]. I then analyzed the arm movement and Figure 1 shows the coefficient of variation on wrist movement during two types of writing tasks. Based on 3-way ANOVA, I found three-way interactions [F (1, 6)= 71.10, p <.001, η^2=.92] showing less stable right wrist movement toward vertical and depth direction in tracing than copying figures. Moreover, students with lower VMI scores showed the unstable writing arm movement. Therefore, analyzing the body movement during handwriting with IoT methods would help us find the students with handwriting difficulties by making objective criteria.

ACKNOWLEDGMENTS

The Waseda University Institutional Review Board approved (No. 2022-513) this study. KAKENHI (grant number JP22K13739).

REFERENCES

[1] Lopez, C., and Vaivre-Douret, L, 2021. Influence of visual control on the quality of graphic gesture in children with handwriting disorders. *Scientific Reports*, *11*, 23537. Doi: https://doi.org/10.1038/s41598-021-02969-7

[2] Lopez, C., & Vaivre-Douret, L, 2023. Concurrent and predictive validity of a cycloid loops copy task to assess handwriting disorders in children. *Children*, *10*, 305. Doi: https://doi.org/10.3390/children10020305

How is our mobility affected as we age? Findings from a 934 users field study of older adults conducted in an urban Asian city

Yi Zhen Tan
yizhen.tan.2021@phdcs.smu.edu.sg
School of Computing and Information Systems, Singapore Management University
Singapore

Thu Tran
ndttran.2019@phdcs.smu.edu.sg
School of Computing and Information Systems, Singapore Management University
Singapore

Sapphire Lin
slin@nus.edu.sg
Centre for Population Health Research and Implementation, SingHealth
Duke-NUS Medical School
Singapore

Fang Zhao
fangzhao11@gmail.com
Singapore University of Technology and Design
Mobile Market Monitor
Singapore

Yee Sien Ng
ng.yee.sien@duke-nus.edu.sg
Duke-NUS Medical School
Singapore General Hospital, SingHealth
Singapore

Dong Ma
dongma@smu.edu.sg
School of Computing and Information Systems, Singapore Management University
Singapore

Jeonggil Ko
jeonggil.ko@gmail.com
School of Computing and Information Systems, Singapore Management University
Singapore

Rajesh Balan
rajesh@smu.edu.sg
School of Computing and Information Systems, Singapore Management University
Singapore

ABSTRACT

In this paper, we analyze the results of a large study involving 934 older adults living in an urban Asian city that collected their mobility patterns, in the form of logged GPS data, along with a multitude of demographic and health data. We show that mobility, in terms of average distance travelled per day, is greatly affected by age and by employment status. In addition, other factors such as type of day, household size, physical and financial conditions and the onset of retirement also play a significant role in determining the mobility of an individual. These results will have high value to any researcher understanding and attempting to transform the lifestyle of older adults.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in collaborative and social computing**; • **Applied computing →** Law, social and behavioral sciences.

ACM Reference Format:
Yi Zhen Tan, Thu Tran, Sapphire Lin, Fang Zhao, Yee Sien Ng, Dong Ma, Jeonggil Ko, and Rajesh Balan. 2024. How is our mobility affected as we age? Findings from a 934 users field study of older adults conducted in an urban Asian city. In *Behavior Transformation by IoT International Workshop (BTIW '24), June 3–7, 2024, Minato-ku, Tokyo, Japan.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3662008.3662016

BTIW '24, June 3–7, 2024, Minato-ku, Tokyo, Japan

ACM ISBN 979-8-4007-0665-3/24/06.
https://doi.org/10.1145/3662008.3662016

1 INTRODUCTION

As projected in a World Health Organization (WHO) report [18], over 16% of the global population will be aged 60 or older by 2030. This demographic shift necessitates a reevaluation of infrastructure design.

There are needs to gain deeper insights into how an older adult population residing in urban settings adapt to diverse socio-economic factors and how these elements shape their daily mobility within their living environments. However, conventional census approaches may fall short in capturing the micro-level mobility patterns of the aging population. We propose that leveraging mobile platforms with their inherent sensing capabilities can serve as a viable solution to this challenge. Given the widespread adoption of mobile devices, these devices can effectively monitor user mobility and engage with users to compile detailed datasets regarding the how and why behind individual mobilization. This data holds the potential to furnish policymakers with valuable insights into evolving urban behaviors.

In this research, we conduct an analysis of data obtained from an extensive study involving more than 934 senior individuals aged 50 and above residing in an urban Asian metropolis. Study participants were requested to install a mobile application on their devices, which recorded their GPS locations over two weeks. We complemented this information with a large amount of demographic details and self-reported data.

The main hypothesis of the paper is that the mobility patterns of older adults change, and in particular their mobility decreases, as they age. We reconfirm this well-known hypothesis and go deeper to understand the factors that affect the mobility of older adults. In particular, what are the factors that make older adults more or less sedentary compared to their peers? Our analysis suggests that older adults that have are younger, employed and married exhibit higher levels of mobility.

Beyond basic demographic factors, we also look at the impact of type of day, household size, retirement and physical and financial conditions as well as mode of transportation has on the mobility of older adults.

We strongly believe that the outcomes of this study can provide value to researchers engaged in the design of cities and living environments tailored to older populations [12]. Specifically, these findings will facilitate a nuanced understanding of the influence that diverse public transportation and housing choices have on the mobility patterns of seniors residing in those areas. Furthermore, these findings carry significant medical implications [12], as a substantial body of prior research has established a direct correlation between sedentary behavior and various medical conditions [10]. By pinpointing the determinants of mobility, our aim is to supply data evidence that can be harnessed to mitigate these factors. This may involve devising initiatives to incentivize retirees to engage in outdoor activities, ultimately enhancing the overall mobility of the older population and, consequently, yielding improved medical outcomes.

2 RELATED WORK

2.1 Mobility patterns from GPS data

Many features of mobility patterns can be extracted from GPS data, including the locations of significance, modes of transport, trajectory patterns and location-based activities [4, 8, 11]. Our study makes use of the average GPS distance travelled as well as modes of transport.

Previous research involved providing participants with a dedicated wearable GPS device to track their mobility patterns during travel [2, 15, 23]. However, this approach encounters challenges related to device management and usage, such as charging or alteration of participants' routines [7, 13]. In comparison, leveraging on smartphone GPS data proves advantageous as most participants habitually charge and carry their smartphones, thereby mitigating the effort and errors associated with data collection challenges [6, 13]. Our study was conducted in a urban setting with dense public transportation coverage and a high rate of smartphone usage, even among older adults. This setting facilitated the non-intrusive collection of comprehensive GPS mobility traces of older adults.

2.2 Mobility of older adults

Research on the mobility of older adults has gained significant attention, leading to the formulation of theoretical and empirical assessment frameworks that reveal its complexity and interdisciplinary nature [14, 16].

Among various travel modes, active travel, such as walking and cycling, has been of special interest due to its direct relevance to the physical activity of older adults. Notably, a robust relationship has been identified between the active travel of older adults and the physical environment of their neighborhoods [1, 17].

Index	Variable	Value	No. of participants (n)	Percentage
1	Gender	Female	621	66.5%
		Male	313	33.5%
2	Age	<60	482	51.6%
		60-70	340	36.4%
		>70	112	12.0%
3	Marital Status	Married	648	69.4%
		Not Married	286	30.6%
4	Employment Status	Full-time/Part-time	501	53.6%
		Not working/retired	433	46.4%
5	Income Adequacy	Some/Much difficulty	413	44.2%
		Just enough	395	42.3%
		Enough with leftover	99	10.6%
		Did not indicate	27	2.9%
6	Frail by any test	None	691	74.0%
		At least 1	243	26.0%
7	Household Members	1	120	12.8%
		2	253	27.1%
		3	214	22.9%
		4	212	22.7%
		>=5	135	14.5%

Table 1: Demographics Information

Hirsch et al. analyzed GPS data from 95 older adults in Canada, investigating how environmental and demographic variables influenced geographical mobility [5]. Larger-scale studies examining the mobility patterns of older adults commonly rely on survey data, such as telephone interviews [9, 19–21]. However, data collected on daily trips using a recall survey method can have limitations due to self-report bias.

Our study leverages the ubiquitous nature of mobile phones and their integrated GPS tracking capabilities to amass a notably large dataset from a diverse population of urban older adults, contributing further empirical evidence to the understanding of previously investigated mobility patterns in older adults.

3 DATASET

The data used in this extensive study combined a range of baseline survey evaluations covering aspects related to health, social and environmental interactions with geo-spatial data collected over a 14-day time period via a mobile GPS tracking application named X-ING [22], installed on the participants' mobile phones – a commercial app that supported both Android and iOS devices. The study was approved by a medical IRB panel run by a large hospital group located in the Asian city the study was conducted in.

3.1 Participants

A non-random sample of community-dwelling adults aged above 50 (N = 934) were recruited through a combination of physical and digital posters distributed across various community locations serving older adults. The demographics and percentage breakdown of the participants is presented in Table 1. We use these attributes in the analysis presented in Section 4.

3.2 Data Collection

Our data collection took place at four community sites across the Asian city from April to September 2022. Participants were asked to engage in a series of three activities, which included answering a questionnaire, performing a series of tests to evaluate their physical condition and given instructions to install and use the X-ING GPS logging application on their mobile phones.

After the session was done, the participants continued with their daily routines while keeping track of their travels for the subsequent 14 days. The attrition rate was less than 1%, with 934 out of the 942 recruited participants successfully completing the study (and their data being used for this paper).

3.3 Travel Diaries

The travel logs captured by the X-ING app recorded the start time, end time, and GPS distance traveled for every trip taken by our participants. In addition, the mode of transport and reason for each trip was also logged. The available options for travel mode are included in Table 2.

Index	Transport Mode	Description
1	Foot	Walk by foot
2	PV Driver	Travel in private vehicle (car or motorcycle) as the driver
3	PV Passenger	Travel in private vehicle (car or motorcycle) as the passenger
4	Taxi/Car Service	Travel in paid taxi or private car hire service
5	Bus	Travel on bus
6	Train	Travel on the railway system
7	Bicycles	Ride with traditional bicycle
8	E-Bike	Ride with electric bicycle. This will be considered "Other" in the analysis due to the small sample size.
9	Wheelchair	Travel with Personal Mobility Device (WheelChair). This will be considered "Other" in the analysis due to the small sample size.
10	E-Scooter	Travel with Personal Mobility Device (E-Scooter). This will be considered "Other" in the analysis due to the small sample size.
11	Other	Any travel that does not match all of the above modes.

Table 2: Travel mode options and description

4 FACTORS AFFECTING MOBILITY

In this section, we analyse the effect that various factors has on the mobility of older adults. These results reinforce conventional wisdom with additional insights provided by our large dataset.

4.1 Base Results

Our base hypothesis is that mobility decreases as age increases. Here we define mobility as the average recorded distance (in kilometers) traveled per day. The left plot in Figure 1 reveals a declining trend in mobility as individuals age. We also observe that those who are not employed exhibit shorter travel distances in comparison to their employed counterparts. These findings align with our initial expectations, as the natural aging process tends to reduce an individual's physical capability, and the absence of regular employment reduces the financial input and eliminates a significant source of daily mobility. In the right plot of Figure 1, we present these two factors combined, corroborating the validity of both age and employment status as distinct factors affecting mobility.

4.2 Weekday mobility vs. Weekend mobility

An individual's mobility frequently follows a weekly pattern, given the common 5-day weekday and 2-day weekend structure. To examine the effect of such a weekly pattern, in Figure 2, we present the average daily GPS distance traveled separated by weekdays and weekends/public holidays, considering the participants' employment status. This distinction is important as employed participants typically have regular work-related travel on weekdays, as opposed to non-employed participants. As depicted in the plots, the disparities in mobility between employed and non-employed participants are more pronounced on weekdays compared to weekends. We observe a slight reduction in mobility during weekends (and public holidays) for the employed group, and not for the non-employed group. Furthermore, an important observation is that even on weekends, the employed population exhibits higher mobility, suggesting that maintaining employment for older adults can serve as an effective strategy to sustain their mobility.

4.3 Household size

We delve deeper into the data using Figure 3, where we depict the average GPS mobility for various age groups, differentiating between weekdays/weekends, and the number of household members.

The plots in Figure 3 reveal that larger households display an increasing trend in mobility, with this effect being more pronounced in the age group below 60. We believe that this is an effect of this group consisting of households with younger members, tending to be more active as a family. Secondly, as participants age, along with a higher likelihood of household members aging correspondingly, the overall travel distance patterns exhibit a gradual decline. Except for some cases with only a small number of samples to analyze (age group >70), we see a common trend in mobility patterns of increased mobility with more household members.

4.4 Marital status

Next, we investigate the effect of marital status and gender on GPS mobility by categorizing participants into different marital status groups, as depicted in Figure 4. The "Not Married" category encompassing individuals who are single, separated, or divorced. Interestingly, the data suggests that, on average, married participants exhibit higher levels of mobility compared to those who are not married.

4.5 Physical and financial conditions

The financial and physical well-being of older individuals can exert a significant influence on their mobility, as activities outside home often entail financial expenses and physical exertion. We looked at the average daily GPS distance traveled with respect to different age groups and their self-reported income adequacy levels, categorized into three levels: some/much difficulty, just enough, and enough with left over. We observed a positive relationship between mobility and income adequacy, indicating that greater financial stability is associated with increased mobility.

Physical motor abilities decreases as people age. We carried out four different frailty tests, Fried, FRAIL, HGS, and GS [3], classifying participants as "frail" if they exhibited frailty in at least one of these tests. The results shown in Figure 5 affirm that physical frailty significantly diminishes participants' mobility, and the degree of decline tends to escalate in higher age groups. This offers strong confirmation that supporting the physical condition of the older population can be highly effective in preserving and promoting the mobility of older individuals in our society.

Another aspect that indirectly influences expected income is the number of years since retirement. In Figure 6 we present the average GPS travel distance with respect to the number of years

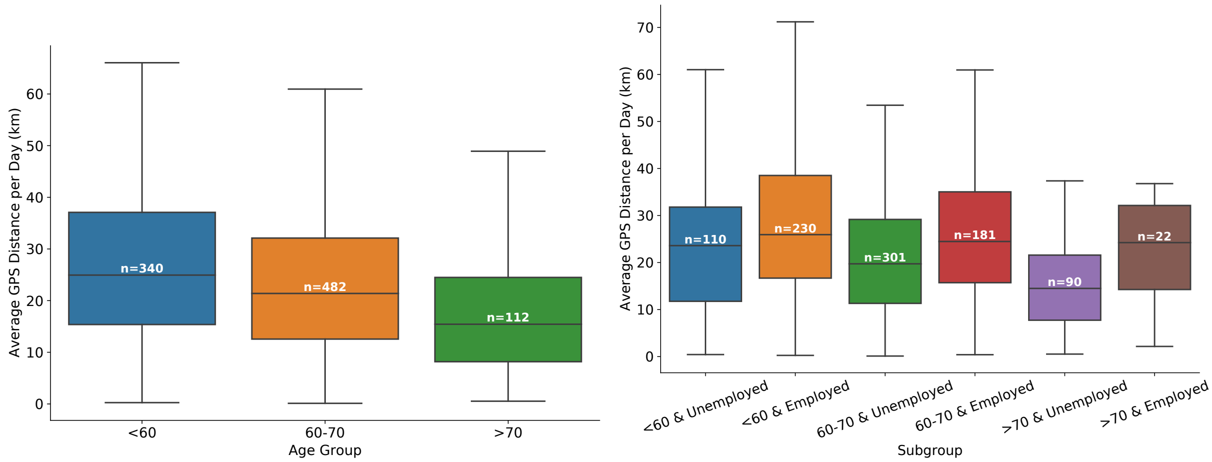

Figure 1: Relationship of Average GPS Distance with Age and Employment

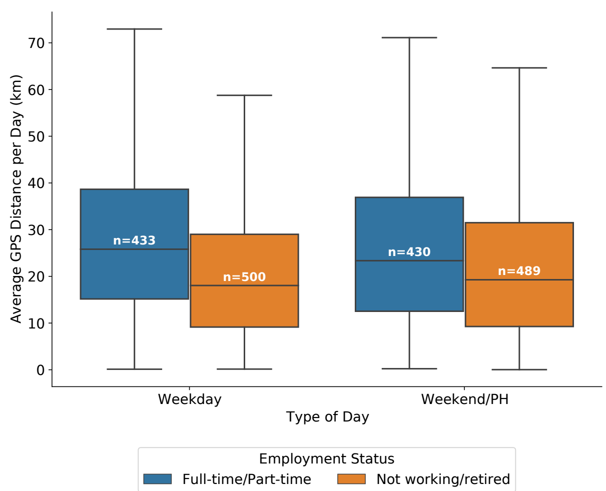

Figure 2: Effect of Type of Day on Average GPS Distance, separated by employment status

since their retirement, only including participants that indicated that they were "retired" (363 subjects). We notice that while the first four years after retirement exhibit similar mobility distances, after passing five years of retirement, there is a noticeable decline in the distance traveled. This suggests that long-retired elders might be less likely to engage in longer travels and this should be taken into consideration by planners.

4.6 Mode of Transport

In a typical urban environment, numerous public transportation methods such as buses and subways coexist with private vehicles and hiring services. The city in this study offers a densely connected bus and subway system where the buses are used for both long distance trips and to connect small neighborhoods with nearby subway stations. The city also offers numerous hired car and taxi services and has a very well deployed road infrastructure.

We found that walking was the most commonly employed mode of transport, with one of the shortest average travel distances. While bicycle trips had longer distances, they were still fairly short compared to other modalities such as the use of personal vehicles or other transportation services. Surprisingly, the average distance traveled by bus was also relatively modest. This might be because buses were frequently used to reach the nearest subway station for longer-distance journeys.

An intriguing finding is that, although the average travel distance is similar for personal vehicles, taxi/car hiring services, and subways/trains, having access to a personal vehicle significantly boosts the number of trips taken by a participant. On average, each of the 283 drivers engaged in 42 driving-based journeys, while subways/trains were only used approximately 11.4 times per participant. Furthermore, it is noteworthy that passengers (not driver) of personal vehicle, despite having access to a personal vehicle, undertook an average of 11.2 journeys, nearly a quarter of the number observed for drivers.

5 CONCLUSION

Mobility patterns, encompassing both transportation modes and distances traveled by individuals, serve as important indicators that reflect people's lifestyles and travel behaviours. Nevertheless, accurately capturing these patterns is challenging, demanding extensive, large-scale, and multi-dimensional data collection.

In this study, we compiled a comprehensive dataset using a mobile application, incorporating GPS tracking information and self-reported trip diaries from 934 seniors residing in an urban Asian metropolis. Through accounting for various factors that related to the senior's living conditions, we unveiled several intriguing findings. Notably, we observed a decline in mobility and a shift in preference towards public transportation among aging individuals, being influenced by various societal and financial factors. These insights provide valuable guidance for potential societal interventions aimed at fostering increased mobility among the elderly.

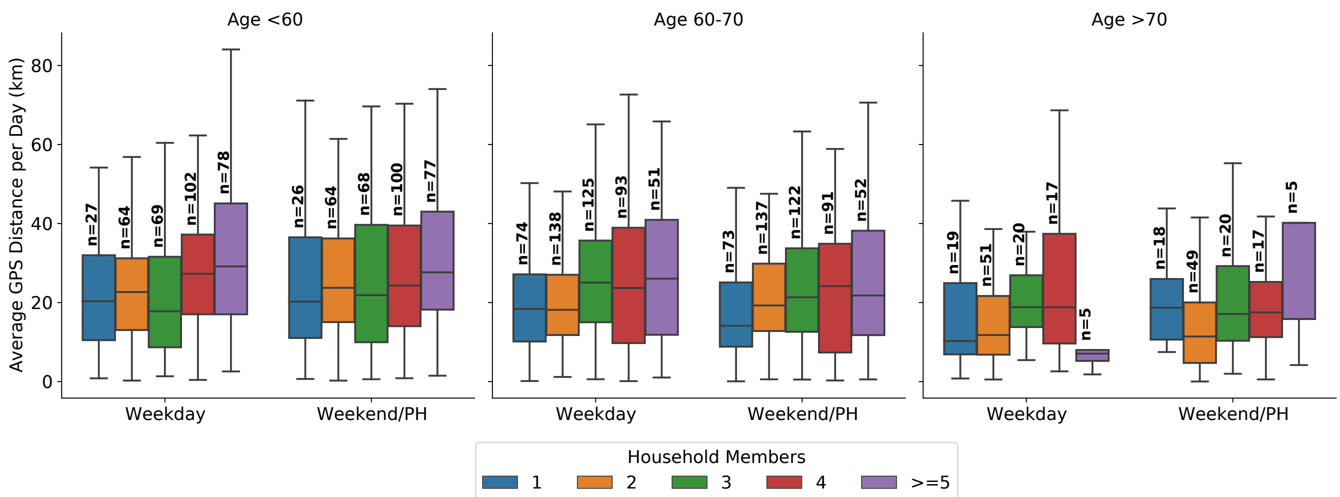

Figure 3: Effect of Type of Day on Average GPS Distance for different number of household members

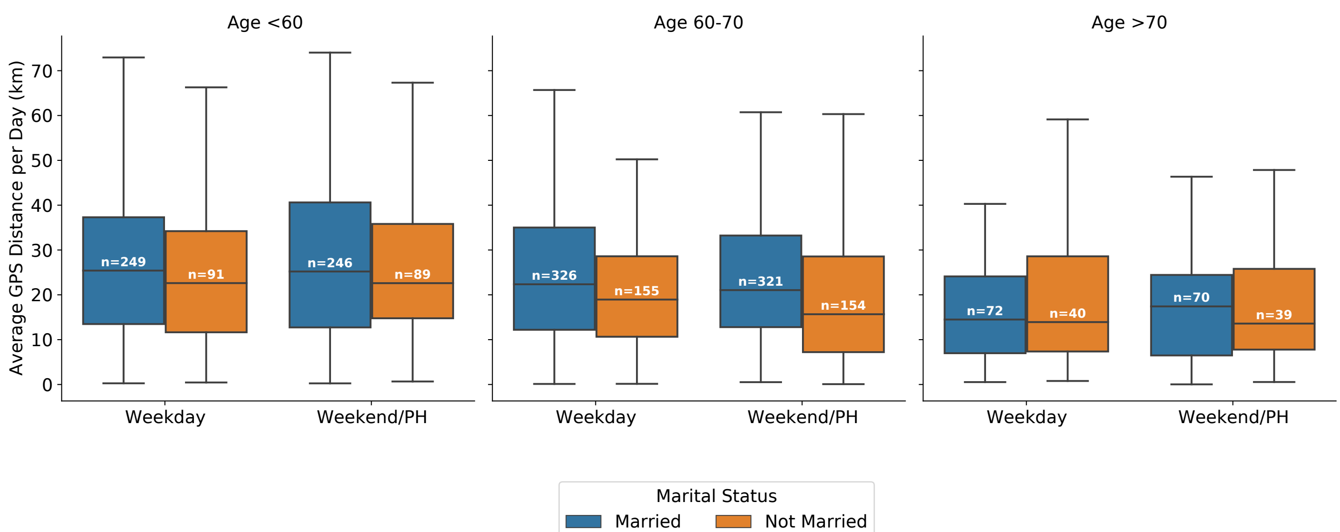

Figure 4: Effect of Type of Day on Average GPS Distance for different marital status

ACKNOWLEDGEMENTS

This research/project is supported by the National Research Foundation, Singapore, and Ministry of National Development, Singapore under its Cities of Tomorrow R&D Programme (CoT Award COT-H1-2020-1).

Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not reflect the views of National Research Foundation, Singapore and Ministry of National Development, Singapore.

This research was supported by the Singapore Ministry of Education (MOE) Academic Research Fund (AcRF) Tier 1 grant.

REFERENCES

[1] E. Cerin, A. Nathan, J. Van Cauwenberg, D. W. Barnett, A. Barnett, C. on Environment, and P. A. C.-O. A. working group. The neighbourhood physical environment and active travel in older adults: a systematic review and meta-analysis. *International journal of behavioral nutrition and physical activity*, 14:1–23, 2017.

[2] A. R. Cooper, A. S. Page, B. W. Wheeler, P. Griew, L. Davis, M. Hillsdon, and R. Jago. Mapping the walk to school using accelerometry combined with a global positioning system. *American journal of preventive medicine*, 38(2):178–183, 2010.

[3] E. Dolenc and D. Rotar-Pavlič. Frailty assessment scales for the elderly and their application in primary care: A systematic literature review. *Slovenian Journal of Public Health*, 58(2):91–100, mar 2019.

[4] F. Giannotti, M. Nanni, D. Pedreschi, F. Pinelli, C. Renso, S. Rinzivillo, and R. Trasarti. Unveiling the complexity of human mobility by querying and mining massive trajectory data. *The VLDB Journal*, 20:695–719, 2011.

[5] J. A. Hirsch, M. Winters, P. Clarke, and H. McKay. Generating gps activity spaces that shed light upon the mobility habits of older adults: a descriptive analysis. *International journal of health geographics*, 13(1):1–14, 2014.

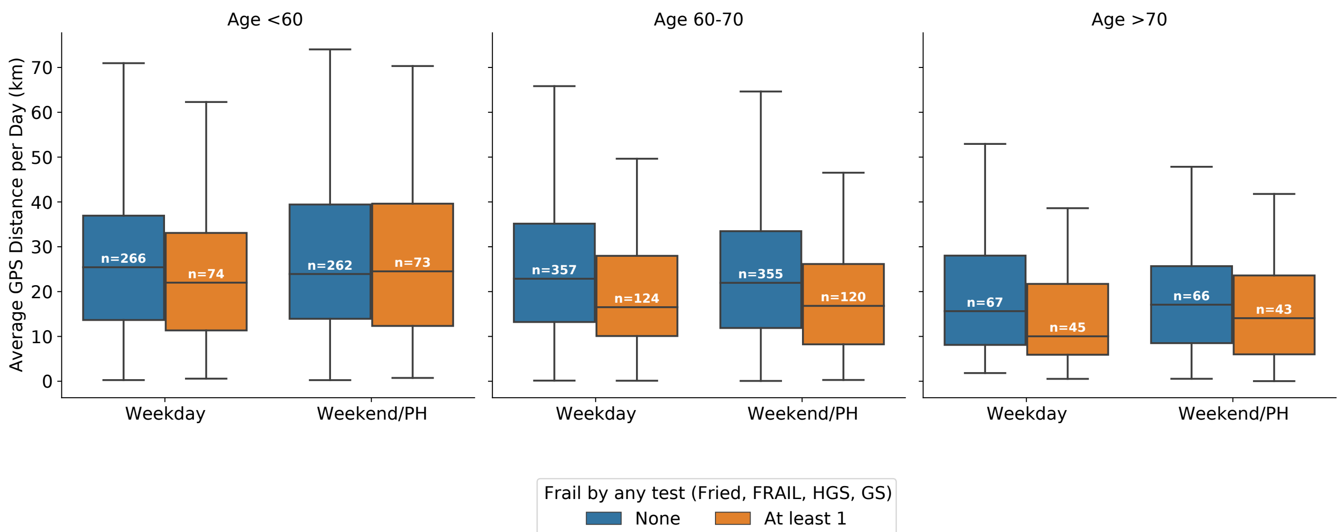

Figure 5: Effect of physical frailty on average GPS distance for different age groups.

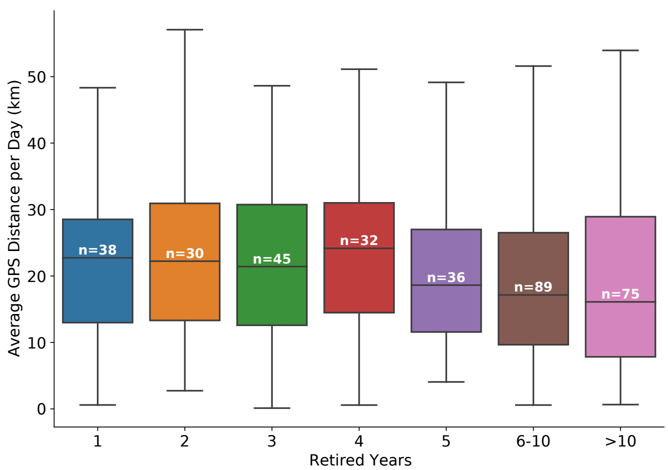

Figure 6: Effect of number of retired years (duration) on average GPS travel distance.

[6] O. D. Incel, M. Kose, and C. Ersoy. A review and taxonomy of activity recognition on mobile phones. *BioNanoScience*, 3(2):145–171, 2013.

[7] P. J. Krenn, S. Titze, P. Oja, A. Jones, and D. Ogilvie. Use of global positioning systems to study physical activity and the environment: a systematic review. *American journal of preventive medicine*, 41(5):508–515, 2011.

[8] M. Lin and W.-J. Hsu. Mining gps data for mobility patterns: A survey. *Pervasive and mobile computing*, 12:1–16, 2014.

[9] Z. Lin, S. Lyu, H. Cao, F. Xu, Y. Wei, H. Samet, and Y. Li. Healthwalks: Sensing fine-grained individual health condition via mobility data. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 4(4):1–26, 2020.

[10] J. H. Park, J. H. Moon, H. J. Kim, M. H. Kong, and Y. H. Oh. Sedentary lifestyle: Overview of updated evidence of potential health risks. *Korean Journal of Family Medicine*, 41(6):365–373, nov 2020.

[11] D. J. Patterson, L. Liao, D. Fox, and H. Kautz. Inferring high-level behavior from low-level sensors. In *UbiComp 2003: Ubiquitous Computing: 5th International Conference, Seattle, WA, USA, October 12-15, 2003. Proceedings 5*, pages 73–89. Springer, 2003.

[12] D. Rainham, I. McDowell, D. Krewski, and M. Sawada. Conceptualizing the healthscape: contributions of time geography, location technologies and spatial ecology to place and health research. *Social science & medicine*, 70(5):668–676, 2010.

[13] A. Rout, S. Nitoslawski, A. Ladle, and P. Galpern. Using smartphone-gps data to understand pedestrian-scale behavior in urban settings: A review of themes and approaches. *Computers, Environment and Urban Systems*, 90:101705, 2021.

[14] R. Soubra, A. Chkeir, and J.-L. Novella. A systematic review of thirty-one assessment tests to evaluate mobility in older adults. *BioMed research international*, 2019, 2019.

[15] S. C. Webber and M. M. Porter. Monitoring mobility in older adults using global positioning system (gps) watches and accelerometers: a feasibility study. *Journal of Aging and Physical Activity*, 17(4):455–467, 2009.

[16] S. C. Webber, M. M. Porter, and V. H. Menec. Mobility in older adults: a comprehensive framework. *The gerontologist*, 50(4):443–450, 2010.

[17] M. Winters, C. Voss, M. C. Ashe, K. Gutteridge, H. McKay, and J. Sims-Gould. Where do they go and how do they get there? older adults' travel behaviour in a highly walkable environment. *Social Science & Medicine*, 133:304–312, 2015.

[18] World Health Organization. https://www.who.int/news-room/fact-sheets/detail/ageing-and-health, Oct 2022.

[19] Y. Yang, Y. Xu, D. A. Rodriguez, Y. Michael, and H. Zhang. Active travel, public transportation use, and daily transport among older adults: The association of built environment. *Journal of Transport & Health*, 9:288–298, 2018.

[20] Y. Zhang, F. Xu, T. Li, V. Kostakos, P. Hui, and Y. Li. Passive health monitoring using large scale mobility data. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 5(1):1–23, 2021.

[21] Y. Zhang, F. Xu, T. Xia, and Y. Li. Quantifying the causal effect of individual mobility on health status in urban space. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 5(4):1–30, 2021.

[22] F. Zhao, A. Ghorpade, F. C. Pereira, C. Zegras, and M. Ben-Akiva. Stop detection in smartphone-based travel surveys. *Transportation Research Procedia*, 11:218–226, 2015.

[23] Y. Zheng, Q. Li, Y. Chen, X. Xie, and W.-Y. Ma. Understanding mobility based on gps data. In *Proceedings of the 10th international conference on Ubiquitous computing*, pages 312–321, 2008.