Estimating Anxiety Intensity of Dementia Patients Using Phrases, Facial Expressions, and Behaviors

Narumi Shigekiyo
Graduate School of Information Science and Technology, Osaka University
Osaka, Japan
n-shigekiyo@ist.osaka-u.ac.jp

Masaaki Yamauchi
Graduate School of Information Science and Technology, Osaka University
Osaka, Japan
m-yamauchi@ist.osaka-u.ac.jp

Hiroshi Tsuji
Graduate School of Engineering, Osaka University
Osaka, Japan
tsuji_hiroshi@arch.eng.osaka-u.ac.jp

Hideyuki Shimonishi
Cyber media center, Osaka University
Osaka, Japan
shimonishi.cmc@osaka-u.ac.jp

Masayuki Murata
Graduate School of Information Science and Technology, Osaka University
Osaka, Japan
murata@ist.osaka-u.ac.jp

Miwa Sugita
Graduate School of Engineering, Osaka University
Osaka, Japan
sugita_miwa@arch.eng.osaka-u.ac.jp

Michihiro Kita
Graduate School of Engineering, Osaka University
Osaka, Japan
kita@arch.eng.osaka-u.ac.jp

***Abstract*—Recently, the number of persons with dementia (PwD) has been steadily increasing. The PwDs live with anxiety derived from the decline in cognitive function, leading to concerns about memory loss and the future. The accumulation of these anxieties causes a state of agitation and Behavioral and Psychological Symptoms of Dementia (BPSD). Dealing with BPSD becomes a burden on the caregivers and the PwDs and violates their wellbeing. It is becoming a social issue. When a caregiver deals with agitation and BPSD, dealing with the early stages gives a smaller load than managing the escalated stage of agitation. Therefore, predicting agitation and BPSD in advance reduces the burden on caregivers and PwD and improves their wellbeing. Several methods have been proposed to predict BPSD using physiological, environmental, and caregiving data. However, there is a lack of research considering the state of feeling anxiety. Also, using a device that touches the skin tends to make PwD stressed, and continuous measurement is difficult. Thus, estimating anxiety from information obtained without contacting devices is necessary. Our research group has developed a metric, the CADATY index. The index is designed to estimate the intensity of anxiety and agitation based on the daily life situations of PwDs. We propose a method for estimating the CADATY index using Bayesian estimation by acquiring multimodal observation information such as phrases, facial expressions, and behaviors as the daily life of PwDs. To evaluate our method, we collect these data by recording video and audio in a nursing home that provides elderly housing with supportive services, i.e., in a daily living environment. We could estimate the CADATY index value in cases where we captured every modality. We found that the information from phrases and behaviors effectively detected signs of agitation and BPSD.**

***Index Terms*—Agitation, Bayesian estimation, BPSD, CADATY index, dementia, emotion estimation, restless.**

I. Introduction

The population of persons with dementia (PwD) has been increasing. The World Health Organization estimates the population of PwDs is around 55 million recently and is increasing at around 10 million per year [1]. The greater the number of PwDs, the greater the burden on caregivers, such as families and nursing home staffs.

One of the major parts of the burden of the PwD and its caregiver is agitation and Behavioral and Psychological Symptoms of Dementia (BPSD). The agitation is a state of restlessness and too heightened excitement. BPSD involves mental and neurological disorders such as heightened excitement, aggression, depression, and apathy. Intense and constant anxieties of PwDs cause them [2]. Due to the decline of cognitive functions of PwDs, they suffer from constant concerns and anxieties about their future and lack of memory. The agitation and BPSD significantly affect the wellbeing of both PwDs and caregivers [3].

Calm down and diverting attention is effective in dealing with agitation and BPSD for their wellbeing. In particular, when a caregiver can deal with the early stages of agitation and BPSD, i.e., feeling anxiety stage, a caregiver and a PwD get a smaller load on the caregiver than managing escalated states of agitation caused by heightened anxiety. Therefore, detecting the state of anxiety before the onset of BPSD and agitation in advance reduces the burden on caregivers and improves the wellbeing of the caregiver and PwD.

Many researchers have conducted several studies to predict BPSD in advance. Specifically, methods to predict BPSD using physiological, environmental, and care data recorded by caregivers have been proposed [4]–[6]. However, there is a lack of research considering the state of feeling anxiety, which is a sign of agitation, as far as we know. Additionally, there are problems that the methods require a substantial amount of data for training estimation models, and many PwDs feel discomfort by wearing contact-type sensors to acquire physiological data. Therefore, we have to collect data using noncontact methods to estimate the state of anxiety in PwDs.

Our research group has been developing the Caregivers

This research was supported by Osaka University NEC Beyond 5G Research Alliance Laboratories.

TABLE I: The Guidance of CADATY Index Values [7].

Value	Explanation	Examples
+3 – +5	Too highly excited	Manic.
+2	Peace towards others	Expressing gratitude towards others.
+1	Reflecting self-peace	The dementia patient feels reassured and relaxed.
0	Neutral	
-1	Self-questioning states	Looking confused or helpless.
-2	Self-blaming or questioning others	Showing irritation towards oneself or seeking confirmation from others.
-3	Depict blaming others	Strongly expressing concerns to others.
-4	Depict blaming others	Uttering offensive language.
-5	Depict blaming others	Violation behaviors towards others.

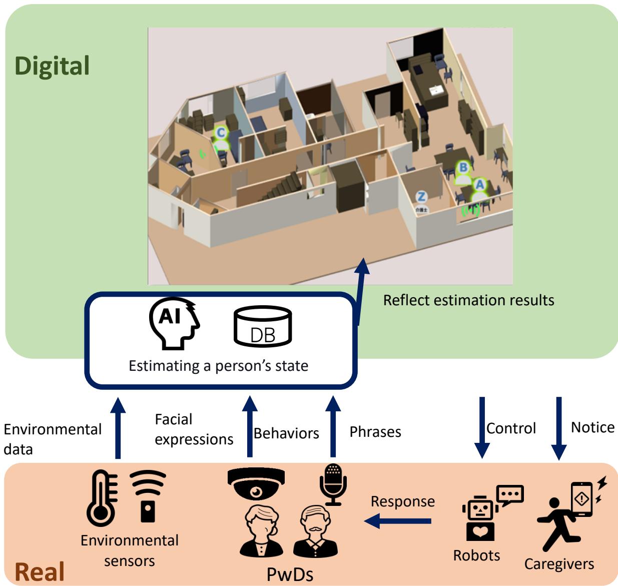


Fig. 1: Envisioned overviews of our alliance laboratory.

Assess Dementia's Anxiety designed by Tsuji and Yamauchi (CADATY) index to evaluate anxiety intensity through the usual appearance observation [7]. We are inspired by the fact that the caregiver who observes the daily appearance of the PwD well can sometimes find a sign of its agitation. The CADATY index is an indicator that assesses and quantifies the level of anxiety and agitation based on daily appearance, such as phrases, facial expressions, and behaviors. Note that the reviewer of the CADATY index is a person who watch the daily appearance of target PwD closely, such as family, friends, and caregivers. The intensity of anxious feelings is rated on a scale from -5 to +5: 0 represents a neutral state, -1 indicates a slightly anxious state, and -5 represents a strongly anxious state. Conversely, positive values indicate a serene emotional state. We explain the guidance of each CADATY index value in Tab. I.

In this paper, we propose a estimation model it learns the CADATY index from multimodal observational information, such as the phrases, facial expressions, and behaviors of PwDs. We aim to verify the possibility of detecting sings of agitation, namely the state of anxiety. Specifically, we can consider that the CADATY index falls below a threshold value, -1 or -1.5, or less is signs of agitation. Our goal is to detect the signs. However, it is necessary to employ methods that can learn individual models even with limited data and constrained modalities. The individual model is required because there are variations in phrases and facial expressions among PwDs. Furthermore, the data we can collect from one PwD is limited. In addition, we consider that we cannot obtain all of the PwD's modalities at any time in a daily living environment; for example, facial expressions are not visible due to the PwD's facing directions. Therefore, the proposed method learns the association of prior probabilities between the CADATY index value and each modality. After learning, the method estimates the current CADATY index using Bayesian inference based on the observed phrases, facial expressions, and behaviors. Then, we evaluate our method to determine whether it can detect the agitation sign, i.e., whether the estimated value is over the threshold. Moreover, we examine the effectiveness of multimodal estimation and each modality.

Our alliance laboratory aims to reduce the burden on residents and caregivers for their wellbeing in a care facility and works to research and develop a digital twin including deploying our CADATY index estimation method. We illustrate the overview of our alliance laboratory in Fig. 1. The digital twin collects and analyzes environmental data such as temperature and atmospheric pressure, as well as changes in facial expressions and conversation content. We also can collect information using the Internet connected consumer electronics. By deploying our method, we can grasp the sign of agitation and notify the sign and request assistance to caregivers or operate AI robots to take care of PwDs.

II. Related Work

A. *Detection and Prediction Methods of BPSD*

Many researchers have tried to detect agitation and BPSD. Studies have focused primarily on the detection of states of heightened excitement, such as violence, verbal aggression, and shouting. Researchers have commonly utilized accelerometer data obtained from actigraphs for detecting heightened excitement [8]. In recent years, research has utilized physiological data, including blood volume pulse, electrodermal activity, and skin temperature, collected from wrist-worn wearable devices to detect agitation [9].

In recent years, several methods have been proposed to predict the signs of agitation and BPSD in advance. HekmatiAthar et al. have proposed a method to predict restlessness within the next 30 minutes using deep learning. They acquired environmental sensor data, specifically brightness, noise level, temperature, humidity, and atmospheric pressure in a home. The recall was high, 84.8%. However, there is a high false positive rate, 48.4%. It is difficult to achieve accurate agitation detection based on only environmental information [4]. Another approach involves combining environmental data with physiological data to detect BPSD. Then, it recommends appropriate care methods when it detects BPSD. This method has reported a reduction of approximately 40% of all agitated behaviors [5]. Additionally, Yamagami et al. have proposed the BPSD prediction method by analyzing environmental data, physiological data, and information from caregiving records. By collaborating with multiple caregiving facilities and analyzing a large volume of collected data, they aim to achieve high-precision predictions [6].

These proposals acquired physiological data using wrist-worn sensors. However, it is well-known that many PwDs resist wearing wrist-worn sensors. The discomfort of wearing sensors would lead to frequent removals and make continuous monitoring difficult. This is a significant problem, and there is some research for data acquisition methods that do not cause discomfort to PwDs [10]. However, it is still difficult to collect physiological data continuously without the stress of touching devices in a daily living environment.

To address these challenges, we utilize data acquired through noncontact methods such as video and audio. Additionally, we construct models capable of learning features from a small amount of data, enabling the creation of personalized models for individuals.

B. Indicator of Anxiety Intensity in Persons with Dementia

Our research group has been developing the CADATY index as an indicator to estimate the level of anxiety of PwDs based on their usual state. The CADATY index is an indicator designed for caregivers and family members who are familiar with the targeted PwD. They assign scores of the CADATY index value by considering the PwD's phrases, facial expressions, and behaviors by evaluating the current state as a match to which CADATY index value and its anxiety level. Additionally, we can use it as an indicator of the current required level of psychological care from caregivers. Based on the index, we can provide care at an early state of anxiety and agitation or BPSD. We anticipate that it reduces the burden and improves the wellbeing of PwDs and caregivers.

The CADATY index uses phrases, facial expressions, and behaviors as usual appearance. This is caused by the recognition that PwDs exhibit distinct characteristics in their behaviors compared to healthy individuals. Previous studies have shown that PwDs and healthy individuals can be distinguished based on activities of daily living on some days [11]. In addition, PwDs and healthy individuals have been determined by facial expressions [12]. Additionally, PwDs use the same words or phrases repetitively, and difficulties recalling words may lead to incorrect language or repeated expressions. Thus, we also focus on the phrase patterns of PwDs, considering these unique features compared to healthy individuals.

On the CADATY index, the transition from anxiety to agitation is categorized into self-questioning, self-blaming, questioning others, and blaming others, rated on a scale from -1 to -5, described in Tab. I. Specifically, scores from -3 to -5 indicate agitation, while scores from -1 to -3 can be interpreted as the preceding anxiety. By providing appropriate care at the stages considered as anxiety, such as -1 and -2, we believe that the progression to agitation, rated -3 to -5, can be prevented.

The scoring of the CADATY index is based on the concept of 0 representing a normal state, with a gradual increase in anxiety from -1 to -5 and a corresponding increase in peaceful emotions from +1 to +5. The scoring criteria are as described in Section I.

III. Anxiety Intensity Estimation Method Based on Bayesian Inference

A. Overview

We illustrate an overview of our method in Fig. 2. First, the proposed method learns prior probability. Second, it gets the target PwD's current phrase, facial expression, and behavior and estimates the anxiety intensity based on the prior probability. We use Bayesian inference because the dataset may include imbalance data and we may lack some modalities.

The proposed method requires recording the PwD's daily behavioral data using network-connected consumer electronics, such as home IoT devices, home cameras, AI speakers, smartphones, and environmental sensors. Particularly in this work, we must collect video and audio data from cameras and microphones to train models. Simultaneously, the CADATY index values should be labeled to train the prior probabilities.

B. Collecting Data of the PwD for the Proposed Method

We have to record the information of the target PwD using video cameras and audio microphones to use it as input data. Then, we extract the features, such as phrases $\alpha(t)$, facial expressions $\beta(t)$, and behaviors $\gamma(t)$ at time t, from the recorded video and audio. Moreover, based on the guideline, a familiar person with the target PwD labels CADATY index value $y(t)$. We treat these values as the training label.

1) Phrases: We transcribe the recorded audio data, including conversations and soliloquies of the subject. Then, we pick up some phrases α_i frequently used by the target PwD. For instance, common phrases such as "arigatou"("thank you." in English) or "gomen"("I am sorry." in English), phrases often uttered by PwD like "wakaranai"("I cannot understand the current situation." in English) or "mou kaeranaito"("I need to go back home." in English), and specified phrase for the target PwD. To pick up the phrases, we should know the background of the target PwD by life-history survey [7].

2) Facial Expressions: We classify the facial expressions into labels β_j as facial expression modality using facial images of the target PwD from the recorded video. The labels are defined by the facial parts of the target. For instance, we combine information such as the angles of eyes and mouth corners and the presence of specific wrinkles. We also define the baseline facial expression as neutral.

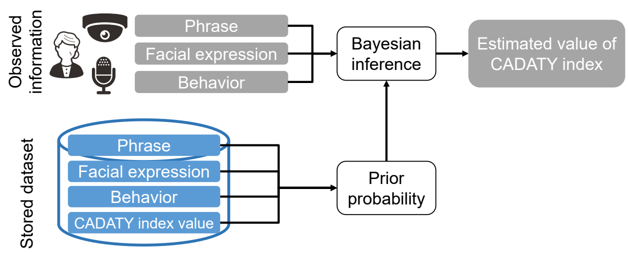


Fig. 2: Overview of the proposed method.

3) Behaviors: We categorize behaviors related to the target PwD from the recorded video. The behavior labels γ_k are used as information on the behavior modality. We define the labels by movements and activities that the target PwD performs. For example, we can list basic actions, such as standing or sitting, and specific ones, such as looking to the right rear or left rear, actions with body parts, and gestures. Also, we can use movements, such as starting point, ending point, and its path.

C. Training of Prior Probability

The proposed method calculates the probabilities $P(\alpha_i|y_c)$, $P(\beta_j|y_c)$, and $P(\gamma_k|y_c)$ of expressing a phrase α_i, facial expression β_j, and behavior γ_k when the CADATY index value is y_c, respectively. We define the calculations in (1), (2), and (3).

$$P(\alpha_i|y_c) = \frac{\sum_{t=0}^{T} F(\alpha_i = \alpha(t) \cap y(t) = y_c)}{\sum_{t=0}^{T} F(y(t) = y_c)}, \quad (1)$$

$$P(\beta_j|y_c) = \frac{\sum_{t=0}^{T} F(\beta_j = \beta(t) \cap y(t) = y_c)}{\sum_{t=0}^{T} F(y(t) = y_c)}, \quad (2)$$

$$P(\gamma_k|y_c) = \frac{\sum_{t=0}^{T} F(\gamma_k = \gamma(t) \cap y(t) = y_c)}{\sum_{t=0}^{T} F(y(t) = y_c)}, \quad (3)$$

where $t = 0, 1, \ldots, T$ is the observation time and $F(equ)$ is a function that returns 1 when the equation *equ* holds true and 0 otherwise. Additionally, we calculate the ratio of the CADATY index being y_c, denoted as $P(y_c)$, by (4).

$$P(y_c) = \frac{\sum_{t=0}^{T} F(y(t) = y_c)}{T}. \quad (4)$$

D. Estimating CADATY Index Value by Bayesian Inference with Prior Probability

We estimate the current CADATY index value by Bayesian inference based on prior probabilities and current observations. We calculate a probability, $P(y_c|\boldsymbol{x_n(t)})$, that the CADATY index is y_c and a sequence of currently observed phrases, facial expressions, and behaviors is $\boldsymbol{x_n(t)}$ during the time from t to $t + \Delta t$ based on the Naive Bayes [13],

$$\begin{aligned} P(y_c|\boldsymbol{x_n(t)}) &= \frac{P(y_c)P(x_1(t), \ldots, x_N(t)|y_c)}{P(x_1(t), \ldots, x_N(t))} \\ &= \frac{P(y_c)\prod_{n=1}^{N} P(x_n(t)|y_c)}{P(x_1(t), \ldots, x_N(t))} \\ &\propto P(y_c)\prod_{n=1}^{N} P(x_n(t)|y_c). \end{aligned} \quad (5)$$

Then, we calculate the expected value of the CADATY index at time t, denoted as $y_{est}(t)$. In other words, we estimate the expected value of y_c at this time using (5) like

$$y_{est}(t) = \frac{\sum_c y_c P(y_c) \prod_{n=1}^{N} P(x_n(t)|y_c)}{C}, \quad (6)$$

where $c = 1, 2, \ldots, C$.

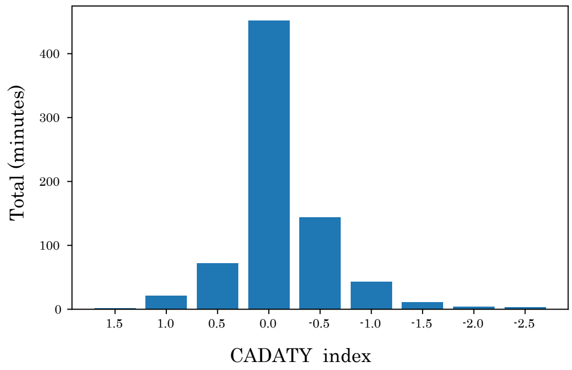


Fig. 3: Distribution of the CADATY index values in the data.

IV. Evaluation

We have received permission to conduct ethical reviews of research in our institutions. We have conducted under the rules [1]. We have considered the privacy and security of gathering datasets with cameras and microphones in the rules.

A. Data Collection in Daily Life Environments

To evaluate the proposed method, we collected datasets by recording the usual state of a PwD living in a nursing home that provides housing with supportive services. A PwD, 90 years old and capable of independent mobility, was selected as a subject from the nursing home residents. We recorded the subject's phrases, facial expressions, and behaviors in shared spaces, such as the communal living room and the hall. An observer, not H. Tsuji, conducted participatory observations and manually recorded the subject's activities using supplemental video and audio captured by multiple cameras and microphones. The label of the CADATY index value, denoted as $y(t)$ at time t, was recorded per minute by H. Tsuji, who regularly observes the subject in the nursing home and has assessed the subject's anxiety intensity well. The CADATY index values were recorded in 0.5 increments, serving as the labeled value for the evaluation. We collected data from 3:00 p.m. to 7:00 p.m. because we thought sundown syndrome could be present during this time period. We collected data for six days and obtained a 752-minute dataset, excluding times when H. Tsuji could not label, e.g., the subject was in its private room. We utilized this dataset for both training and evaluation purposes. We named the data of six days Day1, Day2, ..., Day6.

We illustrated the distribution of the CADATY index values for each minute of the collected data in Fig. 3. They are in the range from -2.5 to 1.5. The times when the CADATY index value is 0.0 are the most prevalent, while 1.5 or below -1.5 are relatively scarce.

[1]The experiment has received two approvals. The approval number is 202305 from the Research Ethics Committees of the Graduate School of Information Science and Technology, Osaka University. The approval number is 4–7–1 from the Research Ethics Committees of the Graduate School of Engineering, Osaka University.

B. Settings and Evaluation Metrics

We train the prior probability using all collected data. We set $\Delta t = 1$ (minute), i.e., we estimate the CADATY index value every minute, $C = 9$, and $T = 752$. When we calculate the estimated CADATY index value $y_{est}(t)$ to asses specific combinations of modalities, we remove the modalities not used for estimation from $\boldsymbol{x_n(t)}$.

Additionally, we set evaluation metrics. First, we use the number of detected agitation signs. We define the agitation sign as a term when the CADATY index label continuously becomes the threshold V or below. Note that when the label is temporarily over the threshold, within one minute, we treat the series as one sign of agitation. When the estimated CADATY index value y_{est} falls the threshold V or below in any timing of the term of agitation signs, we take that the agitation is detected. On the other hand, the estimated value y_{est} does not fall to the threshold V or below in the term of the agitation sign; the sign is overlooked. In another case, the estimated value y_{est} falls to the threshold V or below, but it is not the term of agitation signs based on the labels; we count it as misdetections. Note that, similar to the agitation signs, we count it as one misdetection when the estimated value y_{est} continuously, temporal exceedance of threshold within one minute is considered continuous, falls to the threshold V or below.

We set the threshold V to -1.0 and -1.5. It represents whether we raise an alert of anxiety. Choosing a smaller V means we raise an alert when the tendency of anxiety is relatively intense.

Second, we also assess the predictive error, the difference between the estimated CADATY index value $y_{est}(t)$ and the labeled value $y(t)$ at each time. We calculate the errors' average, median, maximum value, and standard deviation. This provides insights into the accuracy of predictions in states where the labeled value $y(t)$ is not below V, as well as the estimation errors when $y(t)$ is V or below.

C. Estimation Results and Discussions

We show the number of detected agitation signs and the misdetections in Tab. II, the difference between estimated values and labeled ones in Tab. III, and the estimated result of each day in Fig. 4. We omit some days with neither agitation signs in the test data nor detected signs of the estimator.

By using all three modalities, we could estimate the most accurately. It could detect six out of nine and 10 out of 21 agitation signs in Tab. IIa and IIb, respectively. Note that although the results using phrases and behaviors in Tab. IIb detected one more sign, this is caused by an annotation issue of facial expression. Thus, we can consider their accuracy to be similar. In addition, the average, median, and maximum difference are the smallest in Tab. III. We showed a concrete case on Day4 at 3:44 p.m. in Fig. 5. We observed the phrase "aho" ("stupid" in English), a neutral facial expression, and behaviors such as covering ears, shaking head, and turning around right rear; the labeled CADATY index value was -2.0 at this time. When we estimate from the only phrase, we could not determine the current CADATY index value is around zero or -2.0 because there are multiple peaks in Fig. 5a. Note that we think this tendency is not wrong because the subject uses the "aho" phrase in two cases: the subject says it when making fun of someone with dear feelings, and it is angry. Then, combining the other modalities, we could correctly estimate the current CADATY index as -2.0, described in Fig. 5b.

The accuracy was improved not only by using all modalities but also by using two. For example, in Tab. IIb, when we used each modality of phrases and behaviors on Day3, we could detect three and four agitation signs, respectively. Five signs were detected using these two modalities. On Day4, we could reduce one misdetection by using facial information compared to the case using only behaviors. In most cases, we observed a similar tendency.

When a part of the modality was lost, the estimation accuracy worsened. In such cases, the distribution of the estimated probabilities of each CADATY index value tends to have multiple peaks like Fig. 5a. Because our method calculates the expected values, we misdetect or overlook an agitation sign. However, we also can leverage such uncertain information. For example, when we can operate a movable IoT camera to capture facial or behavior information, we can reduce its power consumption by operating only in cases of high uncertainty, i.e., only at the time we want to collect information.

The effect of the phrase modality is crucial, mainly where $V = -1.5$. In the Tab. IIa, the case using only phrase modal detected five out of nine signs, while the result using all three modalities detected six. The prior probability of the phrase modality tends to be biased in a specific CADATY index value. Thus, the phrase information tends to be crucial for detecting CADATY index values of -1.5 or below. However, the phrase is a modality that is difficult to acquire because we could not listen to what the subject said in a case where the subject had a soliloquy.

The behavior modality is the second most important factor in detecting agitation signs. This is primarily due to the practicality of recording behavior modality, which can be done most of the time, regardless of the camera angle or the subject's location. Thus, we could record many behavior data and train the prior probability well. In particular, where $V = -1.0$ in Tab. IIb, whereas the best multimodal case detected 11 agitation signs, the case using behaviors detected 10. Because the many agitation signs where the CADATY index value is -1.0 did not involve phrases, the detected agitation signs by behaviors are larger than by phrases. However, the bias of the prior probability about phrases tends to be minor, and we could not detect many agitation signs using only behaviors where $V = -1.5$. Furthermore, the behaviors have multiple types, but we recorded only typical ones in this experiment. By utilizing other behaviors, we can improve estimation accuracy, particularly in Tab. IIb.

The modality of facial expressions is adapted to estimate the plus values of the CADATY index. In contrast, the result of a single modality of the facial expression could not detect agitation signs, i.e., the minus range of the CADATY index. In Tab. II and III, when we add facial expression information to the phrase and behavior information, we found only one agitation sign, but the average difference becomes small. In addition, we had little facial data because the angles of

TABLE II: Number of Detected Agitation Signs and Misdetections Using Each Combination of Modalities with Threshold V.

(a) V = -1.5.

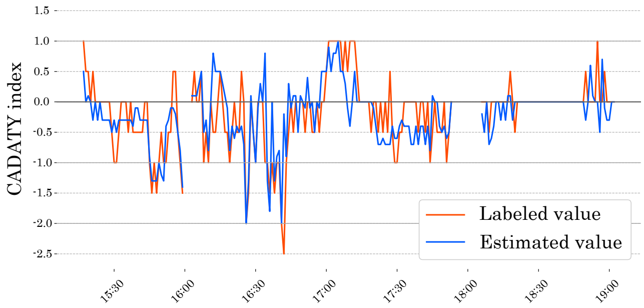
Used modals			Day3		Day4		Day5	
Phr.	***Fac.***	***Bhv.***	***TP***	***FP***	***TP***	***FP***	***TP***	***FP***
✓			3	0	2	0	0	0
	✓		0	0	1	0	0	0
		✓	1	0	2	0	0	0
✓	✓		3	0	3	0	0	0
✓		✓	3	0	2	0	0	0
	✓	✓	1	0	3	0	0	0
✓	✓	✓	3	0	3	0	0	0
Total agitation signs			5		3		1	

Phr.: Phrase. ***Fac.***: Facial expressions. ***Bhv.***: Behaviors.
TP: Detected agitation signs. ***FP***: Misdetections.

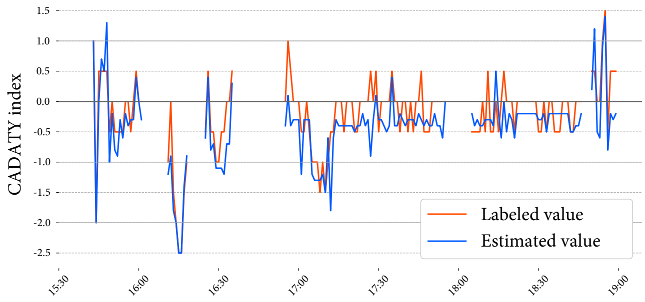
(b) V = -1.0.

Used modals			Day3		Day4		Day5		Day6	
Phr.	***Fac.***	***Bhv.***	***TP***	***FP***	***TP***	***FP***	***TP***	***FP***	***TP***	***FP***
✓			3	0	2	0	0	0	0	0
	✓		0	0	1	1	0	0	0	0
		✓	4	0	4	3	2	0	0	1
✓	✓		3	0	3	2	0	0	0	0
✓		✓	5	0	4	3	2	0	0	1
	✓	✓	4	0	4	2	1	2	0	0
✓	✓	✓	5	0	4	2	1	2	0	0
Total agitation signs			11		4		4		2	

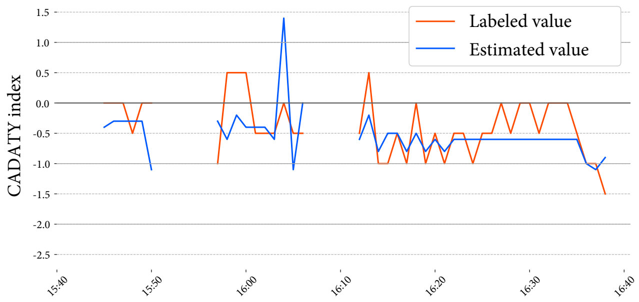
Phr.: Phrase. ***Fac.***: Facial expressions. ***Bhv.***: Behaviors.
TP: Detected agitation signs. ***FP***: Misdetections.



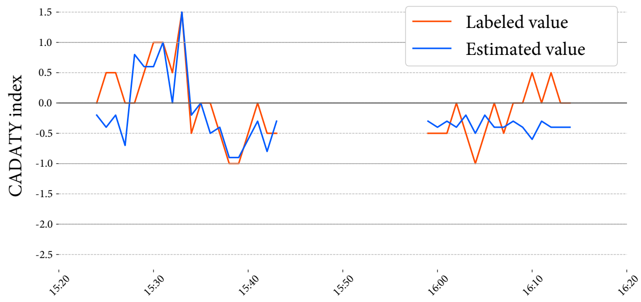
(a) Day3.



(b) Day4.



(c) Day5.



(d) Day6.

Fig. 4: Estimation results of each day using every modal.

TABLE III: Predictive Errors between Estimated CADATY Index Values and Labeled Ones.

Used modals			Avg.	Median	Maximum	Standard deviation
Phr.	***Fac.***	***Bhv.***				
✓			0.40	0.40	1.60	0.34
	✓		0.35	0.30	2.30	0.32
		✓	0.35	0.30	2.30	0.32
✓	✓		0.35	0.30	0.90	0.26
✓		✓	0.36	0.30	1.80	0.38
	✓	✓	0.35	0.30	1.40	0.29
✓	✓	✓	0.29	0.25	0.90	0.31

Phr.: Phrase. ***Fac.***: Facial expressions. ***Bhv.***: Behaviors.

cameras and subjects influenced its collectability. Thus, the trend may change again as the number of data increases.

We could not detect three out of nine signs even if we used all modalities in Tab. IIa. Based on interviews with those who labeled the CADATY index values, our findings suggest a potential solution. By considering environmental effects, such as the room's noise, we can improve our detection capabilities, offering a glimmer of hope in our research.

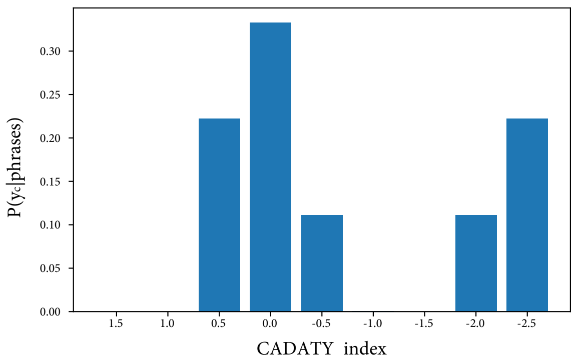
On Day5 and Day6, we could observe limited information about phrases and limited facial expressions. In addition, the subject had behaviors that we did not annotate. As a result, we could not detect agitation signs and some misdetections.

We consider it difficult to estimate the CADATY index value from -0.5 to +0.5 using the three modalities. Even continuous data collection by the sensors touching the skin might be required to capture the subtle changes in this area.

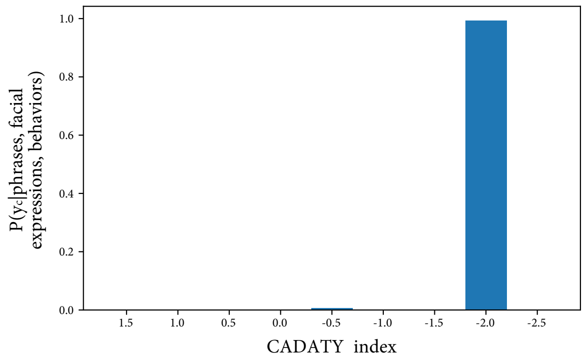
V. Conclusion

We proposed a method to estimate anxiety intensity based on the CADATY index, an indicator of anxiety intensity in PwD, using multimodal information such as phrases, facial expressions, and behaviors.

To evaluate the proposed estimation method, we manually extracted information on phrases, facial expressions, and behaviors from a PwD's living environment. By learning



(a) Estimated by the observed phrase, "aho".



(b) Estimated by the neutral facial expression, the "aho" phrase, and behaviors include covering ears, shaking the head, and turning around the body to the right rear.

Fig. 5: Estimated proportions of the CADATY index on Day4 at 3:44 p.m. values when each modality is used.

prior probabilities and estimating based on Bayesian inference, the method demonstrated its ability to estimate the CADATY index and detect signs of agitation when the estimation method could use phrases, behaviors, and facial expressions. Each modality has characteristics. The phrases were effective in capturing features of relatively strong agitation signs, the CADATY index value is -1.5 or below; the behaviors helped detect signs of relatively minor anxiety, the CADATY index value is -1.0 or below, and the facial expressions demonstrated proficiency in estimating CADATY index values in the positive range. Therefore, phrases and behavior information were deemed crucial for detecting agitation signs.

However, we cannot make correct estimations when some information is lacking. In many such cases, the probability distribution of the estimated CADATY index value had multiple peaks. When the probability distribution is skewed towards a specific value, the model confidently predicts the CADATY index to the value. While this certainty information is not utilized in the current evaluation, incorporating the variability of such probability distributions into the estimation process might contribute to improved detection accuracy.

We evaluated one participant. Therefore, the tendency might be changed by the type of PwDs, such as gender, personality, and the type of dementia. Collecting data from multiple types of PwDs is essential in future work. In addition, the CADATY index approach is based on the personal profiling. Thus, when we apply this method to a new PwD, we must make a new profile at the current stage. In future work, we have to develop a standard model of profiling.

Additionally, the dataset obtained in this experiment does not include data related to evident agitation with a CADATY index of -3.0 or lower. Further verification is necessary to assess the actual capability of detecting clear signs of agitation.

We plan to implement the estimation method in real care-giving environments and verify whether it reduces the burden on PwDs and their caregivers. At that time, verifying whether the threshold should be set at -1.0 or -1.5 is necessary.

References

[1] World Health Organization, "Dementia," https://www.who.int/news-room/fact-sheets/detail/dementia, Mar. 2023, accessed: 2023-10-06.

[2] P. J. Seignourel, M. E. Kunik, L. Snow, N. Wilson, and M. Stanley, "Anxiety in dementia: A critical review," *Clin. Psychol. Rev.*, vol. 28, no. 7, pp. 1071–1082, Oct. 2008.

[3] N. Cloak and Y. A. Khalili, "Behavioral and psychological symptoms in dementia," *StatPearls [Internet]*, Jan. 2024, accessed: 2024-6-1.

[4] S. HekmatiAthar, H. Goins, R. Samuel, G. Byfield, and M. Anwar, "Data-driven forecasting of agitation for persons with dementia: A deep learning-based approach," *SN Comput. Sci.*, vol. 2, no. 326, pp. 1–10, Jun. 2021.

[5] Senior Demantia Institute, "[AI-enabled dementia care IoT service in daily environment: Pilot project report] AI wo katsuyō shita ninchishō taio-gata IoT sabisu jisshō jigyō hōkoku," https://www.mhlw.go.jp/content/10601000/000468143.pdf, Jan. 2019, accessed: 2023-10-18, (in Japanese).

[6] T. Yamagami and H. Yamaguchi, "Review of BPSD prevention/prediction system using information and communication technology," *Tokyo J. Dementia Care Res.*, vol. 5, pp. 8–15, Mar. 2021, (in Japanese).

[7] H. Tsuji, M. Yamauchi, M. Sugita, and M. Kita, "Appearance transcription (CADATY index) and memory book (life-history survey); consideration for a psychological state estimation index for people with dementia," in *Proc. 29th Int. Montessori Congr. - Thailand, Exhib. Showcase*, Aug. 2023, pp. 1–1.

[8] S. S. Khan, B. Ye, B. Taati, and A. Mihailidis, "Detecting agitation and aggression in people with dementia using sensors - a systematic review," *Alzheimers Dementia*, vol. 14, no. 6, pp. 824–832, Jun. 2018.

[9] S. S. Khan, S. Spasojevic, J. Nogas, B. Ye, A. Mihailidis, A. Iaboni, A. Wang, L. S. Martin, and K. Newman, "Agitation detection in people living with dementia using multimodal sensors," in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC) 2019*, Jul. 2019, pp. 3588–3591.

[10] M. Garcia-Constantino, C. Orr, J. Synnott, C. Shewell, A. Ennis, I. Cleland, C. Nugent, J. Rafferty, G. Morrison, L. Larkham, S. McIlroy, and A. Selby, "Design and implementation of a smart home in a box to monitor the wellbeing of residents with dementia in care homes," *Frontiers Digit. Health*, vol. 3, no. 798889, pp. 1–13, Dec. 2021.

[11] P. Urwyler, R. Stucki, L. Rampa, R. Müri, U. Mosimann, and T. Nef, "Cognitive impairment categorized in community-dwelling older adults with and without dementia using in-home sensors that recognise activities of daily living," *Scientific Rep.*, vol. 7, no. 42084, pp. 1–9, Feb. 2017.

[12] Y. Umeda-Kameyama, M. Kameyama, T. Tanaka, B.-K. Son, T. Kojima, M. Fukasawa, T. Iizuka, S. Ogawa, K. Iijima, and M. Akishita, "Screening of alzheimer's disease by facial complexion using artificial intelligence," *Aging*, vol. 13, no. 2, pp. 1765—–1772, Jan. 2021.

[13] I. Rish, "An empirical study of the naive bayes classifier," in *Proc. IJCAI 2001 Workshop Empirical Methods Artif. Intell.*, vol. 3, no. 22, Jan. 2001, pp. 41–46.