

Received 1 August 2024, accepted 14 August 2024, date of publication 19 August 2024, date of current version 11 September 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3445410

## RESEARCH ARTICLE

# Response Style Characterization for Repeated Measures Using the Visual Analogue Scale

SHUNSUKE MINUSA<sup>ID</sup>, (Member, IEEE), TADAYUKI MATSUMURA<sup>ID</sup>,  
KANAKO ESAKI<sup>ID</sup>, (Member, IEEE), YANG SHAO<sup>ID</sup>, CHIHIRO YOSHIMURA<sup>ID</sup>,  
AND HIROYUKI MIZUNO<sup>ID</sup>, (Fellow, IEEE)

Center for Exploratory Research, Research & Development Group, Hitachi, Ltd., Tokyo 185-8601, Japan

Corresponding author: Shunsuke Minusa (shunsuke.minusa.hd@hitachi.com)

This work was supported by the Research & Development Group, Hitachi, Ltd.

This work involved human subjects in its research. Approval of all ethical and experimental procedures and protocols was granted by the internal review board of the Research & Development Group, Hitachi, Ltd., and performed in line with the Declaration of Helsinki.

**ABSTRACT** Self-report measures (e.g., Likert scales) are widely used to evaluate subjective health perceptions. Recently, the visual analog scale (VAS), a slider-based scale, has become popular owing to its ability to precisely and easily assess how people feel. These data can be influenced by the response style (RS), a user-dependent systematic tendency that occurs regardless of questionnaire instructions. Despite its importance, especially in between-individual analysis, little attention has been paid to handling the RS in the VAS (denoted as response profile (RP)), as it is mainly used for within-individual monitoring and is less affected by RP. However, VAS measurements often require repeated self-reports of the same questionnaire items, making it difficult to apply conventional methods used for Likert scales. In this study, we developed a novel RP characterization method for various types of repeatedly measured VAS data. This approach involves modeling RP as distributional parameters  $\theta$  through a mixture of RS-like distributions, and addressing the issue of unbalanced data through bootstrap sampling for treating repeated measures. We assessed the effectiveness of the proposed method using simulated pseudo-data and an actual dataset from an empirical study. The assessment of parameter recovery showed that our method accurately estimated the RP parameter  $\theta$ , demonstrating its robustness. Moreover, applying our method to an actual VAS dataset revealed the presence of individual RP heterogeneity, even in repeated VAS measurements, similar to the findings for Likert scales. Our proposed method enables RP heterogeneity-aware VAS data analysis, similar to Likert-scale data analysis.

**INDEX TERMS** Response style, subjective rating, visual analogue scale, repeated measures, in-the-wild, ecological momentary assessment, experience sampling, subjective health.

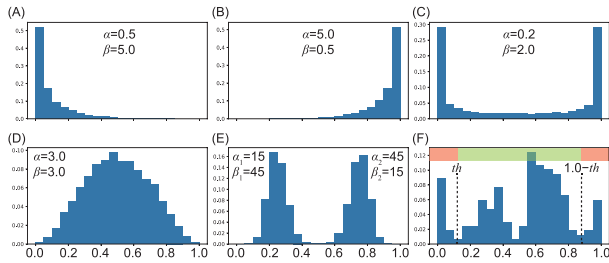
## I. INTRODUCTION

Monitoring human health is essential in various health-related fields. Both objective and subjective monitoring play an important role in examining human health [1]. Self-reports using specific measurement scales have been widely used to monitor subjective feelings such as the degree of fatigue [2], [3], depression [4], pain [5], and emotions [6]. For example,

The associate editor coordinating the review of this manuscript and approving it for publication was Diego Bellan<sup>ID</sup>.

previous studies assessed patients' pain using various scales including Likert, face, numerical rating, and visual analog scales (VAS) [7], [8].

Recently, the use of VAS has become more popular [7]. The VAS consists of a line whose length is typically 100 mm, with anchor descriptors at both ends of the line (unipolar VAS) or both ends and center (bipolar VAS) [9]. Unlike Likert scales, which require the response to fit descriptors in a categorical or discrete manner (e.g., selection from 5- and 6-point), the VAS enables continuous responses to subjective



**FIGURE 1.** Examples of response styles (RSs) observed in histograms with the empirical density distribution of VAS data. (A) dis-acquiescent RS, DRS. (B) acquiescent RS, ARS. (C) extreme RS, ERS. (D) midpoint RS, MRS. (E) bimodal MRS, BiMRS. These data were simulated using Beta distribution with specified parameters. (F) Typical histogram of actual VAS responses of some participants.

health conditions, enabling precise assessment of how people feel [10]. Owing to these characteristics, the VAS has been generally accepted as appropriate for comparing conditions and assessing the time course of how the feelings of specific persons change over time (i.e., within-individual assessment). Furthermore, as smartphones become more readily available, researchers can introduce an electronic VAS (e.g., a slider scale) as a substitute for the conventional paper VAS, enabling the digital collection of large amounts of VAS data [11], [12]. This transition has expanded VAS usage not only among within-individual assessments but also among between-individual comparisons.

One of the issues regarding the application of the VAS to between-individual comparisons is the difference in response styles (RSs) among individuals [13], [14]. RS is the systematic tendency not to depend on the scale instructions, and has been studied extensively for Likert scales (Fig. 1). Among well-known RSs for Likert scales, a popular RS is the dis-acquiescent RS (DRS, Fig. 1A), in which persons respond to the lower/left end of lines regardless of the instructions, and acquiescent RS (ARS, Fig. 1B), vice versa. The other popular RS is the extreme RS (ERS, Fig. 1C), in which individuals tend to preferentially respond at both ends of the scales. Furthermore, some people, especially Asians, tend to show midpoint RS (MRS, Fig. 1D), in which they prefer to select rating items around the center of the scale [14], [15]. RSs similar to those observed for the Likert scale are also observed for the VAS. We often observed an MRS-like response tendency on both sides across the center (hereafter referred to as bimodal MRS, BiMRS, Fig. 1E) upon using the bipolar VAS as a rating scale, which places the neutral position at the center of the scale, and conflicting concepts at both ends [16]. The VAS has often been used to evaluate relative changes among individuals, such as in clinical studies of pain and fatigue [17], [18]. Consequently, there have been cases in which the absolute values of VAS responses were analyzed across patients [19], [20], with little attention paid to RS-like bias in the VAS. However, these individual variations in RS (i.e., individual RS heterogeneity) may affect inter-individual comparisons of subjective health status [21]. Therefore, RS heterogeneity in the VAS is consequential.

Several RS quantization methods based on Likert scales have been proposed [22]. Previous studies, using simple ideas, represented RSs as the ratio of midpoints or extremes to the total number of questionnaire items [13]. This method is easy to use, but cannot be used to separate the RS from the responses. Other studies attempted to separate RSs and psychological constructs using the anchoring vignette method [23], [24]. In anchoring vignette methods, people report their feelings as well as the status of hypothetical individuals or situations based on short descriptions of the hypothesis (referred to as anchoring vignettes). Using both self-reports and reports on others, the anchoring vignette method separates the RS through statistical modeling. This method can separate RS from responses but requires the modification of questionnaires to carefully design anchoring vignette parts, and satisfy response consistency and vignette equivalence. Similarly, several studies have proposed RS elimination methods without modifying the questionnaires [25], [26], [27], [28]. Considering RS emerges independently of questionnaire items, it is possible to distinguish instruction-independent RS parameters using item response theory (IRT). Although this third method can eliminate RS without questionnaire modification, it requires that all people answer the items of various questionnaires in the same manner (i.e., every individual rates all questionnaire items on multiple occasions as instructed) to assume task independence. Thus, three types of RS quantization have been proposed for the Likert scale.

Most previous related studies focused on RS handling methods for Likert scales [29]. One reason for this is that most methods rely on IRT modeling for single or multiple questionnaires, including a fixed number of response items [24], [27]. IRT has been vigorously developed from the perspective of test theory, and generally assumes that responses to the same items of specific questionnaires are obtained from all participants only once. With this restriction, IRT-based RS elimination methods estimate the item-independent effects of the questionnaire as RS parameters. The extension of IRT from items with discrete ordinal scales, such as the Likert scale, to that with continuous interval scales including VAS, have already been proposed [30], [31], [32]. However, repeated VAS measurements, for which the number of responses to items can vary among respondents, and some responses may be omitted, do not satisfy such assumptions. For example, participants #A and #B may have responded to the same scale three and four times, respectively, during the monitoring period. Considering these characteristics, IRT extensions to repeatedly measured VAS are not simple; quantification methods of RS in VAS are required (hereafter referred to as response profile (RP), to distinguish it from RS), especially considering repeated-measures assumptions.

We developed a novel RP characterization method for VAS data. The proposed method models the RP using mixture distributions that represent the RPs using a model selection scheme. We validated its robustness using simulated pseudo-data and evaluated real-world RP characteristics

using actual VAS datasets. A preliminary version of this study was reported as a poster presentation at the 45th annual international conference of IEEE Engineering in Medicine and Biology Society (IEEE EMBC 2023) [33].

## II. PROPOSED METHOD

We developed a novel RP characterization method that can be adapted for both VAS and interval scales. The proposed method is composed of (i) response profile modeling using fitting to mixture distributions of known RS-like distributions, and (ii) handling of data imbalance using bootstrap sampling and aggregation. Because the proposed method does not employ IRT modeling, it can be applied to VAS data, including with repeated responses, and different numbers of responses among individuals. Additionally, by adopting the bootstrap method, the proposed method can relieve the effect of item-dependent response tendencies on RP characterization, which is beyond the definition of RP, by adjusting the volume of each questionnaire item.

### A. RESPONSE PROFILE MODELING

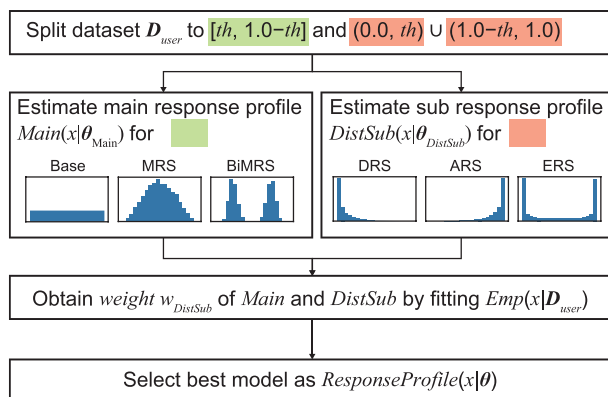
In response profile modeling, we estimate the mixture distribution parameters  $\theta$ , representing the RP of a user  $user$  (Fig. 2). Here, the proposed method estimates  $\theta$  for the user-wise VAS dataset of

$$\mathbf{D}_{user} = \{x_1, \dots, x_n, \dots, x_N\} \subset (0, 1) \quad (1)$$

where  $x$  is the self-reported data obtained using the VAS, normalized to the open interval data between 0 and 1. When  $x$  has extreme values (i.e., 0 and 1), the sample  $x$  of  $\mathbf{D}_{user}$  is converted to open interval data using the following [34] and [35]:

$$x' = \frac{x(N-1) + 0.5}{N}, \quad (2)$$

where  $N$  is the total number of  $\mathbf{D}_{user}$ . The empirical probability density distribution is represented as  $Emp(x|\mathbf{D}_{user})$ .



**FIGURE 2.** Schematic diagram of the proposed response profile modeling.

The proposed method approximates  $Emp(x|\mathbf{D}_{user})$  by the mixture distribution  $ResponseProfile(x|\theta)$ . Generally, the

probability distributions of unimodal continuous variables are modeled using *Gaussian* and *Beta* [36], [37], [38], [39]. However, some actual VAS datasets show a non-unimodal distribution, such as a mixture of ERS and BiMRS distributions, as shown in Fig. 1F. Additionally, when we handle RPs in the following analysis, it was desirable to access knowledge of RS for Likert scales. Regarding widely known RSs, ARS/DRS/ERS and MRS are observed at both ends of, and the center of the scale, respectively. Considering these requirements, the proposed method splits dataset  $\mathbf{D}_{user}$  into sub-datasets  $\mathbf{D}_{user}^{Main} \subset [th, 1.0 - th]$  and  $\mathbf{D}_{user}^{DistSub} \subset (0.0, th) \cup (1.0 - th, 1.0)$ , and fits each dataset to distributions  $Main(x|\theta_{Main})$  and  $DistSub(x|\theta_{DistSub})$  using known RSs, and model  $RP(x|\theta_{user})$  based on these two distributions, where  $th$  is the threshold to split the dataset on the dimension of the VAS scale. By connecting RPs with widely known RSs, we can utilize the various types of knowledge reported in RS studies based on a Likert scale.

### Algorithm 1 Estimate Main Response Profile

1	Construct $\mathbf{D}_{user}^{Main}$ by extracting $[th, 1.0 - th]$ range from $\mathbf{D}_{user} \subset (0.0, 1.0)$
2	<b>for</b> $DistMain = \{\text{Base, MRS, BiMRS}\}$ <b>do</b>
3	Obtain $\theta_{DistMain}$ by fitting $Emp(x \mathbf{D}_{user}^{Main})$ to $DistMain(x \theta_{DistMain})$
4	Calculate fitness indices for $Emp(x \mathbf{D}_{user}^{Main})$ and $DistMain(x \theta_{DistMain})$
5	<b>end for</b>
6	<b>if</b> $\{\text{Bipolar VAS was collected in } \mathbf{D}_{user}\}$ and $\{\text{Dist}(\text{BiMRS}(x \theta_{\text{BiMRS}}) \geq \text{ACCEPT\_BIDIST})$ and $\{\text{BiMRS}(x \theta_{\text{BiMRS}})$ gets best fitness in $Main\}$ <b>do</b>
7	Select $\text{BiMRS}(x \theta_{\text{BiMRS}})$ as $Main(x \theta_{Main})$
8	<b>else if</b> $\{\text{MRS}(x \theta_{\text{MRS}})$ gets better fitness than $\text{Base}(x \theta_{\text{Base}})\}$ <b>do</b>
9	Select $\text{MRS}(x \theta_{\text{MRS}})$ as $Main(x \theta_{Main})$
10	<b>else do</b>
11	Select $\text{Base}(x \theta_{\text{Base}})$ as $Main(x \theta_{Main})$
12	<b>end if</b>
13	<b>return</b> $Main, \theta_{Main}$

The proposed method estimates  $\theta_{Main}$  (Algorithm 1). Here, we assume  $Main(x|\theta_{Main})$  can be modeled by  $\text{MRS}(x|\theta_{\text{MRS}})$  or  $\text{BiMRS}(x|\theta_{\text{BiMRS}})$  if RP exists in  $[th, 1.0 - th]$ . Then, MRS can be represented by an adequate unimodal distribution such as a normal distribution  $Gaussian(x|\mu, \sigma)$ , where  $\mu$  represents the mean, and  $\sigma$  represents the standard deviation, or a beta distribution  $Beta(x|\alpha, \beta)$ , where  $\alpha$  and  $\beta$  are shape parameters. Moreover, BiMRS can be modeled using an adequate mixture of two unimodal distributions. For example, when we select  $Beta(x|\alpha, \beta)$  as a unimodal distribution, then we select  $BetaMixture(x|w_1, \alpha_1, \beta_1, w_2 = 1 - w_1, \alpha_2, \beta_2)$  as a bimodal distribution, where  $w$  denotes the weight of  $Beta_1(x|\alpha_1, \beta_1)$  and  $Beta_2(x|\alpha_2, \beta_2)$ .

**Algorithm 2** Estimate Sub Response Profiles

---

```

1 Construct  $D_{user}^{DistSub}$  by extracting  $(0.0, th) \cup (1.0-th, 1.0)$  range from  $D_{user}$ 
2 for  $DistSub = \{ERS, DRS, ARS\}$  do
3   Obtain  $\theta_{DistSub}$  by fitting  $Emp(x|D_{user}^{DistSub})$  to  $DistSub(x|\theta_{DistSub})$ 
4   Calculate fitness indices for  $Emp(x|D_{user}^{DistSub})$  and  $DistSub(x|\theta_{DistSub})$ 
5 end for
6 return  $\theta_{ERS}, \theta_{DRS}, \theta_{ARS}$ 

```

---

We further fit  $Main(x|\theta_{Main})$  from  $D_{user}^{Main}$  using a model selection scheme with a criterion such Akaike's information criterion (AIC). However, because actual datasets do not always show clear RPs, a simple model selection for a dataset with unclear RPs tends to choose BiMRS, which is more complex in terms of the number of parameters. Thus, in addition to information criterion-based selection, the proposed method also evaluates the degree of separation of the bimodal distribution, which selects the BiMRS with the satisfaction of the smallest information criterion and higher separation than the minimum separation threshold *ACCEPT\_BIDIST*. For example, to calculate the degree of separation, we can use a distance  $Dist(\theta)$  between peaks of each of the fitted mixture distributions. When we approximate BiMRS by *BetaMixture*,  $Dist(\theta)$  can be defined by the difference in the mode of the two Beta distributions, as follows:

$$Dist(\theta_{BiMRS}) = \left| \frac{\alpha_2 - 1}{\alpha_2 + \beta_2 - 2} - \frac{\alpha_1 - 1}{\alpha_1 + \beta_1 - 2} \right| \quad (3)$$

Thus, the proposed method selects  $\theta_{Main}$  from  $\theta_{MRS}$  or  $\theta_{BiMRS}$ , or regards no RPs in  $[th, 1.0-th]$ .

**Algorithm 3** Whole Response Profiles Estimation

---

```

1  $Main, \theta_{Main} \leftarrow$  Estimate main response profile in Algorithm 1
2  $\theta_{ERS}, \theta_{DRS}, \theta_{ARS} \leftarrow$  Estimate sub response profiles in Algorithm 2
3 for  $DistSub = \{ERS, DRS, ARS\}$  do
4   Obtain weight  $w_{DistSub}$  by fitting  $Emp(x|D_{user})$  to  $w_{DistSub} \times DistSub(x|\theta_{DistSub}) + (1-w_{DistSub}) \times Main(x|\theta_{Main})$ 
5   Calculate fitness indices between  $Emp(x|D_{user})$  and  $w_{DistSub} \times DistSub(x|\theta_{DistSub}) + (1-w_{DistSub}) \times Main(x|\theta_{Main})$ 
6 end for
7 If model better than  $Main(x|\theta_{Main})$  exists then
8   Select best model components  $w_{ADE}$  and  $DistSub(x|\theta_{DistSub})$ 
9    $ResponseProfile(x|\theta) \leftarrow w_{ADE} \times DistSub(x|\theta_{DistSub}) + (1-w_{ADE}) \times Main(x|\theta_{Main})$ 
10 else  $ResponseProfile(x|\theta) \leftarrow Main(x|\theta_{Main})$ 
11 end if
12 return  $ResponseProfile(x|\theta)$ 

```

---

The proposed method estimates  $\theta_{DistSub}$  based on the RS knowledge (Algorithm 2). Here, we assume  $DistSub(x|\theta_{DistSub})$  can be modeled by  $ARS(x|\theta_{ARS})$ ,  $DRS(x|\theta_{DRS})$ , or  $ERS(x|\theta_{ERS})$  if RP exists. Because these RSs show monotonic or U-shaped distributions (Fig. 1A-C),  $DistSub(x|\theta_{DistSub})$  is modeled by a Beta distribution  $Beta(\alpha_{ADE}, \beta_{ADE})$  satisfying the shape restriction to represent each RS shape, which can flexibly change its distribution shape. The proposed method obtains  $\theta_{ARS}$ ,  $\theta_{DRS}$ , and  $\theta_{ERS}$ , which are candidates of  $\theta_{DistSub}$ , based on  $D_{user}^{DistSub}$  to fit to  $Beta(\alpha_{ADE}, \beta_{ADE})$ , respectively. For each candidate RP, fixing  $\theta_{Main}$  and candidates of  $\theta_{DistSub}$ , the proposed method calculates the optimal mixture weight  $w_{DistSub}$ , and finally selects the best integration of  $\theta_{Main}$ ,  $w_{ADE}$ , and the candidate of  $\theta_{DistSub}$  in terms of information criterion for  $D_{user}$ . For example, when  $\theta_{ERS}$  obtains the best fit, the proposed method accepts  $\theta_{ERS}$  as  $\theta_{DistSub}$ .

Finally, the proposed method characterizes RPs as  $\theta$  (Algorithm 3). Using the obtained parameters  $\theta_{Main}$ ,  $w_{ADE}$ , and  $\theta_{DistSub}$ , we obtain the mixture distribution representing RPs as

$$ResponseProfile(x|\theta) = w_{ADE} \times DistSub(x|\theta_{DistSub}) + (1 - w_{ADE}) \times Main(x|\theta_{Main}) \quad (4)$$

and the RP parameters  $\theta$  as  $\theta = \{w_{ADE}, \theta_{DistSub}, \theta_{Main}\}$ . For example, in the case of Fig. 1F, with a beta distribution representation, the proposed method characterizes ERS and BiMRS as RPs and obtains

$$\begin{aligned} \theta &= \{w_{ADE}, \theta_{ERS}, \theta_{Main}\} \\ &= \{w_{ADE}, \alpha_{ERS}, \beta_{ERS}, w_1, \alpha_1, \beta_1, w_2 = 1 - w_1, \alpha_2, \beta_2\}. \end{aligned} \quad (5)$$

**B. UNBALANCED DATA HANDLING**

An unbalanced data handling process adjusts the imbalance of the number of responses for each instruction and evaluates the range of RP parameters  $\theta$  (Fig. 3). In contrast to responses to validated questionnaires, in which all users are expected to respond to the same number of questions, the number of responses varies depending on the question in repeated VAS measurements. Because RS and RP are defined as question-independent systematic response tendencies, such an imbalance in the dataset  $D_{user}$  may affect bias in the quantification of RPs by item-dependent tendencies. Specifically, for VAS, the tendency to respond to the instruction center emerges as MRS in monopolar VAS scales, and BiMRS in bipolar VAS scales. To relieve and estimate the impact of such bias, the proposed method uses bootstrapping.

When  $D_{user}$  has a data imbalance, the proposed method constructs a sub-dataset of  $D_{user}$  by bootstrap sampling.  $D_{user}$  comprises  $M$  types of unipolar VAS  $VAS_m^u$  and  $N$  types of bipolar VAS  $VAS_n^b$ . Each type of VAS may include a different number of responses. The proposed method then conducts sampling with replacement by the type of VAS

TABLE 1. Simulation data generated based on beta distributions.

#	Model description	$w_1$	$\alpha_1$	$\beta_1$	$w_2$	$\alpha_2$	$\beta_2$	$w_{ADE}$	$\alpha_{ADE}$	$\beta_{ADE}$
11	ERS	0.0	–	–	–	–	–	1.0	0.1	0.1
12	DRS	0.0	–	–	–	–	–	1.0	1.0	30.0
13	ARS	0.0	–	–	–	–	–	1.0	30.0	1.0
14	MRS-a	1.0	10.0	10.0	–	–	–	–	–	–
15	MRS-b	1.0	15.0	45.0	–	–	–	–	–	–
16	MRS-c	1.0	45.0	15.0	–	–	–	–	–	–
17	BiMRS-a	0.5	15.0	45.0	0.5	45.0	15.0	–	–	–
18	BiMRS-b	0.5	15.0	30.0	0.5	30.0	15.0	–	–	–
19	BiMRS-c	0.5	15.0	20.0	0.5	20.0	15.0	–	–	–
21	ERS-05_MRS-05	1.0	10.0	10.0	–	–	–	0.5	0.1	0.1
22	ERS-03_MRS-07	1.0	10.0	10.0	–	–	–	0.3	0.1	0.1
23	ERS-01_MRS-09	1.0	10.0	10.0	–	–	–	0.1	0.1	0.1
24	DRS-05_MRS-05	1.0	10.0	10.0	–	–	–	0.5	1.0	30.0
25	DRS-03_MRS-07	1.0	10.0	10.0	–	–	–	0.3	1.0	30.0
26	DRS-01_MRS-09	1.0	10.0	10.0	–	–	–	0.1	1.0	30.0
31	ERS-05_BiMRS-05	0.5	15.0	30.0	0.5	30.0	15.0	0.5	0.1	0.1
32	ERS-03_BiMRS-07	0.5	15.0	30.0	0.5	30.0	15.0	0.3	0.1	0.1
33	ERS-01_BiMRS-09	0.5	15.0	30.0	0.5	30.0	15.0	0.1	0.1	0.1
34	DRS-05_BiMRS-05	0.5	15.0	30.0	0.5	30.0	15.0	0.5	1.0	30.0
35	DRS-03_BiMRS-07	0.5	15.0	30.0	0.5	30.0	15.0	0.3	1.0	30.0
36	DRS-01_BiMRS-09	0.5	15.0	30.0	0.5	30.0	15.0	0.1	1.0	30.0

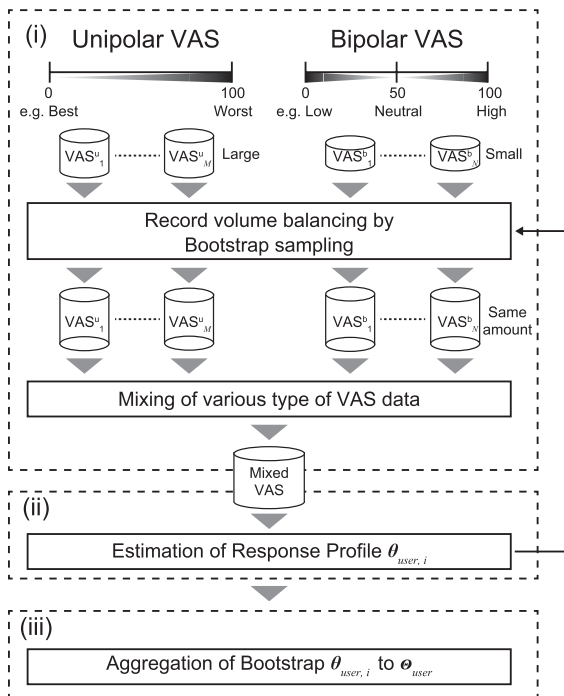


FIGURE 3. Schematic diagram of the proposed response profile quantification method. The proposed method comprises (A) Data volume balancing of several types of VAS data, (B)  $i$ -th response profile  $\theta_{user,i}$  estimation, and (C) Bootstrap aggregation process of  $\Theta_{user} = \{\theta_{user,i}\}$  to aggregated response profile indicator  $\theta_{user}$ .

scale to extract the same number of samples (Fig. 3(i)). When VAS types have a hierarchical structure, such as in the case of both unipolar and bipolar scales, and with dif-

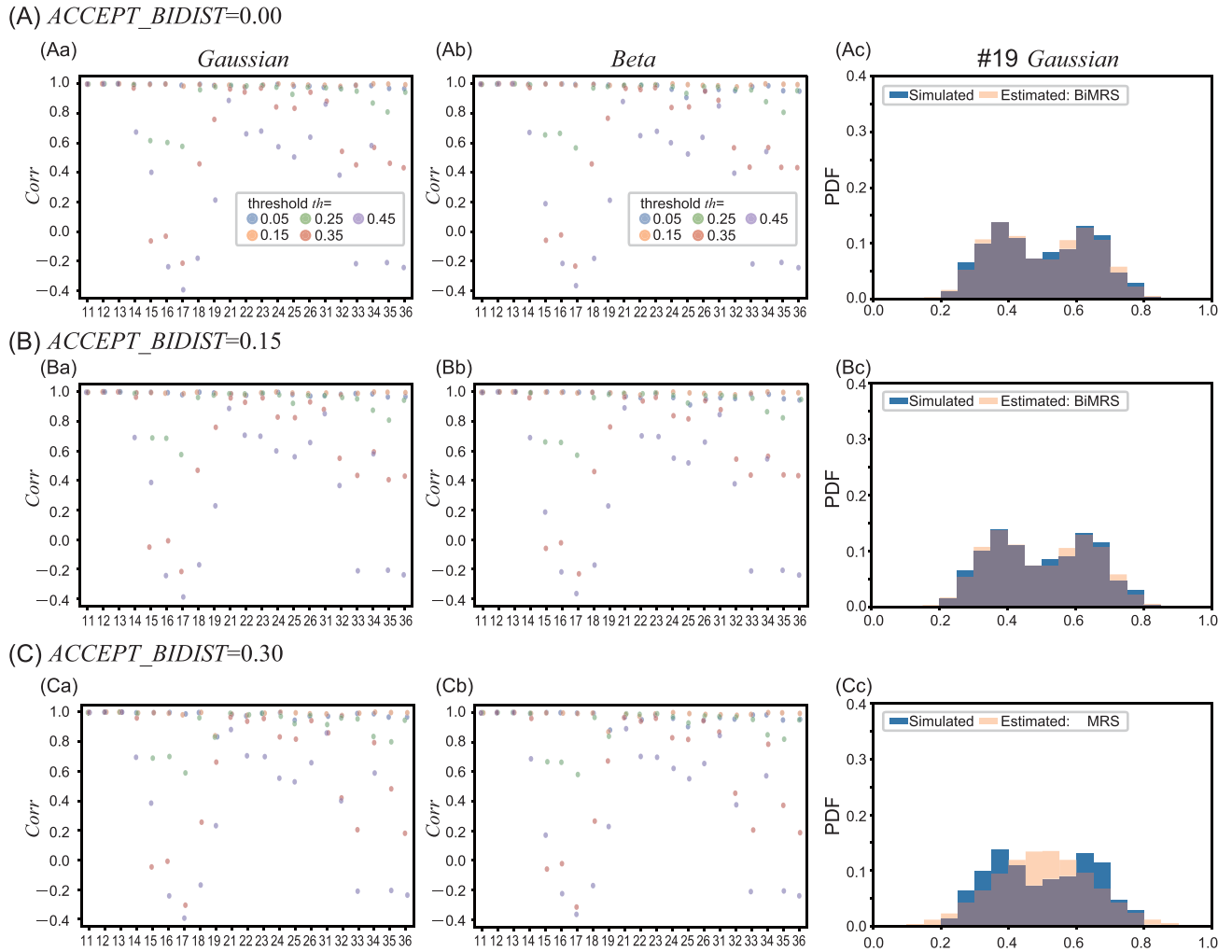
ferent instructions within the scale type, this process can be repeated. After sampling, all the sampled data were used as the dataset for response profile modeling. During sampling, the proposed method constructs an  $i$ -times sampling of the sub-dataset  $D_{user,i}$  from the original dataset  $D_{user}$ . The sub-dataset  $D_{user,i}$  includes the same number of responses by instruction type; therefore, this process can be regarded as the artificial construction of a dataset in which the method can characterize RPs under relieving the effect of item-dependent response tendencies. From the machine learning perspective, this process is also interpreted as class imbalance addressing methods such as oversampling, undersampling, and bagging [40], [41].

Using the sub-dataset  $D_{user,i}$ , the proposed method evaluates the variability of the RP parameters. After constructing  $D_{user,i}$ , we obtain  $i$  th sampling of RP parameters  $\theta$  as  $\theta_{user,i}$  by the response profile modeling process (Fig. 3(ii)). However,  $\theta_{user,i}$  may differ from the actual RP parameters owing to the biased extraction in the sampling process. To assess such variability, the proposed method repeats this sampling and RP modeling cycles  $N$  times as  $\Theta_{user} = \{\theta_{user,i}\}$ , and estimates the characteristics of this bootstrap samples, including ranges and representative statistics (Fig. 3(iii)). Finally, the proposed method obtains RP parameters  $\theta$  by aggregating  $\Theta_{user}$ .

### III. PSEUDO-DATA SIMULATION

#### A. SETUP OF SIMULATION

To assess the robustness of the proposed method, we performed a parameter recovery experiment using simulated



**FIGURE 4.** Parameter recovery results similarity in the shape of the recovered distribution. The similarity is evaluated by Pearson’s correlation coefficient, *Corr*. The main response profile distributions are approximated for (A) Gaussian and (B) Beta distributions. Each color depicts the result for hyperparameter *th*. (a)-(c) represent the effect of hyperparameter *ACCEPT\_BIDIST*. (C) Empirical and estimated response profile distribution for #19 with *th* = 0.05. As shown in the change from (Cb) to (Cc), the selected Main, BiMRS is substituted with MRS.

data. We included various RP types and those mixtures with 21 conditions as  $\theta_{\text{groundtruth}}$  (Table 1), regarded as participants with different RPs. For each condition, 1000 samples of response data were extracted from a beta mixture distribution parameterized by  $\theta_{\text{groundtruth}}$ . The model selection used Akaike’s information criterion. Assuming this simulated data satisfies the task independence (i.e., already balanced), we obtained RP parameters  $\theta_{\text{user}} = \theta_{\text{estim}}$  when the hyperparameters were changed; we changed the data splitting threshold *th* from 0.05 to 0.45 in steps of 0.10, the minimum separation threshold *ACCEPT\_BIDIST* with three conditions (i.e., 0.00, 0.15, 0.30), and unimodal distribution with two conditions (i.e., *Gaussian* and *Beta*).

**B. EVALUATION**

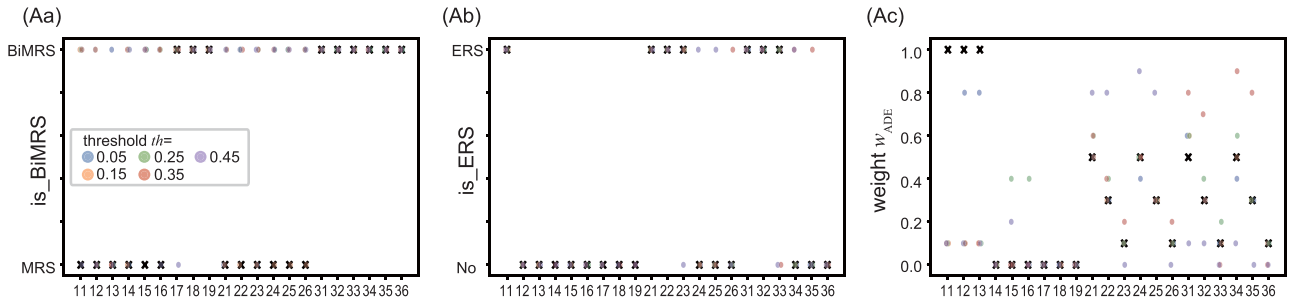
We evaluated the robustness of the metrics by changing the hyperparameters. We further calculated Pearson’s correlation coefficient for the histograms (*Corr*) to assess

how the proposed method approximated the shape of the empirical probability density distribution [42]. For comparison, we obtained empirical probability density distributions using a histogram with a bin width of 0.05. We also analyzed the agreement between the estimated RP parameters  $\theta_{\text{estim}}$  and  $\theta_{\text{groundtruth}}$  with both Pearson’s correlation coefficient *r* and its linear regression. Here,  $\theta_{\text{groundtruth}}$  was defined with beta distributions  $Beta(x|\alpha, \beta)$ , with mean  $\mu$  and standard deviation  $\sigma$  calculated as follows:

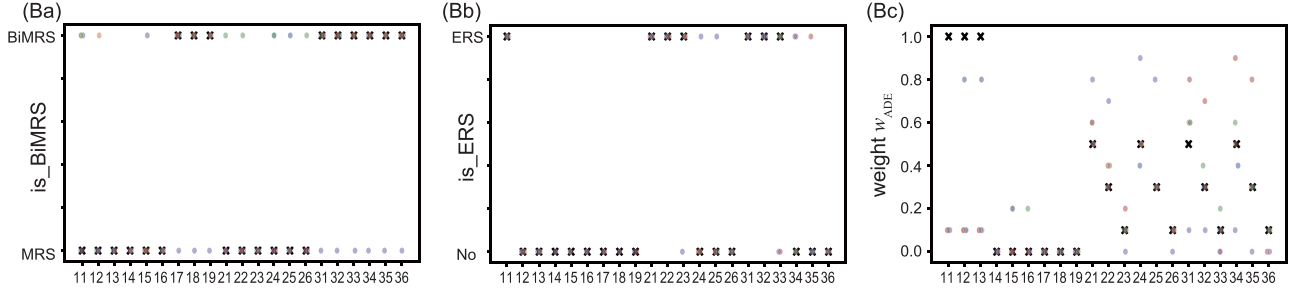
$$\mu = \frac{\alpha}{\alpha + \beta}, \sigma = \sqrt{\frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}} \tag{6}$$

Therefore, in the agreement analysis for the *Gaussian* approximation, we compared *w* and the converted  $\mu$  and  $\sigma$  based on  $\theta_{\text{groundtruth}}$ , and RP parameters  $\theta_{\text{estim}}$  estimated with  $Gaussian(x|\mu, \sigma)$ .

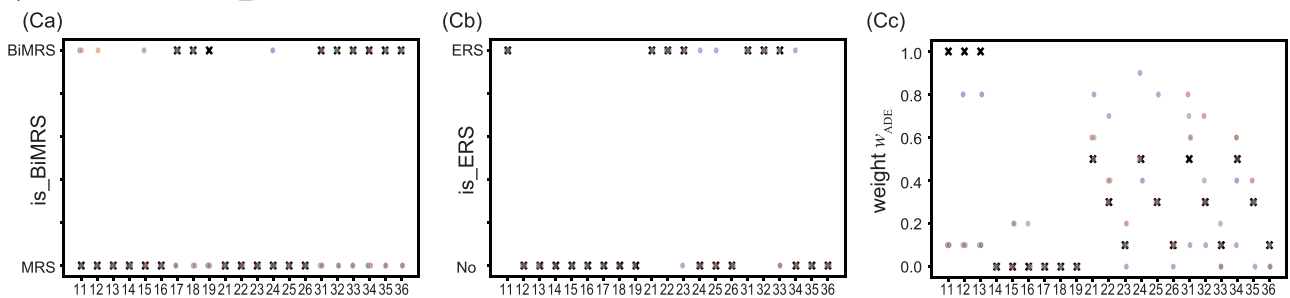
(A) Gaussian, ACCEPT\_BIDIST=0.00



(B) Gaussian, ACCEPT\_BIDIST=0.15



(C) Gaussian, ACCEPT\_BIDIST=0.30



**FIGURE 5.** Parameter recovery results in terms of response profile indicators: (A) *is\_BiMRS*, (B) *is\_ERS*, and (C) weight of sub-response profile distribution  $w_{ADE}$ . Each color depicts the result for hyperparameter  $th$ . (a)-(c) represent the effect of hyperparameter *ACCEPT\_BIDIST*. Black crosses represent the groundtruth values. In these results, the main response profile distributions are approximated using a Gaussian distribution.

**IV. HUMAN EXPERIMENT**

To describe RP characteristics in the real world, we empirically analyzed the RP distribution and the relationship between RPs and participant characteristics using an enhanced version of the experimental dataset called *DailySense*<sup>1</sup> [43]. The experimental design and procedure details have already been reported in [44], from which only the number of participants has been updated. Herein, we briefly describe the experimental information related to this analysis.

**A. EXPERIMENTAL PROCEDURE**

We conducted a consecutive 14-day experiment in a real-life setting composed of a smartphone-based subjective psychological evaluation and physiological sensing. The study protocol was approved by the internal review board of the Research & Development Group, Hitachi, Ltd., and was conducted in accordance with the Declaration of Helsinki. All

participants provided informed consent prior to enrollment in the study.

Thirty-six healthy Japanese adult workers who mainly worked remotely (mean±SD, 35.8±7.5; range, 27–58 years; 21 male and 15 female participants) participated in the experiment through two terms. In the first term, 18 participants joined voluntarily, based on their recruitment using company bulletin boards. In the second term, 18 participants were recruited through the participant pool of a research support company, with rewards depending on the engagement rate in the experiments. One participant (#47) was excluded from analysis because they did not complete the personality questionnaire. Accordingly, we analyzed the data of 35 participants (mean±SD, 35.9±7.6; range, 27–58 years; 20 male and 15 female participants).

The participants responded to various VAS and questionnaires using their smartphones. Before and after the 14-day experimental period, participants reported their demographics and completed seven questionnaires, including the

<sup>1</sup>See: <https://dx.doi.org/10.5281/zenodo.10816004>

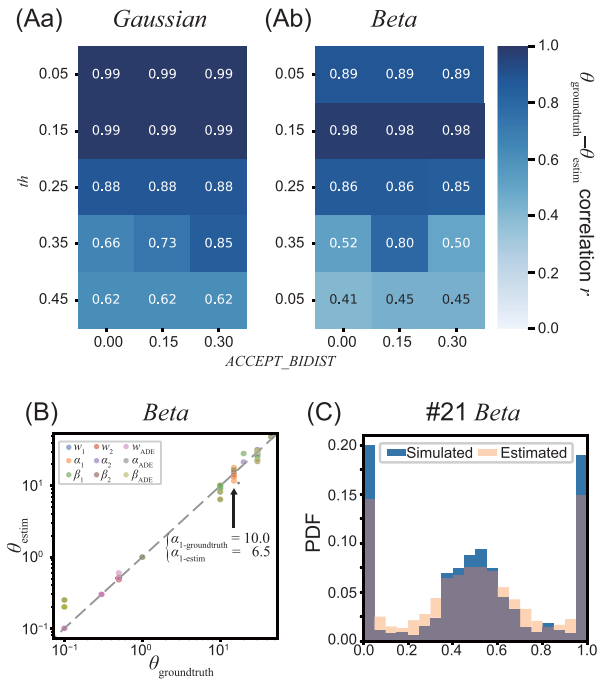
Neuroticism Extraversion Openness Five-Factor Inventory (NEO-FFI), to evaluate Big Five personality traits [45]. Participants were randomly notified of the timing of their responses six times per day using an experience sampling method (ESM), also known as ecological momentary assessment (EMA). They reported the two dimensions of emotions as felt in the 30 minutes immediately before response onset. The dimensional emotions (i.e., valence and arousal) were evaluated using the Affective Slider [46], a bipolar VAS scale with face icon instructions on both ends. Participants also reported their degrees of fatigue, stress, anxiety, depression, and sleeplessness using the day reconstruction method (DRM). These five self-reports were evaluated using unipolar VAS with the instruction of “not at all” on the left end as 0 to “worst” on the right end as 1. In both ESM and DRM self-reports, guidelines for the exact midpoint of VAS were not drawn on the scales; that is, participants had to make their responses without an available exact midpoint reference. Finally, we analyzed 2337 records of two types of bipolar VAS in ESM (on average,  $66.8 \pm 16.4$  [times/person]) and 466 records of five types of unipolar VAS in DRM (on average,  $13.3 \pm 1.2$  [times/person]) (Supplementary Table 1).

**B. EVALUATION**

The proposed method was applied to an actual VAS dataset. Because this dataset was an integration of five and two types of unipolar and bipolar VAS, respectively, and the number of responses varied, we conducted sampling with replacement in two steps to construct  $D_{user,i}$ . Based on the total number of VAS data, we sampled 300 and 1800 records in the first and second steps, respectively. We experimentally performed the response profile modeling process using *Gaussian* with hyperparameters  $th=0.15$  and  $ACCEPT\_BIDIST=0.15$ . Bootstrap sampling was repeated 1000 times. To simplify the calculation, the mixture weight  $w_{ADE}$  was estimated using a 0.1 step resolution (i.e., 0.0 to 1.0 with 0.1 steps).

Note that both valence and arousal were presented as bimodal scales using the Affective Slider [46]. Many studies have treated valence as bimodal, ranging from neutral to either negative or positive, and arousal as unimodal, ranging from low to high. For example, the Self-assessment Manikin (SAM), one of the standard scales for dimensional emotions, treats arousal as unimodal [47]. Conversely, the Affective Slider, which shows a high correlation with SAM, placed two symmetrically mirrored isosceles triangles as an intensity instruction even for arousal, such that arousal could be regarded as bimodal. Our treatment of the arousal scale as bimodal could have led to more unbalanced data collection versus a unimodal treatment. To evaluate the impact on the parameter estimation, we compared both estimation results when arousal scales were treated as bimodal in the bootstrapping process and vice versa.

We evaluated RP characteristics in the real world in two ways. First, we analyzed the stability of the shape of the RP distributions for each bootstrapped  $D_{user,i}$ . We also checked

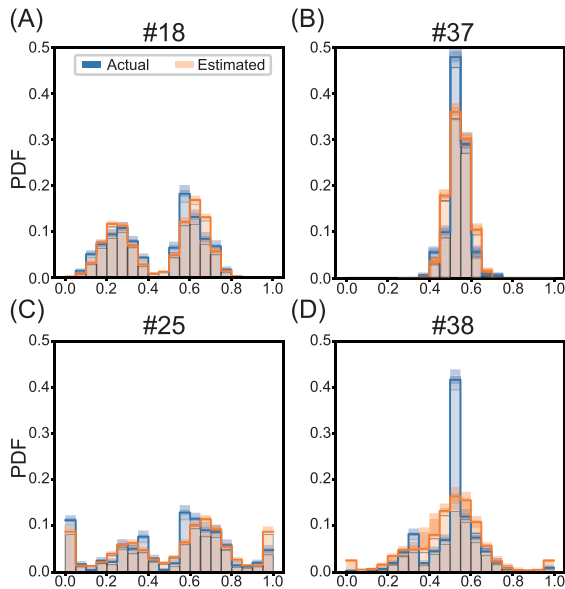


**FIGURE 6.** (A) Parameter agreements in correlation between  $\theta_{groundtruth}$  and  $\theta_{estim}$  approximated using (a) Gaussian and (b) Beta distribution. (B) Representative parameter recovery results with  $th=0.15$ ,  $ACCEPT\_BIDIST=0.15$ , and Beta. Parameter agreements between  $\theta_{groundtruth}$  and  $\theta_{estim}$  are shown. The stars\* highlights underestimation in  $\alpha_1$ . (C) Underestimated example in (B). Blue and orange histograms represent the empirical and the estimated distribution of the whole response profile distribution (i.e., MRS+ERS) for simulation #21, respectively.

whether RPs, which may cause analysis bias, emerged in the actual repeated VAS measurements. Additionally, previous studies reported that subjective perception and its output of feelings differ depending on the heterogeneity of individuality [48], [49], [50]. To clarify whether such RPs are affected by one of the individuality indicators—the participants’ personality—we analyzed the relationships between RP parameters and the Big Five personality traits. Here, RP parameters  $\theta$  comprised eight distributional parameters, which is shown in Eq. 5 except for  $w_2(=1-w_1)$ , and four one-shot features (i.e., whether they have BiMRS or MRS, whether they have ERS, whether they have DRS, and whether they have ARS). As Big Five personality indicators, we used standardized scores of five-dimensional personality as assessed by NEO-FFI as  $\theta_{dem} = \{N_t, E_t, O_t, A_t, C_t\}$ , where each represents neuroticism, extraversion, openness, agreeableness, and conscientiousness, respectively (Supplementary Table 1). Because we cannot assume normality in the relationship between  $\theta$  and  $\theta_{dem}$ , we evaluated this by Spearman’s correlation coefficient  $\rho$ .

All data processing and analyses were performed using Python 3.8.13, including SciPy 1.10.1, scikit-learn 1.0, and opencv-python 4.7.0.68. The beta mixture distribution was fitted using a previously proposed method [51]. Statistical significance considered as  $p < 0.05$ , denoted by an asterisk (\*).





**FIGURE 7.** Response profile distribution by the proposed method for real-world data. (A) BiMRS-like distribution of participant #18, (B) MRS-like distribution of participant #37, (C) Distribution composed of BiMRS-like main RP distribution and ERS-like sub-RP distribution of participant #25, and (D) MRS-like distribution but not fully characterized in terms of its peak concentration of participant #38. Blue and orange histograms represent the empirical distributions of bootstrap samples and estimated RP distributions in each bootstrap trial. Bold lines show the histograms parameterized by the median of bootstrap samples of RP parameters. Boxes around the median lines are 5–95% and 25–75% ranges, respectively.

## V. RESULTS

We confirmed the performance of the proposed method using simulated data, and analyzed the characteristics of the RP of repeated VAS measurements using actual experimental data.

### A. ASSESSMENT OF RP PARAMETER RECOVERY BY PSEUDO DATA SIMULATION

We confirmed the performance of the proposed method by assessing parameter recovery.

First, we compared the effects of the hyperparameters on the shape of the approximated distribution (Fig. 4). Even when we used *Gaussian* (Fig. 4a) or *Beta* (Fig. 4b) distributions, the proposed method obtained similarly shaped RP distributions based on *Corr* with varying thresholds  $th$  between 0.05 to 0.25. This shape similarity gradually decreased when  $th$  was over 0.25. Since larger  $th$  (e.g.,  $th=0.25$ ) includes both the end and the center ranges of the scales, this indicates robustness in terms of the model distribution and the change in  $th$  if *DistSub* only affects the end of the VAS scales. However, when the degree of separation of the bimodal distribution was changed, *Corr* changed significantly, especially when both centers of responses were closed (Fig. 4ab, #19). For example, the *Main* RP of #19 was accurately estimated as BiMRS for  $ACCEPT\_BIDIST=0.15$ , and MRS for  $ACCEPT\_BIDIST=0.30$  (Fig. 4Cbc). This result suggests  $ACCEPT\_BIDIST$  requires appropriate settings based on the data characteristics.

We further assessed the stability of the estimated parameters, particularly for one of the *Main* RP parameters ( $is\_BiMRS$ ), one of the *DistSub* parameters ( $is\_ERS$ ), and mixture weight  $w_{ADE}$ . Because we confirmed the robustness of the estimation results regarding the candidate distribution seed, hereafter, we only show the results of *Gaussian* fitting (Fig. 5), unless otherwise stated. As in the shape similarity analysis, these parameters were correctly estimated even when the hyperparameters were changed with  $th$  smaller than 0.25, indicating robust estimation of the RP parameters.

Finally, we assessed the agreement of  $\theta_{\text{estim}}$  and  $\theta_{\text{groundtruth}}$ . The analysis showed that the agreement was less affected by the approximated distribution (i.e., *Gaussian* and *Beta*) and  $ACCEPT\_BIDIST$ , but changed depending on  $th$  (Fig. 6A). The agreement became best when the threshold  $th$  approximated 0.15, although the agreement gradually decreased when  $th$  became large (see Supplementary Table 2 for details). This result indicates that data thresholding should be adequately considered. For more comprehensive analysis, we focused on the better agreement case (i.e., with *Beta* mixture modeling,  $th=0.15$ ,  $ACCEPT\_BIDIST=0.15$ ; Pearson's correlation coefficient,  $r = 0.99$ ,  $p < 0.001$ ;  $y=1.01x - 0.19$ ,  $R^2 = 0.97$ ; Fig. 6B). This result showed good linear relationships, but several parameters, such as  $\alpha_1$ , seemed to be slightly underestimated. For a detailed analysis, we confirmed the results in #21 as the worst underestimation case (Fig. 6C). When we compared the shape of the RP distribution, the height of ERS and the estimated peak showed a slight difference. However, most of the characteristics of the ERS and unimodality were sufficiently approximated by the proposed method. These results indicate the acceptable RP characterization using the proposed method, unless several parameters are underestimated.

### B. RP CHARACTERIZATION FOR EMPIRICAL STUDY

Using the proposed method, we evaluated the emergence of RP in an empirical study using repeated VAS measurements. Fig. 7 shows the representative types of RP characterizations. Through unbalanced data handling, the proposed method estimated the range of RP depending on each questionnaire response; for example, in #25,  $\mu_1$ , the mean of the first peak, was estimated as 0.66 (0.30–0.67; 95% bootstrap confidence interval, CI). Although the ranges of estimated parameters may be slightly wider, similar RP distribution results were observed based on the estimated parameters (Fig. 7C). Even if the arousal scale was treated as unimodal, most of the ranges of estimated parameters overlapped (e.g., 0.30–0.67 for the case treated in bimodal manner vs. 0.31–0.67 for the case treated in unimodal manner,  $\mu_1$  for #25; see Supplementary Table 3 and 4 for details). These results suggest the robustness of our method by introducing unbalanced data handling. This indicates that RP in the VAS stably emerges, similar to studies using Likert scales. Hereafter, we only show the analysis results in which the arousal scale was treated as bimodal in the unbalanced data handling process.

We comprehensively analyzed each characterization result. Even in a study using both bipolar and unipolar VASs, although most participants (34 of 35) showed BiMRS-like tendencies as *Main* RP, the proposed method clarified the existence of both BiMRS-like and MRS-like RPs (Fig. 7AB). Moreover, some participants had both *Main* and *DistSub* RPs, such as the BiMRS and ERS (Fig. 7C). Regarding *DistSub* RPs, our method quantified 14 participants with an ERS tendency and three with a DRS tendency, but none with an ARS tendency. Despite this success in RP characterization, a few participants showed insufficient fit. For example, for participant #38, the proposed method cannot fully approximate the shape of the response distribution in terms of its peak concentration, which seems to be a limitation of using a *Gaussian* distribution (Fig. 7D). In this case, the proposed method also identified #38 as BiMRS-like RP with larger weight  $w_1 = 0.83$  (CI, 0.16–1.00) but it might appear MRS-like RP based on visual inspection. This tendency was also observed when the CI of  $w_1$  included 1.00 in its range (3 out of 35).

Finally, we analyzed the relationships between the RP parameters and the participants' personalities. Regarding background demographics, the participants exhibited five personalities, as shown in Table 2. The correlation analysis indicated almost no clear relationship between RP parameters and Big-Five parameters (Table 3), except for the mean of the left peak  $\mu_L$  and conscientiousness ( $\rho = -0.39$ ,  $p = 0.020$ ). This result indicates that RP parameters may be unique characteristics as distinct from the respondents' personalities.

## VI. DISCUSSION

We developed a novel method to characterize RSs in VAS measurements (i.e., RPs) consisting of response profile modeling and unbalanced data handling. We further assessed the robustness of the proposed method using a parameter recovery experiment with simulated pseudo data. The proposed method was also applied to evaluate the RP characteristics of an empirical study dataset, clarifying the existence of RP heterogeneity, even in repeated VAS measurements.

We confirmed the characteristics of the proposed method through simulated and actual data analyses. The proposed method has three hyperparameters: degree of separation of the two peaks *ACCEPT\_BIDIST*, data-split threshold  $th$ , and types of candidate distributions. We showed that *ACCEPT\_BIDIST* might change the characterization of *Main* to MRS or BiMRS (Fig. 4C). *ACCEPT\_BIDIST* is a parameter that provides the constraints for selecting a complex bimodal model. Considering its background, this characteristic does not cause severe issues when only one *Main* RP exists (e.g., when the applied dataset is constructed only from unipolar or bipolar VAS responses). However, other cases using both scales, such as the human experiment in this study, require careful setting of the peak separation parameter. In particular, the proposed method sometimes selected BiMRS-like RP different from the visual inspection-based MRS-like envelope in the model selection process, although

the bootstrap CI implied the possibility of MRS-like RP (Fig. 7D). For more stable RP characterization, in future studies we must consider better model selection criteria to distinguish whether RP has MRS or BiMRS.

Moreover, in some cases, the proposed method showed a slightly lower approximation of the RP distribution than that of the empirical existence distribution (Figs. 6C and 7D). This is partly because of the characteristics of the candidate distributions. *Gaussian* and *Beta* distributions are the first choices to represent a unimodal probability distribution. These distributions are limited to specific shape characteristics, especially in the center concentrations, and the length of the tail. Although an improvement in fit is not necessarily needed when only RP classification is required because the proposed method sufficiently captures the RP-related shape characteristics, better fit is desirable when RP parameters themselves are used in the following between-individual comparison, such as for controlling covariates. Regarding the use of these RP parameters, introducing more parameter distributions, such as the Student's  $t$  distribution and Laplace distribution, may improve model fit [52]. Another possible reason for inadequate fit may be the conversion of the dataset using Eq. 2. This conversion was used to omit extreme values and may have particularly affected approximating ERS-like RS as *DistSub*. Updates for better ERS approximation will be a topic of our future studies.

The empirical analysis showed heterogeneity of actual RP emergence in VAS measurements. In our analysis, most participants identified their MRS or BiMRS tendency as *Main* RP, although the intensity of how they preferred midpoint selection differed among participants (Supplementary Table 1). When we interpret BiMRS as a midpoint preference between neutral and the instructions at both ends, this result is consistent with previous studies that showed a clear MRS tendency in Japanese respondents [14], [15]. This heterogeneity in the modality also follows previous studies, in which various emotion measurement datasets using VAS showed bimodality between 20.3–48.9%, regardless of whether the location guidelines of the scales were initially set to the exact midpoint or not [16]. Since midpoint guidelines were not shown on the VAS scales in this experiment, this tendency was likely not a consequence of our data collection procedures, but rather represented participants' characteristics. This result may imply the effectiveness of *ACCEPT\_BIDIST*=0.15, which we experimentally set for this dataset, for other general datasets handling self-reports. In addition to *Main* RP, *DistSub* RP varied among participants. Unlike the conventional VAS usage of within-individual comparisons, RP heterogeneity can distort the results of between-individual VAS comparisons, similar to cross-sectional studies using Likert scales [13], [53]. Moreover, the RP parameters had fewer relationships with the participants' personalities, indicating the difficulty in controlling the effect by introducing demographic covariates (Table 3). Our analysis suggests that we must consider the existence of RP heterogeneity, which can

**TABLE 2. Correlations among big five personality traits.**

	$N_t$	$E_t$	$O_t$	$A_t$	$C_t$
Neuroticism	—	<b>-0.38*</b>	0.09	0.01	-0.33
Extraversion	—	—	-0.04	0.21	0.13
Openness	—	—	—	-0.08	<b>0.36*</b>
Agreeableness	—	—	—	—	0.02
Conscientiousness	—	—	—	—	—

$N_t$ , standardized  $t$ -score of Neuroticism;  $E_t$ , Extraversion;  $O_t$ , Openness;  $A_t$ , Agreeableness; and  $C_t$ , Conscientiousness; \* $p < 0.05$ .

**TABLE 3. Correlations between big five personality traits and RP parameters.**

	$w_1$	$\mu_1$	$\sigma_1$	$\mu_2$	$\sigma_2$	$w_{ADE}$	$\alpha_{ADE}$	$\beta_{ADE}$	$is\_BiMRS$	$is\_ERS$	$is\_DRS$
Neuroticism	0.05	0.18	-0.22	0.18	-0.08	-0.31	-0.07	-0.07	-0.25	-0.30	-0.03
Extraversion	0.13	-0.17	0.06	-0.20	0.16	0.27	0.21	0.23	-0.07	0.29	0.05
Openness	-0.11	-0.14	-0.01	0.13	0.02	0.01	0.07	0.07	0.05	0.01	-0.06
Agreeableness	0.24	0.25	-0.07	-0.13	-0.13	-0.13	-0.04	-0.04	0.19	-0.13	-0.02
Conscientiousness	-0.11	<b>-0.39*</b>	0.26	0.14	0.10	0.18	-0.02	0.00	0.02	0.10	0.07

Note that the results of  $is\_ARS$  are not shown as no participant exhibited  $is\_ARS$ . \* $p < 0.05$ .

degrade inter-individual analysis, even for real-world VAS measurements.

Finally, this study has limitations. In contrast to RS removal methods [25], [26], [27], it is challenging to extend our method directly to RP removal methods. Based on the realistic constraints that users rated various self-report questionnaires with repeated measures, our method assumes that users repeatedly respond to various questionnaire items a different number of times. Owing to this constraint, we cannot simply utilize the task independence of the previously proposed methods, which assume that various users answer fixed questionnaire items the same number of times, especially once per item. Although RP characterization is helpful in terms of using RP parameters as covariates, there is need for future studies to extend this to enable RP removal. Moreover, we applied the proposed method to a single empirical dataset. This *DailySense* dataset was limited to a small Japanese population [43], [44]. To confirm the deleterious effects of RP heterogeneity, the proposed method should be utilized with larger datasets that include participants with diverse cultural backgrounds, namely cross-cultural studies.

## VII. CONCLUSION

This study proposes a novel characterization method for the RSs observed in the VAS (denoted as the response profile, RP). Our proposed method robustly identified RPs in simulated and repeatedly measured VAS datasets. These findings suggest the existence of participant-dependent RP heterogeneity, which can degrade inter-individual analyses. These results elucidate the importance of RP evaluation, even in the analysis of self-reported data obtained by VAS. Future RP-aware analyses of repeatedly measured VAS data will

offer the possibility of RP heterogeneity-less inter-individual analyses, enabling more precise subjective health research.

## ACKNOWLEDGMENT

The authors thank Dr. Hirohiko Kuratsune and Sakura Tatsumi of Osaka Metropolitan University for their helpful suggestions regarding the human experiments. They also thank ASMARQ Company Ltd., for recruiting participants and supporting procedural operations during the second term of the human experiment.

## REFERENCES

- [1] I. McDowell, *Measuring Health: A Guide to Rating Scales and Questionnaires*. London, U.K.: Oxford Univ. Press, 2006, doi: 10.1093/acprof:oso/9780195165678.001.0001.
- [2] M. Tanaka, S. Tajima, K. Mizuno, A. Ishii, Y. Konishi, T. Miike, and Y. Watanabe, "Frontier studies on fatigue, autonomic nerve dysfunction, and sleep-rhythm disorder," *J. Physiological Sci.*, vol. 65, no. 6, pp. 483–498, Nov. 2015, doi: 10.1007/s12576-015-0399-y.
- [3] K. Mizuno, D. Ojio, T. Tanaka, S. Minusa, H. Kuriyama, E. Yamano, H. Kuratsune, and Y. Watanabe, "Relationship between truck driver fatigue and rear-end collision risk," *PLoS ONE*, vol. 15, no. 9, Sep. 2020, Art. no. e0238738, doi: 10.1371/journal.pone.0238738.
- [4] G. Vilagut, C. G. Forero, G. Barbaglia, and J. Alonso, "Screening for depression in the general population with the center for epidemiologic studies depression (CES-D): A systematic review with meta-analysis," *PLoS ONE*, vol. 11, no. 5, May 2016, Art. no. e0155431, doi: 10.1371/journal.pone.0155431.
- [5] G. A. Hawker, S. Mian, T. Kendzerska, and M. French, "Measures of adult pain: Visual analog scale for pain (VAS Pain), numeric rating scale for pain (NRS Pain), McGill pain questionnaire (MPQ), short-form McGill pain questionnaire (SF-MPQ), chronic pain grade scale (CPGS), short form-36 bodily pain scale (SF-36 BPS), and measure of intermittent and constant osteoarthritis pain (ICOAP)," *Arthritis Care Res.*, vol. 63, no. S11, pp. S240–S252, Nov. 2011, doi: 10.1002/acr.20543.
- [6] G. N. Yannakakis, R. Cowie, and C. Busso, "The ordinal nature of emotions: An emerging approach," *IEEE Trans. Affect. Comput.*, vol. 12, no. 1, pp. 16–35, Jan. 2021, doi: 10.1109/TAFFC.2018.2879512.

- [7] G. Z. Heller, M. Manuguerra, and R. Chow, "How to analyze the visual analogue scale: Myths, truths and clinical relevance," *Scandin. J. Pain*, vol. 13, no. 1, pp. 67–75, Oct. 2016, doi: [10.1016/j.sjpain.2016.06.012](https://doi.org/10.1016/j.sjpain.2016.06.012).
- [8] M. D. Reed and W. Van Nostran, "Assessing pain intensity with the visual analog scale: A plea for uniformity," *J. Clin. Pharmacol.*, vol. 54, no. 3, pp. 241–244, Mar. 2014, doi: [10.1002/jcph.250](https://doi.org/10.1002/jcph.250).
- [9] K. Weigl and T. Forstner, "Design of paper-based visual analogue scale items," *Educ. Psychol. Meas.*, vol. 81, no. 3, pp. 595–611, Jun. 2021, doi: [10.1177/0013164420952118](https://doi.org/10.1177/0013164420952118).
- [10] Y.-T. Sung and J.-S. Wu, "The visual analogue scale for rating, ranking and paired-comparison (VAS-RRP): A new technique for psychological measurement," *Behav. Res. Methods*, vol. 50, no. 4, pp. 1694–1715, Aug. 2018, doi: [10.3758/s13428-018-1041-8](https://doi.org/10.3758/s13428-018-1041-8).
- [11] D. Kos, J. Raeymaekers, A. Van Remoortel, M. D'hooghe, G. Nagels, M. D'Haeseleer, E. Peeters, T. Dams, and T. Peeters, "Electronic visual analogue scales for pain, fatigue, anxiety and quality of life in people with multiple sclerosis using smartphone and tablet: A reliability and feasibility study," *Clin. Rehabil.*, vol. 31, no. 9, pp. 1215–1225, Sep. 2017, doi: [10.1177/0269215517692641](https://doi.org/10.1177/0269215517692641).
- [12] C. Escalona-Marfil, A. Coda, J. Ruiz-Moreno, L. M. Riu-Gispert, and X. Gironès, "Validation of an electronic visual analog scale mHealth tool for acute pain assessment: Prospective cross-sectional study," *J. Med. Internet Res.*, vol. 22, no. 2, Feb. 2020, Art. no. e13468, doi: [10.2196/13468](https://doi.org/10.2196/13468).
- [13] H. Baumgartner and J.-B.-E. M. Steenkamp, "Response styles in marketing research: A cross-national investigation," *J. Marketing Res.*, vol. 38, no. 2, pp. 143–156, May 2001, doi: [10.1509/jmkr.38.2.143.18840](https://doi.org/10.1509/jmkr.38.2.143.18840).
- [14] A.-W. Harzing, "Response styles in cross-national survey research: A 26-country study," *Int. J. Cross Cultural Manage.*, vol. 6, no. 2, pp. 243–266, Aug. 2006, doi: [10.1177/1470595806066332](https://doi.org/10.1177/1470595806066332).
- [15] K. Tasaki and J. Shin, "Japanese response bias: Cross-level and cross-national comparisons on response styles," *Japanese J. Psychol.*, vol. 88, no. 1, pp. 32–42, 2017, doi: [10.4992/jjpsy.88.15065](https://doi.org/10.4992/jjpsy.88.15065).
- [16] J. Haslbeck, O. Ryan, and F. Dablander, "Multimodality and skewness in emotion time series," *Emotion*, vol. 23, no. 8, pp. 2117–2141, Dec. 2023, doi: [10.1037/emo0001218](https://doi.org/10.1037/emo0001218).
- [17] J. S. Lee, E. Hobden, I. G. Stiell, and G. A. Wells, "Clinically important change in the visual analog scale after adequate pain control," *Academic Emergency Med.*, vol. 10, no. 10, pp. 1128–1130, Oct. 2003, doi: [10.1197/s1069-6563\(03\)00372-5](https://doi.org/10.1197/s1069-6563(03)00372-5).
- [18] Å. Nordin, C. Taft, Å. Lundgren-Nilsson, and A. Dencker, "Minimal important differences for fatigue patient reported outcome measures—A systematic review," *BMC Med. Res. Methodol.*, vol. 16, no. 1, p. 62, May 2016, doi: [10.1186/s12874-016-0167-6](https://doi.org/10.1186/s12874-016-0167-6).
- [19] S. P. J. Verkleij, T. Hoekstra, R. M. Rozendaal, J. H. Waarsing, B. W. Koes, P. A. J. Luijsterburg, and S. M. A. Bierma-Zeinstra, "Defining discriminative pain trajectories in hip osteoarthritis over a 2-year time period," *Ann. Rheumatic Diseases*, vol. 71, no. 9, pp. 1517–1523, Sep. 2012, doi: [10.1136/annrheumdis-2011-200687](https://doi.org/10.1136/annrheumdis-2011-200687).
- [20] K. L. Druce, G. T. Jones, G. J. Macfarlane, S. M. M. Verstappen, and N. Basu, "The longitudinal course of fatigue in rheumatoid arthritis: Results from the Norfolk arthritis register," *J. Rheumatol.*, vol. 42, no. 11, pp. 2059–2065, Nov. 2015, doi: [10.3899/jrheum.141498](https://doi.org/10.3899/jrheum.141498).
- [21] M. Takagishi, M. van de Velden, and H. Yadohisa, "Clustering preference data in the presence of response-style bias," *Brit. J. Math. Stat. Psychol.*, vol. 72, no. 3, pp. 401–425, Nov. 2019, doi: [10.1111/bmsp.12170](https://doi.org/10.1111/bmsp.12170).
- [22] U. Böckenholt and T. Meiser, "Response style analysis with threshold and multi-process IRT models: A review and tutorial," *Brit. J. Math. Stat. Psychol.*, vol. 70, no. 1, pp. 159–181, Feb. 2017, doi: [10.1111/bmsp.12086](https://doi.org/10.1111/bmsp.12086).
- [23] G. King, C. J. L. Murray, J. A. Salomon, and A. Tandon, "Enhancing the validity and cross-cultural comparability of measurement in survey research," *Amer. Political Sci. Rev.*, vol. 98, no. 1, pp. 191–207, Feb. 2004, doi: [10.1017/s000305540400108x](https://doi.org/10.1017/s000305540400108x).
- [24] D. Hojo and K. Okada, "Classification of response styles using anchoring vignettes," *Jpn. J. Behav.*, vol. 45, no. 1, pp. 13–25, 2018, doi: [10.2333/jbhmk.45.13](https://doi.org/10.2333/jbhmk.45.13).
- [25] G. Tutz, G. Schauberg, and M. Berger, "Response styles in the partial credit model," *Appl. Psychol. Meas.*, vol. 42, no. 6, pp. 407–427, Sep. 2018, doi: [10.1177/0146621617748322](https://doi.org/10.1177/0146621617748322).
- [26] K. G. Jonas and K. E. Markon, "Modeling response style using vignettes and person-specific item response theory," *Appl. Psychol. Meas.*, vol. 43, no. 1, pp. 3–17, Jan. 2019, doi: [10.1177/0146621618798663](https://doi.org/10.1177/0146621618798663).
- [27] S. Kumano and K. Nomura, "Multitask item response models for response bias removal from affective ratings," in *Proc. 8th Int. Conf. Affect. Comput. Intell. Interact. (ACII)*, Sep. 2019, pp. 1–7, doi: [10.1109/ACII.2019.8925539](https://doi.org/10.1109/ACII.2019.8925539).
- [28] C. W. Liu, "Multidimensional item response theory models for testlet-based doubly bounded data," *Behav. Res. Methods*, Nov. 2023, doi: [10.3758/s13428-023-02272-5](https://doi.org/10.3758/s13428-023-02272-5).
- [29] Z. Huang and F. Kämpfen, "Assessing (and addressing) reporting heterogeneity in visual analogue scales (VAS) with an application to gender difference in quality of life," Univ. Pennsylvania, Pennsylvania, PA, USA, Population Center Working Paper 7–30, 2019. [Online]. Available: [https://repository.upenn.edu/psc\\_publications/26](https://repository.upenn.edu/psc_publications/26)
- [30] Y. Noel and B. Dauvier, "A beta item response model for continuous bounded responses," *Appl. Psychol. Meas.*, vol. 31, no. 1, pp. 47–73, Jan. 2007, doi: [10.1177/0146621605287691](https://doi.org/10.1177/0146621605287691).
- [31] T. Shibuya, "An extension of the beta response model for continuous responses and parameter estimation via an EM algorithm," *Jpn. J. Educ. Psychol.*, vol. 68, no. 4, pp. 373–387, 2020.
- [32] M. Kloft, R. Hartmann, A. Voss, and D. W. Heck, "The Dirichlet dual response model: An item response model for continuous bounded interval responses," *Psychometrika*, vol. 88, no. 3, pp. 888–916, Sep. 2023, doi: [10.1007/s11336-023-09924-7](https://doi.org/10.1007/s11336-023-09924-7).
- [33] S. Minusa, T. Matsumura, K. Esaki, Y. Shao, C. Yoshimura, and H. Mizuno, "Characterization of response style in self-reports data measured in visual analogue scale," in *Proc. 45th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Sydney, NSW, Australia, Jul. 2023.
- [34] M. Smithson and J. Verkuilen, "A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables," *Psychol. Methods*, vol. 11, no. 1, pp. 54–71, Mar. 2006, doi: [10.1037/1082-989x.11.1.54](https://doi.org/10.1037/1082-989x.11.1.54).
- [35] M. Manuguerra, G. Z. Heller, and J. Ma, "Continuous ordinal regression for analysis of visual analogue scales: The R package ordinalCont," *J. Stat. Softw.*, vol. 96, no. 8, pp. 1–25, 2020, doi: [10.18637/jss.v096.i08](https://doi.org/10.18637/jss.v096.i08).
- [36] S. Ferrari and F. Cribari-Neto, "Beta regression for modelling rates and proportions," *J. Appl. Statist.*, vol. 31, no. 7, pp. 799–815, Aug. 2004, doi: [10.1080/0266476042000214501](https://doi.org/10.1080/0266476042000214501).
- [37] M. Vuorre. (2019). *How to Analyze Visual Analog (Slider) Scale Data?*. [Online]. Available: <https://vuorre.netlify.app/post/2019/02/18/analyze-analog-scale-ratings-with-zero-one-inflated-beta-models/>
- [38] J. Verkuilen and M. Smithson, "Mixed and mixture regression models for continuous bounded responses using the beta distribution," *J. Educ. Behav. Statist.*, vol. 37, no. 1, pp. 82–113, Feb. 2012, doi: [10.3102/1076998610396895](https://doi.org/10.3102/1076998610396895).
- [39] R. Kubinec, "Ordered beta regression: A parsimonious, well-fitting model for continuous data with lower and upper bounds," *Political Anal.*, vol. 31, no. 4, pp. 519–536, Oct. 2023, doi: [10.1017/pan.2022.20](https://doi.org/10.1017/pan.2022.20).
- [40] B. C. Wallace, K. Small, C. E. Brodley, and T. A. Trikalinos, "Class imbalance, redux," in *Proc. IEEE 11th Int. Conf. Data Mining*, Dec. 2011, pp. 754–763, doi: [10.1109/ICDM.2011.33](https://doi.org/10.1109/ICDM.2011.33).
- [41] H. Kaur, H. S. Pannu, and A. K. Malhi, "A systematic review on imbalanced data challenges in machine learning: Applications and solutions," *ACM Comput. Surv.*, vol. 52, no. 4, pp. 1–36, Jul. 2020, doi: [10.1145/3343440](https://doi.org/10.1145/3343440).
- [42] M. G. Forero, C. Arias-Rubio, and B. T. González, "Analytical comparison of histogram distance measures," in *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications* (Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)), vol. 11401. Cham, Switzerland: Springer, 2019, pp. 81–90, doi: [10.1007/978-3-030-13469-3\\_10](https://doi.org/10.1007/978-3-030-13469-3_10).
- [43] S. Minusa, T. Matsumura, K. Esaki, Y. Shao, C. Yoshimura, and H. Mizuno, 2024, "DailySense: A daily self-reports and physiological signals sensing dataset for subjective health research in the wild," *Zenodo*, doi: [10.5281/zenodo.10816004](https://doi.org/10.5281/zenodo.10816004).
- [44] S. Minusa, C. Yoshimura, and H. Mizuno, "Emodiversity evaluation of remote workers through health monitoring based on intra-day emotion sampling," *Frontiers Public Health*, vol. 11, pp. 1–12, Aug. 2023, doi: [10.3389/fpubh.2023.1196539](https://doi.org/10.3389/fpubh.2023.1196539).
- [45] Y. Shimonaka, K. Nakazato, Y. Gondo, and M. Takayama, "Construction and factorial validity of the Japanese NEO-PI-R," *Jpn. J. Personality*, vol. 6, no. 2, pp. 138–147, 1998, doi: [10.2132/jjpspp.6.2\\_138](https://doi.org/10.2132/jjpspp.6.2_138).

- [46] A. Betella and P. F. M. J. Verschure, "The affective slider: A digital self-assessment scale for the measurement of human emotions," *PLoS ONE*, vol. 11, no. 2, Feb. 2016, Art. no. e0148037, doi: [10.1371/journal.pone.0148037](https://doi.org/10.1371/journal.pone.0148037).
- [47] M. M. Bradley and P. J. Lang, "Measuring emotion: The self-assessment manikin and the semantic differential," *J. Behav. Therapy Experim. Psychiatry*, vol. 25, no. 1, pp. 49–59, Mar. 1994, doi: [10.1016/0005-7916\(94\)90063-9](https://doi.org/10.1016/0005-7916(94)90063-9).
- [48] N. Jaques, S. Taylor, A. Sano, and R. Picard, "Predicting tomorrow's mood, health, and stress level using personalized multitask learning and domain adaptation," in *Proc. IJCAI Workshop Artif. Intell. Affective Comput.*, 2017, pp. 17–33.
- [49] S. Taylor, N. Jaques, E. Nosakhare, A. Sano, and R. Picard, "Personalized multitask learning for predicting tomorrow's mood, stress, and health," *IEEE Trans. Affect. Comput.*, vol. 11, no. 2, pp. 200–213, Apr. 2020, doi: [10.1109/TAFFC.2017.2784832](https://doi.org/10.1109/TAFFC.2017.2784832).
- [50] M. Tsujikawa, T. Kitade, K. Suzuki, and K. Shibuya, "Accurate chronic stress estimation with personalized models based on correlation maximization," in *Proc. 44th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2022, pp. 1761–1765, doi: [10.1109/EMBC48229.2022.9871917](https://doi.org/10.1109/EMBC48229.2022.9871917).
- [51] C. Schröder and S. Rahmann, "A hybrid parameter estimation algorithm for beta mixtures and applications to methylation state classification," *Algorithms Mol. Biol.*, vol. 12, no. 1, pp. 1–12, Dec. 2017, doi: [10.1186/s13015-017-0112-1](https://doi.org/10.1186/s13015-017-0112-1).
- [52] C. E. Galarza, P. Zhang, and V. H. Lachos, "Logistic quantile regression for bounded outcomes using a family of heavy-tailed distributions," *Sankhya B*, vol. 83, no. S2, pp. 325–349, Nov. 2021, doi: [10.1007/s13571-020-00231-0](https://doi.org/10.1007/s13571-020-00231-0).
- [53] S. Dolnicar and B. Grün, "Response style contamination of student evaluation data," *J. Marketing Educ.*, vol. 31, no. 2, pp. 160–172, Aug. 2009, doi: [10.1177/0273475309335267](https://doi.org/10.1177/0273475309335267).



**KANAKO ESAKI** (Member, IEEE) was born in Japan. She received the B.S., M.S., and Ph.D. degrees in mechanical engineering from Waseda University, Tokyo, Japan, in 2010, 2012, and 2024, respectively.

In 2012, she joined the Central Research Laboratory, Hitachi, Ltd., Tokyo. She has been mainly involved in research into robotics and automation systems. Her current research interest includes bio-inspired information processing systems.



**YANG SHAO** was born in China. He received the B.S. degree in mathematics and physics from Tsinghua University, Beijing, in 2007, and the Ph.D. degree in numerical analysis from The University of Tokyo, Japan, in 2013.

Since 2013, he has been a Researcher with the Research & Development Group, Hitachi, Ltd. His current research interests include large language models, reinforcement learning, and artificial intelligence.

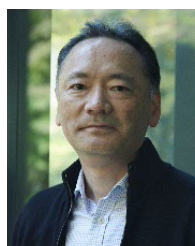
Dr. Shao is a member of the Japanese Society for Artificial Intelligence



**CHIHIRO YOSHIMURA** was born in Japan. He received the B.A. degree in environmental information and the M.A. degree in media and governance from Keio University, Kanagawa, Japan, in 2005 and 2007, respectively.

In 2007, he joined the Central Research Laboratory, Hitachi, Ltd., Tokyo, Japan, where he has been engaged in the research and development of supercomputers and mission critical servers, including vector processors, interconnect switches, and network controllers. Since 2012, he has been engaged in the research of new-paradigm computing including quantum computing and domain specific computing. He is currently a Manager and a Chief Researcher with the Center for Exploratory Research, Research & Development Group, Hitachi, Ltd.

Mr. Yoshimura is a member of the Information Processing Society of Japan.



**HIROYUKI MIZUNO** (Fellow, IEEE) received the M.S. and Dr.Eng. degrees in electronic engineering from Osaka University, Osaka, Japan, in 1993 and 2001, respectively.

In 1993, he joined Hitachi, Ltd., Tokyo, Japan, where he has been engaged in the research and development of high-speed and low-power semiconductor circuits for SRAMs, microprocessors (SuperH), and system-on-a-chip (SH-Mobile).

He is a Pioneer in quantum-inspired computers (CMOS Annealing). From 2002 to 2003, he was a Visiting Scholar with the Department of Computer Science, Stanford University, Stanford, CA, USA. He is currently a Distinguished Researcher with the Center for Exploratory Research (CER) and the Laboratory Manager of the Hitachi-Kyoto University Laboratory, Hitachi, Ltd. His current research interests include quantum computing, artificial intelligence systems using cognitive neuroscientific methods, and cyber-human social system (CHSS). He is a member of ACM. He has served in various roles for international conferences, including the Technical Program Committee of the IEEE International Solid-State Circuits Conference (ISSCC), from 2004 to 2007.



**SHUNSUKE MINUSA** (Member, IEEE) was born in Japan. He received the B.E. and master's degrees in information science and technology from Hokkaido University, Sapporo, Japan, in 2016 and 2018, respectively.

Since 2018, he has been a Researcher with the Research & Development Group, Hitachi, Ltd. His current research interests include human sensing, biostatistics, and psychophysiological data analysis for healthcare applications.

Mr. Minusa is a member of the Behaviormetric Society of Japan and Japan Ergonomics Society.



**TADAYUKI MATSUMURA** was born in Japan. He received the B.S. and M.S. degrees in computer science from Kyushu University, in 2007 and 2009, respectively. Since 2009, he has been a Researcher with the Research & Development Group, Hitachi, Ltd. His current research interests include artificial multi-agent systems, AI ethics, and emotional intelligence for social AI/robot applications.