ICIC Express Letters
Part B: Applications
Volume **15**, Number **1**, January **2024**

ICIC International ©2024 ISSN 2185-2766
pp. **71–81**

PROPOSAL FOR QUANTIFICATION AND ANALYSIS METHOD OF NUANCES IN CONVERSATION RESPONSES USING VISUAL ANALOG SCALE

Naruki Shirahama[1,*], Naofumi Nakaya[2], Kazuhiro Koshi[3]
Keiji Matsumoto[1], Kenji Moriya[4] and Satoshi Watanabe[5]

[1]Department of Creative Engineering
National Institute of Technology (KOSEN), Kitakyushu College
5-20-1, Shii, Kokuraminami-ku, Kitakyushu-shi, Fukuoka 802-0985, Japan
kmatsu@kct.ac.jp
*Corresponding author: naruki@kct.ac.jp

[2]Faculty of Medical Science
Juntendo University
6-8-1, Hinode, Urayasu-shi, Chiba 279-0013, Japan
n.nakaya.ac@juntendo.ac.jp

[3]Department of Human-Oriented Information Systems Engineering
National Institute of Technology (KOSEN), Kumamoto College
2659-2, Suya, Koshi, Kumamoto-shi 861-1102, Japan
kkoshi@kumamoto-nct.ac.jp

[4]Department of Production Systems Engineering
National Institute of Technology (KOSEN), Hakodate College
14-1, Tokura, Hakodate-shi, Hokkaido 042-8501, Japan
moriya@hakodate-ct.ac.jp

[5]College of Bioresource Sciences
Nihon University
1866, Kameino, Fujisawa-shi, Kanagawa 252-0880, Japan
watanabe.satoshi@nihon-u.ac.jp

Received May 2023; accepted July 2023

Abstract. *This paper proposes a method for quantifying and analyzing nuances in conversation responses using a Visual Analog Scale (VAS). VAS is an instrument that has proven effective in various fields for measuring subjective characteristics or attitudes. The authors conducted research on VAS, highlighting its usefulness in capturing human senses and identifying trends in small data samples. This paper presents a proposal for using the VAS to quantify the nuances of responses in conversations. We conducted an experiment using VAS to measure subtle differences in the sentiment information contained in the replies. The scenario was set as two friends conversing using a smartphone chat application. By combining hierarchical clustering and violin plots with the subjective data obtained by the VAS, a typical group of responses can be found.*
Keywords: Visual Analog Scale, Bee swarm plot, Violin plot, Kernel density plot, Hierarchical clustering, Quantification of conversation responses

1. **Introduction.** The intricacies of human conversation and the sentiments embedded within them play an essential role in understanding and interpreting the essence of interactions. Recent studies have shed light on two pertinent areas in the field of information technology: quantifying and analyzing nuances in conversation, and sentiment analysis. This paper elucidates the significant advancements and findings in these two domains by referencing contemporary research works.

DOI: 10.24507/icicelb.15.01.71

1.1. **Quantifying and analyzing nuances in conversation.** Zhou et al. highlighted the disparity in dialogue summarization models, where the preservation of emotion is often overlooked despite its essentiality in capturing the holistic essence of conversations [1]. Recognizing the significance of conversations in our daily lives, Reece et al. introduced a comprehensive multimodal corpus that offers a vast dataset of conversations in spoken English [2]. This corpus aims to pave the way for interdisciplinary research to unravel the complexities of naturalistic conversations. In a related vein, Kim et al. presented evidence underscoring the potent memory-enhancing effect of emotions, revealing that verbatim retention of written material is significantly heightened when laced with emotional nuances [3]. Furthermore, Toubia et al. ventured into the realm of narratives, harnessing natural language processing techniques to uncover the correlation between the shape of stories and their success, and elucidating the intricacies of why particular narratives resonate more with the audience [4].

1.2. **Sentiment analysis.** Sentiment analysis plays a pivotal role in deciphering emotions and sentiments embedded within textual data. Zhang et al. offered an in-depth survey on Aspect-Based Sentiment Analysis (ABSA), elucidating various ABSA tasks that aim to analyze and comprehend sentiments at the aspect level [5]. Their work provides a comprehensive overview of the ABSA landscape, emphasizing the relevance of pre-trained language models and discussing cross-domain and cross-lingual scenarios. Ahmet and Abdullah navigated the burgeoning domain of deep learning-based sentiment analysis, highlighting its superiority in the realm of machine learning [6]. Their comprehensive survey touched upon various deep learning architectures utilized for sentiment analysis, including Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM), and attention mechanisms. Their analysis underscores the potential of CNNs in coarse-grain sentiment analysis, whereas LSTM and attention mechanisms are pivotal in the fine-grain domain, particularly for aspect extraction.

Although these studies provide valuable insights, gaps still exist in the integration and application of these findings to quantify and analyze the nuances of conversational responses. In this study, we propose a method to use VASs to capture the complexity and nuances of conversational responses using a more comprehensive and robust approach. While much research has been conducted on quantifying the nuances of conversation and sentiment analysis, the main objective is generally to analyze the feelings that everyone can agree on. Such research is meaningful, but because feelings and sensations are subjective data, a system that can quantify individual and group differences is required. In addition, a system that enables quantification is required when the sample size is small for conducting experiments. We propose VAS as a tool for measuring individual subjectivity and a combination of visualization and clustering as an analysis method.

The remainder of this paper is structured as follows. The outline and analysis method of the VAS as a subjective evaluation method will be explained. To address the research question of quantifying conversational nuances, we conducted a subjective evaluation experiment on six predefined patterns of responses to a conversation. After visualizing the response data for the results of this experiment, clustering was performed to show that the subjects could be classified into four clusters. Finally, we discuss the experimental results and the potential of the proposed method.

2. **Subjective Evaluation Analysis Using VAS.** Understanding and measuring human subjectivity, especially in conversation responses, poses significant challenges in both scientific and technological fields. The complexity of this task necessitates the development of innovative and robust measurement tools. We performed seminal work on the VAS, highlighting its usefulness in capturing the human sense of brightness and grayscale color [7] and identifying trends in small data samples, notably in gauging user emotions

during interactions with a toy robot [8]. These findings suggest that even with small sample sizes, using the VAS can lead to more accurate sensory evaluation data and emphasize the importance of combining univariate scatter and box plots with bee swarm plots for visual trend analysis. These studies were inspired by the data visualization techniques proposed by Weissgerber et al. [9].

The VAS is a commonly used and straightforward instrument for measuring subjective characteristics or attitudes that cannot be directly measured. Respondents indicated their level of agreement with a statement by marking their position along a continuous line between the two endpoints. Typically, a VAS consists of a horizontal line of a fixed length, usually 10 cm, with verbal descriptors at each end representing the extreme limits of the sensation, feeling, or response to be measured. For example, in pain assessment, one end of the scale might be labeled "No pain" and the other "Worst imaginable pain". Respondents mark on the line the point they feel that represents their perception of their current state. The VAS measurement allows researchers to understand not only whether individuals rate something differently but also the relative size of the differences. Owing to its simplicity and adaptability, the VAS has been widely used in diverse areas, such as pain measurement, mood, sleepiness, health status, and medicine [10, 11, 12, 13]. Despite its numerical appearance, the VAS is technically ordinal rather than an interval or ratio scale; therefore, statistical methods that assume interval or ratio scales should be used cautiously. Consequently, VAS has some complexities in its interpretation. The VAS provides a valuable tool for researchers across many fields, offering a simple and flexible way to measure subjective experiences. Although VAS has some challenges and limitations, careful study design and data analysis can help address these issues.

2.1. **Visualization for VAS data analysis.** In this section, we discuss methods for effectively visualizing the data obtained from the VAS. This section describes the characteristics of the VAS in contrast to the Likert Scale (LS). Data visualization is a way to graphically represent information and data, and data visualization tools provide an easy way to observe and understand trends, outliers, and patterns in the data. The principal distinction between data measured with the LS and those measured with the VAS is that the VAS offers a more intuitively comprehensible set of data. While LS is usually quantified using a 5-point or 7-point scale, VAS allows for the acquisition of data as real numbers in the interval $[0, 1]$. Hence, we believe that it is crucial to plot all measurement data to grasp the trends. In this experiment, we represented the data collected using bee swarm plots. Our objective was to numerically express responses to conversations that are expected to vary because of subjectivity. We overlaid the univariate scatter plot with a violin plot, which visualizes the data probability density to provide a visual representation of the probability density of the data. To date, we have been visualized by combining univariate scatter plots with box plots [7, 8]. A violin plot is a data visualization technique used to represent numerical data. It combines the properties of the box and kernel density plots to provide a richer description of the data. Here is more details about the elements of a violin plot and what they represent.

1) **Kernel Density Plot:** The outer shape or 'violin' part of the plot represents a kernel density estimation of the data. This is essentially a smooth histogram that provides an idea of the data distribution. The width of the 'violin' at any given point corresponds to the density or frequency of the data. The wider parts of the 'violin' represent areas where data are more densely packed (i.e., more common), whereas the narrower parts represent less common values.
2) **Box Plot:** Inside the 'violin', you can often see a simplified box plot. The box plot typically includes a line for the median of the data, a box (or sometimes just lines) extending from the 25th percentile to the 75th percentile (the interquartile range, or

IQR), and potentially 'whiskers' extending out to the minimum and maximum data points that are not considered outliers. Violin plots also indicate individual outliers.

3) **Symmetry:** Violin plots are typically symmetrical, with the shape mirrored on either side of the central axis. This is not a reflection of the data, but merely a visual choice to make the plot more interpretable.

Violin plots are particularly useful for comparing the distribution of variables across categories. They can provide more detailed insights than a standard box plot because they show the full distribution of the data, not just a summary of the central tendency and spread. However, they can be a bit harder to read for people who are not used to seeing them, so they are often used in more exploratory analyses rather than for presentations to non-technical audiences.

3. **Experimental Methods.** This section describes the experimental methods employed. The design of this experiment focused on the ability to measure subtle differences in sentiment information carried by words using a VAS. The scenario was set as follows: two friends conversed through a smartphone chat application. One friend invited another person to dine. The invitee responds to whether they will go for a meal. For clarity, refer to the invitations as Friends A and B. It is anticipated that the degree to which Friend B truly wants to go to a meal may vary subtly. The aim of this experiment was to investigate whether these differences could be measured using the VAS. Figure 1 depicts this hypothetical chat scenario, reminiscent of a screenshot from a chat application. The grey character icon on the left represents Friend A. We established a set of potential replies to be placed in the final blank-speech bubble at the lower right, as shown in Table 1. The first option, "I will go" was set as a simple choice indicating the decision to go. This response did not invoke any particular context, and we set this choice as the standard. The second choice implies that there might be the possibility of not going if circumstances are not allowed. While the likelihood of going may vary depending on these circumstances, this is a commonly used expression, and we consider the degree of certainty in "going" in response to this choice to be a settled sentiment. The third choice suggests that they will probably be able to go because there are no particular issues preventing them. We expected a high desire to respond to this question. The fourth choice implies that they will go as long as there are no issues with their schedules. If they really wanted to go, they checked or adjusted their schedules. Therefore, this choice may be considered to have a low likelihood of going. The fifth choice, "If I remember, I will show up", does not involve recording the plan, leaving it solely to memory. This response may have confused the

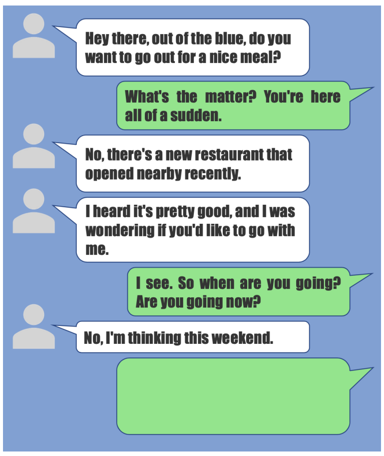

FIGURE 1. A representation styled after a chat application screenshot, illustrating an example conversation under the scenario set for the experiment

Table 1. Candidate replies for the last empty speech bubble in Figure 1

No.	Reply example
Reply 1	I will go.
Reply 2	If it works out, I will go.
Reply 3	I can probably make it.
Reply 4	If it fits in my schedule, I am there.
Reply 5	If I remember, I will show up.
Reply 6	If I manage to wake up, I will roll out.

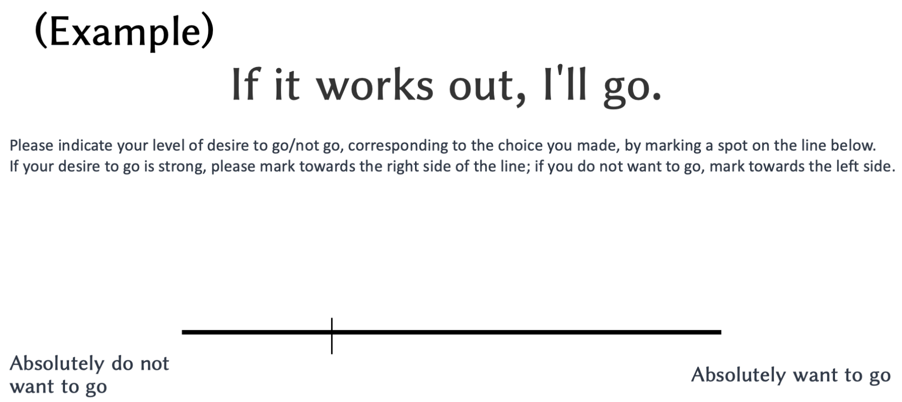

Figure 2. VAS sheet

inviting person, indicating a low likelihood of going. The final choice was the possibility of oversleeping. This option can also be considered to have a low likelihood of going.

Figure 2 presents a sample of the VAS evaluation sheets for the experimental participants. After explaining the scenario in which Friend A invites Friend B to a meal and presents the chat application screenshot in Figure 1, we provide a VAS evaluation sheet, as shown in Figure 2 for each of the six response examples. Six VAS evaluation sheets were presented to the participants in random order.

The experiment was conducted in a student class. The number of valid responses was set to 40. The students' ages ranged from 17 to 18 years, and they regularly used social networking services, making such a scenario commonplace.

We predicted the tendencies of the responses. Specifically, we anticipated that Reply 1 and Reply 3 would receive high evaluations, whereas Reply 4 and Reply 5 would receive lower scores. We expected that the evaluations of Reply 2 and Reply 6 would be further divided. Responses other than Reply 1 and Reply 3 imply going to a meal if certain criteria are met. It is anticipated that evaluations might diverge based on impressions of these conditions. Therefore, although we can expect to observe some general trends, it is also entirely possible for evaluations to vary based on individual interpretations. We hope that by using the VAS for subjective evaluation, we can detect these nuanced differences among individuals.

4. Experimental Results.

4.1. **Understanding trends from the distribution of the entire data.** Using the measurement data obtained from this experiment, we conducted several analyses to explore the response trends. First, we calculated the mean and standard deviation of the desire to go based on the responses of the 40 participants to each of the six questions. Figure 3 shows a graph of the mean response ± standard deviation for each reply. The highest mean value was 0.92, corresponding to Reply 1. This was followed by Reply 3,

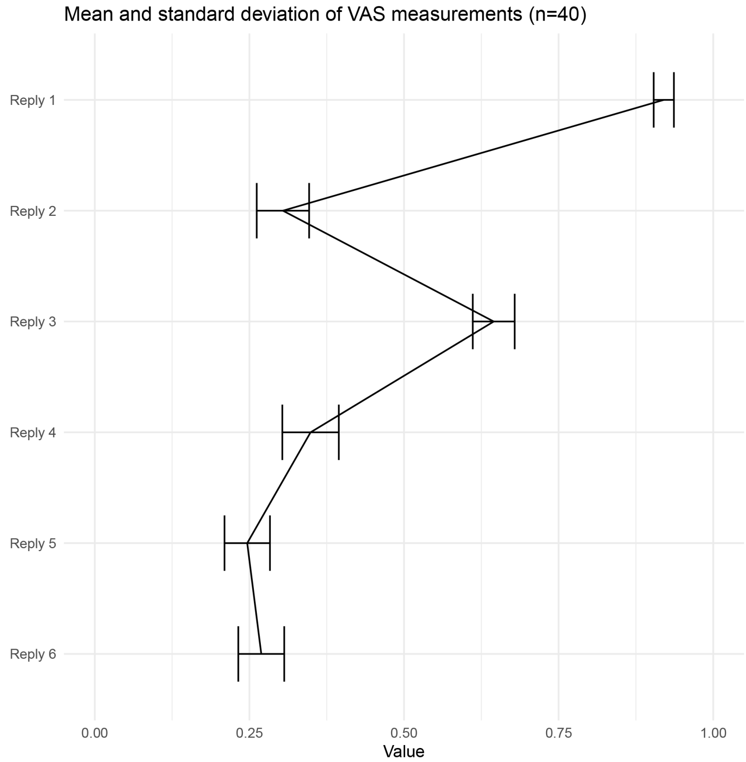

FIGURE 3. The mean ± standard deviation of subjective evaluation data by VAS for each reply

TABLE 2. The descriptive statistic values of subjective evaluation data by VAS for each reply

No.	max	min	mean	median	var	sd	se
Reply 1	1.000	0.541	0.920	0.965	0.011	0.103	0.016
Reply 2	0.953	0.000	0.304	0.241	0.072	0.268	0.042
Reply 3	0.994	0.000	0.645	0.700	0.046	0.214	0.034
Reply 4	0.994	0.000	0.349	0.275	0.083	0.289	0.046
Reply 5	0.859	0.000	0.246	0.174	0.054	0.232	0.037
Reply 6	0.982	0.000	0.269	0.235	0.055	0.235	0.037

with a mean value of 0.645. All other replies had values below 0.35, which is in line with our initial prediction that only Replies 1 and 3 would have a high likelihood of going, while the rest would most likely not go. Reply 1, which firmly states, “I will go”, understandably has a high value. For Reply 3, which includes the condition “If it fits in my schedule”, it seems that the intent to go was judged to be lower because of this caveat.

Replies 2 and 4 are relatively standard responses that reserve judgement, expressing that the respondents would go if they knew their schedule or other details. On the other hand, Replies 5 and 6 present conditions that could be considered somewhat impolite to the other party, such as “If I remember” and “If I am awake”. Considering these factors, the results seem reasonable, with Replies 2 and 4 having mean values in the 0.3 range and Replies 5 and 6 having mean values in the 0.2 range.

Table 2 presents the descriptive statistics for each response, including the maximum and minimum values, mean, median, variance, standard deviation, and standard error. Looking at the variance, Reply 1 had the smallest variance, whereas Replies 2 and 4 had higher variances. This suggests that the perception of wanting to go was relatively stable for Reply 1, whereas there was some variability for Replies 2 and 4.

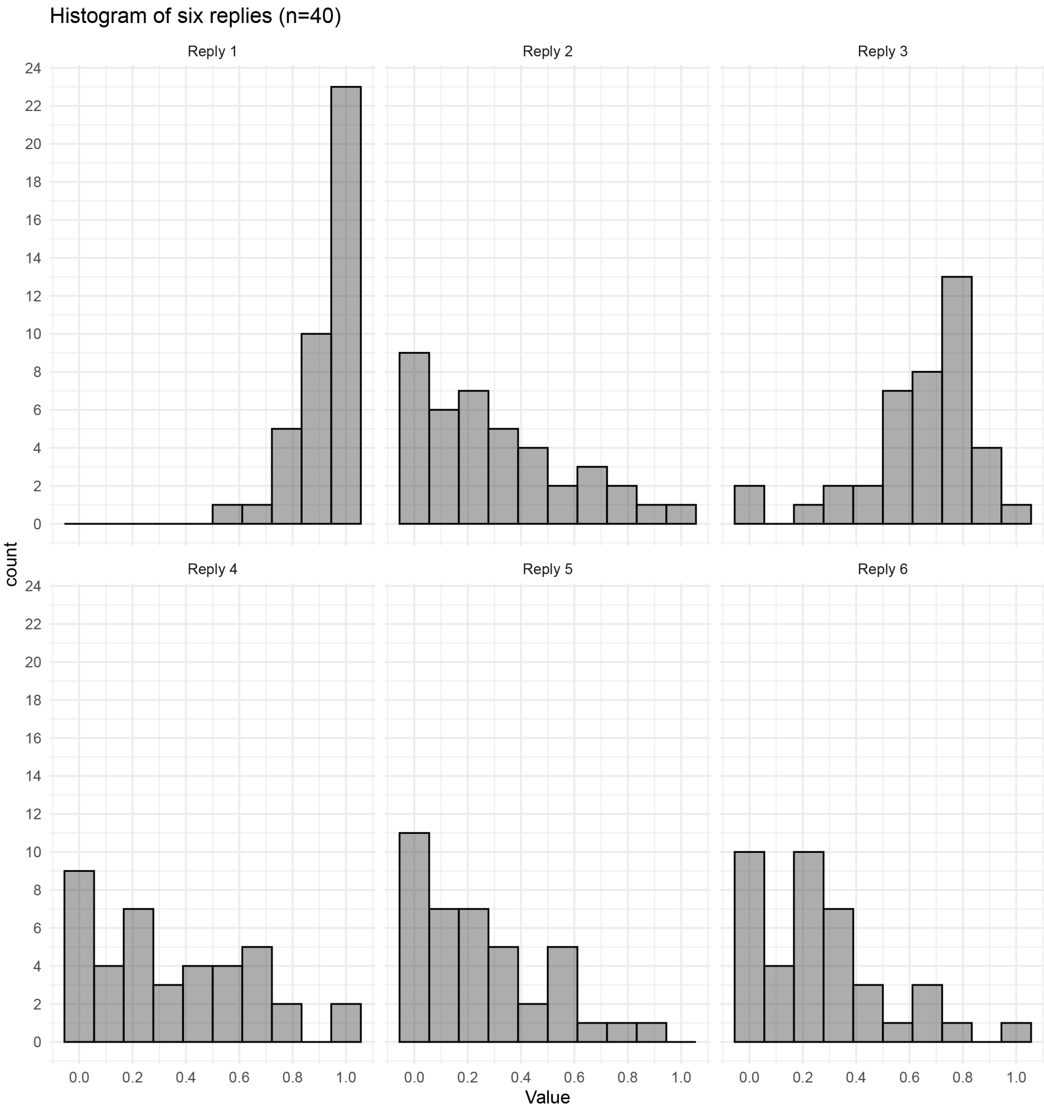

Figure 4. The histogram of subjective evaluation data by VAS for each reply

Next, we verified how the evaluations of the 40 participants were distributed for each reply using the histogram shown in Figure 4. Reply 1 clearly indicated the intent to go. This is followed by Reply 3, which shows the tendency to want to go. As mentioned above, the histogram confirms that Replies 2 and 4, as well as Replies 5 and 6, exhibit similar trends.

We analyzed subjective evaluation results using basic descriptive statistics, such as mean, variance, and histograms, to explore trends. Our findings are as follows:

- Reply 1 and Reply 3 are more likely to identify as "going";
- Replies 2, 4, 5, and 6 are more likely to identify as "not going".

We strive to avoid relying solely on condensed data. Our objective was to acquire a comprehensive understanding of the data by visualizing them in their entirety. As shown in Figure 5, the use of bee swarm, violin, and box plots effectively represents all data that were evaluated subjectively using the VAS.

By visualizing all the data in a bee swarm plot, we were able to identify patterns that could not be discerned by simply looking at the mean or variance. While the end result of the discussion is a binary choice between "Go" or "Not go", the bee swarm plot reveals that many experimental collaborators answered with a maximum value of 1 for "Go" in Reply 1. Conversely, the bee swarm plot also showed that many experimental

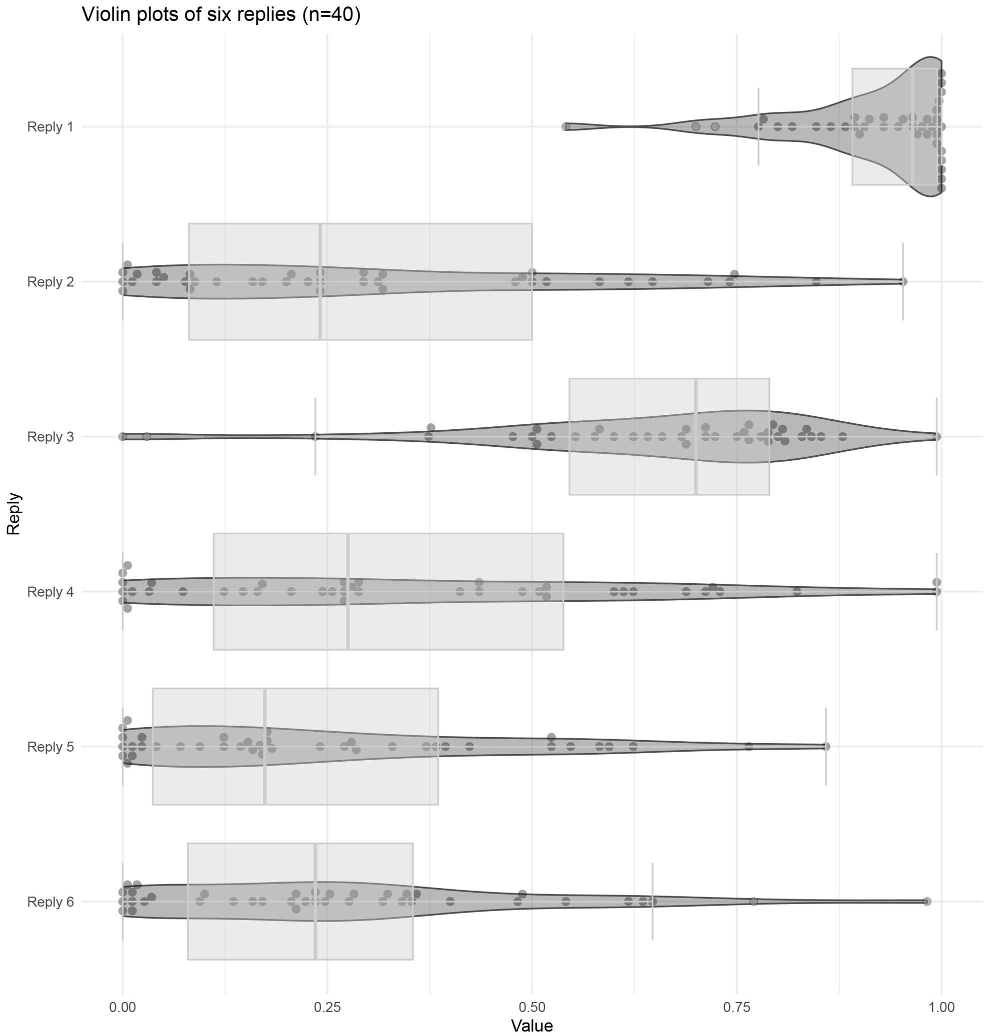

FIGURE 5. The plots of subjective evaluation data by VAS for each reply using bee swarm plots, violin plots, and box plots

collaborators answered with a minimum value of 0, indicating "Not go", in Replies 2, 4, 5, and 6.

Furthermore, it is worth noting that some responses fell outside the first and third quartiles of the box plots. Although proper understanding of box plots can prevent such misunderstandings, it can be argued that violin plots are more suitable for this purpose. Kernel density estimation in the violin plot extrapolates population data, offering a more nuanced representation. Although the box plot box was defined by the first, median, and third quartiles, it may be more appropriate to depict it as a probability density function, as demonstrated in the violin plot. Given the subjective nature of the conversation's response, it is crucial to consider the varying perspectives and opinions of the respondents towards the population.

4.2. **Visualization of groups with specific tendencies through cluster analysis.** According to the analysis presented, it was discovered that the subjective evaluations of conversation responses vary greatly depending on each individual's personal bias. We hypothesized that "specific groups with certain response tendencies can be classified" and aimed to confirm this hypothesis through cluster analysis. Owing to the unknown optimal number of divisions, we conducted hierarchical clustering. Figure 6 depicts a dendrogram plot of hierarchical clustering applied to subjective evaluation data evaluated using the

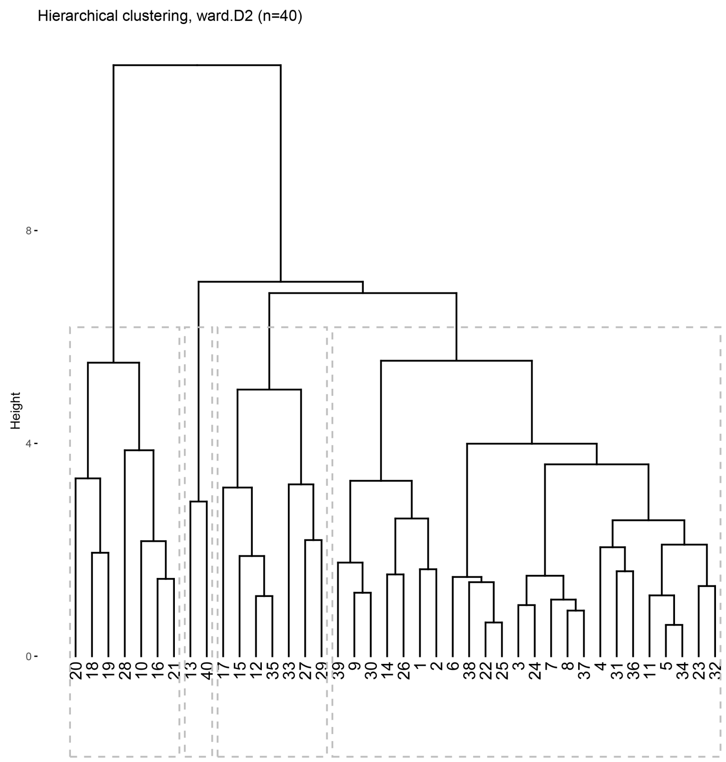

Figure 6. Dendrogram plot of hierarchical clustering applied to subjective evaluation data evaluated by VAS

Table 3. Classification results by experimental collaborator when the number of clusters is 4

No.	n	Experimental collaborator (Anonymous ID)
Cluster 1	24	39, 9, 30, 14, 26, 1, 2, 6, 38, 22, 25, 3, 24, 7, 8, 37, 4, 31, 36, 11, 5, 34, 23, 32
Cluster 2	7	20, 18, 19, 28, 10, 16, 21
Cluster 3	7	17, 15, 12, 35, 33, 27, 29
Cluster 4	2	13, 40

VAS. Up to this point, we have been conducting an analysis in order to understand common trends among experimental collaborators, assuming that such trends exist. However, we have observed evaluation data, such as responses in conversations, that possess certain characteristics, including subjectivity, ambiguity, polysemy, and context-sensitivity. As a result, we hypothesized that the perception of these responses can be grouped into several distinct categories. This hypothesis led us to perform clustering. Using hierarchical clustering, we constructed clusters in a step-by-step manner, represented as a tree diagram, without having to determine the number of clusters in advance. After careful consideration, we decided to adopt four clusters for interpretability and explanation. It is worth noting that it is possible to interpret the results with fewer or more clusters, but we chose the number that we felt provided the most appropriate interpretation given our understanding. The classification results are presented in Table 3.

To effectively analyze this cluster classification, we suggest using a combination of violin and bee swarm plots along with box plots. Figure 7 shows the plots for each of the four clusters. Cluster 1, plotted in the top left of Figure 7, represents the majority of 24 experimental participants. Reply 1 strongly indicated their intention to go, followed by Reply 3. Replies 2, 4, and 5 are negative and indicate that they will not go. The distinctive feature is Cluster 2, plotted at the top right of Figure 7, where seven participants belong. All replies, except for Reply 6, were interpreted as intending to go, indicating

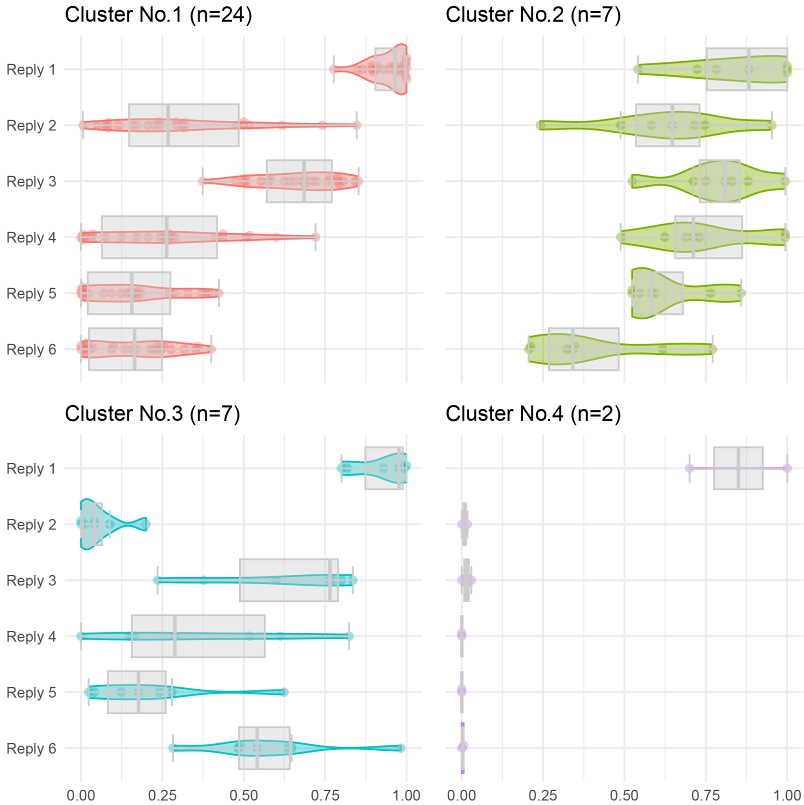

Figure 7. 4 cluster violin plot

that respondents in this cluster may have a positive disposition. Cluster 3, plotted in the bottom left of Figure 7, is characterized by seven people who indicated in Reply 2 that they will not go. Finally, the smallest group was Cluster 4, in which two people belonged. Apart from Reply 1, they interpret all other replies as equivalent to "not going".

5. **Conclusions.** Initially, we conducted an analysis to gain insight into the general trend by confirming descriptive statistics values, histograms, bee swarm plots, violin plots overlaid with each other, and a basic analysis. To quantify subjectivity, we utilized a tool known as the VAS, which is more effective than the LS. In addition, we analyzed the data based on the assumption that the respondents could be categorized into distinct groups and discovered intriguing findings. As a result, hierarchical clustering was performed, and it was possible to discover distinctive groups of respondents through classification into four clusters.

By focusing on clustering, we believe that the quantification of nuances in conversation has been achieved to some extent by classifying them into distinctive groups of respondents. As responses in conversation have subjective, ambiguous, polysemous, and context-dependent properties, exploratory data analysis-oriented analysis methods seem appropriate. We propose an experimental style that applies various analytical methods after effectively visualizing the VAS and its data to grasping the characteristics. We will continue to work on improving the proposed methodology.

REFERENCES

[1] Y. Zhou et al., Evaluating emotional nuances in dialogue summarization, *arXiv Preprint*, arXiv: 2307.12371, 2023.
[2] A. Reece et al., The CANDOR corpus: Insights from a large multimodal dataset of naturalistic conversation, *Science Advances*, vol.9, no.13, DOI: 10.1126/sciadv.adf3197, 2023.

[3] Y. Kim et al., Emotional nuance enhances verbatim retention of written materials, *Frontiers in Psychology*, vol.12, DOI: 10.3389/fpsyg.2021.519729, 2021.
[4] O. Toubia et al., How quantifying the shape of stories predicts their success, *Proc. of the National Academy of Sciences*, vol.118, no.26, DOI: 10.1073/pnas.2011695118, 2021.
[5] W. Zhang et al., A survey on aspect-based sentiment analysis: Tasks, methods, and challenges, *arXiv Preprint*, arXiv: 2203.01054, 2022.
[6] A. Ahmet and T. Abdullah, Recent trends and advances in deep learning-based sentiment analysis, in *Deep Learning-Based Approaches for Sentiment Analysis. Algorithms for Intelligent Systems*, B. Agarwal, R. Nayak, N. Mittal and S. Patnaik (eds.), Singapore, Springer, DOI: 10.1007/978-981-15-1216-2_2, 2020.
[7] N. Shirahama, K. Matsumoto, K. Moriya, K. Koshi, N. Nakaya and S. Watanabe, Sensory amount measurement experiment to validate the accuracy of the VAS measurement method, *Journal of the Institute of Industrial Applications Engineers*, vol.10, no.4, pp.90-96, DOI: 10.12792/JIIAE.10.90, 2022.
[8] N. Shirahama, S. Watanabe, K. Moriya, K. Koshi and K. Matsumoto, A new method of subjective evaluation using visual analog scale for small sample data analysis, *Journal of Information Processing*, vol.29, pp.424-433, DOI: 10.2197/ipsjjip.29.424, 2021.
[9] T. L. Weissgerber, N. M. Milic, S. J. Winham and V. D. Garovic, Beyond bar and line graphs: Time for a new data presentation paradigm, *PLOS Biology*, vol.13, no.4, DOI: 10.1371/journal.pbio.1002128, 2015.
[10] O. Karcioglu et al., A systematic review of the pain scales in adults: Which to use?, *The American Journal of Emergency Medicine*, vol.36, no.4, pp.707-714, DOI: 10.1016/j.ajem.2018.01.008, 2018.
[11] F. X. Lesage et al., Clinical stress assessment using a visual analogue scale, *Occupational Medicine (Oxford, England)*, vol.62, no.8, pp.600-605, DOI: 10.1093/occmed/kqs140, 2012.
[12] Y. D. Alqurashi et al., A visual analog scale for the assessment of mild sleepiness in patients with obstructive sleep apnea and healthy participants, *Annals of Thoracic Medicine*, vol.16, no.2, pp.141-147, DOI: 10.4103/atm.ATM_437_20, 2021.
[13] T. Kimura et al., How valid is the self-administered visual analogue scale for assessing the health status of elderly people?, *Hiroshima Journal of Medical Sciences*, vol.57, nos.3-4, pp.85-92, 2008.