“Ah-aloud”: Method for Evaluating Cognitive Processes Occurring During Tasks from Vocal Information

Takuya Kawashima
Frontier Media Science Program,
Meiji University
Nakano, Tokyo, Japan
cs212007@meiji.ac.jp

Keita Watanabe
Department of Frontier Media Science,
Meiji University
Nakano, Tokyo, Japan
keita_w@meiji.ac.jp

2022 8th International HCI and UX Conference in Indonesia (CHIuXiD) | 978-1-6654-7664-5/22/$31.00 ©2022 IEEE | DOI: 10.1109/CHIuXiD57244.2022.10009797

***Abstract*—The think-aloud method, in which users verbalize thoughts generated during a task, provides insight into users' task-directed cognitive processes. However, verbalization during a task is burdensome and reactive, and the words used and amount of speech vary between users. To overcome this challenge, we propose the ah-aloud method in this study, wherein cognitive processes are expressed solely using “ah” and evaluated through vocal volume and intonation. Experiments compared to think-aloud using video game tasks with Japanese-speaking participants showed negligible difference in the reduction of verbalization burden and reactivity. It seems that people are very different regarding whether think-aloud or ah-aloud is easier. However, we found that ah-aloud can express various feelings, so we will establish ah-aloud position by comparing it with experimental methods to evaluate feelings, such as physiological signal measurement. We also discuss the recommended methods for obtaining think-aloud data. Additionally, we surveyed the literature pertaining to English interjections and discussed the possibility of extending this method to other languages.**

Keywords—think-aloud, cognitive processes, experimental method, evaluation

I. INTRODUCTION

The think-aloud method [1] is an experimental method that allows users to verbalize thoughts generated during a task. Concurrent think-aloud is the only method that can completely reflect thoughts and mental states generated during task execution because retrospective verbalization may cause falsification and loss of recollection [2, 3]. The think-aloud method has been applied in diverse research fields, including usability testing [4], body and strategy perception in sports [5], and translation studies [3].

However, this method burdens verbalization and reactivity and verbalizing while performing a task is challenging. As speaking while thinking is an unusual behavior that is not usually performed, the amount of speech may be extremely small for some participants [6]. The solution to this problem is to have the participants practice thinking aloud in advance [7] or prepare a closed environment for them [8]. However, such a practice often does not work in a limited time, and preparing a strictly closed environment is difficult. The collected speech data are often analyzed qualitatively; however, differences in the amount of speech and words spoken by the participants render the analysis difficult. Some studies have reported the reactivity of the think-aloud method. It influences performance and task completion time, particularly in tasks with a high cognitive load and complexity [9, 10].

Fig. 1. “Ah-aloud”: Experimental method in which one talks about what comes to mind while performing a task using only “ah.”

To overcome these shortcomings, we propose ah-aloud method in this study. This method helps determine the cognitive processes that occur during a task, while reducing the burden of verbalization and reactivity. This is a method in which all thoughts that occur during task execution are expressed using solely the sound “ah.” We believe that restricting participants to a single sound mitigates the aforementioned challenges. In Japanese, the interjection “ah” is used as subjective expression, such as feelings and thought processes, e.g., “*Ah, naruhodo!*” (“*Oh, I see!*”) or “*Ahh, yuutsu da...*” (“*Ugh, I’m depressed...*”) [11]. Sudo [12] classified functions according to accent and intonation, stating that descending tones imply approval or remembrance, whereas descending to slightly ascending tones implies discouragement or dissatisfaction. It has many other meanings and functions, including doubt and discovery. Therefore, the sound “ah” is suitable for expressing the cognitive processes.

We conducted an experiment with Japanese-speaking participants to demonstrate the viability of the proposed strategy as a novel experimental method for evaluating cognitive processes. Based on the results of this experiment, we discuss its validity as an experimental method for evaluating cognitive processes. In addition, we examined interjections, particularly the English “ah,” which can express a wide range of cognitive processes.

Fig. 2. Experimental Environment: The experimenter is on the other side of the white partition during the task so that the participants can speak easily.

978-1-6654-7664-5/22/$31.00 ©2022 IEEE 19 November 2022

II. Related Works

A. Experimental Methods to Evaluate Cognitive Processes

In this section, we introduce experimental methods other than the think-aloud method to evaluate cognitive processes, and show the differences between these methods and the proposed method.

In contrast to the think-aloud method, in which participants talk to themselves, there is a "question-asking protocol" [6], wherein the experimenter asks questions to the participants and collects speech data as if it were a dialogue. Although this method naturally encourages speech, it is difficult to appropriately set the content and timing of the experimenter's questions. Boren et al. [13] suggested speech communication, in which the experimenter utters acknowledgement tokens. This is difficult to implement correctly, because it requires an appropriate response from the experimenter. Another experimental method that encourages participants to speak while interacting with each other is the dialogue method (Taiwa Ho) [8]. This is not a dialogue between an experimenter and a participant but a dialogue between multiple participants. It is necessary to consider the relationship between the experiment, participants, and setting the goal of the task, as it is difficult to obtain the appropriate speech data for which the experimenter aims. This method facilitates the experimental setup and can be used to obtain spontaneous speech data from participants.

A retrospective method is described at the beginning of this section. This is also called aided subsequent verbal protocol [14]. In this method, the participants are recorded on video while performing a task and are asked to talk about their intentions and thoughts while watching the video after the task is completed. Because the participants were not asked to speak during the task, there was no concern about affecting their task performance. Haak et al. [9] compared speech concurrent with the task (CTA: concurrent think-aloud) and speech while reflecting on the task (RTA: retrospective think-aloud) in a usability test and established the high reactivity of CTA. However, after performing a task that requires more than 10 s, it is possible that all but some of the information transferred from short-to long-term memory is forgotten. In such cases, posterior verbalization is difficult, and the data are known to be incomplete. This is a new method for retrieving information from short-term memory with low reactivity.

III. Proposal: ah-aloud

We propose a new method that can completely reflect the cognitive processes during task execution. As shown in Fig. 1, this is the method in which all thoughts and feelings that occur during task execution are expressed using solely the sound "ah." This may lead to task-directed cognitive processes such as discovery, approval, and incomprehension of the task, as in think-aloud.

We proposed the following two patterns for the use of "ah" during a task.

Single pattern: The participants were instructed to represent their what come to mind solely using "ah."
Continuous pattern: The participants were instructed to vocalize "ah~" continuously.

Fig. 3. Recorded using Multi View Recorder: screen of video game (top left), frequency domain of participant's voice(top right), participant's face(bottom left), and participant's hand operating the controller(bottom right): The participants are asked to watch and reflect on what comes to mind.

The single pattern increases the burden of converting any word that comes to mind into "ah," which may shift the attention of the user from the task to vocalization. Therefore, we propose a continuous pattern that enables users to vocalize their feelings subconsciously. Cognitive processes are evaluated by changes in loudness and tone that occur during continuous vocalization.

IV. Methodology

We experimented with three different patterns: the two proposed in Chapter III, plus the think-aloud pattern, on Japanese-speaking participants to evaluate whether they could reduce the burden of verbalization and reactivity, and what they could express with "ah". As detailed in the next section, we used two video games as the tasks. The participants experimented with three patterns in two tasks for a total of six trials. In the continuous pattern, the participants were allowed to pause for breath as necessary.

Think-aloud pattern: Participants were instructed to verbalize their what come to mind.

Twelve male and female participants, university students between the ages of 21 and 24, were chosen. Because we wanted to acquire a diverse experience, we did not limit their personalities, skills in video games, and experience with this task.

978-1-6654-7664-5/22/$31.00 ©2022 IEEE 19 November 2022
2022 8th International HCI and UX Conference in Indonesia (CHIuXiD)

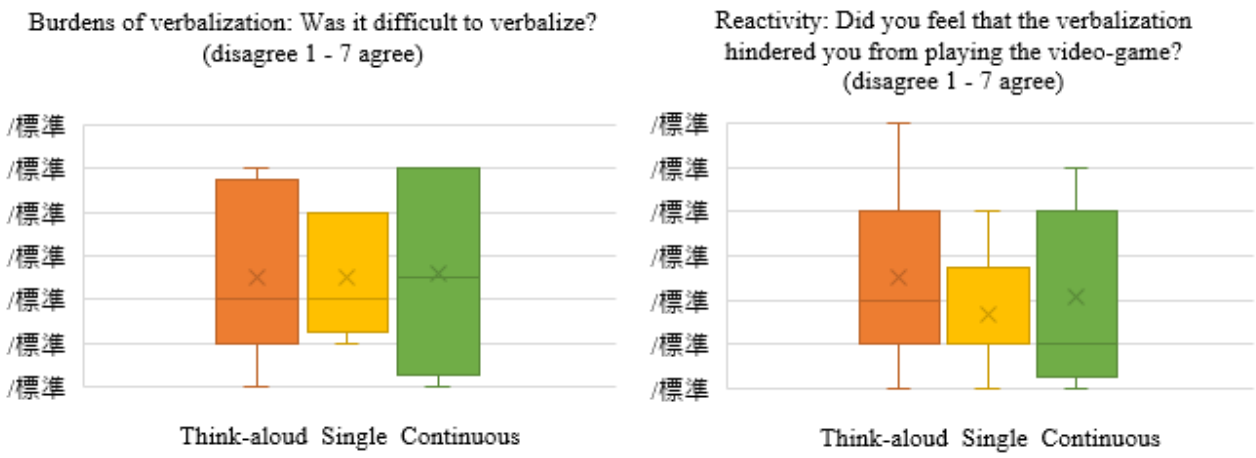

Fig. 4. Burdens of verbalization(left) and reactivity(right): questionnaire results

A. Selection of Tasks

The think-aloud method has been used for relatively self-paced tasks, such as the usability evaluation of application concepts and navigation and strategy analysis in sports. However, reducing the burden of verbalization may be effective for tasks with a high cognitive load, such as playing video games or sports, where the pace of the task may be removed. In this study, we used a video game task. Two video games, Sliding Penguin and Minimum Tennis from Open Video Game Library[1], were used. Escape Fish was used to practice the utterances of each pattern.

Sliding Penguin is a driving game in which penguins slide on ice. It was selected because it is thrilling, and feelings easily come to mind. Minimum Tennis is a simple tennis game. It was selected because the keyboard operation method is complex, and thoughts about the operation and game strategy easily come to mind. To avoid the influence of fatigue caused by a continuous pattern, we set the number of games to a minimum of one.

B. Procedure

The experimental procedure for the participants was as follows. The order of turning trials was counterbalanced. To facilitate speech, a partition was set up, as shown in Fig. 2, to make participants unaware of the experimenter.

1) Receive an explanation of the experiment and how the data to be collected will be handled, and sign the experimental consent form.
2) Receive an explanation of the three patterns to be tested.
3) Receive an explanation of the rules and operations of the video game.
4) Practice each of the three patterns once using the video game for practice.
5) Receive an explanation of the rules and operations of the video game.
6) Practice the video game.
7) The experimental task was performed using a video game for each of the three patterns. After the trials of single and continuous patterns, watch the recording of the experiment using Multi View Recorder[2] (Fig. 3) and reflect on what was on your mind.
8) Repeat 5) - 7) with another video game.
9) Answer the questionnaire using Google Forms.

C. Results

Upon completing the tasks, participants answered a subjective questionnaire on the burden of verbalization and reactivity using a seven-point Likert scale. Fig. 4 shows the questions and box-and-whisker diagram of their answers. The questionnaire results were analyzed using two-way ANOVA.

No significant differences were found for either of the questionnaire items. In other words, ah-aloud did not reduce the burden of verbalization and reactivity compared with think-aloud. Especially in the continuous pattern, the variance was large, and opinions were divided among the participants. In the open-ended descriptions in the questionnaire, the participants claimed about the continuous pattern that "*I could concentrate on the play because I could just say 'ah,'* " "*By continuously saying it out loud, I was able to generate it in response to changes in my feelings*," and "*I didn't have to think about words, so it was easier to play than the think-aloud pattern*." The effectiveness of the continuous pattern was evaluated as hypothesized. Whereas, the breathing problem was a major issue: "*It was physically burdensome to breathe*," "*I was a little tired in terms of breathing and vocalization*," and "*It was very difficult to keep saying things all the time*." The participants claimed about the single pattern that "*I often say 'ah' when I make an operational missing or when I am surprised, so it was not that difficult. In fact, because I vocalized naturally, I felt like it was the most natural of the three patterns*," "*I did not know what it would be like to convert the word to 'ah' at first, but once I tried it, it went smoothly*," and "*I did not feel burdened because I just said 'ah' at my leisure*." On the other hand, negative claim were also obtained: "*I was a little confused about when to say 'ah,'* " "*When I was concentrating on the game or thinking about the next development, I sometimes forgot to say 'ah' and became silent*," and "*While it was easy to respond, it was difficult to express what I was thinking about*," and "*It was burdensome to convert every word that came to mind into 'ah'.* "

Furthermore, in both patterns, the differences in the time and tone of "ah" could express various thoughts and feelings. The participants reflected on their thoughts and feelings after

[1] Library with open video-games and experimental support tools for researchers (https://open-video-game-library.github.io/info/)

[2] https://open-video-game-library.github.io/multi-view-recorder/

978-1-6654-7664-5/22/$31.00 ©2022 IEEE 19 November 2022
2022 8th International HCI and UX Conference in Indonesia (CHIuXiD)

the task, and in many cases, the experimenter could determine the meaning without asking for the participants' opinions. Joy was expressed in various ways, some in small increments and others in up and down tones. However, by analyzing the game screen and facial expressions together, the experimenter was able to easily determine their feelings. The tone was often lowered to express relief; thus, it was easy to detect the emotional state of relief using intonation alone. The tone while expressing sadness increased as if to say "*I've done it*," and conversely, the tone while expressing sadness decreased as if in despair. The magnitudes of these feelings varied but were easier to understand when analyzed in conjunction with the game situation. Surprise and impatient tones often rose sharply in tone and loudness. Some of the same rapid increases in tone expressed interrogation because of a lack of understanding of the game specifications. Some of the awareness tone responses were nuanced, as if they were satisfied, whereas others contained a hint of surprise. Whereas, the contemplation tone was often constant and could not be determined from voice information alone. In some cases, the participants did not vocalize their thoughts, even though they were thinking about something. In particular, when the concentration was high, vocalizations seemed to be difficult to produce. Even during continuous vocalization, some participants experienced vocalizations that were interrupted during important scenes in the video game. Although the subjective evaluation questionnaire suggested a reduction in reactivity, some participants showed reactivity. There were instances of "ah" vocalizations regardless of the thoughts and feelings at the time, such as using "ah" to measure the timing of hitting the ball and adjusting the vocalizations to avoid forceful manipulation. In addition, some participants experienced real tennis, and their habit of vocalizing when hitting the ball was observed in the tennis game. Because these vocalizations affected the task, it is necessary to consider countermeasures in the future.

There was no significant difference in the amount of speech between the patterns. As an example, a spectrogram of the speech produced by Ocenaudio[3] for one participant while playing Sliding Penguin is shown in Fig. 5. The upper panel shows a single pattern, and the lower panel shows a think-aloud pattern. Although it was not possible to compare the amount of speech because of the differences in the content of speech and the duration of play, the amount of speech did not become extremely small for either the participant or the pattern.

D. Discussion

The reason the burden of verbalization and reactivity could not be reduced compared to think-aloud is thought to be a problem with the instructions. In the single pattern, the participants were instructed to represent their what come to mind solely using "ah." We wanted the participants to express their cognitive processes, such as agreement and questioning, as interjections used on a daily basis, but they tried to convert every word into "ah." We need to make what we want the participants to express clear and find an appropriate instruction, for example, "*Please represent 'ah' as if you give back-channel feedback to the task.*" First, it may not be suitable for obtaining data targeted by think-aloud. It is necessary to clarify the type of data that should be appropriately obtained. This is discussed in detail in Section V. In addition, in Minimum Tennis, "ah" was used to figure right time when hitting a ball. Such use of "ah" does not provide meaningful data, so we should had instructed in advance not to use it.

The task design is also problematic. We chose two video games because thoughts and feelings easily come to mind, and participants completed a questionnaire after playing both video games. We should have experimented separately to investigate what kind of task aloud was suitable. Sliding Penguin was more likely to express feelings, and Minimum Tennis was more likely to express thoughts, but it was difficult to determine the specific thoughts based on "ah" alone. From this experiment, it appears that the continuous pattern is suitable for tasks that require high concentration or are in a short period of time, whereas the single pattern is suitable for other tasks. Unlike the continuous pattern, the single pattern makes it easier to understand the triggers of thoughts and feelings; therefore, it is better to use a single pattern. Additional experiments should be conducted in the future to clarify the tasks that are suitable for this method.

In this study, the analysis was conducted by subjective questionnaire evaluation; however, as in previous studies [9], an objective analysis of the task data may show reductions in the burdens of verbalization and reactivity. Burdens of verbalization can use data such as the completion time of a task or, in the case of usability testing, the number of problems detected. However, the task completion time was preferred not to compare because it depended more on video game skills than on vocalization skills. Reactivity was highlighted by the retrospective speech data in this study; in the future, the think-aloud pattern will also be analyzed. Thus, objective analysis might produce results that would reduce the burden of verbalization and reactivity.

V. Limitation and Future Work

A. The method to obtain feelings rather than thoughts

We expected to obtain task-directed cognitive processes, such as discovery, approval, and incomprehension, but we obtained far more feelings. Essentially, feelings are excluded in order to analyze cognitive processes [1], but this was not done in this study. Ah-aloud is more like a method that evaluates feelings from physiological signals, such as electrocardiograms, skin temperature variation, and EEG frequencies (e.g., [15, 16]), rather than thoughts that are closely tied to cognitive processes. Ah-aloud is different from physiological signals in that it prompts spontaneous action (vocalization), but may be able to capture trivial feelings that do not appear in physiological signals. We will clarify the position of this method by comparing it with experimental and analytical methods that use physiological signals.

B. Combination with other methods

As mentioned in Chapter IV, many feelings could be obtained, but if you want to know more about the cognitive processes, you may want to combine retrospective think-aloud, as we did in our experiment. Inherently, retrospective methods may lose vital information that was being considered during the task [2]. However, they greatly depend on the stimuli that participants receive to help them recall their thoughts [9]. Therefore, "ah" may provide stimuli for recalling the cognitive process during the task. Its effectiveness as a stimulus should be verified in future studies.

[3] https://www.ocenaudio.com/

978-1-6654-7664-5/22/$31.00 ©2022 IEEE 19 November 2022
2022 8th International HCI and UX Conference in Indonesia (CHIuXiD)

Eye-tracking measurements may be also valid. Eye tracking has already been validated in combination with think-aloud [17], which provides additional information. In ah-aloud, Eye-tracking could reveal what the vocalization "ah" is in reaction to.

C. Potential Application to English

Because this method utilizes the Japanese "ah" to express various feelings, it is unknown whether this method can be applied to participants whose native language is not Japanese. However, the limitation of expression to a single sound reduces the burden of verbalization and reactivity; therefore, a single sound must be selected for each language. Although "ah" in English is similarly pronounced as "ah" in Japanese, it is unclear whether English speakers naturally express various feelings with this sound.

In his attempt to classify interjections in English by meaning, Jovanović [18] stated that "ah" indicates delight, pain, or surprise, "aargh" indicates disgust, and "ach" indicates delight. There are other examples of expressing awareness and changing nuances through intonation in English [19]. Although this strategy is applicable to English-speaking countries, further experiments with English speakers are necessary for verification.

VI. Conclusion

Here, we proposed the ah-aloud method in which users were asked to represent what comes to their minds with "ah," and the cognitive processes were thus evaluated using the phonetic information. We conducted the experiment using a video game as the task and analyzed a questionnaire to see if ah-aloud could reduce burdens of verbalization and reactivity compared to the think-aloud method, and found that it could not. This may be due to the poor selection of tasks and instructions for the participants, so it is necessary to seek an appropriate experimental environment through further experiments. We found that rich feelings could be evaluated. In the future, it will be necessary to clarify where ah-aloud stands compared to experimental methods that can assess feelings, such as physiological signal measurement. We will also experiment with English-speaking participants to investigate whether the method is valid for them.

References

[1] K. A. Ericsson and H. A. Simon, "Protocol analysis: verbal reports as data (revised edition)," The MIT Press, April 1993.

[2] J. E. Russo, E. J. Johnson, and D. L. Stephens, "The validity of verbal protocols," Memory & Cognition, vol. 17, no. 6, pp. 759-769, November 1989, doi: https://doi.org/10.3758/BF03202637.

[3] S. Bernardini, "Think-aloud protocols in translation research: Achievements, limits, future prospects," Target. International Journal of Translation Studies, vol. 13, no. 2, pp. 241-263, January 2001, doi: https://doi.org/10.1075/target.13.2.03ber.

[4] J. Nielsen, "Usability engineering," Morgan Kaufmann, Octorber 1994.

[5] D. W. Eccles and G. Arsal, "The think aloud method: what is it and how do I use it?," Qualitative Research in Sport, Exercise and Health, vol. 9, no. 4, pp. 514-531, May 2017, doi: https://doi.org/10.1080/2159676X.2017.1331501.

[6] T. Kato, "What "question-asking protocols" can say about the user interface," International Journal of Man-Machine Studies, vol. 25, no. 6, pp. 659-673, December 1986, doi: https://doi.org/10.1016/S0020-7373(86)80080-3.

[7] G. M. Olson, S. A. Duffy, and R. L. Mack, "Thinking-out-loud as a method for studying real-time comprehension processes," New Methods in Reading Comprehension Research, pp. 253-286, Routledge, April 2018, doi: http://dx.doi.org/10.4324/9780429505379-11.

[8] H. Kaiho and E. Harada, "[Introduction to Protocol Analysis -What to Read from Speech Data-] Purotokoru bunseki nyumon -hatsuwa deta kara nani wo yomuka-," Shinyosha, November 1993, Japanese.

[9] M. van den Haak, M. D. Jong, and P. J. Schellens, "Retrospective vs. concurrent think-aloud protocols: testing the usability of an online library catalogue," Behaviour & Information Technology, vol. 22, no. 5, pp. 339-351. doi: https://doi.org/10.1080/0044929031000.

[10] M. Hertzum and K. D. Holmegaard, "Thinking aloud in the presence of interruptions and time constraints," International Journal of Human–Computer Interaction, vol. 29, no. 5, pp. 351–364, January 2012, doi: https://doi.org/10.1080/10447318.2012.711705.

[11] Y. Yao, "[On word recognition of the 'a' type interjections in Japanese] 'A' kei kandoshi ni okeru go no ninntei ni tsuite," Bulletin of the Graduate School of Literature of Waseda University, vol. 66, pp. 209-220, March 2021, Japanese.

[12] J. Sudo, "[An attempt to describe the tone of Japanese interjections] Nihonngo kandoshi no oncho kijutsu no kokoromi," Kinki Society of Phonetics "Studies in Phonetics and Speech Communication," vol. 6, pp. 29-52, December 2008, Japanese.

[13] T. Boren and J. Ramey, "Thinking aloud: reconciling theory and practice," IEEE Transactions on Professional Communication, vol. 43, no. 3, pp. 261-278, October 2000, doi: https://doi.org/10.1109/47.867942.

[14] J. M. Hoc and J. Leplat, "Evaluation of different modalities of verbalization in a sorting task," International Journal of Man-Machine Studies, vol. 18, no. 3, pp. 283-306, March 1983, doi: https://doi.org/10.1016/S0020-7373(83)80011-X.

[15] S. Koelstra, C. Muhl, M. Soleymani, J. S. Lee, et al., "DEAP: A Database for Emotion Analysis; Using Physiological Signals," IEEE Transactions on Affective Computing, vol. 3, no. 1, pp. 18-31, January-March 2012, doi: https://doi.org/10.1109/T-AFFC.2011.15.

[16] K. H. Kim, S. W. Bang, and S. R. Kim, "Emotion Recognition System Using Short-term Monitoring of Physiological Signals," Medical and Biological Engineering and Computing, vol. 42, pp. 419–427, May 2004, doi: https://doi.org/10.1007/BF02344719.

[17] A. Ruckpaul, T. Fürstenhöfer, and S. Matthiesen, "Combination of eye tracking and think-aloud methods in engineering design research," In Design Computing and Cognition '14, pp. 81-97, Springer, Cham, 2015, doi: https://doi.org/10.1007/978-3-319-14956-1_5.

[18] V. Ž. Jovanović, "The form, position and meaning of interjections in English," Facta Universitatis: Series Linguistics and Literature, vol. 3, no. 1, pp. 17-28, June 2004.

[19] D. Bolinger and D. L. M. Bolinger, "Intonation and its uses: Melody in grammar and discourse," Stanford University Press, August 1989.

978-1-6654-7664-5/22/$31.00 ©2022 IEEE 19 November 2022