Hands-Free Interface using Breath for Robot-Assisted Operation

Atsuhiro Imai and Daigo Misaki

***Abstract*— In this paper, we propose a novel hands-free interface for robot-assisted operation. Surgeons use both hands during surgery, making difficult for them to operate the display. However, by using a modality other than the hands to operate the display, the surgeon can check the necessary information for the operation in real time. Therefore, we developed a system that recognizes breath from the temperature change on the mask surface. Then, we compared the proposed system with a directly recognized breath method. Consequently, the proposed method had the same performance as directly recognized breath method and minimized the risk of contamination. We believe that the application of our developed system will contribute to the improvement of safety and efficiency in the medical field.**

I. Introduction

Recent advances in robot-assisted technologies have enabled more precise and efficient procedures. In this context, hands-free interfaces play a crucial role in providing medical professionals with a method to effectively control robots while minimizing unnecessary movements and risks. Among the various approaches proposed, hands-free interfaces using HMDs (Head-mounted displays) have gained significant attention [1].

The problem in hands-free interfaces using HMDs is switching between "pointing" that involves moving the cursor over an object and "selection" that involves placing an object in a selected state. Most systems use eye or head movements for pointing and gazing or blinking for selection; this may induce the Midas touch problem and false input owing to physiological reactions [2] [3]. A method using breath for selection has been proposed, and it is expected to provide intuitive and environmentally independent inputs [4]. Recently, a method to recognize breath by incorporating a sensor into a mask has been proposed [5]; however, its practicality in the medical field and other considerations have not been fully addressed. The disadvantages of conventional methods include an increased risk of infection and contamination, reduced operability, sensor degradation, and setup complexity.

A. Imai is with the Graduate School of Engineering, Mechanical Engineering Program, Kogakuin University, Tokyo, Japan (e-mail: am23011@g.kogakuin.jp).

D. Misaki is Faculty of Engineering Department of Mechanical System Engineering, Kogakuin University, Tokyo, Japan (e-mail: misaki@cc.kogakuin.ac.jp).

In this study, we developed a contactless breath recognition system that can recognize the breath without contact with a mask (Figure 1). Because the system does not require contact with the mask, it is expected to prevent saliva droplets at the mask performance level. The system uses an infrared sensor to recognize temperature changes on the mask surface owing to breathing and judges the input when the temperature exceeds a reference level. We implemented a hands-free interface using head movements and breath input by combining an HMD with a contactless breath recognition system. Furthermore, we compared the performance and usability of "contact-type interface" using a directly recognized breath method and "contactless-type interface" using the proposed breath recognition method in a virtual reality (VR) environment.

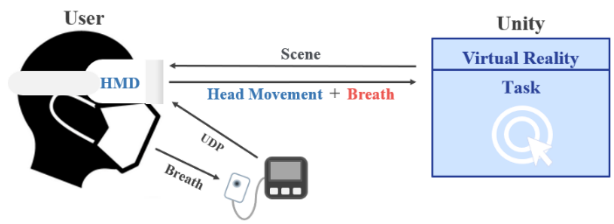

Figure 1. Hands-free interface with head movements and breath inputs

II. Related Works

A. Surgical Hands-Free Interfaces

Several studies have focused on developing hands-free interfaces using HMDs for surgical robots. One study developed a wearable hands-free human–robot interface for robotized flexible endoscopes [6]. Another study proposed a hands-free user interface for VR headsets based on in situ facial gesture recognition [7]. A third study developed a hands-free interface for surgical procedures based on foot pressure sensors [8]. A fourth study proposed a hands-free user interface for VR devices that recognizes facial gestures [9]. Finally, a fifth study developed VR simulation of novel hands-free interaction concepts for surgical robots [10].

B. Pointing for Hands-Free Operations

Eye and head movements are the main pointing methods for hands-free operations. Qian et al. [11] compared eye-only, head-only, and eye + head pointing (selection: keyboard) in 3D space in a Fitts' law task using an HMD. The results showed that head-only inputs performed significantly better than eye-only and eye +

head inputs in terms of throughput, error rate, and movement time. Hansen et al. [12] compared head and eye pointing (selection: dwell time) and mouse input on a 2D plane in a Fitts' law task using an HMD. The results showed that the throughput was the highest for the mouse, followed by the head and eyeballs. These results suggest that the use of head movement for pointing in a hands-free interface enables stable input.

C. *Breath Recognition for Selection*

Breath recognition methods include the recognition of intake sounds using microphones mounted on PCs, tablet terminals, and HMDs [13] [14] [15]. However, voice is easily affected by the environment, and noise such as ambient noise can affect recognition accuracy. Lee et al. [16] used an infrared camera to recognize residual heat from breathing on an object to determine the input. Onishi et al. [4] developed a system that recognized breathing using an infrared camera and used intentional changes in rhythm as input. They also proposed a hands-free interface that could solve the Midas touch problem by combining it with eye movements. In this study, we recognized breath temperature using an infrared sensor based on the aforementioned studies.

D. *Interaction using Head Movements and Breathing*

Many interaction systems using head movements and breathing have been proposed. For example, one method combines head movement recognized by a PC and breath recognized by a piezoelectric sensor [17]. Another method improves the gaming experience in VR using four command patterns, where breathing is regarded as a specific input command [18]. Furthermore, breathing can be utilized in a wide variety of ways, such as breathing inputs for people with severe motor disabilities [19]. Therefore, a comprehensive survey of interaction methods using breathing has been conducted [20], and many studies using breathing to enhance meditation and VR experiences have been mentioned.

E. *Hands-Free Operations using a Mask*

In recent years, research using masks as interfaces has attracted attention. Tatzgern et al. [21] proposed a head-mounted interface that uses breathing as input/output using an HMD and airflow sensors. In recent years, research has also been conducted on the recognition of mouth movements by attaching capacitance sensors to commercially available masks [22] and on interface operation by recognizing inhalation and exhaust movements associated with breathing [5]. Although various interfaces using masks have been proposed, as aforementioned, few studies have considered the hygienic safety associated with the use of interfaces such as that in this study.

III. Principle of Contactless Breath Recognition System

A. *Breath Recognition*

The principle of a contactless breath recognition system is as follows. Because the fabrics used for masks generally have low thermal conductivity, the amount of heat transferred to the mask depends on the breathing duration. Therefore, an intentionally high-temperature breath was used as the input. Because the temperature during breathing varies depending on individual differences, usage conditions, and each breath, the input criteria must be updated each time the system is used. In this system, the previous breath was measured for a certain period of time, and the highest temperature at that time was used as the reference temperature for input judgment. Radiant heat was recognized using an NCIR unit (MLX90614) and M5Stack Core to recognize breath temperature.

B. *Measurement Time of the Reference Temperature*

The measurement time of the reference temperature was investigated in five participants using a preliminary survey. In this study, the participants were seated at a distance of 3cm between the mask and the sensor, and nasal breathing was measured for 30s. This was performed three times for each participant, and the time at which the maximum temperature was measured was determined from the recorded temperature changes. The interquartile range of the measurement time was 3.58–18.4s, and the quartile range of the maximum temperature was 27.8–29.6°C. In this system, reference temperature measurement time was set to 20s to minimize leakage, considering the quartile range.

IV. Response Time of the Contactless Breath Recognition System

The purpose of this study was to measure the response time for breath recognition using a contactless breath recognition system and examine its usefulness. The times required to turn the breath input on and off were evaluated.

A. *Participants*

Five male university students in their 20s were used as participants, and the experiment was conducted with them wearing masks.

B. *Equipment*

The M5Stack Core and NCIR unit (MLX90614) were used for the contactless breath recognition system. In addition, a non-woven cloth mask (FACE MASK Disposable) was used.

C. *Procedure*

First, the participant was seated at a distance of 3cm between the mask and the sensor. Next, the temperature change during breathing was measured for 20s and a reference temperature was set. Subsequently, mouth breathing was input for only 1s at the time specified by the timer displayed on the PC. The response time for judging whether to input or stop the breath was then examined.

D. *Results*

The median response time for breath recognition was 565ms (interquartile range: 323–809ms) for input and 1239ms (interquartile range: 1066–1239 ms) for stopping. The distribution of the measured response times is shown in Figure 2. The median response time at the time of breath input was 565ms, but differences were observed among the participants. Similarly, the response time at the time of stopping was also distributed and might have been affected by the amount of breathing and airflow velocity of the individual. As shown in Figure 2, the response at the time of breath-holding was slower than that at the time of breath input. This might be because the temperature of the mask surface continued to increase during the one second of breath input, and the temperature took time to decrease below the reference temperature. This suggested that the response time for breath recognition was delayed more at the time of stopping than at the time of input.

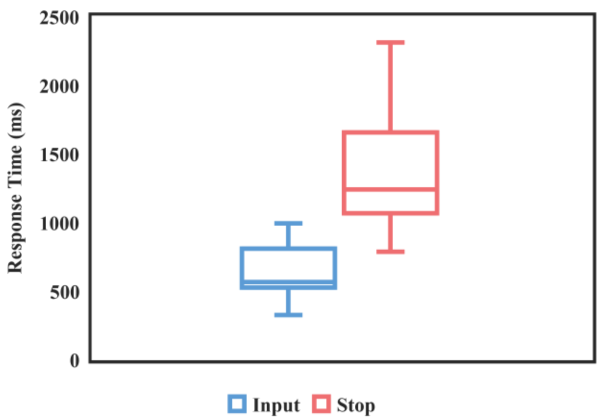

Figure 2. Response time when breath is input and stopped

V. EVALUATION OF CONTACTLESS-TYPE INTERFACE

A. *About the Interface*

A contactless-type interface is a hands-free interface that enables pointing by head movements and selection by breath. Head movements can be selected by recognizing the temperature of the mask surface using the Meta Quest2 contactless breath recognition system that enables breath input without saliva splashing on the sensor. In this study, to verify the usefulness of the contactless-type interface, we conducted a comparison experiment with a contact-type interface. Both used the same sensor, but the contact-type sensor was attached to a mask with a hole in it and detected breath directly inside the mask. This allowed us to determine whether the performance between direct and indirect breath detection differed. Figure 3 shows the contactless-type interface, and Figure 4 shows the contact-type interface.

Figure 3. Contactless-type interface

Figure 4. Contact-type interface

B. *Evaluation by the Fitts' Law Task*

To evaluate the performance and usability of the interface, we constructed a Fitts' law task in a VR environment based on a previous study [24]. In this study, a sphere with a diameter of 3px was placed in a circle with a radius of 10px, and the task progressed in the order of moving around the circumference by selecting across the circle diagonally. The starting position of the breath input was Start (green sphere), and the ending position was Target (red sphere). When the selection was successful, the current position changed to Start and a new Target was displayed on the diagonal. Each participant performed 13 trials in a set, using two different interfaces. A pointer (white sphere) displayed at the center of the viewpoint was used for pointing, and the color of the sphere dimmed when it was held. Figure 5 shows the Fitts' law task built in a VR environment.

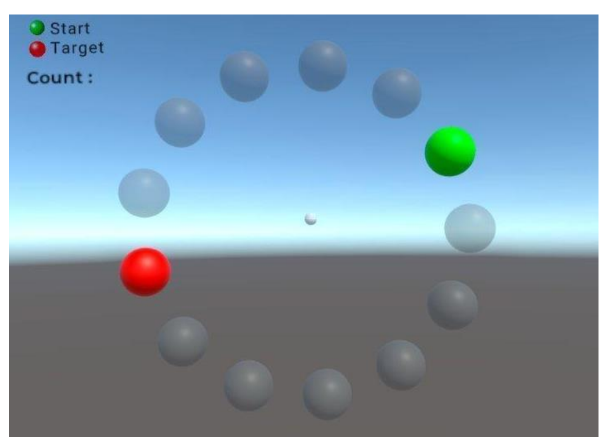

Figure 5. Fitts' law task constructed in a VR environment

C. Participants

Eight male university students in their 20s were used as participants.

D. Equipment

The M5Stack Core, NCIR unit (MLX90614), and Meta Quest2 were used as the interfaces. A nonwoven cloth mask (FACE MASK Disposable) was used.

E. Procedure

First, the content of the task was explained to the seated participants, and they were requested to wear the breath recognition interface used in the first set of experiments. Next, the interface was set up (reference temperature setting), and selecting with mouth breathing was practiced 2–3 times. Immediately before the experiment, the subjects were instructed to perform the task as quickly and accurately as possible. During the experiment, the reaction time and distance traveled were recorded for each trial to evaluate performance. The questionnaire included items on the mental effort required to operate the device, physical effort required to operate the device, and smoothness of operation (including movement and input) that were rated on a seven-point Likert scale. After answering the questionnaire, the participants wore the interface for the second set of experiments and performed the same tasks as those in the first set. At the end of the second set, the subjects were asked questions regarding the concept of contactless breath recognition and problems with the system.

F. Results

The performance results for the Fitts' law task are presented. As shown in Figure 6, the average throughput was 1.32 bps (95% confidence interval: 1.21–1.43 bps) for the contact-type interface and 1.33 bps (95% confidence interval: 1.26–1.40 bps) for the contactless-type interface. As shown in Figure 7, the mean error rates were 11.5% (95% confidence interval: 11.4–11.7%) for the contact-type interface and 12.5% (95% confidence interval: 12.4–12.6%) for the contactless-type interface. The average travel time per trial is shown in Figure 8.

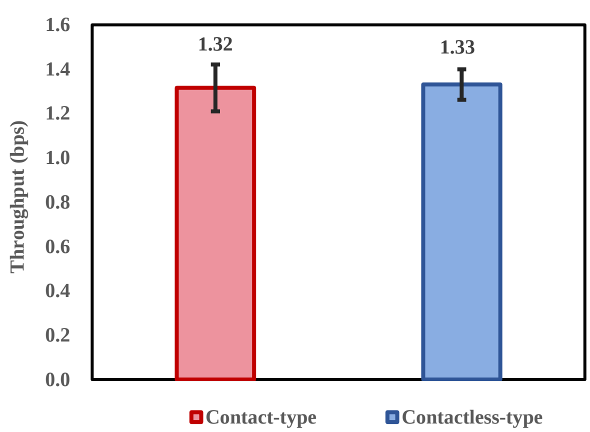

Figure 6. Throughput (95% confidence interval error bars)

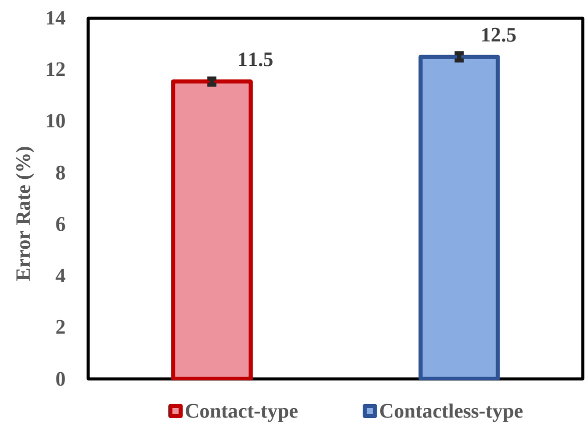

Figure 7. Error rate (95% confidence interval error bars)

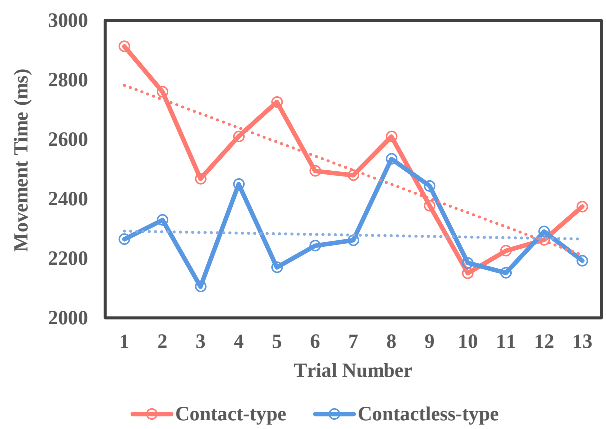

Figure 8. Travel time per trial

The questionnaire results for the Fitts' law task are as follows. The median mental effort required for operation was 4 (interquartile range: 2.25–5.75) for the contact-type interface and 3.5 (interquartile range: 2.25–4.75) for the contactless-type interface. The median physical effort required for operation was 5 (interquartile range: 5–6) for the contact-type interface and 4.5 (interquartile range: 3.25–6) for the contactless-type interface. The median smoothness of operation was 5 (interquartile range: 3–6.75) for the contact-type interface and 5.5 (interquartile range: 3.5–7) for the contactless-type interface. The distributions of the mental effort required for operation, physical effort required for operation, and smoothness of operation are shown in Figures 9, 10, and 11, respectively.

All participants were positive about the usefulness of the contactless breath recognition interface for the mask. However, some participants commented on the problems with the system such as "the amount of breath required is too much," "there is a noticeable delay," and "I wondered if the system would respond well even if there was moisture on the surface inside the mask due to breathing or other influences."

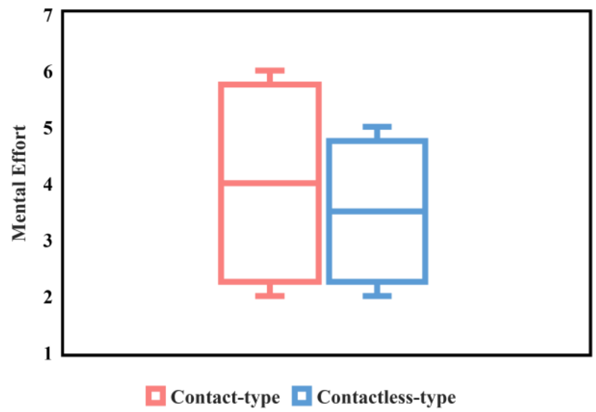

Figure 9. Mental effort required to operate

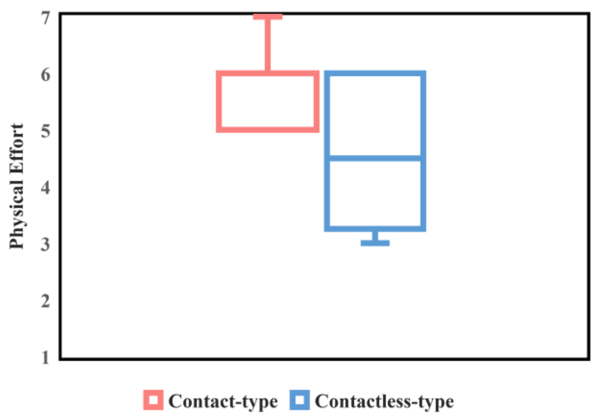

Figure 10. Physical effort required to operate

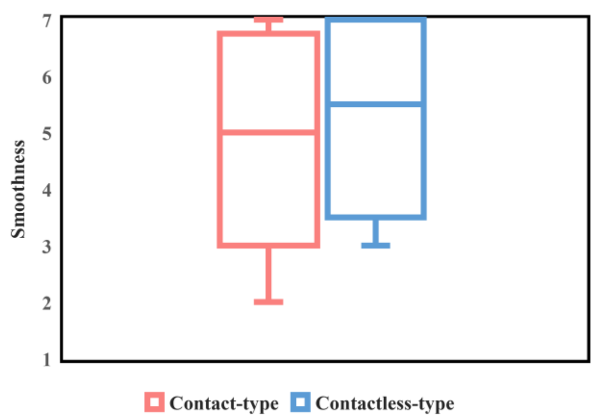

Figure 11. Smoothness of operation

VI. DISCUSSION

A. *Performance*

The throughput and error rate were 1.32 bps and 11.5%, respectively, for the contact-type device and 1.33 bps and 12.5%, respectively for the contactless-type device. A study by Soukoreff et al. at the University of Toronto [23] noted that the throughput and error rates for existing devices were approximately 4 bps and 20%, respectively, for mouse, 3 bps and 32%, respectively, for laser pointers, and 2 bps and 4%, respectively, for touchpads. These results suggested that, although the processing performance was inferior to that of the mouse, laser pointer, and touchpad, selection errors were less likely to occur. This might be due to the differences in the body parts used for pointing. Mouse and laser pointers are agile because they use fingers, but they may induce input errors because the fingers and the displayed image are located far from each other. In contrast, the proposed method used the head that was not as agile as a mouse. However, the head and the displayed image were relatively close to each other; therefore, false inputs were unlikely to occur. The positional relationship between the body part and the displayed image was assumed to affect the throughput and error rate values of the touchpad. In addition, the throughput and error rate of the touchpad were compared with those of existing devices, suggesting that little difference existed between touchpads and contactless-type devices. This suggested that almost no difference existed between the recognition of the temperature change inside the mask and that on the mask surface, and that it can be used as an input to the interface.

A linear approximation of the travel time per trial showed that the travel time of the contact-type interface tended to decrease with the number of trials, whereas that of the contactless-type interface remained relatively constant irrespective of the number of trials. Because the conditions for the travel distance and order of the experiments were the same in this study, the decrease in the travel time of the contact-type interface was due to the characteristics of the interface. We discuss the causes of travel time in detail based on the results of the questionnaire.

B. *Questionnaire*

The physical effort required for the operation concentrated on a higher score for the contact-type interface when the quartile ranges were compared. This might be due to the difference in the breath recognition methods used for input judgment. The contact-type interface recognized temperature changes inside the mask; therefore, the reference temperature increased owing to the heat retained in the mask, making it difficult to recognize the breath at the time of input. However, because the contactless-type interface recognized the temperature of the mask surface, the reference temperature decreased by the mask medium, and the breath was easily recognized.

The aforementioned results suggested that the decrease in the movement time of the contact-type interface in each trial might be due to the effect of habituation to the recognition accuracy of the sensor.

VII. CONCLUSION

In this study, we developed a surgical interface that used head movements and breathing inputs, without the need for the use of hands. Comparing the results with those of a conventional contact-type interface that used masks, the new interface had the same processing ability and a high possibility of reducing the scattering of saliva on the sensor. This makes it suitable for use in surgical environments. We believe that by conducting experiments in environments with good communication conditions, we can optimize the response time of the system and reduce the amount of breathing required. We will continue to improve the system based on the feedback from participants and explore the possibility of integrating it into a surgical interface.

REFERENCES

[1] R. Rahman, et al., "Head-mounted display use in surgery: A systematic review," Surg. Innov., vol. 27, no. 1, pp. 88-100, 2020.

[2] X. Meng, W. Xu, and H.-N. Liang, "An exploration of hands-free text selection for virtual reality head-mounted displays," 2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), 2022.

[3] H. Sato, K. Abe, S. Ohi, and M. Ohyama, "A text input system based on information of voluntary blink and eye-gaze using an image analysis," Electronics and Communications in Japan, vol. 101, no. 2, pp. 9–22, 2018.

[4] R. Onishi et al., "GazeBreath: Input method using gaze pointing and breath selection," Augmented Humans, pp. 1-9, 2022.

[5] T. Kusabuka and T. Indo, "IBUKI: Gesture input method based on breathing," Adjunct Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology, 2020.

[6] S. Zuo, T. Chen, X. Chen and B. Chen, "A wearable hands-free human-robot interface for robotized flexible endoscope," IEEE Robotics and Automation Letters, vol. 7, no. 2, pp. 3953-3960, April 2022, doi: 10.1109/LRA.2022.3149303.

[7] J. Kim, J. Cha, and S. Kim, "Hands-free user interface for VR headsets based on in situ facial gesture sensing," Sensors (Basel), vol. 20, no. 24, p. 7206, Dec. 16, 2020, doi: 10.3390/s20247206. PMID: 33339247; PMCID: PMC7766087.

[8] T. Kawai, M. Fukunishi, A. Nishikawa, Y. Nishizawa, and T. Nakamura, "Hands-free interface for surgical procedures based on foot movement patterns," Annu Int Conf IEEE Eng Med Biol Soc, pp. 345-8, 2014, doi: 10.1109/EMBC.2014.6943600. PMID: 25569968.

[9] J. Kim, J. Cha, and S. Kim, "Hands-free user interface for VR headsets based on in situ facial gesture sensing," Sensors, vol. 20, no. 24, p. 7206, 2020.

[10] F. You et al., "VR simulation of novel hands-free interaction concepts for surgical robotic visualization systems," Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part III, vol. 23, Springer International Publishing, 2020.

[11] Y. Y. Qian and R. J. Teather, "The eyes don't have it: An empirical comparison of head-based and eye-based selection in virtual reality," Proceedings of the 5th Symposium on Spatial User Interaction (SUI '17), New York, ACM, pp. 91–98, 2017.

[12] J. P. Hansen et al., "A Fitts' law study of click and dwell interaction by gaze, head and mouse with a head-mounted display," in Proc. Workshop on Communication by Gaze Interaction, 2018.

[13] S. N. Patel and G. D. Abowd, "Blui: low-cost localized blowable user interfaces," in Proc. 20th Annual ACM Symp. User Interface Software and Technology, 2007.

[14] J.-H. Kim and J. Lee, "Controlling your contents with the breath: Interactive breath interface for VR, games, and animations," PLoS One, vol. 15, no. 10, Art. no. e0241498, 2020.

[15] K. Yamano et al., "VR blowgun: respiratory rehabilitation using breath detection," in Proc. Entertainment Computing Symposium 2021, pp. 322-328, 2021. (in Japanese)

[16] K. Lee, S. H. Lee, and J.-I. Park, "Hands-free interface using breath residual heat," in Human Interface and the Management of Information. Interaction, Visualization, and Analytics: 20th Int. Conf., HIMI 2018, Held as Part of HCI Int. 2018, Las Vegas, NV, USA, July 15-20, 2018, Proceedings, Part I, vol. 20, Springer International Publishing, 2018.

[17] E. M. Campos et al., "Evaluating alternative interfaces based on puff, electromyography and dwell time for mouse clicking," in Proc. 18th Brazilian Symp. Human Factors in Computing Systems, 2019.

[18] M. Sra, X. Xu, and P. Maes, "Breathvr: Leveraging breathing as a directly controlled interface for virtual reality games," in Proc. 2018 CHI Conf. Human Factors in Computing Systems, 2018.

[19] O. Hassan, S. Shamsir, and S. K. Islam, "Machine learning based hardware model for a biomedical system for prediction of respiratory failure," in 2020 IEEE Int. Symp. Medical Measurements and Applications (MeMeA), IEEE, 2020.

[20] M. Prpa, E. R. Stepanova, T. Schiphorst, B. E. Riecke, and P. Pasquier, "Inhaling and Exhaling: How Technologies Can Perceptually Extend our Breath Awareness," in Proc. 2020 CHI Conf. Human Factors in Computing Systems, pp. 1-15, 2020.

[21] M. Tatzgern et al., "Airres mask: A precise and robust virtual reality breathing interface utilizing breathing resistance as output modality," in Proc. 2022 CHI Conf. Human Factors in Computing Systems, 2022.

[22] T. Yamazaki, B. Shizuki, and S. Takahashi, "A hands-free input method with masked interfaces," in IPSJ Interaction 2018, pp. 2B24, 2018. (in Japanese)

[23] R. W. Soukoreff and I. S. MacKenzie, "Towards a standard for pointing device evaluation, perspectives on 27 years of Fitts' law research in HCI," Int. J. Human-Computer Studies, vol. 61, no.

[24] Y. S. Pai, T. Dingler, and K. Kunze, "Assessing hands-free interactions for VR using eye gaze and electromyography," Virtual Reality, vol. 23, no. 2, pp. 119–131, June 2019.