Journal of Data Intelligence, Vol. 3, No. 4 (2022) 421–440

NASAL BREATH INPUT: EXPLORING NASAL BREATH INPUT METHOD FOR HANDS-FREE INPUT BY USING A GLASSES TYPE DEVICE WITH PIEZOELECTRIC ELEMENTS

RYOMA OGAWA
College of Info. Sci. and Eng., Ritsumeikan University
Shiga, Japan [a]
ryoma.ogawa@iis.ise.ritsumei.ac.jp

KYOSUKE FUTAMI
College of Info. Sci. and Eng., Ritsumeikan University
Shiga, Japan
Digital Spirits Teck
Kyoto, Japan
futami@fc.ritsumei.ac.jp

KAZUYA MURAO
College of Info. Sci. and Eng., Ritsumeikan University
Shiga, Japan
Japan Science and Technology Agency, PRESTO
Saitama, Japan
murao@cs.ritsumei.ac.jp

Research on hands-free input methods has been actively conducted. However, most of the previous methods are difficult to use at any time in daily life due to using speech sounds or body movements. In this study, to realize a hands-free input method based on nasal breath using wearable devices, we propose a method for recognizing nasal breath gestures, using piezoelectric elements placed on the nosepiece of a glasses-type device. In the proposed method, nasal vibrations generated by nasal breath are acquired as sound data from the devices. Next, the breath pattern is recognized based on the factors of breath count, time interval, and intensity. We implemented a prototype system. The evaluation results for 10 subjects showed that the proposed method can recognize eight types of nasal breath gestures at 0.82% of F-value. The evaluation results also showed that the recognition accuracy is increased to more than 90% by limiting gestures to those with a different breath count or different breath interval. Our study provides the first glasses type wearable sensing technology that uses nasal breathing for hands-free input.

Keywords: Glasses, Nose breath, Wearable device, Vibration sensing, Hands-free input, Piezoelectric elements

1. Introduction

In recent years, research on simple hands-free input methods has been actively conducted as devices for various purposes have become widespread. Devices, including information terminals and peripheral devices, are integrated into our lives. Many activities of daily living are performed with the assistance of devices, and people are always using some type of device. However, problems arise when users cannot operate these devices because their hands are

[a]College of Info. Sci. and Eng., Ritsumeikan University, 1-1-1 Nojihigashi, Kusatsu, Shiga 525-8577, Japan

occupied, such as holding a package in our hands. Currently, most devices require display touch input, keyboard input, or button input by finger or hand movements. To solve this problem, research on simple hands-free input methods has been actively conducted since devices are now a vital and integral part of our everyday lives.

One of the hands-free input methods is speech input [25]. For example, artificial intelligence assistants for voice input (e.g., Amazon's Alexa, Apple's Siri) are in widespread use and enable hands-free operation. There are also methods that allow hands-free input by moving various parts of the body, such as the face[5][7], head[8][15], mouth[2][3], and eyes[6][10].

However, for some people, the previous methods are difficult to use at any time in daily life. For example, input methods using speech sounds or body movements yield problems in situations where others are present due to some reasons, such as the input act is unnatural and conspicuous, and the contents of the input act are known to others. In addition, input methods using speech sounds have the problem of degraded recognition accuracy due to external noise (e.g., environmental noise, the voice of other people). Therefore, if a simple hands-free input method using wearable devices without speech sounds or body movements can be realized, it will be useful on a daily basis and reduce such problems.

In this study, to realize a simple hands-free input method based on nasal breath using wearable devices, we propose a method for recognizing nasal breath gestures using piezoelectric elements placed on the nosepiece of a glasses-type device. In the proposed method, nasal vibrations generated by nasal breathing are acquired as sound data from small piezoelectric devices placed on the nasal surfaces. Next, the breath pattern is recognized based on the factors of breath count, time interval, and intensity. We implemented a prototype system in which piezoelectric elements were placed on both nose pads of a glasses-type device. We conducted three types of evaluations. In Evaluation 1, we verified whether the proposed method could recognize eight types of breath gestures, which are combinations of three factors: breath count, time interval, and intensity. In Evaluation 2, we evaluated the proposed method under multiple conditions that may reduce the recognition accuracy. In Evaluation 3, we evaluated whether the recognition accuracy of the proposed method can improve by limiting the breath gestures.

In this study, we focused on nasal breathing as a gesture that can be performed at all times in daily life, and a glasses-type device was adopted to sense nasal breathing. Although many studies have been conducted on devices for detecting breath and hands-free input methods, few studies use nasal breath and a glasses-type device. Our proposed method can be applied to various wearable devices containing the installation part of the nose, such as glasses, head mounted display, and masks.

The remainder of this paper is organized as follows. Section 2 introduces the related research. Section 3 explains the proposed method, and Section 4 describes the implementation. Section 5 shows Evaluation 1 with eight gestures. Section 6 shows Evaluation 2 under multiple conditions. Section 7 shows Evaluation 3 with limited gestures. Section 8 describes the discussion and future work. Finally, Section 9 summarizes our study.

Note that we published the concept of the proposed method in the short paper [27]. The differences between this paper and the previous one are as follows. This paper improved Experiment 1 in Section 5. Although Experiment 1 is the same as the previous paper, the number of subjects increased from eight to 10. Therefore, this paper provides more reliable results than the previous paper. In addition, this paper added Experiment 2 in Section 6 and Experiment 3 in Section 7.

2. RELATED WORK

2.1. *Hands-free input method*

To solve the problems in device operation, a lot of research has been done on hands-free input as a new device operation method. Lyons et al. [2] proposed MouthType, as an input method using the shape of the mouth and a numeric keypad, by recognizing the shape of the mouth captured by a small camera. The error rate for English input was 3.1% and that for Japanese input was 8.7%. Although the error rate was higher than that of keyboard input on cell phones, MouthType was faster. Similarly, Oguchi et al. [3] proposed a method of device operation with the use of commands inferred from vowel input sequences, using Japanese vowel lip shape recognition. In addition, these lip shape recognition technologies have the disadvantage that the mouth must be visible in the camera, so they cannot be used when wearing something that hides the mouth, such as a mask. Nakao et al. [5] have proposed a mask-type wearable device, Make-a-Face, which enables hands-free operation using tongue, mouth, and cheek gestures by recognizing muscle movements in the lower half of the face from electrodes attached around the mouth.

Manabe et al. [12] proposed an earphone that recognizes gaze input gestures using an electrooculography (EOG) sensor. Amesaka et al. [26] proposed a method that recognizes gestures of facial expression by using active acoustic sensing and the change of the shape of the ear canal. EarFieldSensing [13] is a method that recognizes gestures of the movements of the face by using electrodes for electromyography (EMG) attached to earphones and the changes in the electric field of the ear canal. Taniguchi et al. [14] proposed a method that recognizes the movement of the tongue by using an infrared distance sensor attached to the tip of a canal type earphone and the movement of the bottom of the ear canal. CanalSense [15] is a method that recognizes the movements of the face by using a barometer attached to earphones. TongueInput [16] is a method that recognizes tongue gestures by using a mouthpiece with an infrared distance sensor.

When compared to these methods, this study proposes a method using nasal breath without speech sounds or body movements.

2.2. *System using breath recognition*

Breath information has attracted much attention as important biological information, and many studies have focused on it. Chauhan et al. [17] proposed BreathPrint, a behavioral ecological authentication system that recognizes individuals based on speech information obtained from breath gestures. There are also studies aimed at breath training, using breath recognition. Shih et al. [18] proposed Breeze, which is an application for biofeedback-guided breath training through a game. There are also methods that use breaths to control games [19][20] and methods that use breathing as an input interface [21][22][23] for various devices. Although previous research senses breath from heat, flow rate, and expiratory pressure obtained from heat sensors and flow sensors, this research proposes an approach to sense nasal

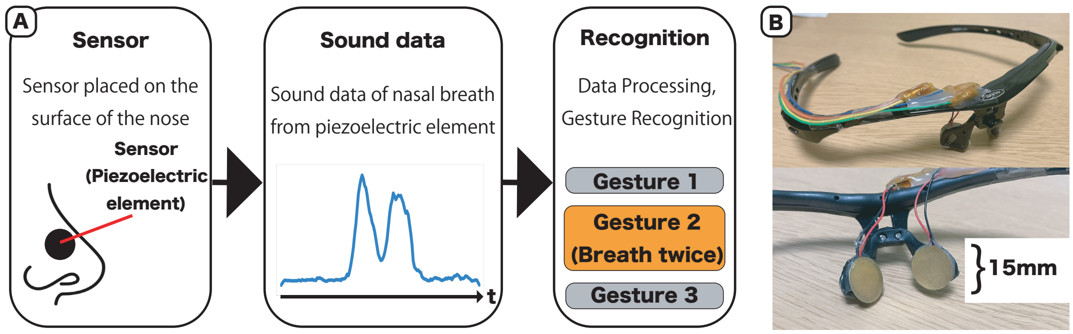

Fig. 1. (A) Flow of the proposed method. (B) Glasses-type device with piezoelectric element.[27]

breath gestures from a piezoelectric element attached to the nose. In addition, for an application of nasal breath, this study proposes a hands-free input method that can be used in daily life using an eye wear device.

3. PROPOSED METHOD

The outline of the proposed method is shown in Figure 1 (A). The implemented glasses-type device is shown in Figure 1 (B). We placed piezoelectric elements on both nasal surfaces of an eyeglass-type device. The two piezoelectric elements touching the left and right nasal surfaces are used as microphones. Then, the vibration of the breath sound generated by nasal breath is acquired from each piezoelectric element as a sound signal. In the sound processing phase, the volume is extracted as a feature amount from the sound signals. In the recognition phase, breath gestures are recognized based on breath count, time interval, and intensity.

The reason for placing the piezoelectric element on the nose plate of the glasses is the following. V-Speech was proposed by Hector et al. [24]. By acquiring the vibration of the nasal bone during speech using the piezoelectric element placed on the nose pad of smart glasses, it becomes easier to recognize the spoken voice even in noisy conditions. Based on the results of this previous study, we hypothesized that it is possible to acquire the vibration of nasal breath instead of voice by placing piezoelectric elements at similar sensor locations.

3.1. *Processing of sound data*

In sound data processing, the volume is extracted as a feature from the acquired sound data and the volume of the piezoelectric elements on both sides of the nose is averaged to obtain a one-dimensional value. At this time, Root Mean Square (RMS) is performed. Next, a moving average is performed to smooth the time series data of the intensity, and resampling is performed to reduce the number of samples. Since the sampling frequency was 48,000 [Hz] and there were 48,000 samples per second, we performed a moving average every 4,800 samples and resampled data by 1/1000 times.

3.2. *Gesture recognition*

The system recognizes the user's nasal breath gesture based on the change in the volume

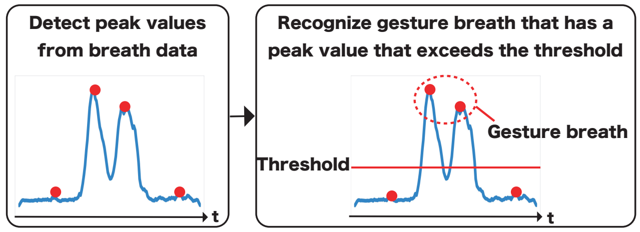

Fig. 2. Recognition of breath gesture[27]

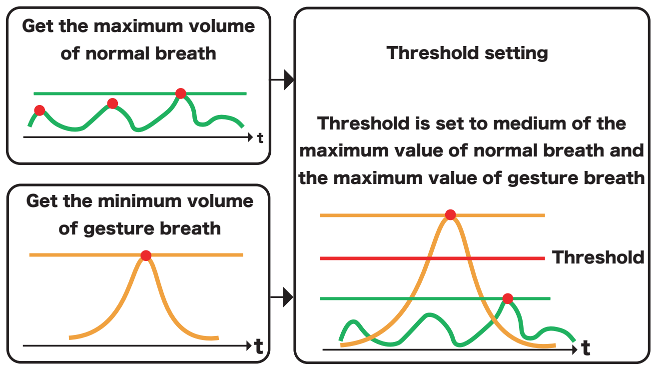

Fig. 3. Threshold Setting[27]

obtained from the vibration of the piezoelectric element. The volume goes up at the beginning of the nasal breath and goes down at the end of the nasal breath. This flow appears as the waveform. We call this flow a single breath. As a result, when you breathe through your nose, a peak of volume is created as shown in the sound data part of Figure 1.

3.2.1. *How to recognize breath as a gesture*

The method of detecting a single breath as a gesture is shown in Figure 2. (1) First, we find the maximum value (i.e., the peak value) of the waveform of the volume for a single breath. The gradient value of the waveform is obtained from the increase/decrease of the differential value. Based on this gradient value, we found the peak value of the breath waveform. In Figure 2, the peak value is indicated by the red dot. (2) Next, among the points detected as the maximum value, the points that exceed the threshold are recognized as intentional breath (i.e., a single breath as a gesture). In Figure 2, the threshold value is represented by the red line. (3) If there is no peak value that exceeds this threshold, it is determined to be normal breathing.

3.2.2. *How to set the threshold to recognize breath for a gesture:*

The explanation of the threshold setting is shown in Figure 3. This threshold was set so that a single breath as a gesture would not be misrecognized as normal breathing. First, both normal breathing data and breath data as a gesture are prepared for each user. Then, the threshold is set to the median value (red line in Figure 3) of the maximum value of normal breathing (green line in Figure 3) and the peak value of a breath gesture (orange line in Figure

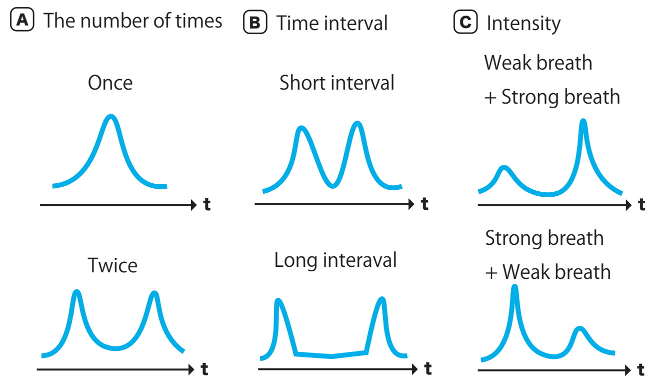

Fig. 4. Types of elements for nasal breath gestures[27]

3). When the breath data as a gesture contains multiple breath gestures, the minimum value of the peak value of multiple breath gestures is used for the peak value of a breath gesture (orange line in Figure 3). The threshold is set for each user. These settings are determined before using the system.

3.2.3. *Recognition of gestures based on several elements*

This algorithm recognizes the following three different breath gestures. The algorithm of each recognition is shown below. Note that the algorithm is created for a gesture consisting of a maximum of two gesture breaths.

1. Recognition of gestures with different breath counts: This algorithm recognizes breath gestures based on the difference of breath counts (e.g., one or two) in a certain time range. An example is shown in Figure 4 (A). It is recognized from the following steps. (1) The number of single breaths as a gesture in a certain time range is counted. The single breath as a gesture is obtained from the detection method described above "How to recognize breath as a gesture." In the evaluation described later, the certain time range was set to four seconds.

2. Recognition of gestures with different time intervals between breath: This algorithm recognizes breath gestures based on the difference of the time intervals of breaths as shown in Figure 4 (B). For example, this algorithm distinguishes between gestures with short time intervals and gestures with long time intervals. (1) If two single breaths as a gesture are detected in a certain time range, the time interval between the maximum values of the single breaths is calculated. (2) If this time interval is above a certain threshold, this breath gesture is defined as a gesture with a long time interval. If this time interval is less than a certain threshold, this breath gesture is defined as a gesture with a short time interval. In the evaluation described later, the certain threshold is set to one second.

3. Recognition of gestures with different breath intensities: This algorithm recognizes breath gestures based on the change in the intensity of successive breaths as shown in Figure 4 (C). For example, this algorithm distinguishes three patterns of a breath gesture consisting of a strong breath followed by a weak breath, a breath gesture consisting of a weak breath followed by a strong breath, and a breath gesture consisting of successive breathing

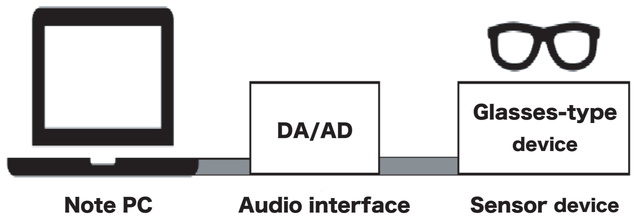

Fig. 5. System configuration[27]

with no difference in intensity. It is recognized from the following steps. (1) If two single breaths as a gesture are detected in a certain time range, the quotient of the maximum value of the next breath divided by the maximum value of the previous breath is calculated. (2) If the quotient is above or less than the threshold, the gesture is defined to be successive breaths with different intensity values. In the evaluation described later, the threshold was set to the following. If the quotient was 1.8 or higher, this breath gesture is determined that the order was a strong breath followed by a weak breath. If the quotient was 5/9 or lower, this breath gesture is determined that the order was a weak breath followed by a strong breath. Otherwise, this breath gesture is defined as one consisting of successive breathing with no difference in intensity.

4. Implementation

4.1. *Hardware*

The implemented prototype system consists of a glasses-type device with two piezoelectric elements placed on the left and right nose pads, an audio interface (Steinberg UR22mk), a laptop (Lenovo ThinkPad X1 Carbon), and software indicated in Subsection 4.2. The system configuration is shown in Figure 5. The glasses-type device and a laptop are connected via an audio interface. The audio interface is used as an analog to digital converter to convert the vibration obtained from the piezoelectric element into digital data. The piezoelectric element (FGT-15T-6.0A1W40, UNIVERSAL(CHANGZHOU) ELECTRONICS CO.) has a diameter of 15 mm and a thickness of 0.3 mm or less. The conductor in the audio cable and the piezoelectric element are connected; the piezoelectric element is installed on both nose pads, and the cable is run along with the frame of the glasses. Two piezoelectric elements were used to respond to changes in breath volume due to alternating nasal congestion. The glasses-type device is shown in Figure 1 (B).

4.2. *Software*

We used two types of software in this study. The first is Audacity(made from the Audacity Team), which is audio editing software for acquiring audio from the device, and the second is software that recognizes processing processes and breathing gestures. This software was implemented in Python language.

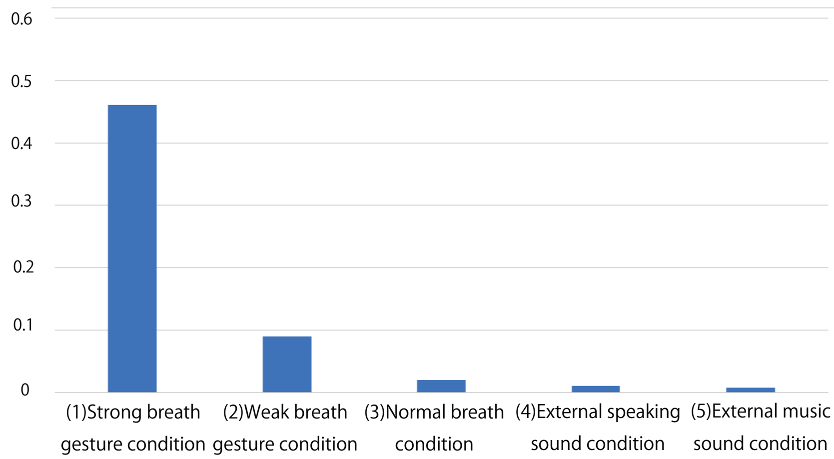

Fig. 6. Preliminary Experiment. The volume for five conditions obtained from the implementation system

4.3. *Preliminary Experiment*

We conducted a simple preliminary experiment to verify whether the volume of nasal breathing gestures could be obtained from the implemented system. In this preliminary experiment, the volume for five conditions was obtained from the implementation system by one subject. The results of the volume for the following five conditions are shown in Figure 6. (1) Strong breath gesture condition: This is the average of the volume of three trials of a single strong breath gesture. (2) Weak breath gesture condition: This is the average value of the volume of three trials of a single weak breath gesture. (3) Normal breath condition: This is the volume of normal breathing. This is the average of 10 trials of normal breathing that are not breathing gestures. (4) External speaking sound condition: This is the average value of the volume when the speaker plays a human conversation voice of two people talking on the radio for 60 seconds. The volume of the speaker was set to be 80db near the device. (5) External music sound condition: This is the average value of the volume when the speaker plays popular music for 60 seconds. The volume of the speaker was set to be 80db near the device.

The evaluation results indicated that (1) the volume of the nasal breathing gesture can be obtained from the implementation system, and (2) the volume of the nasal breathing gesture obtained from the implementation system can be distinguished from other sounds, such as external sounds and normal breathing sounds. We consider that sounds other than the nasal breathing gesture are unlikely to cause the proposed method to malfunction accidentally.

5. Evaluation 1

We verified the effectiveness of the proposed method to recognize breath gestures. The number of subjects was 10 (nine males and one female). They were Asian college students, and the average age was 22 years (max: 23 years, min: 19 years).

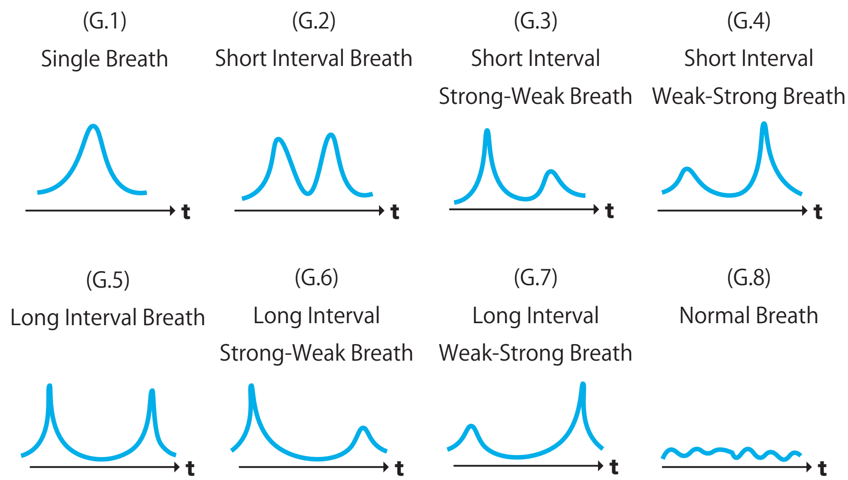

Fig. 7. Types of Breath Gestures[27]

5.1. *Breath gesture types*

We prepared eight types of nasal breath gestures as shown in Figure 7. This time, in order to confirm which breath element is recognizable and suitable for gestures, we constructed it using three elements: breath frequency, breath time interval, and breath intensity. Eight types of nasal breath gestures are shown in Figure 7. G.1(Single Breath) is a gesture defined when the subject breathes once. G.2((Short Interval Breath) is a gesture defined as when the subject breathes twice with similar intensity and short intervals. G.3(Short Interval Strong-Weak Breath) and G.4(Short Interval Weak-Strong Breath)) are gestures defined when the subject breathes twice with different intensities and short intervals. G.3 consists of a strong breath followed by a weak breath, and G.4 consists of a weak breath followed by a strong breath. G.5(Long Interval Breath)) is a gesture defined when the subject breathes twice with similar intensity and long intervals. G.6(Long Interval Strong-Weak Breath) and G.7(Long Interval Weak-Strong Breath) are gestures defined when the subject breathes twice with different intensities and long intervals. G.6 consists of a strong breath followed by a weak breath, and G.7 consists of a weak breath followed by a strong breath. G.8(Normal Breath) is normal nasal breathing, which is not a breath gesture. The threshold of the time interval between gestures consisting of short time intervals (G.2, G.3, G.4) and gestures consisting of long time intervals (G.5, G.6, G.7) was 1 second. These breath gesture terms were explained to the subjects during the preparatory phase of the experimental procedure described later.

5.2. *Experimental procedure*

First, the gestures to be performed were explained. Specifically, the subjects practiced the breath gestures with the explanation and guidance of the experimental supervisor. At this time, the subjects listened to the sound of the breath gesture that was prerecorded by the experimental supervisor as an example of the model gesture to imitate. Practice time was set to a maximum of 30 seconds per gesture. Next, the procedure was the following. First, the subjects performed 10 trials of normal nasal breathing (i.e., G.8) for six seconds. Next,

Table 1. Evaluation 1. Classification result. (R: Recall, P: Precision, F: F-value)

subjects	R	P	F
Sub.1	0.94	0.94	0.94
Sub.2	0.38	0.35	0.31
Sub.3	0.80	0.82	0.79
Sub.4	0.92	0.93	0.92
Sub.5	0.89	0.90	0.89
Sub.6	0.86	0.87	0.86
Sub.7	0.97	0.98	0.97
Sub.8	0.86	0.87	0.86
Sub.9	0.89	0.90	0.89
Sub.10	0.81	0.82	0.81
Average	0.83	0.84	0.82

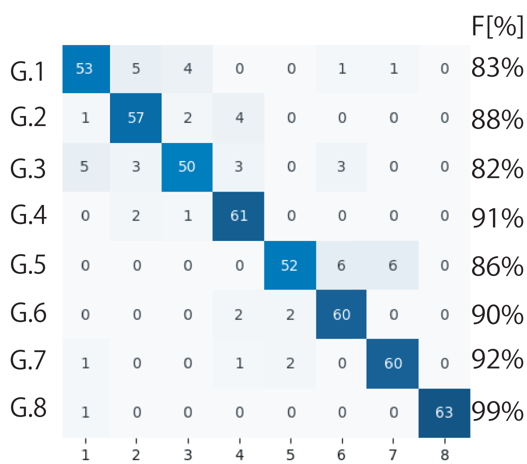

Fig. 8. Evaluation 1: Confusion matrix

the subjects performed 10 trials of each gesture from G.1 to G.7. At this time, the subjects were seated on a chair. Finally, we conducted an interview survey of the gestures that were difficult to perform.

The data set consisted of eight types of gestures for 10 trials, and thus the number of breath gesture samples was 80 for each subject. Two of the 10 trials were used as training data to determine the threshold of normal nasal breathing and breath gestures. Then, the recognition accuracy of the proposed method was evaluated for the data of eight trials. The time width of the data for G.8 was set to six seconds. The time width of the data of the other gestures was set to five seconds because each gesture would be completed within two seconds.

5.3. *Results*

The recognition results for each subject are shown in Table 1, which shows the F-value, Precision, and Recall. Figure 8 shows the confusion matrix and the F-values for each gesture.

Our recognition algorithm achieved an F-value of 82% on average for all subjects. This result indicates that the nasal breathing gesture can be recognized using the proposed method.

In addition, since the accuracy of normal breathing in G.8 is high, it is considered that gesture breathing and normal breathing can be distinguished.

The results of Figure 8 showed that some of the gestures were difficult to perform. First, short time interval breaths tend to be less accurate than long time interval breaths. This tendency can be seen from the following results. 90% of F-values of G.6(Long Interval Strong-Weak Breath) was higher than 82% of F-values of G.3(Short Interval Strong-Weak Breath). In addition, 92% of F-values of G.7(Long Interval Weak-Strong Breath) was higher than 91% of F-values of G.4(Short Interval Weak-Strong Breath). Note that 86% of F-values of G.5(Long Interval Breath) was not higher than 88% of F-values of G.2(Short Interval Breath), which was not the case for the trend of higher F-value for longer interval breath gestures. It seems that breath gestures with long time intervals are easier to perform. Second, the order of strong breaths and weak breaths tends to be less accurate than the order of weak breaths and strong breaths. This tendency can be seen from the following results. 91% of F-values of G.4(Short Interval Weak-Strong Breath) was higher than 82% of F-values of G.3(Short Interval Strong-Weak Breath). 92% of F-values of G.7(Long Interval Weak-Strong Breath) was higher than 90% of F-values of G.6(Long Interval Strong-Weak Breath). It seems that breath gestures with the order of weak breaths and strong breaths are easier to perform. From this tendency, the reason for the significantly lower accuracy of G.3 could be the combination of short time intervals and the order of strong and weak breathing. In fact, some subjects commented that G.3 was the most difficult to perform.

6. Evaluation 2

We evaluated the recognition accuracy of the proposed method under multiple conditions that may reduce the recognition accuracy. This time, we adopted two conditions of the reattachment condition and the body vibration condition. The number of subjects was 10, and they were the same as in Evaluation 1.

6.1. *Condition 1. Reattachment condition*

This condition evaluated the accuracy of the proposed method due to the deviation of the device reattachment. From this condition, it can be seen whether it is necessary to reacquire the learning data when using the proposed method after reattaching the sensor device.

Experimental procedure and evaluation environment: The subjects were those who participated in Evaluation 1 and acquired data once. In Evaluation 2, reattachment of the device was done twice, and data was acquired twice. The procedure was the following. After Evaluation 1, subjects removed the device and reattach it. Then, the subjects performed 5 trials of each gesture. Then, the subjects reattached the device again. Then, the subjects performed 5 trials of each gesture. As described above, reattachment was performed twice, and data for a total of 10 trials were obtained. During this experiment, the subject sat on a chair.

In the verification, since the subjects were the same as in Evaluation 1, the threshold of gesture recognition for each subject was the same as that in Evaluation 1. In other words, the threshold was created from the data of Evaluation 1. Then, the recognition accuracy of the proposed method was evaluated for the data of 10 trials of Evaluation 2.

Table 2. Evaluation 2. Classification result. (R: Recall, P: Precision, F: F-value)

	Reattachment			Body vibration		
	R	P	F	R	P	F
Sub.1	0.89	0.86	0.87	0.95	0.95	0.95
Sub.2	0.62	0.67	0.57	0.63	0.63	0.61
Sub.3	0.76	0.84	0.75	0.78	0.78	0.75
Sub.4	0.85	0.90	0.84	0.88	0.90	0.87
Sub.5	0.97	0.97	0.96	0.90	0.92	0.90
Sub.6	0.73	0.80	0.76	0.58	0.46	0.51
Sub.7	0.94	0.95	0.93	0.73	0.78	0.70
Sub.8	0.92	0.94	0.91	0.75	0.78	0.76
Sub.9	0.93	0.94	0.93	0.93	0.94	0.92
Sub.10	0.94	0.95	0.94	0.88	0.91	0.87
Average	0.86	0.88	0.84	0.80	0.81	0.78

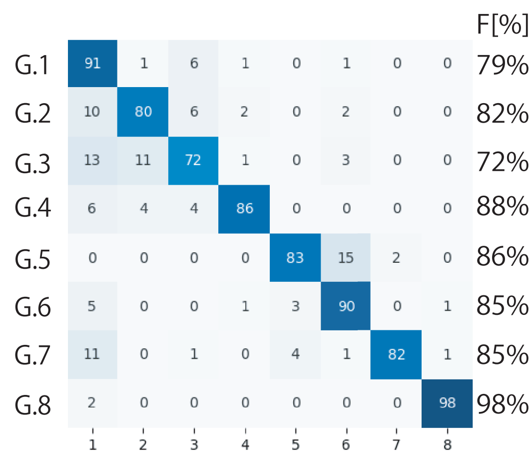

Fig. 9. Evaluation 2: Confusion matrix (Reattachment condition)

6.2. *Condition 2. Body vibration condition*

This condition evaluated the change in the accuracy of the proposed method due to body vibration. From this condition, it can be seen whether the learning data at rest can be used when using the proposed method in a scene where body vibration occurs (e.g., walking).

Experimental procedure and evaluation environment: The subjects were those who participated in Evaluation 1 and acquired data once. In Experiment 2, data was acquired while walking.

The procedure was the following. Similar to Evaluation 1, we explained each gesture before the experiment, and the device was attached. Then, each gesture was performed 5 times while walking on the spot (i.e., stepping). The subjects replicated their normal walking speed. As described above, data for 5 trials was obtained. Since the subjects were the same as in Evaluation 1, the threshold of gesture recognition for each subject was the same as that in Evaluation 1. Then, the recognition accuracy of the proposed method was evaluated for the data of 5 trials while walking.

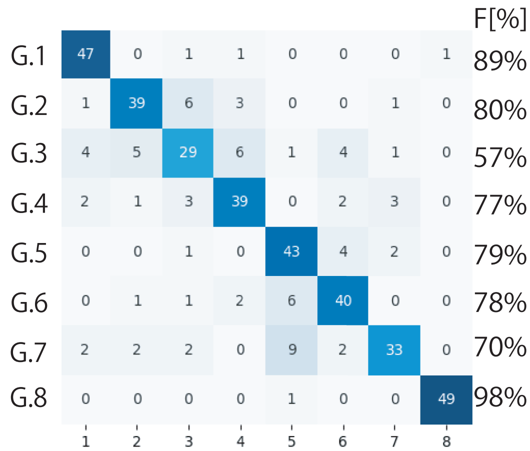

Fig. 10. Evaluation 2: Confusion matrix (Body vibration condition)

6.3. *Results*

First, the result of the reattachment condition is described. The recognition results for each subject are shown on the left of the table 2. The table shows the Recall, Precision, and F-value. Figure 9 shows the confusion matrix for all subjects. The proposed method achieved an average F-value of 84% in all subjects. Although the accuracy in each subject was slightly different from the result of Evaluation 1, the difference from Evaluation 1 was small in the overall average accuracy. This result shows that the proposed method can recognize the nasal breathing gesture even if there is a deviation due to reattachment.

Next, the results of the body vibration condition are described. The recognition results for each subject are shown on the right side of Table 2. Table 2 shows the Recall, Precision, and F-value. Figure 10 shows the confusion matrix for all subjects. The proposed method achieved an average F-value of 78% in all subjects. This indicates that the proposed method can be used even while the vibration of the walking level is occurring. However, the F-value is lower than that of Evaluation 1, which is conducted in the sitting state. We considered that this reason is the difficulty of changing the breathing rhythm while moving the legs. For the gestures, the F-values of the gestures of similar intensity (G.2 and G.5) were relatively higher than those of the other evaluation experiments. These results also suggest that when the body is moving, such as walking, changing the rhythm and intensity of breathing is more difficult than sitting. In fact, there was an opinion from the subjects that it is difficult to breathe while moving the legs.

The results of Figure 9 of Reattachment condition showed that some of the gestures were difficult to perform. As in Evaluation 1, short time interval breaths tend to be less accurate than long time interval breaths. This tendency can be seen from the following results. 86% of F-values of G.5 (Long Interval Breath) was higher than 82% of F-values of G.2 (Short Interval Breath). In addition, 85% of F-values of G.6 (Long Interval Strong-Weak Breath) was higher than 72% of F-values of G.3 (Short Interval Strong-Weak Breath). Note that 85% of F-values of G.7 (Long Interval Weak-Strong Breath) was not higher than 88% of F-values of G.4 (Short Interval Weak-Strong Breath), which was not the case for the trend of higher

F-value for longer interval breath gestures. It seems that breath gestures with long time intervals are easier to perform. As in Evaluation 1, the order of strong breaths and weak breaths tends to be less accurate than the order of weak breaths and strong breaths. This tendency can be seen from the following results. 88% of F-values of G.4 (Short Interval Weak-Strong Breath) was higher than 72% of F-values of G.3 (Short Interval Strong-Weak Breath). Note that 85% of F-values of G.7 (Long Interval Weak-Strong Breath) was equal to 85% of F-values of G.6 (Long Interval Strong-Weak Breath). It seems that breath gestures with the order of weak breaths and strong breaths are easier to perform. Also, as with Evaluation 1, G.3 was the most difficult to recognize.

7. Evaluation 3

One way to improve the accuracy of the proposed method is to use only gestures that are easy to perform. In this evaluation experiment, eight types of breathing gestures were prepared, but since hands-free input of various applications is possible with about five types of gestures, evaluation using five or six types of gestures in the previous research [26] is being performed. Therefore, we evaluated whether the recognition accuracy of the proposed method can improve by limiting the breath gestures.

7.1. *Gesture patterns*

We prepared the three sets of gestures based on three elements of breath counts, intensity, and interval. (1)The first gesture set is three types of gestures consisting of G.1(Single Breath), G.2(Short Interval Breath), and G.8(Normal Breath). Figure 11 shows the first gesture set. These were selected based on breath counts. These recognize the difference in breath counts of zero, one, and two times. (2)The second gesture set is five types of gestures consisting of G.1, G.5(Long Interval Breath), G.6(Long Interval Strong-Weak Breath), G.7(Long Interval Weak-Strong Breath), and G.8. Figure 12 shows the second gesture set. These were selected based on breath intensity. These recognize three types of differences in breath intensity: no difference in intensity, order of strong and weak, and order of weak and strong. G.1 and G.8 were included regardless of the perspective of breath intensity. We adopted breath gestures with a long breathing interval because we considered that breath gestures with a long breathing interval tend to be easier than those with a short breathing interval from Evaluation 1 and Evaluation 2. (3)The third gesture set is four types of gestures consisting of G.1, G.2, G.5, and G.8. Figure 13 shows the third gesture set. These were selected based on breath intervals. G.1 and G.8 were included regardless of the perspective of breath intervals. These recognize the difference between two types of breath intervals: breathing with a long breathing interval and breathing with a short breathing interval. The evaluation is performed for each set. The data-set was the same as in Evaluation 1 and Evaluation 2.

7.2. *Results*

First, the results of the first gesture set about breath count are described. Note that the result of a resting state condition used the data from Evaluation 1. The recognition results for each subject are shown in Table 3. The table shows the Recall, Precision, and F-value. The average F-value was 92% under the resting state condition, 94% under the reattachment state

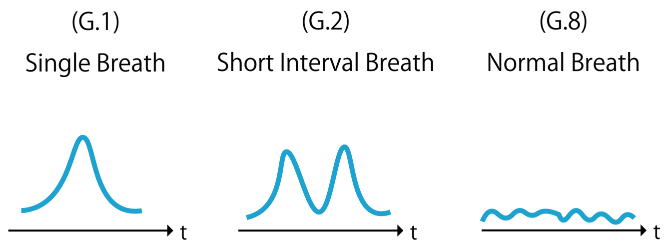

Fig. 11. The first gesture set

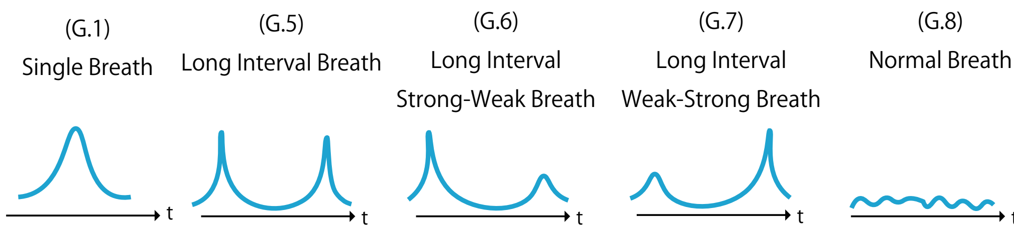

Fig. 12. The second gesture set

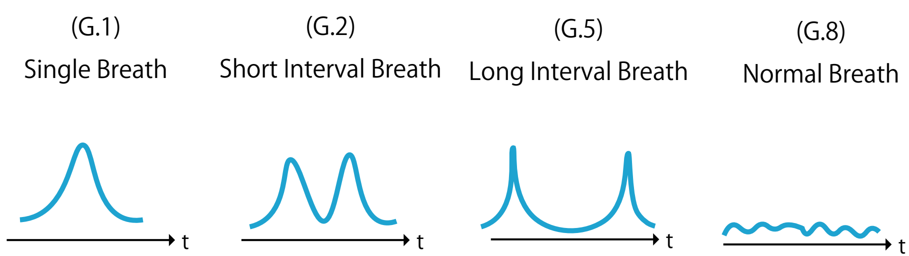

Fig. 13. The third gesture set

Table 3. Evaluation 3. Classification result of the first gesture set about breath counts. (R: Recall, P: Precision, F: F-value)

	Resting state			Reattachment			Body vibration		
	R	P	F	R	P	F	R	P	F
Sub.1	1.00	1.00	1.00	0.93	0.94	0.93	1.00	1.00	1.00
Sub.2	0.75	0.80	0.74	0.93	0.94	0.93	0.40	0.25	0.28
Sub.3	0.79	0.87	0.77	0.97	0.97	0.97	0.87	0.90	0.86
Sub.4	1.00	1.00	1.00	0.93	0.94	0.93	1.00	1.00	1.00
Sub.5	0.96	0.96	0.96	1.00	1.00	1.00	0.93	0.94	0.93
Sub.6	1.00	1.00	1.00	0.90	0.90	0.90	0.53	0.65	0.48
Sub.7	1.00	1.00	1.00	0.87	0.90	0.86	0.87	0.90	0.87
Sub.8	0.96	0.96	0.96	0.90	0.90	0.90	0.93	0.94	0.93
Sub.9	0.75	0.86	0.71	1.00	1.00	1.00	1.00	1.00	1.00
Sub.10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Average	0.92	0.92	0.92	0.94	0.95	0.94	0.85	0.87	0.85

Table 4. Evaluation 3. Classification result of the second gesture set about breath intensities. (R: Recall, P: Precision, F: F-value)

	Resting state			Reattachment			Body vibration		
	R	P	F	R	P	F	R	P	F
Sub.1	0.93	0.93	0.92	0.96	0.96	0.96	0.84	0.91	0.83
Sub.2	0.50	0.52	0.44	0.66	0.60	0.59	0.28	0.19	0.19
Sub.3	0.73	0.75	0.72	0.60	0.50	0.51	0.88	0.93	0.88
Sub.4	0.95	0.95	0.95	0.66	0.80	0.65	0.52	0.42	0.44
Sub.5	0.93	0.93	0.92	0.70	0.85	0.66	0.92	0.94	0.92
Sub.6	0.58	0.45	0.48	0.54	0.44	0.46	0.48	0.34	0.39
Sub.7	0.70	0.81	0.70	0.58	0.47	0.49	0.64	0.58	0.56
Sub.8	0.78	0.77	0.76	0.94	0.95	0.94	0.48	0.40	0.40
Sub.9	0.93	0.94	0.92	0.66	0.67	0.60	0.68	0.88	0.67
Sub.10	0.80	0.81	0.80	0.88	0.90	0.88	0.68	0.76	0.66
Average	0.78	0.80	0.78	0.72	0.80	0.70	0.64	0.75	0.62

Table 5. Evaluation 3. Classification result of the third gesture set about breath time intervals. (R: Recall, P: Precision, F: F-value)

	Resting state			Reattachment			Body vibration		
	R	P	F	R	P	F	R	P	F
Sub.1	1.00	1.00	1.00	0.95	0.96	0.95	1.00	1.00	1.00
Sub.2	0.72	0.82	0.73	0.95	0.96	0.95	0.35	0.39	0.27
Sub.3	0.84	0.90	0.83	0.98	0.98	0.98	0.90	0.93	0.90
Sub.4	1.00	1.00	1.00	0.93	0.94	0.93	1.00	1.00	1.00
Sub.5	0.97	0.97	0.97	1.00	1.00	1.00	0.95	0.96	0.95
Sub.6	1.00	1.00	1.00	0.93	0.92	0.92	0.65	0.74	0.61
Sub.7	0.97	0.97	0.97	0.90	0.93	0.90	0.90	0.93	0.90
Sub.8	0.97	0.97	0.97	0.93	0.93	0.93	0.80	0.86	0.79
Sub.9	0.81	0.89	0.78	1.00	1.00	1.00	0.95	0.96	0.95
Sub.10	1.00	1.00	1.00	1.00	1.00	1.00	0.95	0.96	0.95
Average	0.93	0.93	0.93	0.96	0.96	0.96	0.85	0.87	0.85

condition, and 85% under the body vibration condition. These average F-values increased about 10% compared to those in Evaluation 1 and Evaluation 2 that use eight gestures. These results showed that the recognition accuracy of the proposed method is improved by limiting gestures to those with breath count.

Secondly, the results of the second gesture set about breath intensity are described. The recognition results for each subject are shown in Table 4. The table shows the Recall, Precision, and F-value. The average F-value was 78% under the resting state condition, 70% under the reattachment state condition, and 62% under the body vibration condition. These average F-value decreased about 5 to 10% compared to those in Evaluation 1 and Evaluation 2 that use eight gestures. These results showed that the recognition accuracy of the proposed method decreased by limiting gestures to those with different breath intensity.

Thirdly, the results of the second gesture set about the breath interval are described. The recognition results for each subject are shown in Table 5. The table shows the Recall, Precision, and F-value. The average F-value was 93% under the resting state condition, 95%

under the reattachment state condition, and 85% under the body vibration condition. These average F-values increased about 10% compared to those in Evaluation 1 and Evaluation 2 that use eight gestures. These results showed that the recognition accuracy of the proposed method increased by limiting gestures to those with different breath intervals.

These results showed that the recognition accuracy is increased by limiting gestures to those with a different breath count or different breath interval. These also indicate that the proposed method can recognize those two factors. On the other hand, these results showed the recognition accuracy is decreased by limiting gestures to those with the breath intensity. This also indicates that it is difficult for the proposed method to recognize the different breath intensities. From these results, we considered that gestures with different breath intensities would be less reproducible since they are difficult to perform. The breath count and the breath intervals can be explained for subjects with a clearer standard, such as once, twice, 0.5 seconds, and 1 second, compared with the breath intensity, such as strong breath or weak breath. Due to this point, it was difficult for the subjects to reproduce the gestures with different breath intensities.

8. Discussion and Future Work

The evaluation results showed that nasal breathing gestures can be recognized by using the piezoelectric element of the nasal pad of the glasses. We also confirmed the feasibility of the proposed method. The proposed method can be widely used for various wearable devices containing nose pads, such as eyeglasses for vision correction, VR head-mounted displays, smart glasses such as AR glasses. If technology that can easily recognize nasal breathing gestures using eyewear devices becomes widespread, it can be expected to be applied to new services using nasal breathing gestures, such as simple hands-free input and a reflection of nasal breathing motion on VR a avatar.

The evaluation results showed that a few subjects have difficulty using the proposed method. The recognition accuracy of Sub.2 tended to be lower than the other subjects. This is because Sub.2 was not good at the nasal breathing gesture because Sub.2 commented that "I am not good at breathing gestures as a whole." in the questionnaire. We considered that the skill of breathing control, such as adjusting the strength and interval of breath, is different for each individual depending on any experience such as sports. Therefore, we plan to investigate the difference between those who can use the proposed method and those who cannot.

Since our prototype device cannot adjust the position of the nose pad for each individual, some subjects commented that they felt uncomfortable with the position of the nose pad. We plan to design devices for installing sensors in appropriate positions for each individual in the future. Such a device is expected to obtain a clearer volume by adjusting appropriate positions of the nose pad and to improve the recognition accuracy of the proposed method.

Since the attributes and number of subjects in this experiment were limited, we plan to evaluate the proposed method for a larger number of subjects with greater diversity in the future. We consider that the effects of gender and nationality differences will be almost nonexistent, except for the device-related problems mentioned earlier. However, we consider that aging can affect breathing gestures due to any factor, such as the decrease in breathing volume that occurs with aging. In addition to aging, the recognition accuracy of the proposed method may be changed by some factors, such as health state (e.g., breath disease) and physical fitness. We plan to investigate these points in the future.

9. Conclusion

We proposed a simple hands-free input method using nasal breath information and wearable devices. We implemented a glasses-type device, using a piezoelectric element and designed eight types of nasal breath gestures using breath count, time interval, and intensity. The evaluation results for ten subjects showed that the proposed method can recognize eight types of nasal breath gestures at 82% of F-value. The evaluation results also showed that the recognition accuracy of the proposed method does not decrease with device reattachment but decreases with body vibration. The evaluation results also showed that the recognition accuracy is increased to more than 90% by limiting gestures to those with a different breath count or different breath interval. Our study provides the first wearable sensing technology that uses nasal breathing for hands-free input.

Acknowledgements

This research was supported in part by JSPS(Japan Society for the Promotion of Science) KAKENHI Grant Number JP19K20330.

References

1. Azh, Maryam and Zhao, Shengdong, LUI: Lip in Multimodal Mobile GUI Interaction, Proceedings of the 14th ACM international conference on Multimodal interaction, pp. 551–554.(2012).
2. Lyons, M. J., Chan, C. H., and Tetsutani, N, MouthType: text entry by hand and mouth, CHI EA '04: CHI '04 Extended Abstracts on Human Factors in Computing Systems, pp. 1383–1386 (2004).
3. Koguchi, Y., Oharada, K., Takagi, Y., Sawada, Y., Shizuki, B., and Takahashi, S, A Mobile Command Input Through Vowel Lip Shape Recognition, In International Conference on Human-Computer Interaction, pp. 297–305 (2018).
4. Masai, K., Kunze, K., Sakamoto, D., Sugiura, Y., and Sugimoto, M, Face Commands - User-Defined Facial Gestures for Smart Glasses, 2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), pp. 374–386 (2020).
5. Takuro Nakao, Yun Suen Pai, Megumi Isogai, Hideaki Kimata, and Kai Kunze, Make-a-face: a hands-free, non-intrusive device for tongue/mouth/cheek input using emg, In ACM SIGGRAPH 2018 Posters, pp. 1–2 (2018).
6. Augusto Esteves, Eduardo Velloso, Andreas Bulling, and Hans Gellersen, Orbits: Gaze Interaction for Smart Watches using Smooth Pursuit Eye Movements, In Proceedings of the 28th Annual ACM Symposium on User Interface Software and Technology, pp. 457–466 (2015).
7. Katsutoshi Masai, Kai Kunze, Yuta Sugiura, Masa Ogata, Masahiko Inami, and Maki Sugimoto, Evaluation of Facial Expression Recognition by a Smart Eyewear for Facial Direction Changes, Repeatability, and Positional Drift, ACM Transactions on Interactive Intelligent Systems (TiiS), 7(4), pp. 1–23 (2017).
8. Yan, Y., Yu, C., Yi, X., and Shi, Y, Headgesture: Hands-free input approach leveraging head movements for hmd devices, Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2(4), pp. 1–23 (2018).
9. Ando, T., Kubo, Y., Shizuki, B., and Takahashi, S, Canalsense: Face-related movement recognition system based on sensing air pressure in ear canals, In Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology, pp. 679–689 (2017).

10. Lutz, O. H. M., Venjakob, A. C., and Ruff, S, SMOOVS: Towards calibration-free text entry by gaze using smooth pursuit movements, Journal of Eye Movement Research, 8(1), (2015).
11. Amesaka, T., Watanabe, H., and Sugimoto, M, Facial expression recognition using ear canal transfer function, In Proceedings of the 23rd International Symposium on Wearable Computers, pp. 1-9 (2019).
12. Manabe, H., Fukumoto, M., and Yagi, T, Conductive rubber electrodes for earphone-based eye gesture input interface. Personal and Ubiquitous Computing, 19(1), pp. 143–154 (2015).
13. Matthies, D. J., Strecker, B. A., and Urban, B, Earfieldsensing: A novel in-ear electric field sensing to enrich wearable gesture input through facial expressions. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, pp. 1911–1922 (2017).
14. Taniguchi, K., Kondo, H., Kurosawa, M., and Nishikawa, A. Earable, TEMPO: a novel, hands-free input device that uses the movement of the tongue measured with a wearable ear sensor, Sensors, 18(3):733 (2018).
15. Ando, T., Kubo, Y., Shizuki, B., and Takahashi, S, Canalsense: Face-related movement recognition system based on sensing air pressure in ear canals, In Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology, pp. 679–689 (2017).
16. Hashimoto, T., Low, S., Fujita, K., Usumi, R., et al, TongueInput: Input Method by Tongue Gestures Using Optical Sensors Embedded in Mouthpiece. In 2018 57th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE), pp. 1219–1224 (2018).
17. Chauhan, J., Hu, Y., Seneviratne, S., Misra, A., Seneviratne, A., and Lee, Y, BreathPrint: Breathing acoustics-based user authentication, In Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services, pp. 278–291 (2017).
18. Shih, C. H., Tomita, N., Lukic, Y. X., Reguera, Á. H., Fleisch, E., and Kowatsch, T, Breeze: Smartphone-based acoustic real-time detection of breathing phases for a gamified biofeedback breath training, Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 3(4), pp. 1–30 (2019).
19. Tennent, P., Rowland, D., Marshall, J., Egglestone, S. R., Harrison, A., Jaime, Z., ... and Benford, S, Breathalising games: understanding the potential of breath control in game interfaces, In Proceedings of the 8th international conference on advances in computer entertainment technology, pp. 1–8 (2011).
20. Sra, M., Xu, X., and Maes, P, Breathvr: Leveraging breathing as a directly controlled interface for virtual reality games, In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, pp. 1–12 (2018).
21. Evreinov, G., and Evreinova, T, " Breath-Joystick"-Graphical Manipulator for Physically Disabled Users, Proc. of the ICCHP2000, pp. 193–200 (2000).
22. Yamamoto, M., Ikeda, T., and Sasaki, Y, Real-time analog input device using breath pressure for the operation of powered wheelchair, In 2008 IEEE International Conference on Robotics and Automation, pp. 3914–3919, (2008).
23. Marshall, J., Rowland, D., Rennick Egglestone, S., Benford, S., Walker, B., and McAuley, D, Breath control of amusement rides, In Proceedings of the SIGCHI conference on Human Factors in computing systems, pp. 73–82 (2011).
24. Héctor A. Cordourier Maruri, Paulo Lopez-Meyer, Jonathan Huang, Willem Marco Beltman, Lama Nachman, and Hong Lu, V-Speech: Noise-Robust Speech Capturing Glasses Using Vibration Sensors, Proc. ACM Interact. Mob. Wearable Ubiquitous Technol, 2(4), pp. 1–23 (2018).
25. He, Jibo and Chaparro, Alex and Nguyen, Bobby and Burge, Rondell and Crandall, Joseph and Chaparro, Barbara and Ni, Rui and Cao, Shi, Texting while driving: Is speech-based texting less risky than handheld texting, Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, pp. 124–130 (2013).
26. Amesaka, T., Watanabe, H., and Sugimoto, M, Facial expression recognition using ear canal transfer function, In Proceedings of the 23rd International Symposium on Wearable Computers, pp. 1–9 (2019).
27. Ryoma Ogawa, Kyosuke Futami, and Kazuya Murao, NasalBreathInput: A Hands-Free Input

Method by Nasal Breath Gestures using a Glasses Type Device, In The 23rd International Conference on Information Integration and Web Intelligence, pp. 620-–624 (2021).