Response to the reviewers’ comments

**Paper Number: 21-Q004e  
Authors: Atsuhiro Fujii, Kazuya MURAO  
Tile: User Identification Method based on Head Shape using Pressure Sensors embedded in a Helmet**

First of all, we would like to thank the reviewers again for their useful comments, suggestions, and criticisms. They have played a vital role in improving the paper quality. Below, we provide responses to each comment and the improvements done in the revised version of the paper. The revised parts in the paper are shown in red.

**Meta-Reviewer:**

[1] It is difficult to understand the contribution of the paper from the abstract because of the lack of the description of the proposed method. Please add the description about how to identify and authenticate the user briefly. [R2-1]

Answer: As the reviewer pointed out, the abstract lacks the description of the proposed method. The number of participants is nine. We obtained sensor data for 2 seconds 20 times. The accuracy was evaluated using 5-fold cross-validation, and we achieved 100% accuracy with five sensors and 92% with two sensors for user identification and an average EER of 0.076 with 32 sensors for user authentication. We added the description of the proposed method and evaluation experiment in the abstract. **Please see the abstract.**

[2] Please describe the evaluation environment clearly and discuss the robustness to the individual and the environmental differences. It is expected that size/shape/hairstyle of the subject affects the accuracy of identification and authentication. The size/shape/weight of the helmet may also affect the performance. [R1-1][R2-4]

Answer: The description of the evaluation environment for helmets and users was insufficient. First of all, the nine subjects were all males with an average head shape of about 60 cm in circumference (around the temples). **Photographs of the nine subjects' heads were added to the paper (Fig. 10).** If the subjects have a distinctive hairstyle and head shape, the sensor data will be distinctive and the performance will be improved. However, we cannot deny the influence of hair growth and haircut. Investigation of the effect of hairstyle will be a future work. **We have added the information about the subjects in Section 4.1, and the discussion of the effects of different head shapes and hairstyles on the performance of the proposed method in Section 5.1.**

As for the helmet, the shape is 260 mm in height, 213 mm in width, and 282mm in depth. **We added a figure to clarify the length of the helmet (Fig. 9).** The weight of the helmet is 1456 g. The helmet used is an inexpensive general helmet. In this paper, one full-face helmet was used, and we distinguish between full-face and half-helmets by the position of the sensor used. **The information of the helmet is added in Section 3.2.**

As for the relationship between helmets and wearers, when multiple people share multiple helmets (e.g., 10 people use 10 helmets freely), if the 10 helmets have the same shape and the same sensor arrangement, we can collect data for each wearer with any one helmet. However, if the helmets are different, it is necessary to collect data from all helmets. If one person occupies one helmet, such as a motorcycle helmet, the system only needs to collect data from that helmet. In other words, different helmets cannot share the training data, which is a limitation of the proposed method. **This point is described in Section 5.1.**

As for R2-4, the sex of all 9 subjects was male. Regarding the point that some sensors do not respond, there are some sensors that do not respond due to lack of pressure depending on the shape of the head. **We have added a time series of 32 sensor data for subjects A and B as Figure 11.** The figure shows the voltage value output by each sensor, where the pressure is zero at 5V and increases as it approaches 0V. As you can see from the figure, more than 10 sensors responded to both subjects, and the evaluation results showed that about 4 of them had high enough discrimination performance. Therefore, we do not have a situation where only 4 sensors are responding as the reviewer is concerned.

In this experiment, all subjects used the same helmet to collect data. In the case of shared helmets, fitting is generally not performed; in the case of multiple sizes such as S/M/L, it is necessary to register data with the helmet of the size used by the subject. In the case of a personal helmet, we should prepare helmets that fit each wearer individually. However, in this experiment, we were only able to evaluate the proposed method with one type of helmet. If we used the most suitable helmet for each subject, we can obtain accurate data of the head shape of a registered user, while we cannot obtain accurate data of the head shape of a non-registered user whose head size does not fit the helmet (i.e., many sensors are tightly fitting the head or not fitting the head). Experiments will be done in the future. **A discussion on helmet fitting is given in Section 5.2.**

[3] Please add the discussion about the effect of the sensor position to the performance. Also, the performance difference between full helmet and half helmet should be described. (Generally, half helmet is more suitable for many scenes.) [R1-2][R2-3]

Answer: With respect to the comment [R1-2], inner shape of the helmet and sensor positions would affect the performance, therefore sharing the machine learning model among multiple types of helmet is difficult. If the user uses multiple helmets, we assume the environment that the helmet and sensor positions are identical. If the user uses different types of helmet, training the model of the helmets one by one is needed.**This point is discussed in section 5.1.**

With respect to the commnet [R2-3], both full-face and half-face helmets are used depending on the situation, so we tried both. As the reviewer pointed out, half-helmets are used more often than full-helmets, but full-helmets are used for riding motorcycles and other vehicles, and full-helmets can be equipped with more sensors. As a result, we confirmed that the same accuracy as that of a full-face helmet could be obtained with only the sensors installed in a half-helmet. Tables 1 and 2 show the accuracies of the full-face and half-helmet models, with the half-helmet model showing slightly higher accuracies in some cases, but this is probably due to the random division of the cross validation set, which resulted in slight differences in accuracy between the full-face and half-helmet models. **The usage of full-face helmets is discussed in Section 1, and discussed in Section 4.2.2.**

As for sharing recognition models with other helmets, as we answered in [2], if helmets differ in type, shape, and sensor placement, we need to build a model for each helmet. **This point is discussed in section 5.1.**

The thickness of the urethane sponge is 20 mm. The size of the helmet is 260 mm in height, 213 mm in width, and 282 mm in depth, as described in [2]. The weight of the helmet is 1456 g. **We added the length of each part to Figure 9.**

The sensors 26-31 are placed around the cheeks. Although they are in contact with the face, they do not adhere strongly to each other for some people, so the effect is small. For the sensors 20-25, The sensors 24 and 25 are shown as effective sensors in Table 1. In fact, no difference in performance exists between 20-25, but they may have been selected because they were particularly close to the subject’s head. **This point is discussed in section 4.2.2.**

As for **Figure 3, we've included photos taken from both sides of the helmet** so that both sides can be clearly seen. **The caption was changed to “Appearance of the prototype device”**. In **Fig. 6, the sensor numbers 20-23 were difficult to identify, so the sensor numbers were added** to the illustrations on both sides. In order to clarify the difference between the sensors used in half-helmets and full-face helmets, **the numbers in Fig. 6 have been color-coded.**

[4] Please discuss the limitation of the proposed method. Is the method applicable to all gender and ages? Also, please make the target and the restriction of the proposed method clear. [R1-3][R2-2][R2-5]

Answer：With respect to the comment R1-3, in this experiment, only nine males in their twenties were tested, and it was not possible to investigate whether the results obtained would be applicable to female users and users of a wide range of ages. The environment assumed in this study is a relatively high-risk work environment where users use helmets, and it is thought that there are many male users. Therefore, we conducted the experiment with only male subjects. However, since the proposed method is not limited to males, we would like to conduct evaluation experiments with a wider range of subjects such as women, children, and the elderly as future work. **We have added the description on this point to Section 5.1.**

In reference provided by R2-2, although the camera-based method shows high performance, it may be affected by water droplets and dust. In addition, existing authentication methods (e.g. PIN, vein, etc.) require an action for authentication, but a helmet is superior in that it only needs to be worn. Bone conduction and sound-based methods are also affected by vibration and environmental noise, so pressure is also considered to be advantageous in this respect. We believe that it is also difficult to apply authentication using movements such as taking out a smartphone to wearing a helmet due to its low reproducibility. This is not a criticism of existing research, but rather an assertion that it may not be possible to operate stably in an environment with water droplets, dust, vibration, and noise, and that an approach that uses a pressure sensor to acquire head shape may solve this problem. In this paper, we propose an approach that uses a pressure sensor to acquire the head shape. **We have revised the description in Section 2.1 on this point.**

As for the comment R2-2 about the GPS, we assume that it can be used outdoors.

As for the comment about the camera at the mouth in R2-2, a small camera can be embedded in the mouth area of the helmet, and it can be illuminated in the dark. However, one of the weaknesses of using a camera is the noise in the captured image caused by water droplets and dust, so I would argue that this is a problem. **We have revised the description in Section 2.1 on this point, and deleted the term “using in the dark”.**

As for the comment R2-2 about fingerprint authentication, it is true that fingerprint authentication is widely used even though its vulnerability has been pointed out. Our argument is that fingerprint authentication, which is not necessary for using a helmet, is itself cumbersome, and that authentication by simply putting on a helmet without thinking about it is more advantageous. **We have revised the description in Section 2.1 on this point.**

As for the comment R2-2 about the various situations in which we wear a helmet, we think that the helmet-wearing motion is not highly reproducible. However, the helmet is worn with one hand, two hands, or in different directions, and there are many different motions between picking up the helmet and putting it on. Therefore, various training data are necessary to recognize individuals from their helmet-wearing behavior. **We have revised the description in Section 2.1 on this point.**

As for the comparison with the existing studies using pressure sensors [16]~[19] in comment R2-2, **we compared the accuracy in [19] with the proposed method in Section 4.3.2** because only [19] showed EER in the literature.

As for the research on facial expression recognition in comment R2-2, as the reviewer pointed out, it was out of the scope of this research, so **we deleted it.**

As for the comment R2-2 about EEG, **we deleted the aforementioned research** on facial expression recognition, so there will be no mention of EEG in the revised paper.

In the paper by Zhuang et al. presented in R2-5, they reported that they could classify 1,169 people with 90% accuracy using 50 data points obtained from 3D head data. Therefore, if we use a large number of pressure sensors, we may be able to classify nearly 1,000 people. According to the following documents in Japanese, since the number of people working at a large construction site is about 1,000, the proposed method can be applied to a wide range of environments if 32 or more sensors are used. **We have added the description on this point to Section 5.3.**

みんなの建設業Q&A50, The associated general contractors of tokyo, p. 85, <http://www.token.or.jp/book/pdf/book130919_05.pdf>

福島第一原子力発電所廃炉作業の至近の状況について, TEPCO, p. 15, <https://www.tepco.co.jp/decommission/information/committee/kenminkaigi/pdf/2018/k180903_04-j.pdf>

With respect to the comment R2-5, for comparison with existing authentication technologies, a **comparison with the accuracy of general fingerprint authentication is described in Section 5.4.**

[5] Please discuss how to determine the threshold of user authentication. The FRR/FAR crossing points are slightly different between subjects in Fig.9. In the actual scene, it seems to be difficult to set the threshold automatically to the optimal value.

Answer: We believe that the determination of the threshold value depends on the application. Stricter thresholds reduce the probability of successful access by others, but also increase the probability that the correct user will not be authenticated. On the other hand, if the threshold is loosened, the probability that the correct user will be rejected will be reduced, and the stress will decrease, but the probability that others will be able to break through will also increase. In general, we use a threshold where the values of FAR and FRR are the same. If data of multiple users is obtained in advance in the development stage, the threshold value that FAR=FRR can be calculated when the helmet user's data is registered multiple times. If you want to reject other people almost without fail, you can use a stricter threshold value, even if it reduces the usability of the system. **The method for determining the threshold is described in Section 5.5.**