CanalSense: Face-Related Movement Recognition System based on Sensing Air Pressure in Ear Canals

Toshiyuki Ando, Yuki Kubo, Buntarou Shizuki, and Shin Takahashi
University of Tsukuba
Tennodai 1-1-1, Tsukuba, Ibaraki 305-8573, Japan
{ando,kubo,shizuki,shin}@iplab.cs.tsukuba.ac.jp

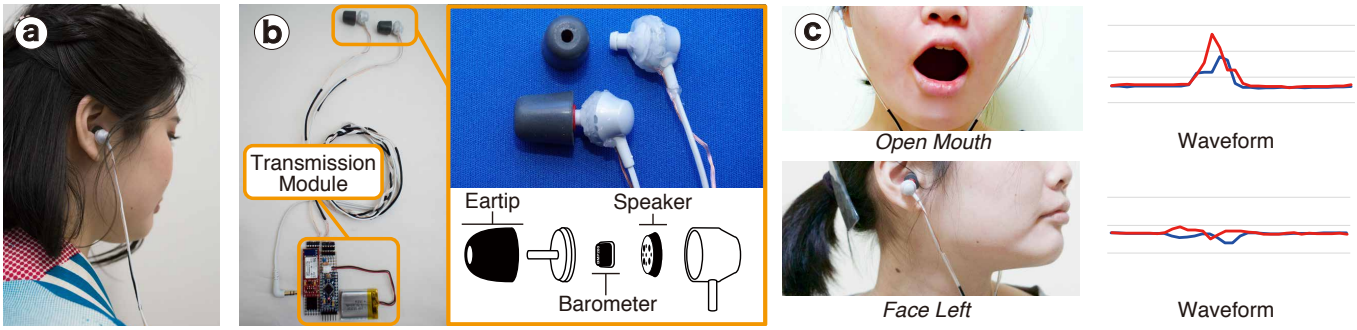

Figure 1: CanalSense. a) A user can use this system simply by wearing its earphone-type barometers in the same manner as wearing canal-type earphones. b) Earphone-type barometers of CanalSense, each of which consists of a canal-type earphone and a small embedded barometer and a transmission module. c) Two examples of face-related movements and the observed air pressure values inside the ear canals during the movements.

ABSTRACT

We present a jaw, face, or head movement (face-related movement) recognition system called CanalSense. It recognizes face-related movements using barometers embedded in earphones. We find that face-related movements change air pressure inside the ear canals, which shows characteristic changes depending on the type and degree of the movement. We also find that such characteristic changes can be used to recognize face-related movements. We conduct an experiment to measure the accuracy of recognition. As a result, random forest shows per-user recognition accuracies of 87.6% for eleven face-related movements and 87.5% for four *Open Mouth* levels.

ACM Classification Keywords

H.5.2. Information Interfaces and Presentation (e.g. HCI); User Interfaces –Input devices and strategy.

UIST 2017, October 22–25, 2017, Quebec City, QC, Canada

DOI: https://doi.org/10.1145/3126594.3126649

Author Keywords

Jaw movement; mouth movement; facial movement; head movement; barometer; hands-free; eyes-free; earphones; outer ear interface; wearable computing.

INTRODUCTION

Among wearable listening devices available on the market, earphones and headphones, which are used to listen to sounds anytime and anywhere, are the most popular wearable device worn around the head. Between the two of them, earphones are used the most in public. A typical scenario in which earphones are used is listening to music, where the user connects a pair of earphones to a smartphone. However, to do this, the user usually needs to use at least one hand and look at the smartphone's screen to take the smartphone out of the pocket and control its music player.

To address this issue, some commercial earphones have adopted sensors to allow users to operate the connected smartphone [9, 15]. For example such earphones, they have an accelerometer, a photo-reflector, and a touch-sensor for detecting a head/touch gestures (e.g., a nod for confirming a command, a shake of the head for cancelling a command, and a touch of the earphone for starting/stopping music). However, the number of head gestures is narrowly limited, and touching is not hands-free.

In this paper, we present CanalSense, a system for recognizing a jaw, face, or head movement (face-related movement).

The system concept emerged from our finding that air pressure inside the ear canals reflects characteristic changes casually dependent upon the type and degree of face-related movement. These changes can be measured using a barometer embedded in the earphone (earphone-type barometer) or by two earphone-type barometers in both ears. Therefore, we estimate face-related movements by employing machine-learning or pattern-matching to characterize the motions.

CanalSense is an easy-to-use outer ear interface (OEI) system. A user can wear earphone-type barometers in the same manner as wearing canal-type earphones, which are manufactured using commercially available earphones and eartips. Thus, CanalSense is a wearable system that can be used in everyday situations, like conventional wearable glass-shaped [7] and headphone-shaped eye trackers [23].

Through our work, we provide the following findings and contributions:

- We find that movements of the jaw, face, or head can be recognized based on changes in the air pressure inside the ear canals, measured using barometers.
- We present an implementation method to recognize face-related movements. We also show an implementation called CanalSense based on the proposed method.
- Our experiment showed that CanalSense achieves per-user recognition accuracies of 87.6% for eleven face-related movements, and 87.5% for four *Open Mouth* levels using random forest (RF).

RELATED WORK

CanalSense is a system that recognizes a user's face-related movements based on changes in the air pressure inside the ear canal. This section reviews prior work on OEIs, face-related movement recognition and recognition methods using barometers.

Outer Ear Interface

Several OEIs that recognize gestures and/or movements by attaching various sensors, earphones, and headphones to the outer ears have been proposed. For example, Manabe and Fukumoto [24] presented a technique called Headphone Taps that recognizes taps to headphones using built-in speakers with no additional hardware. EarPut [20] is an OEI that uses augmented earphones with an attached touch sensor array, which recognizes touch gestures to the user's ear. Universal Earphones [26] use a proximity sensor to detect which ear (i.e., left or right) an earphone is worn on. Septimu^2 [13] performs heart-rate monitoring, fine grained posture detection, and external sound source localization and classification by embedding a 3-axis accelerometer, a gyroscope, IR sensors, and two microphones into the earphones. Heartphones [30] provide non-intrusive continuous heart rate monitoring by integrating photoreflectors into earphones. Tayama et al. [35] developed a biometric sensor with a pressure transducer. When the ear is sealed, it detects the user's heartbeat by monitoring the air pressure inside the ear canal. SweepSense [19] uses the earphone's built-in speakers and microphones to recognize how a user wears them, based on the reflected sound in the ear canals. Manabe et al. [25] implemented earphones that recognizes eye movements by capturing an electrooculogram; they utilized the recognized eye movements as gesture inputs. CanalSense is also an OEI and can recognize face-related movements by using earphone-type barometers.

Similar to CanalSense, some OEIs that recognize face-related, mouth, and/or tongue movements have been proposed. For example, Bedri et al. [3] developed an OEI with three proximity sensors, one of which measures the degree of deformation in an ear canal. This OEI recognizes heart rate, tongue activities, jaw motion, and eye blinking. Moreover, Bedri et al. [5] attempted to recognize eating activities by using both the OEI and a 3D gyroscope placed on a hat. Bedri et al. [4] recognizes silent speech using a magnet attached to the user's tongue, magnetometers mounted near the cheek to track the 3D position of the magnet, and a pair of proximity sensors mounted near the user's ear canals. InEar BioFeedController [27] recognizes head nodding and shaking, eye winking, and ear wiggling by attaching a gyroscope and electroencephalography (EEG) sensors to each earphone. By contrast, CanalSense recognizes face-related movements by using only barometers.

Face-Related Movement Recognition

There has been some research, other than the above, that attempted to recognize face-related movements for achieving universal access to computer systems.

Mouth and/or Tongue

MouthType [22] is one example of research on mouth and/or tongue movement recognition. It uses hand and mouth movements to register text-entry, and a camera to recognize the shape, area, and aspect ratio of an *Open Mouth*. Bitey [2] recognizes tooth clicking sounds using one bone-conduction microphone just above an ear, and utilizes the results for eyes-free input. The system by Cheng et al. [8] recognizes tongue movements by attaching a textile-type pressure sensor array to the user's cheek. Tongue-in-Cheek [11] recognizes face-related gestures caused by tongue movements by using three X-band Doppler radars placed in front of the face. In addition, there is research on facial expression recognition with a head mounted display with photo-reflective sensors mounted on its interior [34].

Jaw

Some methods to recognize jaw movements have been proposed. For example, Kato [17] proposed a method that recognizes mastication by measuring changes in the shape of a tube inserted into the ear canal, using a pressure sensor. Nemirovski [29] proposed a method for recognizing movements by observing otoacoustic emissions caused by a user's movements. Wakamoto [36] measured chewing counts by inserting a pressure sensor into the ear canal. Aoki and Mine [1] proposed a mastication count system that recognizes biting sounds using a microphone. Sakai [33] developed a mastication sensor that can be inserted by the user into an ear canal, and which includes an air chamber that changes its shape during mastication.

Head

Recognition of head movement has been extensively researched, and their usability in various kinds of fundamental tasks in human-computer interaction is already being explored. For example, LoPresti et al. [21] detected neck movements with an ultrasonic-based approach, and used the results for icon selection and tracking tasks. Crossan et al. [10] proposed an input method for a smartphone where the user can move a cursor by tilting the head, which is measured by a 3-axis accelerometer and a magnetometer mounted on a cap. Rigas and Komogortsev [32] explored eye movement biometrics and showed these possible directions for the future research on eye movement-based recognition. Jacob et al. [16] mounted a motion tracker on the user's head to investigate how head movements can efficiently serve as a method to change the viewpoint in 3D applications.

In contrast to these works, CanalSense recognizes face-related movements, such as jaw, mouth, or head movements using earphones with embedded barometers.

Recognition Method using Barometers

Barometers are used for context-awareness and movement recognition in various research areas. For example, Wu et al. [37] recognized whether doors were open or closed using a built-in barometer in a smartphone. Ye et al. [38] proposed a floor localization system using a barometer, which can estimate the floor level in a multi-floor building by measuring changes in elevation. Emoballoon [28] recognizes touch gesture inputs on a balloon containing a barometer and microphone. Hyuga et al. [14] proposed a localization method that estimates the user's motion state and location in the subway using only a barometer built-into a smartphone.

CanalSense uses barometers to recognize face-related movements.

SENSING PRINCIPLE OF CANALSENSE

A portion of the musculoskeletal system used to move the jaw, face, and head related to the ears is illustrated in Figure 2. When a user performs a face-related movement, the musculoskeletal system changes, affecting the shape of the ear canals [6].

Specifically, when the jaw (i.e., mandible) is moved, the shape of the left ear canal changes depending on the positional relationship between the ear canal and the left mandibular condyle [18], which is the protrusion located at the left end of the mandible (the same is true on the right side). When the face or head is moved, the sternocleidomastoid muscle, which is large and close to the ear canal (see Figure 2c) and connects the back of the ear to the clavicle, either relaxes or contracts. This relaxation/contraction respectively expands/compresses the ear canal, changing its shape.

When the shape changes due to those factors, the volume of the space within the ear canal also changes. Therefore, when the ear canal is sealed using a material object such as an earphone, the changing volume inside the ear canal causes the internal air pressure to change.

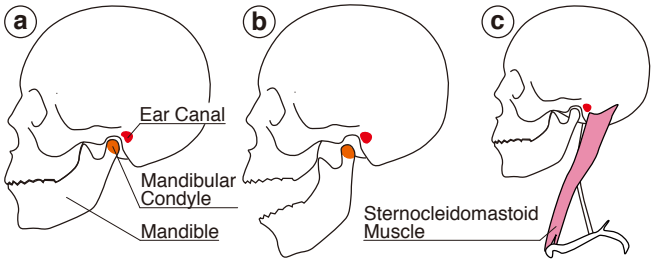

Figure 2: Principle of air pressure changes.

The air pressure inside the ear canal shows characteristic changes depending on the type and the degree of movement. To illustrate this, examples of face-related movements and examples of waveforms of the air pressure of both canals corresponding to the movements are shown in Figure 3. As this figure shows, the waveforms differ from each one other. For example, in *Open Mouth*, both the left and right air pressure values ascend, while in *Close Mouth*, both values descend. Interestingly, some symmetrical movements show symmetrical changes (e.g., *Slide Jaw Left* and *Slide Jaw Right*) while other symmetrical movements show asymmetrical changes (e.g., *Face Left* and *Face Right*, *Tilt Head Left* and *Tilt Head Right*). This is due to the difference between the left and right muscles, and individual idiosyncrasies in the movements of the face and head. Note that the degrees of change with *Slide Jaw Left* or *Slide Jaw Right* are smaller than *Open Mouth* and *Close Mouth*; this is consistent with [31].

By using a barometer inserted into the ear canal, changes in the air pressure can be measured. Therefore, employing machine learning or pattern matching to characterize changes in the air pressure enables the estimation of the movements.

IMPLEMENTATION

Our implementation consists of hardware, including earphone-type barometers, and software that employs an algorithm for recognizing the face-related movements from the changes in barometer values. Figure 4 shows the overview of this implementation.

Hardware

We implemented earphone-type barometers to measure the air pressure changes inside the ear canals (Figure 1a). We embedded a tiny barometer (Bosch BMP280 in our current implementation) in a commercially available canal-type earphone (Panasonic RP–HJE260). Moreover, because it is important to improve the airtightness of the ear canals to measure the air pressure precisely, we sealed the vent holes of the canal-type earphones with hot glue. Furthermore, the user was made to choose suitably-fitting eartips with no vent holes, which were used in combination with the canal-type earphones to keep the ear canal airtight. In our experiment, described later, we used *foam eartips*, which are earphone tips whose surfaces are made of soft, thick rubber. We asked the participants to squeeze the foam eartips enough before inserting them into the ear canals so as to make them fit and expand into the ear canals tightly.

Figure 3: Face-related movements and waveforms of the air pressure (hPa) of both ear canals corresponding to the movements.

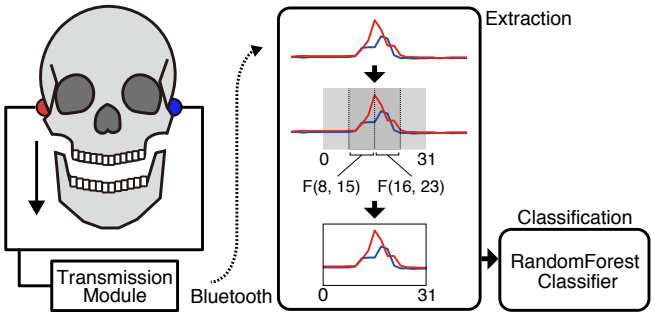

Figure 4: Overview of the implementation.

We also built a mobile transmission module (Figure 1b) consisting of a microcomputer (Arduino Pro Mini) and a Bluetooth module (SparkFun BlueSMiRF Silver) to wirelessly transmit the barometers' values to the computer that hosts our software. In our current implementation, the transmission module samples the air pressure at 32 Hz in each ear (64 values per second in total) using the barometers, and transmits the values to the computer. This module also has a rechargeable battery (we used a lithium polymer battery, DTP 502535, the battery drove this module for four hours in our experiment).

Software

Our recognition algorithm is composed of two processes. The first extracts the waveforms of the barometer values during a face-related movement. The second classifies the movements.

Extraction

To extract waveforms of the barometer values during a face-related movement, the recognition algorithm intake barometer values and tries to find a peak or valley with a gradient steeper than a threshold. If such a peak or valley is found, the barometer values at and around the respective peak or valley are saved as a waveform of a face-related movement.

To do this, the algorithm collects 32 barometer values g_k (k = 0, ..., 31; 31 is the most recent) per ear in every frame; these values represent the air pressure changes in one second. Then, the algorithm divides the 32 values into four parts (first part: 0–7; second part: 8–15; third part: 16–23; fourth part: 24–31) and uses the mean gradient (Equation 1) of the second and third parts to find the waveform of a steep peak or a valley.

$$F(i,j) = \frac{g_j - g_i}{j - i} \tag{1}$$

Equation 1 calculates the gradient between g_i and g_j. The algorithm uses this equation to calculate the gradient of the second part $F(8,15)$ and the third part $F(16,23)$. After this calculation, if the second and the third parts in either ear satisfy Equations 2 or 3, the algorithm saves the 32 barometer values of both ears as the waveform of a movement:

$$F(8,15) > \alpha \wedge F(16,23) < -\alpha \tag{2}$$

$$F(8,15) < -\alpha \wedge F(16,23) > \alpha \quad (3)$$

Moreover, because these equations fail to extract fast movements whose barometer values oscillate greatly within 16 frames, the algorithm also saves the barometer values that satisfy Equation 4:

$$max(g_k) - min(g_k) > 4\alpha \quad (4)$$

where $max(g_k)$ and $min(g_k)$ are the maximum and minimum values of g_k, respectively, and the range of k is $8 < k < 23$.

Note that the constant α in these equations depends on the airtightness inside the ear canal and the volume within the earphones. To detect small face-related movements, α should be set to a small value. In our current implementation, we use a constant of 0.02 hPa as α, which we derived empirically.

Classification

We used RF for classifying the saved waveforms to recognize the face-related movements. This process first extracts feature vectors from each of the saved waveforms. A feature vector consists of the following 90 values: the difference between each barometer value and its previous one (31 values from each of the left and right barometers, 62 values in total); the amplitude spectrum obtained by applying a fast Fourier transform algorithm to the differences between the left and right barometer values (16 values); the standard deviation (one value from each, two values in total), the maximum to minimum (one value from each, two values in total); the difference between the last and first values (one value from each, two values in total); the difference between the maximum and first value (one value from each, two values in total); the difference between the minimum and first value (one value from each, two values in total); and the number of values that are $\alpha \times 4$ hPa or more distant from the mean (one value from each, two values in total). Because atmospheric pressure changes from day to day, and time to time, we designed these features without using the raw barometer values directly. In the evaluation of using one ear, we used a feature vector consisting of 37 values: we removed the amplitude spectrum from 90 values and used the 37 values for evaluating each of the left and right barometers.

EVALUATION

We conducted an experiment to measure the accuracy of movement recognition.

Participants

We recruited 12 participants (P1–P12, eight males and four females) ranging in age from 21 to 23 years old (SD = 0.7); we collected participants of similar age, whose growth of the jaw and muscle would be in the same degree to control the experimental condition. All were full-time undergraduate students at the local institution. All had used canal-type earphones; eight used them usually (P1, P3, P5, and P7–P11). Two had suffered from a temporomandibular disorder, but both had already been cured (P6 and P11).

Experimental Environment

The experiment was held in a room in which all doors and windows were closed. The room's atmospheric pressure ranged from 996.6 hPa to 1013.3 hPa (M = 1002.5 hPa); its temperature ranged from 21.8 °C to 24.6 °C (M = 23.6 °C); its humidity ranged from 20% to 31% (M = 23%).

Procedure

The experiment was held under a sitting condition. Before performing the tasks, we asked the participants to complete a consent form and a questionnaire assessing demographics information and their experience with earphones. We also instructed the participants to prepare themselves to wear the earphone-type barometers, because some participants had never used the foam eartips.

We then asked the participants to perform Task A and then Task B. In each task, we first explained each of the face-related movements of the task and asked the participants to perform them. In a trial, one instruction was displayed in a random order in both tasks. Each instruction was the name of a face-related movement (e.g., *Face Left*) along with its Japanese translation, because the native language of all participants was Japanese. For one trial, we collected barometer values for three seconds (96 values from each barometer, 192 values in total). We asked the participants to take a break after every round and to re-wear (i.e., remove then reattach) the earphones after each round.

After the participants completed Tasks A and B, we asked them to answer the questionnaire regarding the ease of the movements. The experiment took approximately 70 minutes per participant. All participants were paid JPY 820 (approximately USD 7.40) for their time.

Task A

The purpose of Task A was to record the barometer values of multiple kinds of face-related movements. We asked the participants to perform each movement illustrated in Figure 3 as quickly and widely as possible. Because it was necessary to keep the mouth open before the *Close Mouth* movement, we asked participants expressly to do so to allow us to record only the barometer values of *Close Mouth*.

As a trial, the participants performed one of the eleven movements. In one round, they performed each of the eleven movements. In this task, each participant performed 12 rounds. In total, we collected the barometer values of 1584 trials (12 participants × 11 kinds of face-related movements × 12 rounds).

Task B

The purpose of Task B was to record the barometer values of multiple levels of *Open Mouth*. Specifically, we asked the participants to perform four levels of *Open Mouth* (i.e., *Keep Mouth Closed*, *Open Mouth Slight*, *Open Mouth*, and *Open Mouth Wide*). Because there are individual differences in the range of jaw movement, the distinction of the four levels was left to the participants' interpretation.

As a trial, the participants performed one of the four levels of *Open Mouth*. In one round, they performed each of

ⓐ	RandomForest			DTW+kNN		
	L	R	L+R	L	R	L+R
P1	69.7	78.8	90.2	72.0	73.5	88.6
P2	64.4	40.9	80.3	74.2	37.9	81.8
P3	66.7	51.5	87.9	80.3	60.6	89.4
P4	78.0	83.3	87.9	75.6	75.6	93.2
P5	78.8	87.1	92.4	88.6	84.1	99.2
P6	77.3	65.9	87.1	81.8	65.9	87.1
P7	65.2	62.1	81.1	77.3	62.9	90.9
P8	69.7	73.5	91.7	74.2	77.2	90.9
P9	63.6	58.3	80.3	75.0	75.0	87.9
P10	73.5	75.8	97.7	78.0	80.3	98.5
P11	73.5	73.5	88.6	83.3	77.2	90.2
P12	69.7	65.2	86.4	75.0	74.2	90.9
Mean	70.8	68.0	87.6	77.9	70.4	90.7
SD	5.4	13.4	5.2	4.8	12.4	4.7

ⓑ	RandomForest			DTW+kNN		
	L	R	L+R	L	R	L+R
P1	77.1	72.9	87.5	79.2	85.4	81.3
P2	93.8	66.7	91.7	95.8	54.2	93.6
P3	95.8	85.4	97.9	91.7	77.1	91.7
P4	93.8	87.5	91.7	93.8	83.3	89.6
P5	79.2	91.7	87.5	75.0	81.3	79.2
P6	83.3	77.1	91.7	79.2	77.1	77.1
P7	72.9	89.6	87.5	79.2	89.6	93.8
P8	62.5	79.2	79.2	68.3	72.9	75.0
P9	75.0	72.9	79.2	72.9	72.9	83.3
P10	75.0	77.1	79.2	91.7	83.3	93.8
P11	85.4	85.4	87.5	79.2	83.3	83.3
P12	62.5	91.7	89.6	75.0	91.7	95.8
Mean	79.7	81.4	87.5	81.8	79.3	86.5
SD	11.2	8.3	5.8	9.1	9.9	7.4

Table 1: Accuracies of movement recognition. L uses only the left-ear values; R uses only the right-ear values; L+R uses both the left- and right-ear values. a) Eleven face-related movements. b) Four *Open Mouth* levels.

four levels of *Open Mouth*. In this task, a participant performed 12 rounds. In total, we collected the barometer values of 576 trials (12 participants $\times$ four levels of *Open Mouth* $\times$ 12 rounds).

RESULTS

We measured the recognition accuracy of the eleven face-related movements and the four levels of *Open Mouth* using the saved waveforms in Task A and Task B, respectively (Table 1.) We used a leave-one-out cross-validation to measure the accuracy. In this measurement, we mainly used RF provided by Weka [12] with its default parameters for classification. In addition, we tested the combination of Dynamic Time Warping and *k*-nearest neighbor (DTW+*k*NN) with k = 1, since each face-related movement shows a different shape as shown in Figure 3, which we thought DTW cloud classify well. With DTW+*k*NN, we used the raw data of the saved waveforms (unlike RF). However, because the results of DTW+*k*NN were similar to those of RF and DTW is computationally intensive and thus real-time classification is difficult in a casual computational platform, we only discuss RF hereafter.

Through all the evaluations, there were no significant differences between participants who had suffered from a temporomandibular disorder and the others.

Recognizing Eleven Face-Related Movements

In order to measure the accuracy of recognition of eleven face-related movements, we built and executed per-user classifiers using the data from Task A. We trained a per-user classifier with only that user's data and performed cross-validation using the same user's data. Table 1a shows the accuracy for each participant. The confusion matrix of this classification is shown in Table 2.

The overall accuracy was 87.6% (SD = 5.2). There were significant differences between using both barometers (i.e., L+R) and using a single barometer (i.e., L or R) with a Kruskall-Wallis test (L and L+R: $p = 0.00 < 0.05$, R and L+R: $p = 0.00 < 0.05$). This suggests that there is an advantage of using both barometers for recognizing eleven face-related movements.

In Table 1a, the accuracy of P2's Right was fairly low compared to the others. In the post hoc questionnaire, P2 wrote that he thought his ear canals were so small that the earphones could not fit his ears well during the experiment. On the other hand, the accuracies of the other participants who wrote the same comment were not as low as those of P2. Furthermore, unpaired two-tailed t-tests showed no significant difference between the group using earphones usually and the others, which includes P2 (L: $p = 0.52$, R: $p = 0.47$, and L+R: $p = 0.32$). From these results, the low accuracy of P2's Right would be an issue peculiar to P2's right ear canal, which should be further investigated.

In addition, since the accuracy largely varies between movements, we performed a Kruskall-Wallis test. The test showed that there was a significant effect of movements on accuracy ($p = 0.02 < 0.05$); a post-hoc Tukey test showed that the accuracy of *Keep Mouth Closed* was higher than *Face Right* ($p = 0.00 < 0.05$, *Face Left* ($p = 0.00 < 0.05$), and *Tilt Head Right* ($p = 0.04 < 0.05$).

Recognizing Four Levels of Open Mouth

In order to measure the accuracy in recognizing the four levels of *Open Mouth*, we built and executed per-user classifiers using the data of Task B. Table 1b shows the accuracy for each participant. The confusion matrix of this classification is shown in Table 3.

The overall accuracy was 87.5% (SD = 5.8). In contrast to recognizing the eleven face-related movements, there were no significant differences between using both barometers (i.e., L+R) and using a single barometer (i.e., L or R). This suggests that a single barometer would be enough to recognize the four levels of *Open Mouth*.

In this result, as the confusion matrix shows, the most confused pair of levels was *Open Mouth* and *Open Mouth Slight*.

Actual Movements \ Predicted Movements	OM	CM	SJL	SJR	KMC	FL	FR	FU	FD	THL	THR
Open Mouth (OM)	**93.1**	0.7	0.0	0.7	0.0	0.7	0.7	2.1	2.1	0.0	0.0
Close Mouth (CM)	0.0	**89.6**	0.0	0.7	0.0	0.7	1.4	2.8	3.5	1.4	0.0
Slide Jaw Left (SJL)	0.0	0.7	**88.2**	2.1	0.7	0.7	0.7	0.0	0.0	4.2	2.8
Slide Jaw Right (SJR)	2.1	0.7	0.0	**91.0**	0.7	0.7	0.0	0.7	1.4	0.7	2.1
Keep Mouth Closed (KMC)	0.0	0.0	0.0	0.0	**99.3**	0.7	0.0	0.0	0.0	0.0	0.0
Face Left (FL)	0.0	0.7	1.4	0.7	2.1	**86.8**	0.7	2.8	0.0	4.2	0.7
Face Right (FR)	1.4	0.0	1.4	0.7	0.7	3.5	**79.9**	5.6	0.0	1.4	5.6
Face Up (FU)	2.1	2.8	0.0	0.7	0.7	2.8	5.6	**83.3**	0.0	0.0	2.1
Face Down (FD)	1.4	0.7	0.7	0.7	0.0	2.8	1.4	0.0	**87.5**	1.4	3.5
Tilt Head Left (THL)	1.4	0.0	1.4	2.8	0.0	2.8	1.4	2.1	0.7	**85.4**	2.1
Tilt Head Right (THR)	0.0	0.7	1.4	1.4	0.7	1.4	3.5	3.5	2.1	1.4	**84.0**

Table 2: Confusion matrix for recognizing the eleven face-related movements shown in Figure 3 using RF.

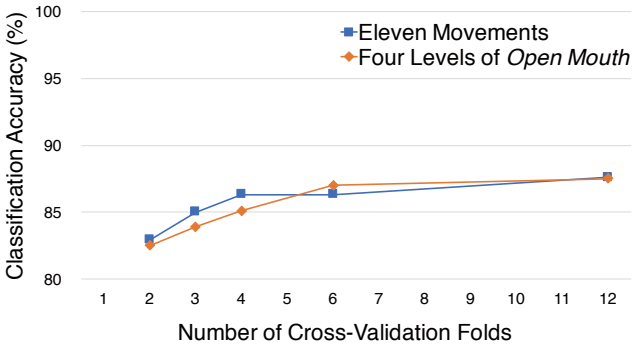

Figure 5: Accuracies in each cross-validation fold using RF.

This would be because these need to adjust the degree of openness of the mouth more finely than *Open Mouth Wide*, which only needs to open the mouth as widely as possible; therefore, some participants could not keep the degree the same in all rounds when performing them. In addition, 11.8% of *Open Mouth Slight* was mis-recognized as *Keep Mouth Closed*. We observed some participants opened their mouth by only sticking our her/his upper and lower lips forward in *Open Mouth Slight*, which is similar to the movement of kiss. In this case, the bones and/or muscles were not moved and thus *Open Mouth Slight* was mis-recognized as *Keep Mouth Closed*.

In addition, we investigated the relation between the amount of data and the accuracies. To do so, two-, three-, four-, six-, and twelve-fold cross-validations were performed as shown in Figure 5. For the eleven movements, the accuracies in the two-, three-, four-, six-, and twelve-fold cross-validations were 82.9%, 85.0%, 86.3%, 86.3%, and 87.6%, respectively. For the four *Open Mouth* levels, the accuracies in the two-, three-, four-, six-, and twelve-fold cross-validations were 82.5%, 83.9%, 85.1%, 87.0%, and 87.5%, respectively.

EXAMPLE APPLICATIONS

We developed two example applications. One is a hands-free and eyes-free music player. The other is a hands-free content reader application that the user can use it with dirty hands, such as when cooking.

Actual Levels \ Predicted Levels	OMW	OM	OMS	KMC
Open Mouth Wide (OMW)	**91.7**	6.3	2.1	0.0
Open Mouth (OM)	5.6	**81.3**	13.2	0.0
Open Mouth Slight (OMS)	0.7	8.3	**79.2**	11.8
Keep Mouth Closed (KMC)	0.0	0.0	2.1	**97.9**

Table 3: Confusion matrix for recognizing four levels of *Open Mouth* using RF.

Hands-Free and Eyes-Free Music Player

The user can use this music player even when her/his hands are occupied, such as when carrying bags in both hands. The face-related movements allow the user to perform several actions, such as play/pause (*Open Mouth* and *Close Mouth* instantly), and skip to the previous or next song (*Jaw Slide Left* or *Jaw Slide Right*) in a hands-free and eyes-free manner.

Hands-Free Content Reader

This content reader is used when the user cannot use her/his hands for touch operations (e.g., the hands are dirty, occupied such as when cooking, injured, or disabled). By using the *Slide Jaw Left/Right* movements, the user can flip pages back or forward. Similar to these operations, *Tilt Head Left/Right* show the previous or next section.

DISCUSSION AND LIMITATIONS

Low Accuracy of Face Right

While there were no significant differences between *Face Right* and the other movements except for *Keep Mouth Closed*, *Face Right* tends to exhibit low accuracy. Specifically, as shown in Table 1, *Face Right* was confused with *Tilt Head Right* and *Face Up*. The confusion with *Tilt Head Right* would be because it involves similar muscle movements to *Face Right* since both are movements in the right direction (this also appears in *Face Left* and *Tilt Head Left*). By contrast, the confusion with *Face Up* would be because it is a

movement which involves subtle air pressure changes in the ear canals as shown in Figure 3 and thus tends to be confused with various movements including *Face Right*.

Influence by Various Factors

Although the results of the experiment shows that CanalSense can recognize face-related movements in high accuracies, we conducted the experiment only in a controlled lab environment. Therefore, in order to evaluate whether CanalSense actually works, it is necessary to examine various factors that could influence the accuracy.

Physical Condition

Changes in physical condition would affect the recognition because they will cause air pressure changes inside the ear canals or make the air pressure changes different from usual without the user's intention. Examples of such changes include having a cold or a stuffy nose, and changes in the body temperature (e.g., by drinking or performing intense exercise). Therefore, further experiments are necessary to investigate the effect of physical conditions (e.g., experiments to examine whether inserting earplugs will affect the accuracy or not).

User's Posture/Movement/Action

While our experiment was held under the sitting condition, the user's posture would make the air pressure changes different. Moreover, user's movements/actions also cause air pressure changes inside the ear canals. Especially, movements using the whole body, such as walking, running, biking, or riding, do cause air pressure changes inside the ear canals. For example, we observed that walking generates waveforms similar to *Face Up* and *Face Down*, since the user's head moves up and down while walking. Actions around the face, such as talking, relieving ear pressure, and swallowing, would cause erroneous detection, since the bones and muscles around the mouth move by such actions. In addition, we observed casual small facial expression changes such as wink or nose twitching have no influences on waveforms while large or strong facial expressions changes such as bursting into laughter or squeezing the eyes tightly do have influences.

In order to reduce erroneous detection, we will collect data during such user's movements/actions under various postures conditions, including standing and lying, to improve our system.

Environmental Factors

Some environments, such as on an elevator, aboard an airplane, and in bad weather, would affect the air pressure inside the ear canals. In our implementation, we did not use raw barometer values considering atmospheric pressure changes from day to day, but the short-time pressure changes were out of consideration. We will verify the barometer value changes in these environments and revise our implementation to accommodate the factors.

Sound Factors

The influence on air pressure by the sound played by the earphones should be considered. Because sound vibrates the air inside the ear canals, there is a possibility that the air pressure will be changed by the sound. As a pilot study to evaluate the influences, we conducted a small experiment to compare the accuracy in recognizing seven face-related movements (i.e., *Open Mouth*, *Close Mouth*, *Slide Jaw Left*, *Slide Jaw Right*, *Keep Mouth Closed*, *Open to Close Mouth*, and *Read Sentences*). We recruited six participants (four males and two females) ranging in age from 20–23. We compared two conditions: music and no-music. In the music condition, Symphony No. 5 by Ludwig van Beethoven was played. The RF results showed no significant difference between the two conditions (using a paired t-test; music condition: accuracy 89.9%; no-music condition: accuracy 87.7%; $p=0.38>0.05$). In future, we plan to conduct experiments to verify the influences of other types of sound.

Airtightness

CanalSense requires airtightness inside the ear canals. Therefore, for a user who dislikes to seal the ears, our system is difficult to use. The user whose ear canals are too small or large would not be able to use CanalSense because of the loss of airtightness, although the user can try various sizes of eartips or make personal fitted earphones/eartips to use CanalSense.

In addition, placement changes of the earphones would affect the airtightness. In the experiment, participants re-wore (i.e., removed and reattached) the earphones after each round. Therefore, our system would be robust to re-attachment. However, it is necessary to examine the effect of the placement changes due to long-time uses (e.g., four hours) and inserted conditions (deep or shallow).

Comfort

We used the commercial eartips that were carefully designed for comfort. Although we used only one size of eartips in the experiments for unifying the experimental environment, a user should use eartips that match the user's ear size. We obtained two comments about comfort. P2, who wrote his ear canals were small in his demographic information, commented "Smaller eartips would be better." This problem prevented the eartips from being inserted well into his ear canals, and thus the airtightness was low, which reduced the recognition rate. By contrast, P10, who wrote nothing about his ear canals, commented "The earphones were too tight." This would mean that the airtightness of P10 was kept high, resulting in a high recognition rate. However, comfort in long-time use should be explored as future work.

Verification of Individual Difference

There is a possibility that the recognition rate is influenced by the age of the participant. Therefore, further experiments with other age-groups are needed. In particular, experiments with subjects aged 20 years or younger is required because one's mandibular condyle grows until age 20 [18]. Moreover, to better understand the feasibility of CanalSense, we plan to conduct experiments with participants with various characteristics, such as the ones with temporomandibular joint derangement.

FUTURE WORK

The face-related movements shown in Figure 3 can be combined into a compound movement (e.g., performing *Open Mouth* and *Tilt Head Right* simultaneously). In our experiment, we did not evaluate such combinations. To explore this, we plan to observe the air pressure changes during compound movements and will attempt to recognize the compound movements in the future experiments.

In our implementation, user calibration for specific contexts was not carried out. However, to remedy effects of the user's condition (e.g., having a cold or not having a cold), we can use calibration by training the model using the data under both conditions. Similarly, by sensing the environment (e.g., using an elevator or boarding an airplane), it would be certainly possible to improve accuracy by canceling the linear increasing/decreasing components in the barometer values.

The experiments in this study were carried out in a controlled environment. It is necessary to verify the environmental factors (e.g., using an elevator, boarding an airplane, and enduring bad weather) in subsequent experiments. In addition, we must conduct experiments with participants of diverse demographics (e.g., wider age-ranges or those with temporomandibular disorder), because the experiments we conducted targeted participants who were similar in age, and whose jaw and muscle growth would be in the same degree of development.

CONCLUSION

We presented a novel face-related movement recognition system called CanalSense. It recognizes face-related movements using barometers embedded in earphones. The system recognizes face-related movements based on air pressure changes inside the ear canal. The results of our experimental evaluation showed that the recognition accuracies for eleven face-related movements were 87.6%, and the accuracies for four levels of *Open Mouth* were 87.5%, using random forest.

ACKNOWLEDGMENTS

This research has been supported in part by Takahashi Industrial and Economic Research Foundation.

REFERENCES

1. Ken Aoki and Masashi Mine. 2010. Mastication Frequency Detecting Device. (2010). JP-2010154985-A, 2010. (In Japanese).
2. Daniel Ashbrook, Carlos Tejada, Dhwanit Mehta, Anthony Jiminez, Goudam Muralitharam, Sangeeta Gajendra, and Ross Tallents. 2016. Bitey: An Exploration of Tooth Click Gestures for Hands-free User Interface Control. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '16)*. ACM, New York, NY, USA, 158–169. DOI:http://dx.doi.org/10.1145/2935334.2935389
3. Abdelkareem Bedri, David Byrd, Peter Presti, Himanshu Sahni, Zehua Gue, and Thad Starner. 2015a. Stick It in Your Ear: Building an In-ear Jaw Movement Sensor. In *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers (UbiComp/ISWC '15 Adjunct)*. ACM, New York, NY, USA, 1333–1338. DOI: http://dx.doi.org/10.1145/2800835.2807933
4. Abdelkareem Bedri, Himanshu Sahni, Pavleen Thukral, Thad Starner, David Byrd, Peter Presti, Gabriel Reyes, Maysam Ghovanloo, and Zehua Guo. 2015b. Toward Silent-Speech Control of Consumer Wearables. *Computer* 48, 10 (Oct. 2015), 54–62. DOI: http://dx.doi.org/10.1109/MC.2015.310
5. Abdelkareem Bedri, Apoorva Verlekar, Edison Thomaz, Valerie Avva, and Thad Starner. 2015c. Detecting Mastication: A Wearable Approach. In *Proceedings of the 17th ACM International Conference on Multimodal Interaction (ICMI '15)*. ACM, New York, NY, USA, 247–250. DOI: http://dx.doi.org/10.1145/2818346.2820767
6. Henry S. Brenman, Robert C. Mackowiak, and M. H. F. Friedman. 1968. Condylar Displacement Recordings as an Analog of Mandibular Movements. *Journal of Dental Research* 47, 4 (1968), 599–602. DOI: http://dx.doi.org/10.1177/00220345680470041501
7. Andreas Bulling, Daniel Roggen, and Gerhard Tröster. 2008. It's in Your Eyes: Towards Context-awareness and Mobile HCI using Wearable EOG Goggles. In *Proceedings of the 10th International Conference on Ubiquitous Computing (UbiComp '08)*. ACM, New York, NY, USA, 84–93. DOI: http://dx.doi.org/10.1145/1409635.1409647
8. Jingyuan Cheng, Ayano Okoso, Kai Kunze, Niels Henze, Albrecht Schmidt, Paul Lukowicz, and Koichi Kise. 2014. On the Tip of My Tongue: A Non-invasive Pressure-based Tongue Interface. In *Proceedings of the 5th Augmented Human International Conference (AH '14)*. ACM, New York, NY, USA, Article 12, 4 pages. DOI:http://dx.doi.org/10.1145/2582051.2582063
9. Sony Corporation. 2016. Xperia Ear. (2016). http://www.sonymobile.com/us/products/smart-products/xperia-ear/ (accessed 2017-4-4).
10. Andrew Crossan, Mark McGill, Stephen Brewster, and Roderick Murray-Smith. 2009. Head Tilting for Interaction in Mobile Contexts. In *Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '09)*. ACM, New York, NY, USA, Article 6, 10 pages. DOI: http://dx.doi.org/10.1145/1613858.1613866
11. Mayank Goel, Chen Zhao, Ruth Vinisha, and Shwetak N. Patel. 2015. Tongue-in-Cheek: Using Wireless Signals to Enable Non-Intrusive and Flexible Facial Gestures Detection. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 255–258. DOI: http://dx.doi.org/10.1145/2702123.2702591

12. Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. 2009. The WEKA Data Mining Software: An Update. *SIGKDD Explorations Newsletter* 11, 1 (Nov. 2009), 10–18. DOI:http://dx.doi.org/10.1145/1656274.1656278

13. Pan Hu, Guobin Shen, Xiaofan Jiang, Shao-fu Shih, Donghuan Lu, Feng Zhao, Dezhi Hong, Qiang Li, Shahriar Nirjon, Robert Dickerson, and John A. Stankovic. 2012. Septimu^2 - Earphones for Continuous and Non-intrusive Physiological and Environmental Monitoring. In *Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems (SenSys '12)*. ACM, New York, NY, USA, 387–388. DOI:http://dx.doi.org/10.1145/2426656.2426722

14. Satoshi Hyuga, Masaki Ito, Masayuki Iwai, and Kaoru Sezaki. 2015. Estimate a User's Location using Smartphone's Barometer on a Subway. In *Proceedings of the 5th International Workshop on Mobile Entity Localization and Tracking in GPS-less Environments (MELT '15)*. ACM, New York, NY, USA, Article 2, 4 pages. DOI:http://dx.doi.org/10.1145/2830571.2830576

15. Apple Inc. 2016. AirPods. (2016). http://www.apple.com/airpods/ (accessed 2017-4-4).

16. Thibaut Jacob, Gilles Bailly, Eric Lecolinet, Géry Casiez, and Marc Teyssier. 2016. Desktop Orbital Camera Motions using Rotational Head Movements. In *Proceedings of the 2016 Symposium on Spatial User Interaction (SUI '16)*. ACM, New York, NY, USA, 139–148. DOI: http://dx.doi.org/10.1145/2983310.2985758

17. Takashi Kato. 2008. Manducating Motion Detection Apparatus. (2008). JP-200848791-A, 2008. (In Japanese).

18. Yojiro Kawamura. 1972. About Occlusion Physiology. *The Nippon Dental Review* 359 (1972), 25–33. (In Japanese).

19. Gierad Laput, Xiang 'Anthony' Chen, and Chris Harrison. 2016. SweepSense: Ad Hoc Configuration Sensing using Reflected Swept-Frequency Ultrasonics. In *Proceedings of the 21st International Conference on Intelligent User Interfaces (IUI '16)*. ACM, New York, NY, USA, 332–335. DOI: http://dx.doi.org/10.1145/2856767.2856812

20. Roman Lissermann, Jochen Huber, Aristotelis Hadjakos, Suranga Nanayakkara, and Max Mühlhäuser. 2014. EarPut: Augmenting Ear-worn Devices for Ear-based Interaction. In *Proceedings of the 26th Australian Computer-Human Interaction Conference on Designing Futures: The Future of Design (OzCHI '14)*. ACM, New York, NY, USA, 300–307. DOI: http://dx.doi.org/10.1145/2686612.2686655

21. Edmund LoPresti, David M. Brienza, Jennifer Angelo, Lars Gilbertson, and Jonathan Sakai. 2000. Neck Range of Motion and Use of Computer Head Controls. In *Proceedings of the Fourth International ACM Conference on Assistive Technologies (ASSETS '00)*. ACM, New York, NY, USA, 121–128. DOI: http://dx.doi.org/10.1145/354324.354352

22. Michael J. Lyons, Chi-Ho Chan, and Nobuji Tetsutani. 2004. MouthType: Text Entry by Hand and Mouth. In *Proceedings of the ACM Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '04)*. ACM, New York, NY, USA, 1383–1386. DOI: http://dx.doi.org/10.1145/985921.986070

23. Hiroyuki Manabe and Masaaki Fukumoto. 2006. Full-time Wearable Headphone-type Gaze Detector. In *Proceedings of the ACM Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '06)*. ACM, New York, NY, USA, 1073–1078. DOI: http://dx.doi.org/10.1145/1125451.1125655

24. Hiroyuki Manabe and Masaaki Fukumoto. 2012. Headphone Taps: A Simple Technique to Add Input Function to Regular Headphones. In *Proceedings of the 14th International Conference on Human-computer Interaction with Mobile Devices and Services Companion (MobileHCI '12)*. ACM, New York, NY, USA, 177–180. DOI: http://dx.doi.org/10.1145/2371664.2371703

25. Hiroyuki Manabe, Masaaki Fukumoto, and Tohru Yagi. 2013. Conductive Rubber Electrodes for Earphone-based Eye Gesture Input Interface. In *Proceedings of the 2013 International Symposium on Wearable Computers (ISWC '13)*. ACM, New York, NY, USA, 33–40. DOI: http://dx.doi.org/10.1145/2493988.2494329

26. Kohei Matsumura, Daisuke Sakamoto, Masahiko Inami, and Takeo Igarashi. 2012. Universal Earphones: Earphones with Automatic Side and Shared Use Detection. In *Proceedings of the 2012 ACM International Conference on Intelligent User Interfaces (IUI '12)*. ACM, New York, NY, USA, 305–306. DOI: http://dx.doi.org/10.1145/2166966.2167025

27. Denys J.C. Matthies. 2013. InEar BioFeedController: A Headset for Hands-free and Eyes-free Interaction with Mobile Devices. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems (CHI EA '13)*. ACM, New York, NY, USA, 1293–1298. DOI: http://dx.doi.org/10.1145/2468356.2468587

28. Kosuke Nakajima, Yuichi Itoh, Yusuke Hayashi, Kazuaki Ikeda, Kazuyuki Fujita, and Takao Onoye. 2013. Emoballoon: A Balloon-Shaped Interface Recognizing Social Touch Interactions. In *Proceeding of the 10th International Conference on Advances in Computer Entertainment (ACE 2013)*. Springer-Verlag New York, Inc., New York, NY, USA, 182–197. DOI: http://dx.doi.org/10.1007/978-3-319-03161-3_13

29. Guerman G. Nemirovski. 2001. System and Method for Detecting an Action of the Head and Generating an Output in Response Thereto. (2001). WO 200139662-A2.

30. Ming-Zher Poh, Kyunghee Kim, Andrew D. Goessling, Nicholas C. Swenson, and Rosalind W. Picard. 2009. Heartphones: Sensor Earphones and Mobile Application for Non-obtrusive Health Monitoring. In *Proceedings of the 2009 International Symposium on Wearable Computers (ISWC '09)*. IEEE Computer Society, Washington, DC, USA, 153–154. DOI: http://dx.doi.org/10.1109/ISWC.2009.35

31. JunRong Qi. 2016. *Cross-correlation between Mandibular Condylar Movements and Distortion of External Auditory Meatus*. Ph.D. Dissertation. Matsumoto Dental University. (In Japanese).

32. Ioannis Rigas and Oleg V. Komogortsev. 2017. Current Research in Eye Movement Biometrics. *Image and Vision Computing* 58 (Feb. 2017), 129–141. DOI: http://dx.doi.org/10.1016/j.imavis.2016.03.014

33. Masahiko Sakai. 1999. Chewing Sensor. (1999). JP-11318862-A, 1999. (In Japanese).

34. Katsuhiro Suzuki, Fumihiko Nakamura, Jiu Otsuka, Katsutoshi Masai, Yuta Itoh, Yuta Sugiura, and Maki Sugimoto. 2016. Facial Expression Mapping Inside Head Mounted Display by Embedded Optical Sensors. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology (UIST '16 Adjunct)*. ACM, New York, NY, USA, 91–92. DOI: http://dx.doi.org/10.1145/2984751.2985714

35. Ai Tayama, Hiromasa Yamashita, Tomoo Sato, Gontaro Kitazumi, Toshio Chiba, and Akira Toki. 2014. Development and Accuracy of A Miniature Earphone-Type Bilogical Information Sensor. *Journal of the Showa University Society* 74, 1 (2014), 60–66. DOI: http://dx.doi.org/10.14930/jshowaunivsoc.74.60 (In Japanese).

36. Shigeo Wakamoto. 1995. Mastication Count Meter. (1995). JP-07213510-A, 1995. (In Japanese).

37. Muchen Wu, Parth H. Pathak, and Prasant Mohapatra. 2015. Monitoring Building Door Events using Barometer Sensor in Smartphones. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '15)*. ACM, New York, NY, USA, 319–323. DOI: http://dx.doi.org/10.1145/2750858.2804257

38. Haibo Ye, Tao Gu, Xianping Tao, and Jian Lu. 2014. B-Loc: Scalable Floor Localization using Barometer on Smartphone. In *11th IEEE International Conference on Mobile Ad Hoc and Sensor Systems (MASS '14)*. IEEE Computer Society, Washington, DC, USA, 127–135. DOI:http://dx.doi.org/10.1109/MASS.2014.49