IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 27 March 2023, accepted 24 April 2023, date of publication 1 May 2023, date of current version 10 May 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3271990

RESEARCH ARTICLE

Distinguishing Orthodontic Experts From Laypersons Through Gaze Analysis

JUNZO KAMAHARA[1,7], (Member, IEEE), TAKASHI NAGAMATSU[1], (Member, IEEE), KYOKO ITO[2,3], MAMORU HIROE[1,7,8], (Member, IEEE), HARUKI SAO[1,9], SAIZO AOYAGI[4], JUNKO NAGATA[5], AND KENJI TAKADA[6]

[1]Graduate School of Maritime Sciences, Kobe University, Kobe 658-0022, Japan
[2]Faculty of Engineering, Kyoto Tachibana University, Kyoto 607-8175, Japan
[3]Graduate School of Information Science and Technology, Osaka University, Osaka 565-0871, Japan
[4]Faculty of Global Media Studies, Komazawa University, Tokyo 154-8525, Japan
[5]Faculty of Medicine, University of Miyazaki, Miyazaki 889-1692, Japan
[6]Ask-Prof Oral Health Consulting, Osaka 530-0001, Japan
[7]Faculty of Data Sciences, Osaka Seikei University, Osaka 533-0007, Japan
[8]Graduate School of Medicine, Kobe University, Kobe 650-0017, Japan
[9]Nomura Research Institute Ltd., Tokyo 100-0004, Japan

Corresponding author: Junzo Kamahara (kamahara@g.osaka-seikei.ac.jp)

This work was supported in part by the Japan Society for the Promotion of Science (JSPS) KAKENHI under Grant 20H01748 and Grant 20H04229.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Administration Committee of the Graduate School of Maritime Sciences at Kobe University and the Medical Ethics Committee, University of Miyazaki.

ABSTRACT Visual inspection is an important process conducted as an initial diagnostic step in medical examinations. It is assumed that the gaze movements of an orthodontist (expert) differ from those of a layperson. In this study, to examine whether the degree of proficiency in conducting a visual examination can be estimated from gaze movement, we conducted a gaze measurement experiment in which facial images (frontal and lateral images of three patients) were viewed by ten experts and ten laypersons. The performance in discriminating whether a subject was an expert or layperson exhibited a certain improvement when applying an aggregation method for the gaze data, that is, the grid gaze frequency and AOI gaze frequency. We examined whether proficiency levels could be determined using machine learning techniques. The results demonstrated that our method distinguished experts and laypersons relatively effectively using gaze frequency based on the grid and area of interest set by an expert for each face part.

INDEX TERMS Expertise, eye tracking, gaze, machine learning, orthodontist.

I. INTRODUCTION

Visual inspection is an important process conducted as the initial diagnostic step in medical examinations. An expert doctor may determine the adequate course of treatment based on the first visual inspection. However, lack of skill in such visual inspection may result in misdiagnosis or prolonged treatment.

Gaze movement in the observation of a target is considered to depend on the interests and implicit knowledge of a person [1]. It was assumed that the gaze movement of an expert doctor differs from that of a layperson with insufficient experience or training.

Education regarding how to conduct visual examinations is necessary. Therefore, it is necessary to understand how skilled and unskilled practitioners provide visual examination. If we can determine proficiency level from one's gaze movement, we can prepare educational content according to the proficiency level of the learner. As an initial step, this study examined the differences in gaze between experts and laypersons.

In this study we developed methods for measuring proficiency using gaze movements to distinguish between experts and laypersons in providing orthodontic diagnoses.

The associate editor coordinating the review of this manuscript and approving it for publication was Aasia Khanum.

We defined an expert as being fully trained and experienced in orthodontics, and a layperson as having little or no knowledge of orthodontics. We conducted experiments, that included measurements of gaze as the subject browsed through frontal and lateral facial images of orthodontic patients.

The contributions of this study are as follows: 1) a frequency analysis of gaze positions using an interest grid and the area of interest (AOI); 2) the introduction of a machine-learning method into gaze analysis; and 3) confirmation that gaze analysis can distinguish between the proficiency of an expert and that of a layperson. Furthermore, as an innovation of this study, the gaze of the orthodontist while looking at the face for diagnosis was measured. Although there have been studies on measuring gaze while looking at a radiograph, few have addressed situations in which the orthodontist looks at the face for diagnostic purposes.

II. RELATED WORK

Several studies [1], [2], [3] have demonstrated that the gaze movement of observer when viewing the same object is completely different depending on the interest and intention of observation. Nakayama and Harada [3] showed that the type of background knowledge possessed by viewers affects their gaze movements while reading program code. Moreover, studies of gazing at 3D objects (bonsai [4] and pieces of pottery [5]) have revealed that novices' viewing patterns differ from those of experienced observers. According to the results of these studies, it can be assumed that the ''seeing'' behavior of experienced observers differs from that of novice observers and that a type of spatial order or pattern is exhibited.

Some medical studies have used gaze tracking in pathology [6], radiology [7], [8], and anesthesia [9]. In the medical field, machine learning has recently been used to analyze eye movements [10], [11], [12], [13], [14]. Kollias et al. [10] reviewed studies adapting machine learning and eye tracking technology in autism spectrum disorder research. This review introduces various studies that have attempted to identify patients by comparing them to subjects with typical developments. In contrast, our study attempted to discriminate between skilled orthodontists with expertise and laypersons using eye movement analysis. Hosp et al. presented a model for classifying surgeons into three levels of expertise using only eye movements when applying a support vector machine model [12]. Castner et al. compared the scan paths of five semesters of dental students when viewing orthopantomograms (OPT), which are radiographic images specifically used for dentistry, to distinguish sixth-tenth-semester students [13]. Recently, Castner et al. distinguished the saccadic behavior of dental experts during an OPT inspection using LSTMs [14].

The importance of visual inspection skills has been recognized in the medical field. Effective teaching methods for diagnostic skills based on rational evidence in clinical education and methods for measuring the degree of acquisition of expert knowledge are necessary. If the gaze features of an expert in the medical field can be determined, they can be applied for the development of medical skills. To the best of our knowledge, no method has been proposed to distinguish between the gaze of experts and laypersons when conducting a diagnosis in orthodontics. Furthermore, because the number of experts with similar skills has been limited, it was difficult to cooperate in the data collection process.

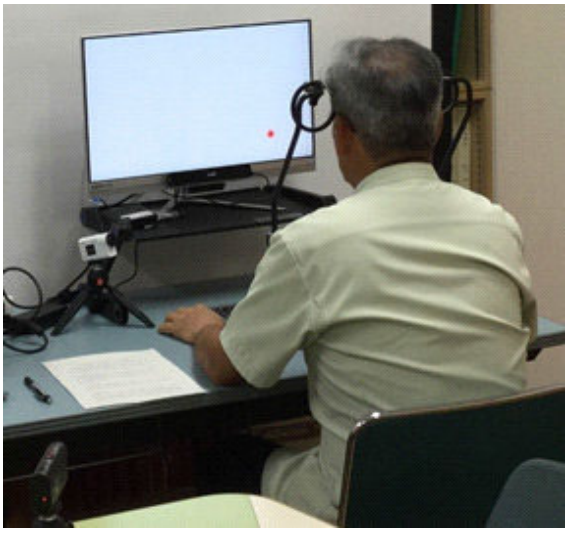


FIGURE 1. Experimental scene.

In this study, we investigated whether eye-gaze data can be used to distinguish between experts and laypersons as a measure of proficiency in visual inspection.

III. EXPERIMENTS FOR GAZE POSITION DATA ACQUISITION

A. SUBJECTS

The subjects included 10 experts (8 males and 2 females, aged 29-67 years) and 10 laypersons (8 males and 2 females, aged 21-25 years).

B. EQUIPMENT

Eye movement data were measured using Tobii Pro X2-30 [15] (Tobii AB), an instrument for measuring eye movements (sampling rate of 30 Hz, precision of 0.4°, and accuracy of 0.32°). Tobii Pro X2-30 outputs data, such as the recorded timestamp, local timestamp, and gaze point (x, y) in the screen coordinate system. The recorded timestamp was the time (in ms) at the start of gaze measurement. The local timestamp is the time (in ms) from the start of data recording. In this study, we used the gaze points in the screen coordinate system. The sampling rate was 30 Hz; for example, 300 eye-position samples were recorded within a 10-s period.

In the experimental environment, Tobii Pro X2-30 was placed at the bottom center of the display (screen resolution of 1920 × 1080 pixels), and the distance between the display and eyeball was approximately 63 cm. Figures 1 and 2 show the experimental setup and hardware settings, respectively.

C. DATA COLLECTION

All the subjects were presented with three sets of facial images. One set consisted of frontal and lateral facial images

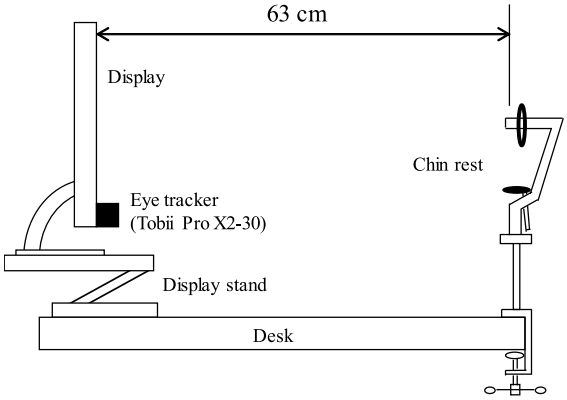


FIGURE 2. Experimental hardware setting.

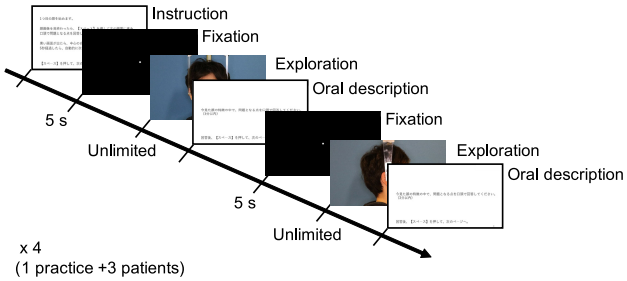


FIGURE 3. Outline of experimental session.

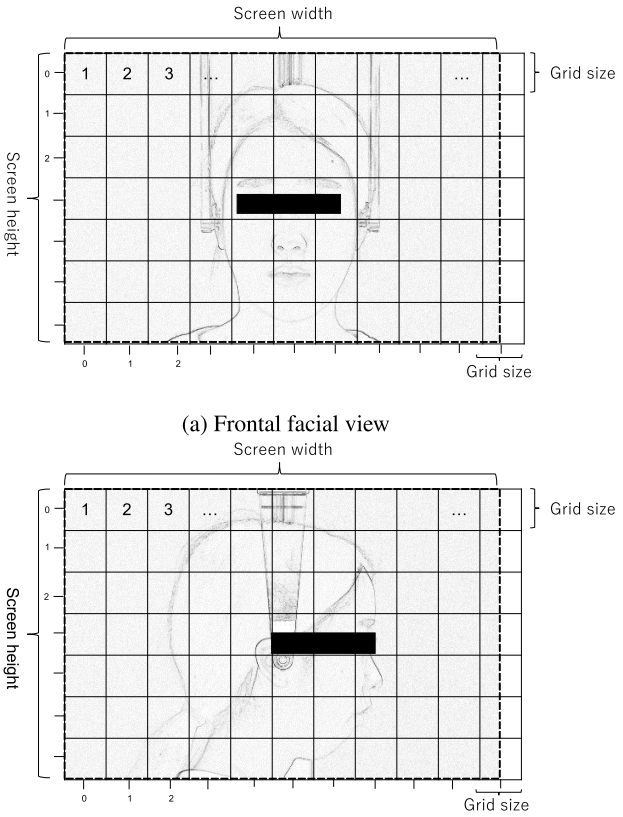


FIGURE 4. Gridding of screen (sample).

of one patient requiring orthodontic treatment. Patients 1 and 2 had mandibular protrusions, whereas Patient 3 had maxillary protrusions due to a hypoplastic mandible. The display order of the stimuli was leveled across the subjects by adjusting the display order of the three patients.

After calibration of the eye tracker, the following instructions were provided to the subjects: "These are facial images of patients who visited to the clinic for treatment. Two facial images (frontal and lateral views) constitute a single display set. Please look at them individually. After examining the facial features, please describe the problems faced by each patient."

Figure 3 illustrates the experimental protocol for a single session. Following the instructions, the subjects gazed at a fixation circle for a 5-s period. Thereafter, the session was changed to the exploration phase and a frontal facial view was displayed. The subjects spent as much time as they needed and pressed the space key to proceed to the next phase. In the oral description phase, the instructions indicate that the subject should loudly describe the problem of each patient. Subsequently, each subject focused on a fixation circle for 5 seconds, explored the lateral facial image, and described it orally. The sessions were repeated three times by changing the patient images. Prior to the first session, the subjects underwent preliminary sessions with illustrated facial images.

The experiment was conducted with approval from the Research Ethics Review Committee of the Graduate School of Maritime Sciences at Kobe University. The use of patient images was approved by the Ethics Committee of the University of Miyazaki.

IV. EXTRACTION OF GAZE POSITION DISTRIBUTIONS

There are two main ways to process a recorded gaze based on movement and position. There are various types of Movement Measures, including direction, amplitude, duration, velocity, acceleration, shape, AOI order, and scan path comparison [16, Chapter.10]. The scan path is a sequence of gazes in space. By contrast, Position Measures are measures of the stillness of a gaze in one or more positions [16, Chapter.11]. There are several varieties of position measures, such as basic position, dispersion, similarity, duration, and pupil dilation. We used position duration, which measures how long the gaze remains at a position and can be thought of as the frequency over a unit of time.

In this study, a preliminary analysis was conducted using the scan paths. Because scan path methods consider the order in which objects are viewed, they can be applied to objects with regularity in their viewing order, such as "writing text." However, in our visual inspection study, a specific order for a part of the face could not be assumed. Our preliminary analysis determined that it is difficult to classify experts and laypersons using scan-path methods. Therefore, to measure

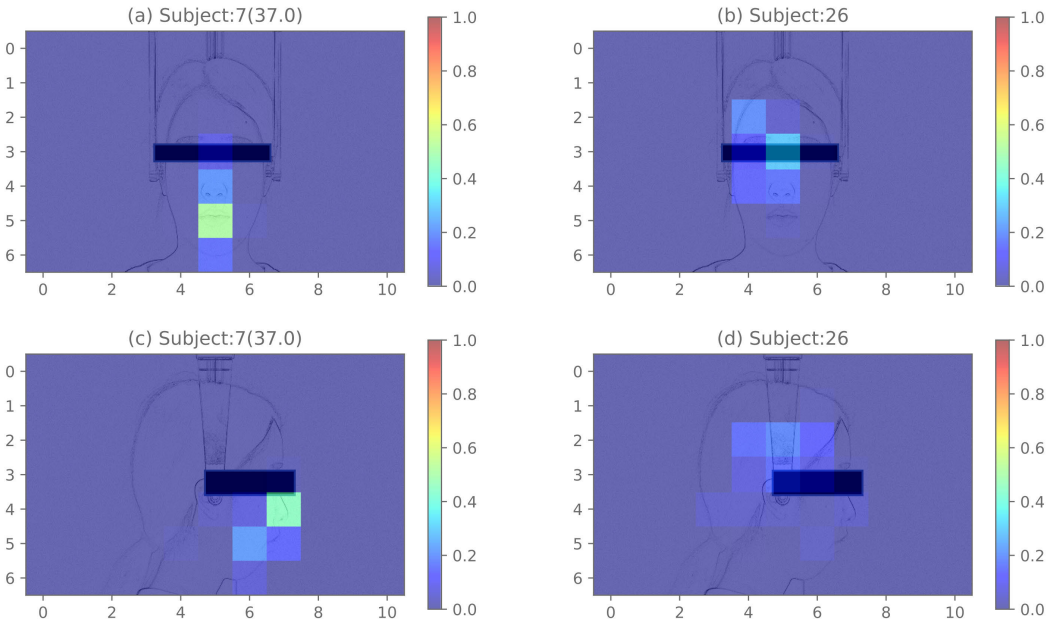


FIGURE 5. Examples of grid frequency maps. (a) Frontal view with expert, (b) frontal view with layperson, (c) lateral view with expert, and (d) lateral view with layperson.

the expertise of experts, we analyzed gaze frequency in an area based on gaze position as the duration of the gaze fixation.

In this analysis, we examined two methods to measure gaze frequency: dividing the screen into square regions (grids) at equal intervals (grid gaze frequency), as described in detail in Section IV-A, and measuring the gaze frequency of the AOI, as described in detail in Section IV-B.

A model of the fixation area is important for measuring the gaze fixation duration. As setting the AOI based on facial parts requires the supervision of an orthodontist, we also considered a model that can be measured more simply by setting the grid size.

A. GRID GAZE FREQUENCY

When we counted the grid gaze frequency, the screen on which the target faces were displayed was divided into square regions (grids) (Figs. 4 (a) and (b)). The grids were numbered from the top left to the bottom right. The gaze point at each time point is located on a single grid. Therefore, each grid counted the frequency of gaze positions. The frequency of each grid was dependent on the grid size and target face size. Therefore, we analyzed the performance of distinguishing gaze while changing grid size. The grid size used is described in Section V-D.

Figures 5 (a) and (c) for the experts, and (b) and (d) for the laypersons show the normalized frequency maps when the grid size was set to 180 pixels (grid = 180, where grid = <Number> indicates that the grid has <Number> square pixels) for the gaze data (the normalization is described in Section V-A). The subject <number> in each caption is the subject number. The numbers in parentheses after the subject number in Figs. 5 (a) and (c) indicate the years of experience of the expert. In the frequency map, locations with low frequencies are indicated in blue, and as the frequency increases, they became reddish.

Before displaying the facial images in the experiments, the user was required to look at the center of the screen. At the start of the gaze measurement, the gaze started from the center; therefore, the grid frequency at that point was high for both the experts and laypersons.

A comparison of the frequency maps of the experts and laypersons revealed that the gaze frequencies of several laypersons were dispersed, indicating that they browsed the facial images without adequate intention (Figs. 5 (b) and (d)). Nevertheless, the frequency maps of the experts showed that the frequency tended to be concentrated from the center to the lower center because the target facial images were obtained from patients who required treatment of the lower jaw. It should be noted that the laypersons frequently gazed at the same grid. There were cases in which it was difficult to distinguish between experts and laypersons by considering only grid gaze frequency.

B. AOI GAZE FREQUENCY

The AOI is defined as a specific region, or area of interest. In this study, we defined AOIs as polygonal regions containing each facial part, based on the experience of one of the co-authors, who is an orthodontist. The AOIs in this study are set as shown in Figs. 6 and 7. The AOIs were numbered

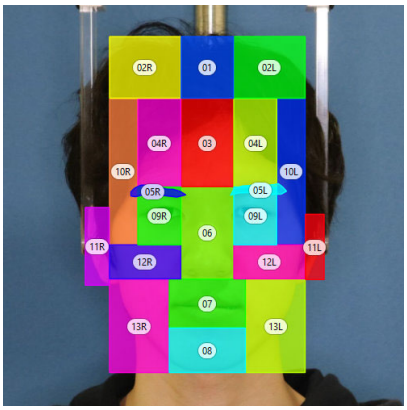


FIGURE 6. AOI of frontal view.

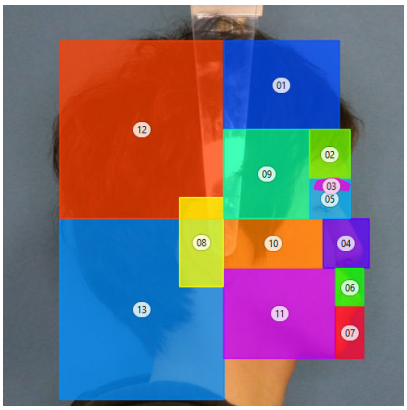


FIGURE 7. AOI of lateral view.

in order, and the AOIs corresponding to the same regions on the left and right sides were marked with R for the right side of the frontal facial image and L for the left side of the frontal facial image. The numbers used for the frontal and lateral facial images were independent and the same numbers did not indicate the same part of the face.

Because the coordinates of the AOI were set along each part of the facial image, they differed for each facial image.

Figure 8 shows examples of the normalized AOI frequency maps for Patient 3 (retrognathic mandible). As in the previous section, subject <number> in each caption is the subject number. The numbers in parentheses after the subject numbers indicate the years of expert experience. Subjects without parentheses are laypersons. In the frequency map, locations with low frequencies are indicated in blue, and as the frequency increased, they became reddish. The out-of-AOI area is shown as a white background.

A comparison of Figs. 8 (a) and (b) reveals that the experts focused on the lips (frontal face AOI Region 07). However, laypersons looked at various parts of the face. Similarly, in Figs. 8 (c) and (d), the experts focused on the area around the lips (lateral AOI region 06), whereas the laypersons focused on the back of the head and area above the ear (lateral AOI regions 12 and 09). These examples demonstrate the differences between experts and laypersons. However, other laypersons focused on the lips in lateral images. Therefore, it was difficult to distinguish between experts and laypersons by considering only AOI gaze frequency.

V. DISCRIMINATION OF PROFICIENCY

We were interested in differentiating between experts and laypersons. Our hypothesis was that the gaze fixation duration of the subjects would vary by facial area depending on their expertise. Differences were observed in the frequency of gaze per area. These frequency differences also depend on the patient's case. A simple method of discrimination based on a threshold placed on the frequency of an area was not applicable. Therefore, we propose a method for discriminating between experts and laypersons by considering the frequency of an area as a vector and determining its boundary in the projected vector space using machine learning.

First, to compare our proposed methods in terms of accuracy, we define a random selection method (RANSEL) as the base criterion. RANSEL randomly selects an individual from a group of subjects as an expert or layperson. In several trials, the average accuracy of the RANSEL converged to 0.5. In this experiment, the average accuracy of the RANSEL model over 100 trials was 0.5065. In this study, we compared our analysis methods (grid and AOI) to RANSEL.

A. FREQUENCY NORMALIZATION

As mentioned in Section III-C , because the browsing time of the facial image varied depending on the subject, the frequency of each grid was normalized by dividing it by the sum of the frequency values of all grids, which means the total browsing time of the subject. The normalized frequency vector $\mathbf{F}_{\text{grid}}(s,t)$ for subject s and target image t is denoted as follows:

$$\mathbf{F}_{\text{grid}}(s,t) = \frac{1}{\sum_{i=1}^{N_{\text{grid}}} f_{s,t}(i)} \left(f_{s,t}(1), \cdots, f_{s,t}(N_{\text{grid}})\right), \quad (1)$$

where N_{grid} is the number of grids based on the grid size and $f_{s,t}(i)$ is the gaze frequency of subject s on the $i(1, \cdots, N_{\text{grid}})$th grid.

Similarly, the normalized frequency vector $\mathbf{F}_{\text{AOI}}(s,t)$ for subject s and target image t is expressed as follows:

$$\mathbf{F}_{\text{AOI}}(s,t) = \frac{1}{\sum_{j=1}^{N_{\text{AOI}}} g_{s,t}(j)} \left(g_{s,t}(1), \cdots, g_{s,t}(N_{\text{AOI}})\right), \quad (2)$$

where j is a serial number converted from the AOI number with R and L, N_{AOI} is the number of areas based on the AOI, and $g_{s,t}(j)$ is the gaze frequency value of subject s on the $j(1, \cdots, N_{\text{AOI}})$th grid.

B. LINEAR DISCRIMINANT ANALYSIS

As mentioned above, using machine learning to configure a projection of the frequency vector that can distinguish experts from laypersons at the learned boundary within the vector space of the frequency of the gaze area, such a projection can be used to identify the level of proficiency with a new gaze vector. In preliminary studies, we applied several

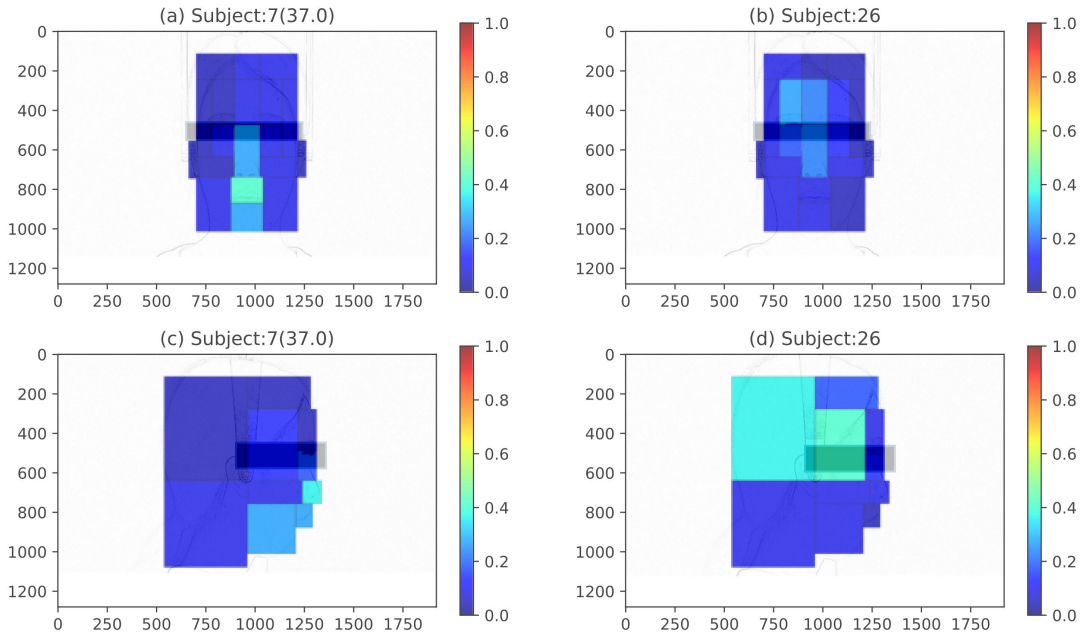


FIGURE 8. Examples of AOI frequency maps. (a) Frontal view with expert, (b) frontal view with layperson, (c) lateral view with expert, and (d) lateral view with layperson.

machine learning methods to discriminate between experts and laypersons. After examining several methods, such as the multidimensional composition scale and principal component analysis, it appeared that adequate classification could be achieved using linear discriminant analysis (LDA).

LDA, which is a supervised learning method, determines a projection that minimizes the variance within a class while maximizing the difference between the averages of the two classes.

We divide $\mathbf{F}_{\text{grid}}(s, t)$ into a two-class set:

$$C_{\text{grid},k}(t) = \{\mathbf{F}_{\text{grid}}(s, t)|s \in S_k\} \quad (k = 0, 1)$$

where S_k is a set of subjects belonging to class k, which means that the subjects are experts at $k = 0$ and the subjects are laypersons at $k = 1$.

Similarly, for $\mathbf{F}_{\text{AOI}}(s, t)$, we define

$$C_{\text{AOI},k}(t) = \{\mathbf{F}_{\text{AOI}}(s, t)|s \in S_k\}(k = 0, 1).$$

To apply LDA, we assumed $\mathbf{x} = \mathbf{F}_m(s, t)$ $(m = \{\text{grid, AOI}\})$ and obtained the following projection, $\mathbf{y}$:

$$\mathbf{y} = \mathbf{w}^{\mathsf{T}}\mathbf{x}.$$

The linear transformation vector $\mathbf{w}$ can be solved as the following eigenvalue problem:

$$\Sigma_{\text{b}}\mathbf{w} = \lambda\Sigma_{\text{w}}\mathbf{w},$$
$$\mathbf{w}^{\mathsf{T}}\Sigma_{\text{w}}\mathbf{w} = 1,$$

where Σ_{b} and Σ_{w} are between- and within-class covariance matrices, respectively.

For discrimination, the classification consisted of two classes, with laypersons indicated by one and experts by 0. The analysis tools used were Scikit-learn 0.21.3, and Python 3.7. LDA was performed using the Scikit-learn LDA module.

C. CROSS-VALIDATION

In the field of machine learning, cross-validation (CV) is a popular analysis method that avoids overfitting and cherry picking for a stable evaluation. In the following sections, we present the accuracy score as the average score of 10-fold CV. The K-fold CV divides data into k data blocks and uses $k - 1$ data blocks for learning when applying specified method. The remaining data block was used to test the accuracy of learned model. Because k data blocks exist, the validation test is repeated k times by changing the data block, and thus, the number of calculated accuracies is equal to k. Therefore, the accuracy that was determined using CV in this study was the average value of 10 tests in 10-fold CV. We selected the stratified extraction method [17] of K-fold CV, which is a stratified sampling method for test data extraction.

In the accuracy evaluation using 10-fold CV, 20 subjects were too few to allow the accuracy to be properly evaluated. Therefore, the cases were processed into one group of 60 subjects (20 × 3 cases). Facial orientations (frontal and lateral views) were still divided.

D. LDA FOR GRID GAZE FREQUENCY

When the grid gaze frequency is learned through LDA, the linear transformation vector $\mathbf{w}$ corresponding to each grid is

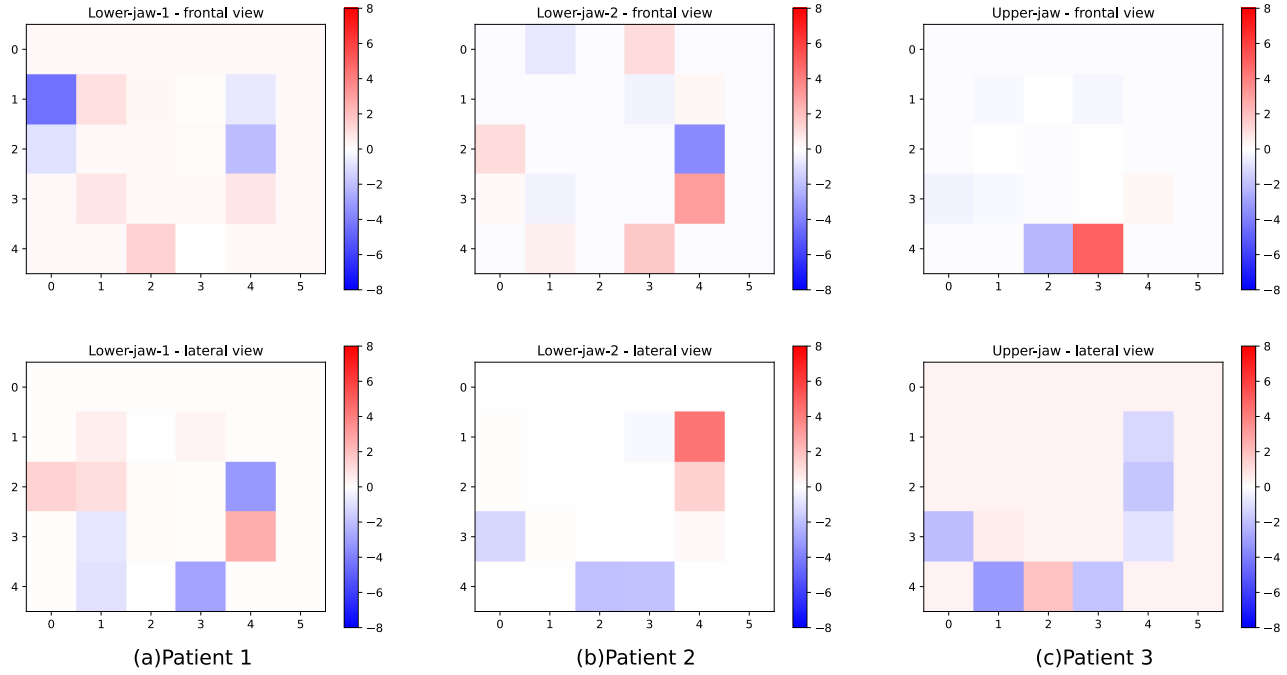


FIGURE 9. Learned transform vector w using LDA (grid size = 350 pixels).

TABLE 1. Accuracy according to various grid sizes.

Grid	100	150	200	250	300	350	400	450	500	550
Frontal view	0.617	0.533	0.6	0.533	0.667*	0.667*	0.55	0.667*	0.667*	0.65*
Lateral view	0.7*	0.55	0.667*	0.867*	0.767*	0.767*	0.767*	0.7*	0.733*	0.733*
Harmonic mean	0.656	0.541	0.632	0.66	0.714	0.714	0.641	0.683	0.698	0.689

* There were significant differences with RANSEL.

the output of the heat map, as illustrated in Fig. 9 (grid = 350). The red portion indicates a large positive effect, and the blue portion indicates a large negative effect. The average value was standardized to zero, as shown in Fig. 9. The column in the figure shows the target patients (in the caption of Fig.9, lower jaw 1 and 2 indicate that the patients had a mandibular protrusion and the upper jaw indicates that the patient had a retruded mandible). The upper part of the figure shows the frontal view, and the lower part shows the lateral view. This figure shows learning with a group of 20 subjects.

For grid gaze frequency, the discrimination accuracy changed depending on the grid size. As it was impractical to consider all grid sizes, only certain sizes were selected. Table 1 presents the values of the frontal view, lateral view, and harmonic mean when the grid size (pixels) is changed from 100 to 550 in steps of 50 pixels. The table presents the combined average accuracy of the three cases using 10-fold CV. The highest values are underlined in each row of the tables. We examined the significant difference between each result for various grid sizes and RANSEL based on the Mann-Whitney U rank test. Certain grid sizes exhibited significant differences.

The highest accuracy values for the frontal view were obtained for four sizes (300, 350, 450, and 500). The highest lateral view accuracy (0.867) was obtained with a size of 250. For the harmonic mean of the accuracy, grids 300 and 350 achieved the highest values.

We present confusion matrices of the frontal view for grid size 250 and 350 (Table 2). The results indicated that when the grid was equal to 250 (Tables 2 (a) and (b)), the estimation of the experts was appropriate for the lateral view. However, the estimation of laypersons in the frontal view could not be differentiated (Table 2 (a)). When the grid size was 350, the number of cases in which an expert was erroneously determined as a layperson decreased (Tabl. 2 (c)). Moreover, the number of cases in which a layperson was correctly determined was greater than that in which the grid was equal to 250 (Table 2 (c)). However, the proportion of laypersons who were estimated to be experts remained high. Table 2 (d) shows that the expert estimation ratio is 40:20 (expert vs. layperson). Unbalanced estimation is unsuitable for classification.

E. LDA FOR AOI GAZE FREQUENCY

The results of the AOI frequency were learned using LDA and the transformation vector was calculated. Tabl. 3 presents

TABLE 2. Confusion matrix for grid gaze frequency.

(a) Frontal view: grid size = 250

	Est. experts	Est. laypersons
Experts	22	8
Laypersons	20	10

(b) Lateral view: grid size = 250

	Est. experts	Est. laypersons
Experts	30	0
Laypersons	8	22

(c) Frontal view: grid size = 350

	Est. experts	Est. laypersons
Experts	24	6
Laypersons	14	16

(d) Lateral view: grid = 350

	Est. experts	Est. laypersons
Experts	28	2
Laypersons	12	18

TABLE 3. Accuracy for AOI gaze frequency.

Frontal view	0.617
Lateral view	0.734*
Harmonic mean	0.67

* There were significant differences with RANSEL.

TABLE 4. Confusion matrix for AOI gaze frequency.

(a) Frontal view

	Est. experts	Est. laypersons
Experts	18	12
Laypersons	9	21

(b) Lateral view

	Est. experts	Est. laypersons
Experts	22	8
Laypersons	8	22

the combined average accuracy of the three cases using 10-fold-CV. We examined the significant differences in each result of the AOI accuracy and RANSEL based on the Mann-Whitney U rank test. The results of the lateral view showed a significant difference, whereas those of the frontal view did not ($p = 0.0504$).

Similarly, we confirmed the confusion matrix using LDA based on AOI gaze frequency, as shown in Tabl. 4. The number of estimations for the experts was lower than that for grid gaze frequency. However, the number of estimations for laypersons was equal to or better than the grid gaze frequency. Furthermore, Table 4 (b) shows that the estimated numbers of experts and laypersons were balanced (i.e., 30:30).

F. ESTIMATING EXPERTS BY MAJORITY VOTE

The majority vote is a well-known method for using multiple results with the same target in the field of machine learning and is a type of ensemble learning that combines multiple classifiers. In this study, three estimations (three patients) for the frontal view and three for the lateral view were obtained for each subject using our analytical methods. The estimation of experts by majority vote means that a subject was judged as an expert when estimated as an expert two or three times. The same procedure was applied to the lateral-view case.

TABLE 5. Confusion matrix for a grid by majority vote (size = 250).

(a) Frontal view

	Est. experts	Est. laypersons
Experts	9	1
Laypersons	8	2

(b) Lateral view

	Est. experts	Est. laypersons
Experts	10	0
Laypersons	3	7

TABLE 6. Confusion matrix for a grid by majority vote (size = 350).

(a) Frontal view

	Est. experts	Est. laypersons
Experts	6	4
Laypersons	1	9

(b) Lateral view

	Est. experts	Est. laypersons
Experts	6	4
Laypersons	0	10

TABLE 7. Confusion matrix for AOI by majority vote.

(a) Frontal view

	Est. experts	Est. laypersons
Experts	7	3
Laypersons	2	8

(b) Lateral view

	Est. experts	Est. laypersons
Experts	9	1
Laypersons	1	9

Tables 5 and 6 present the confusion matrices for the grids (size = 250 and size = 350). The Grid (size = 250) resulted in 55% accuracy for accurate responses, with 11 out of 20 subjects in the frontal view, and 85% accuracy for correct responses with 17 out of 20 subjects in the lateral view. The harmonic mean of the majority vote is 66.8% when the grid size is 250. Additionally, in the frontal view, the grid (size = 350) resulted in 75% accuracy for correct responses in 15 out of 20 subjects, whereas it attained 80% accuracy for accurate responses in 16 out of 20 subjects in the lateral view. The harmonic mean of the majority vote is 77.4% when the grid size is 350.

Table 7 shows the confusion matrices for the AOI when the same person was judged as an expert if it was determined to be an expert in two or more of the three cases and as a layperson if they were judged to be an expert in one or fewer cases. The results for the frontal view were correct for 15 of the 20 subjects and those for the lateral view were correct for 18 of the 20 subjects; thus, the accuracies were 75% and 90%, respectively. The harmonic mean of the majority vote is 81.8%.

The harmonic mean of the AOI gaze frequency was better than that of the grid gaze frequency (grid size = 250,350) because the AOI result of 81.8% was greater than the grid results of 66.8% and 77.4%.

VI. DISCUSSION AND LIMITATIONS

We aim to establish a high-performance method that distinguishes between experts and laypersons. The performance in

discriminating whether a subject was an expert or layperson exhibited a certain improvement when applying an aggregation method for the gaze data, that is, the grid gaze frequency and AOI gaze frequency. We found that the estimation accuracy was improved compared to that of the RANSEL method. With the majority vote using AOI gaze frequency, the misjudgment rate of discrimination was balanced between experts or laypersons.

The grid gaze frequency does not require manual operation, such as having an expert set the AOIs in advance, and its performance can be varied by adjusting the grid size. However, it is unrelated to the facial regions because it simply divides the screen coordinates into equal squares. However, in the case of AOI gaze frequency, the cost of advanced preparation is high, because at least one expert must set the AOI by viewing each facial image. However, the difference between important and unimportant areas may lead to stable discrimination. The polygon region, as the AOI for each patient can be automatically extracted by recognizing facial parts from the facial images.

The results showed that the gaze of experts and laypersons can often be classified more appropriately for lateral facial images than for frontal facial images. During an actual visual inspection, the dentist can recognize the depth direction by simply observing it from the front. Photographs were used for this experiment. Therefore, it may be difficult to recognize the 3D depth from frontal facial images. A lateral view displays the depth information of the face, and the manner in which the depth information is handled may cause differences in the eye movements of experts and laypersons. Therefore, a lateral view can be efficiently distinguish experts from laypersons. In the future, we will examine whether the lateral view is more suitable for distinguishing between experts and laypersons through interviews with experts.

The limitation of this study is that the number of subjects in the experiment may be insufficient to reveal the effectiveness of our proposed methods; however, we used the Mann-Whitney U rank test for significant differences, which is valid for small sample sizes. Owing to the small number of subjects, orthodontists as experts are rare, and there is a low chance of cooperation from those who are busy with their daily work. As an example of this number of subjects, Shahimin and Razali [18] conducted an AOI analysis of the difference in gaze between an ophthalmologist and an optometrist when diagnosing digital fundus photographs. The subjects were eight ophthalmologists and eight optometrists. In addition, a review paper on the skill assessment of surgeons based on gaze measurements [19] showed that there are many studies in which the number of subjects in each group was 10 or less.

VII. CONCLUSION AND FUTURE WORK

To determine proficiency in orthodontic skills, we first examined an analytical method to determine whether experts and laypersons could be distinguished based on their gaze. In the experiment, we collected data from images of patients who required treatment. These images were the faces of three patients (frontal and lateral facial images of each patient), constituting a total of six images. The results of the analysis revealed that our methods distinguished between an expert and layperson using gaze frequency from RANSEL as the baseline.

There were no obvious differences in the grid frequency and AOI frequency analyses. However, the results of AOI frequency contributed to a well-balanced estimation. Furthermore, the majority vote method, which integrates the LDA results for different patients, increases distinguishing accuracy. In the future, the definition of each facial AOI should be set automatically through facial region recognition.

REFERENCES

[1] A. L. Yarbus, *Eye Movements and Vision*. Boston, MA, USA: Springer, 1967.

[2] S. Brams, G. Ziv, O. Levin, J. Spitz, J. Wagemans, A. M. Williams, and W. F. Helsen, "The relationship between gaze behavior, expertise, and performance: A systematic review," *Psychol. Bull.*, vol. 145, no. 10, pp. 980–1027, Oct. 2019.

[3] M. Nakayama and H. Harada, "Eye movement features in response to comprehension performance during the reading of programs," in *Proc. ACM Symp. Eye Tracking Res. Appl. (ETRA)*. New York, NY, USA: Association for Computing Machinery, Jun. 2020, pp. 1–5, Art. no. 55.

[4] T. Miura, "Eye movements in apprehension of bonsais: The effect of knowledge and experience," in *Proc. 16th Congr. Int. Assoc. Empirical Aesthetics*, 2000, pp. 95–96.

[5] Y. Tokitsu, "Visual scanning patterns of skilled archaeologists when observing pottery," (in Japanese), *Jpn. J. Cognit. Psychol.*, vol. 1, no. 1, pp. 75–84, 2004.

[6] E. A. Krupinski, A. A. Tillack, L. Richter, J. T. Henderson, A. K. Bhattacharyya, K. M. Scott, A. R. Graham, M. R. Descour, J. R. Davis, and R. S. Weinstein, "Eye-movement study and human performance using telepathology virtual slides. Implications for medical education and differences with experience," *Hum. Pathol.*, vol. 37, no. 12, pp. 1543–1556, Dec. 2006.

[7] H. L. Kundel, C. F. Nodine, and L. Toto, "Searching for lung nodules: The guidance of visual scanning," *Investigative Radiol.*, vol. 26, no. 9, pp. 777–781, Sep. 1991.

[8] D. V. Beard, R. E. Johnston, O. Toki, and C. Wilcox, "A study of radiologists viewing multiple computed tomography examinations using an eyetracking device," *J. Digit. Imag.*, vol. 3, no. 4, pp. 230–237, Nov. 1990.

[9] C. M. Schulz, E. Schneider, L. Fritz, J. Vockeroth, A. Hapfelmeier, T. Brandt, E. F. Kochs, and G. Schneider, "Visual attention of anaesthetists during simulated critical incidents," *Brit. J. Anaesthesia*, vol. 106, no. 6, pp. 807–813, Jun. 2011.

[10] K.-F. Kollias, C. K. Syriopoulou-Delli, P. Sarigiannidis, and G. F. Fragulis, "The contribution of machine learning and eye-tracking technology in autism spectrum disorder research: A systematic review," *Electronics*, vol. 10, no. 23, p. 2982, Nov. 2021. [Online]. Available: https://www.mdpi.com/2079-9292/10/23/2982

[11] M. E. Król and M. Król, "A novel machine learning analysis of eye-tracking data reveals suboptimal visual information extraction from facial stimuli in individuals with autism," *Neuropsychologia*, vol. 129, pp. 397–406, Jun. 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S002839321930106X

[12] B. Hosp, M. S. Yin, P. Haddawy, R. Watcharopas, P. Sa-Ngasoongsong, and E. Kasneci, "Differentiating surgeons' expertise solely by eye movement features," in *Proc. Companion Publication Int. Conf. Multimodal Interact.*, New York, NY, USA, Oct. 2021, pp. 371–375.

[13] N. Castner, E. Kasneci, T. Kübler, K. Scheiter, J. Richter, T. Eder, F. Hüttig, and C. Keutel, "Scanpath comparison in medical image reading skills of dental students: Distinguishing stages of expertise development," in *Proc. ACM Symp. Eye Tracking Res. Appl. (ETRA)*. New York, NY, USA: Association for Computing Machinery, Jun. 2018, pp. 1–9, Art. no. 39.

[14] N. Castner, J. Frankemölle, C. Keutel, F. Huettig, and E. Kasneci, "LSTMs can distinguish dental expert saccade behavior with high 'plaque-urracy,'" in *Proc. Symp. Eye Tracking Res. Appl.*, New York, NY, USA, Jun. 2022, pp. 1–7.
[15] Tobii AB. *Tobii Pro X2-30*. Accessed: Feb. 26, 2021. [Online]. Available: https://www.tobiipro.com/ja/product-listing/tobii-pro-x2/
[16] K. Holmqvist, M. Nyström, R. Andersson, R. Dewhurst, H. Jarodzka, and J. van de Weijer, *Eye Tracking: A Comprehensive Guide to Methods and Measures*. Oxford, U.K.: OUP, Sep. 2011.
[17] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Proc. 14th Int. Joint Conf. Artif. Intell. (IJCAI)*, vol. 2. San Francisco, CA, USA: Morgan Kaufmann, 1995, pp. 1137–1143.
[18] M. M. Shahimin and A. Razali, "An eye tracking analysis on diagnostic performance of digital fundus photography images between ophthalmologists and optometrists," *Int. J. Environ. Res. Public Health*, vol. 17, no. 1, p. 30, Dec. 2019.
[19] T. Tien, P. H. Pucher, M. H. Sodergren, K. Sriskandarajah, G.-Z. Yang, and A. Darzi, "Eye tracking for skills assessment and training: A systematic review," *J. Surgical Res.*, vol. 191, no. 1, pp. 169–178, Sep. 2014.

JUNZO KAMAHARA (Member, IEEE) received the B.E., M.E., and Ph.D. degrees in engineering from Osaka University, in 1992, 1994, and 1999, respectively. He joined the Kobe University of Mercantile Marine, as an Assistant Professor, in 1996, and became a Lecturer and an Associate Professor with Kobe University, in 2000 and 2003, respectively. He is currently a Professor with Osaka Seikei University, Japan. He is involved in research on multimedia technology and information retrieval, among other areas.

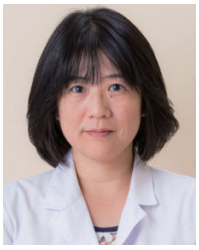
TAKASHI NAGAMATSU (Member, IEEE) received the B.S. and M.S. degrees in engineering and the Ph.D. degree in energy science from Kyoto University, in 1994, 1996, and 2004, respectively. He joined Mitsubishi Heavy Industries Ltd., in 1996. He was a Research Associate with Kyoto University, in 1999. He joined the Kobe University of Mercantile Marine, in 2000. He is currently a Professor with the Graduate School of Maritime Sciences, Kobe University. His research interest includes human–computer interaction, particularly gaze-based interactions.

KYOKO ITO received the B.S., M.S., and Ph.D. degrees in energy science from Kyoto University, in 1999, 2001, and 2004, respectively. In 2004, she joined Osaka University, as an Assistant Professor, engaged in research and education for human–computer interaction and computer-supported communication systems. From 2011 to 2012, she was with the Department of Computer Science and Engineering, University of California at San Diego, San Diego, CA, USA. She is currently a Professor with Kyoto Tachibana University.

MAMORU HIROE (Member, IEEE) received the Ph.D. degree in engineering from Kobe University, Japan, in 2022. He is currently an Assistant Professor with the Faculty of Data Science, Osaka Seikei University, Japan. He is also a Postdoctoral Fellow with the Graduate School of Medicine, Kobe University. His research interests include machine learning and gaze tracking.

HARUKI SAO received the B.S. degree in maritime sciences from Kobe University, in 2018. In 2018, he joined Nomura Research Institute Ltd. His research interest includes human–computer interaction.

SAIZO AOYAGI received the B.S., M.S., and Ph.D. degrees in energy science from Kyoto University, in 2008, 2012, and 2013, respectively. In 2017, he joined Toyo University, as an Assistant Professor. He is currently a Lecturer with Komazawa University. His research interest includes human–computer interaction, particularly communication enhancement.



JUNKO NAGATA received the D.D.S. and Ph.D. degrees from Kagoshima University, in 1990 and 1994, respectively. She is currently an Assistant Professor with the Faculty of Medicine, University of Miyazaki. She is also a Clinical Professor and the Director of the Advanced Cleft Lip and Palate Center, University Hospital. Her research interests include morphology and function of the maxillofacial complex.

KENJI TAKADA received the D.D.S. and Ph.D. degrees in dental science from Osaka University, in 1973 and 1981, respectively. He was a Visiting Assistant Professor with The University of British Columbia, from 1981 to 1983. Since 1996, he has been a Professor with Osaka University. In 2012, he honorably left Osaka University, from which he was granted the title of Professor Emeritus, and moved to the National University of Singapore, as a Visiting Professor. In 2013, he was promoted to Professor with tenure. He retired from the National University of Singapore, in 2019. He is currently with Ask-Prof Oral Health Consulting.