DEVELOPMENT AND VALIDATION OF THE KOKORO RESEARCH CENTER (KRC) FACIAL EXPRESSION DATABASE

Yoshiyuki UEDA1, Masato NUNO1, and Sakiko YOSHAKAWA1

1Kyoto University, Japan
2Seisen University, Japan

We developed a new high-resolution facial expression image database with multiple face directions and multiple gaze directions. This database was developed for usage as materials for psychological experiments and included six basic expressions (angry, disgusted, fearful, happy, sad, and surprised) and neutral expressions. Facial expressions in the database were formed in accordance with the instructions based on previous studies. To obtain the most expressive faces, models checked their facial expressions via a prompter, and we recorded them with video clips. After shooting, we extracted the most expressive images from the video clips. The size and location of the faces and lighting were perfectly controlled. With these procedures, we obtained 4,736 images in total. As a validation procedure, images in the database were presented to naive observers: some of them judged the intensity of each facial expression and others categorized the images as basic facial expressions. Furthermore, personality traits of models were evaluated and principal components of East Asian face evaluations were obtained. This database, including the results of validation experiments, is available to researchers for non-commercial academic use only.

Key words: facial expression, Japanese face, database, face evaluation, principal component analysis

Faces are remarkably meaningful objects for human beings. Unlike other objects around us, components such as the eyes, nose, and mouth have largely fixed locations, but their fine shapes and spatial relationships are slightly different depending on the person, which gives each face a personal identity. Interestingly, these shapes and spatial relationships of the components dynamically change to express the person’s emotions, which are called facial expressions. The literature has shown that, to recognize facial identity and facial expressions efficiently, faces are processed holistically rather than in a part-based manner (Tanaka & Farah, 1993), and this process is mainly conducted in the
fusiform gyrus, superior temporal sulcus, and inferior temporal gyrus in the brain (Haxby, Hoffman, & Gobbini, 2000). These mechanisms have been specifically equipped for faces, and are not observed for other objects, at least at early ages or in non-experts regarding particular objects. Based on these properties, we can recognize the states of persons around us and engage in prosocial behaviors.

The most essential material to investigate in facial processing and person-to-person social interactions is facial images. To date, various facial images have been used for this purpose and some of them are contained in databases that others can access. One of the most famous collections of facial images is Pictures of Facial Affect (POFA) by Ekman and Friesen (1978). POFA consists of 110 grayscale photographs of basic facial expressions, including angry, disgusted, fearful, happy, sad, and surprised faces. It has been widely used even in cross-cultural and neuropsychological studies. In addition, the FERST, AR, XM2VTS, Cohn-Kanade, MMI, Yale B, and Multi-PIE databases have been developed over the last decades (for a review, see Gross, 2005).

To investigate facial processing with Japanese participants, we should use a Japanese face database because there is an own-race effect (also referred to as the own-race bias, or other-race/cross-race effect), which indicates a tendency that people can more easily identify or discriminate faces of their own race than that of other races (Meissner & Brigham, 2001). However, previous databases have mainly included Caucasian faces. Some databases, such as the CMU Multi-PIE (Gross, Matthews, Cohn, Kanade, & Baker, 2008), the MR2 (Strohminger et al., 2016), and the BU-3DFE Database (Yin, Wei, Sun, Wang, & Rosato, 2006; for other database see Okubo & Takahashi, 2015), consist of huge images, including East Asian faces, but most of them are from countries other than Japan. Therefore, these faces are a little different from today’s Japanese faces.

For Japanese faces, the Japanese Female Facial Expression (JAFFE) database, the ATR DB99, the facial information norm database (FIND), the Matsumoto and Ekman’s Japanese and Caucasian facial expressions of emotion (JACFEE; Biehl et al., 1997) have been developed. In JAFFE, 10 female models showed neutral expressions as well as three or four examples of each of the six basic emotions, and 219 images in total were included (Lyons, Akamatsu, Kamachi, & Gyoba, 1998). In the ATR DB99 (developed by ATR-Promotions), 10 models, including four women, displayed 10 emotional expressions (neutral, happy with mouth open, happy with mouth closed, sad, surprised, angry with mouth open, angry with mouth closed, disgusted, contemptuous, and fearful) with multiple face directions and gaze directions. Although they were collected in well-controlled environments, the number of models was small (both included 10 models). Therefore, it is difficult to use them for memory experiments or for grasping broad tendencies in impressions extracted from Japanese faces. Although FIND included 150 Japanese neutral faces (Watanabe et al., 2007), this database is now unavailable. JACFEE included relatively larger number of individuals (32 Caucasian and 32 Japanese) as compared to JAFFE and ATR DB99, but each model showed only one expression (Biehl et al., 1997). Therefore, it is difficult to systematically discriminate the effects of identity and facial expressions using this database. Recently, a database with high
resolution and multi viewpoints was developed (Fujimura & Umemura, 2018), but this database also included a small number of models. For another strategy, computer modeling was used to manipulate personality trait impressions from faces (e.g., Oosterhof & Todorov, 2008; Todorov, Said, Engell, & Oosterhof, 2008); however, it is also difficult to compare the results of computer modeling with those of actual face images for Japanese people because there is no appropriate face database. In addition, the quality of cameras has improved incredibly during the last decade, and therefore, image resolution should also be improved.

To overcome such problems and provide a database for psychological experiments with Japanese observers, we developed a new facial expression database. The development was based on these concepts: first, we collected images in well-controlled environments (i.e., backgrounds, lighting, face direction, gaze direction, instructions on how to form facial expressions, etc.). Furthermore, we asked models to wear a hairband to prevent their hair from occluding their faces and to crop their face out from the background easily if we so desired. Second, we used high resolution cameras and strong lights to take good images without shadows on the face. Third, we asked models to display seven facial expressions (six basic expressions plus a neutral expression), and they were taken by three cameras located in different directions (as van der Schalk, Hawk, Fischer, & Doosje, 2011 did). Moreover, models moved their eyes in three different directions to enable us to manipulate gaze direction. Gaze direction is a vital signal to understand at which direction people transmit information, and person perception (e.g., attractiveness and likability), object perception, and attention can be modulated by it (Bayliss, Paul, Cannon, & Tipper, 2006; Frischen, Bayliss, & Tipper, 2007; Jones, DeBruine, Little, Conway, & Feinberg, 2006; Kampe, Frith, Dolan, & Frith, 2001; Mason, Tatkow, & Macrae, 2005). Hence, the database includes seven facial expressions × three face directions × three gaze directions. Fourth (but the most distinctive method in this database as compared to existing ones), to improve the quality of facial expressions, models could always check their own expression via a prompter, which had a one-way mirror, and behind it, the model’s face was recorded with a video recorder (see Fig. 1A). Moreover, we recorded models’ faces with video recorders, rather than still cameras, and extracted the best image with the instructed expression after recording (offline). After collecting images, we conducted experiments to examine their validity.

**Recording Methods**

*Photo Models*

Seventy-four Japanese students, including 50 women, from several universities located in Kyoto, Japan, participated as photo models. Written informed consent (purpose, methodology, risks, the right to withdraw, handling of individual information, and the voluntary nature of participation) was obtained from all models prior to taking photographs. Informed consent included a statement of whether models agreed to their photographs being shown for any psychological experiments. It also included a statement of whether they agreed to their photographs being shown for academic purposes including psychological experiments...
Fig. 1. Schema of apparatus for video recording. A) Prompter. Video recorder behind a one-way mirror can record the model’s face (1) and output this via a display. The model can see and check their face while recording (2). B) Setting of the room for recording.

(e.g., shown in publications, or in conference/lecture documents). Models were paid 2500 JPY for participation.

Apparatus
We used three video recorders (HXR-MC1 made by Sony Business Solution Corporation, https://www.sony.jp/products/Professional/mamecam/products/index.html) to record models’ facial expressions as video clips, with 1920 x 1080-pixel resolution and 30 Hz refresh rate. A prompter (MPL-32R-20 made by LIFE-ON Co., Ltd., http://www.life-on.jp/prompter/mpl-32e/) connected to one video recorder was placed in front of the models. Using this device, models could see their own facial expressions while recording them (see Fig. 1A). This enabled models to check whether they had made facial expressions accurately by themselves during recording. Another two video recorders were placed beside the prompter to record the model’s face from 45 degrees (see Fig. 1B).

Models sat on a chair during recording. The height of the prompter changed depending on each model to record their faces from the front. Moreover, the video recorders’ zoom was adjusted for each model to show the same face size in the monitor \(^1\). The model’s face was centered in the monitor against a white background. To make their face clear and remove shadows, two photographic lights (500 W Flood Photoreflector Lamp for daylight color by TOSHIBA Corp.) illuminated each model’s face from 45 degrees.

Procedure
Before recording, models were asked not to wear full makeup. In the recording room, models sat on a chair in front of the prompter, and they were asked to make angry, disgusted, fearful, happy, neutral, sad, and surprised facial expressions facing the prompter. They were given one example of each expression from Ekman and Friesen (1978) and instructed which action unit of a face should be moved to make such an expression.

First, models practiced making all expressions and checked their expressions via the prompter by themselves. In addition, the experimenter advised them how to move each action unit to form each expression. After practice, the video recorder began recording the model’s face, and the model showed the expression. Then, if the expression was not sufficiently formed, the experimenter asked them to show it again. To record the expression with averted gaze, models were asked to keep their expression and the direction of face and to move just their eyes toward the left and right video recorders, which was located 45 degrees from the front recorder. The experimenter showed a pen to models as a gaze cue and moved it slowly to each video recorder to prevent models from changing their facial direction to follow their gaze direction. Following this procedure, models gazed at the left and right video recorders twice (i.e., the order

\(^1\) The zoom control was adjusted by a very small amount. Another way to equalize face sizes was to change the distance between the model and the video recorders; however, we did not employ this strategy because changing recorder locations may have led to deviation of the angle across models.
was front, right, front, left, front, right, front, left, and front). For neutral expressions, faces with closed eyes were also recorded. After completing recording one expression, models practiced forming the next expression. The duration of recording was approximately 90 minutes.

Making Images From Video Clips
After video recording, still images were extracted from video clips for the database. The authors and assistants checked video clips of the video recorder located in front of the models, found a moment when facial muscular activities related to the emotion were maximum, and extracted this image. Then, the same moment was found from video clips of video recorders located on the left and right sides, and those images were also extracted. The procedure was conducted with all expressions and all gaze directions (i.e., front, right, and left).

Extracted images were confirmed regarding whether models showed correct facial expressions by at least two persons (either authors or an author and an assistant). Then, they were resized to a width of 851 pixels and a height of 993 pixels, adjusting the location of the middle point of the eyes to be identical to the center of the image.

With these procedures, we obtained 4,736 images in total, which included 1,628 images from the front view, 1,554 images from the right 45 degrees, and 1,554 images from the left 45 degrees (see Fig. 2 for examples). Images of each facial direction consisted of seven facial expressions (angry, disgusted, fearful, happy, neutral, sad, and surprised) × three gaze directions (direct, left-verted, and right-verted) × 74 models (50 women and 24 men). In addition, 74 neutral images with closed eyes (one for each model) were included in the front view.

VALIDATION EXPERIMENT

Although these images were checked by multiple authors or assistants, they were not presented to naïve observers. To examine the validity of the database, we conducted two experiments: one was a rating experiment, in which participants were asked to rate how strongly each image displayed each facial expression; the other was a discrimination experiment, in which participants were presented each image and asked to judge which of seven facial expressions (angry, disgusted, fearful, happy, neutral, sad, and surprised) models expressed. Furthermore, in the discrimination experiment, participants judged the attractiveness of each image. These two experiments used frontal view images with direct gaze and were conducted separately with different participants.

Participants
Thirty undergraduate and graduate students from Kyoto University participated in each validation experiment. The sample size was based on the Experiments of Ueda and Yoshikawa (2018). No participants participated in both experiments. Written informed consent (purpose, methodology, risks, the right to withdraw, handling of individual information, and the voluntary nature of participation) was obtained from all participants prior to the test. Participants were paid 1,000 JPY for participating in a one-hour experiment.

Stimuli
Five hundred and eighteen images were used in both experiments. These images
Fig. 2. Examples of images for one individual in the KRC facial expression database. The database includes seven facial expressions (anger, disgust, fear, happiness, neutral, sadness, and surprise) \times three gaze directions (direct, left-averted, and right-averted) \times three face directions (front, left-oriented, and right-oriented), and includes neutral expressions with closed-eyes images only for the front view. In total, there was 64 images for each model.
included 74 models’ frontal view images of seven facial expressions (angry, disgusted, fearful, happy, neutral, sad, and surprised) with direct gaze from the database.

**Procedure**

For the rating experiment, one image was presented to participants and they rated how strongly the model showed six facial expressions (anger, disgust, fear, happiness, sadness, and surprise) on a 7-point Likert scale \((1 = \text{very weak}, 7 = \text{very strong})\). Note that regardless of which expression the model showed, participants were asked to rate the intensity of the six facial expressions: for example, participants rated the intensity of anger, disgust, fear, happiness, sadness, and surprise even when the model displayed a happy expression. The experiment was divided into six sessions: in each session, participants rated the intensity of one expression, and the order of rating intensities and the order of presented images were randomized across participants. There were 3,108 trials in total (i.e., 518 trials in each session).

For the discrimination experiment, one image was presented to participants and they judged which of seven facial expressions (angry, disgusted, fearful, happy, neutral, sadness, and surprised) models expressed. Feedback about correct/incorrect responses was not provided to participants. The order of presented images was randomized across participants. After completing the discrimination task, each image was presented to participants again, and they were asked to rate its attractiveness on a 9-point Likert scale \((1 = \text{not very attractive}, 9 = \text{very attractive})\). The order of presented images was also randomized across participants. Therefore, there were 1,036 trials in total.

**Results**

Table 1 shows the mean and standard deviations of participants’ ratings for each facial expression image in the rating experiment. To examine whether observers felt a strong intensity of emotions that each face expressed, a 7 (facial expression of image) \(\times 6\) (intensity of each expression) repeated measure analysis of variance (rmANOVA) was conducted. It showed significant main effects of facial expression of image and intensity of each expression, \(F(6, 438) = 759.30, p < .0001, \eta^2_p = .91\) and \(F(5, 365) = 389.16, p < .0001, \eta^2_p = .84\), respectively. Furthermore, a significant interaction between them was also significant, \(F(30, 2190) = 579.97, p < .0001, \eta^2_p = .89\), indicating that observers felt different emotions depending on emotions that each face expressed. Follow-up analyses using a 6 (intensity of each expression) rmANOVA showed significant main effects in disgusted, happy, and surprised expressions, \(F(5, 365) > 629.57, ps < .0001, \eta^2_p > .90\). Multiple comparisons using Shaffer’s method indicated that their own expressions were rated significantly higher than any other expressions (adjusted \(ps < .0001\)). For fearful and sad expressions, rmANOVA showed significant main effects, \(F(5, 365) > 304.00, \eta^2_p > .81\), and multiple comparisons indicated that the intensity of their own expressions was the highest (adjusted \(ps < .0001\)), but there were no significant differences from the intensity of surprised and disgusted expressions, respectively (adjusted \(p = .31\) and \(.10\)). For angry expressions, rmANOVA showed a significant main effect, \(F(5, 365) = 802.84, p < .0001, \eta^2_p = .92\), and multiple comparisons indicated that
Table 1. Means and Standard Deviations of How Strongly Each Facial Expression Image Showed Six Basic Facial Expressions

| Rated as     | Anger       | Disgust     | Fear        | Happiness   | Sadness     | Surprise    |
|--------------|-------------|-------------|-------------|-------------|-------------|-------------|
|              | Mean ± SD   | Mean ± SD   | Mean ± SD   | Mean ± SD   | Mean ± SD   | Mean ± SD   |
| Anger        | 4.22 ± 0.53 | 4.49 ± 0.50 | 2.75 ± 0.31 | 1.51 ± 0.14 | 3.38 ± 0.52 | 2.11 ± 0.25 |
| Disgust      | 4.37 ± 0.60 | 4.86 ± 0.42 | 2.90 ± 0.38 | 1.42 ± 0.12 | 3.41 ± 0.71 | 2.15 ± 0.19 |
| Fear         | 3.71 ± 0.52 | 4.29 ± 0.51 | 4.48 ± 0.51 | 1.53 ± 0.16 | 3.49 ± 0.72 | 4.56 ± 0.80 |
| Happiness    | 1.55 ± 0.24 | 1.51 ± 0.24 | 1.49 ± 0.20 | 4.59 ± 0.69 | 1.44 ± 0.20 | 1.60 ± 0.15 |
| Neutral      | 2.35 ± 0.33 | 2.38 ± 0.35 | 2.03 ± 0.21 | 2.02 ± 0.22 | 2.26 ± 0.30 | 1.68 ± 0.24 |
| Sadness      | 3.27 ± 0.47 | 4.05 ± 0.49 | 3.42 ± 0.64 | 1.55 ± 0.15 | 4.19 ± 0.67 | 2.37 ± 0.47 |
| Surprise     | 2.59 ± 0.34 | 2.86 ± 0.33 | 3.70 ± 0.54 | 2.15 ± 0.34 | 2.27 ± 0.25 | 5.32 ± 0.67 |

Note. Cells with bold numbers indicate the same ratings as the expressions in the images shown.

The intensity of angry expressions was the second highest but significantly lower than that of disgusted expressions (adjusted $p < .0001$), indicating that angry expressions were confused with disgusted expressions.

The absolute values of ratings of each expression was 4.2–5.3 using 7-point Likert scales. This suggested that the intensity of facial expressions in the database was moderate to slightly strong. For neutral expressions, the ratings were under 2.4 for all expressions, indicating that participants perceived little facial expressions in neutral faces.

Table 2 shows the mean accuracy of participants’ discrimination for each facial expression image in the discrimination experiment. The results showed that accuracy levels were significantly greater than chance in all facial expressions (anger = 38.38%, $t(73) = 17.91, p < 2.2 \times 10^{-16}, r = .90$; disgust = 40.99%, $t(73) = 25.58, p < 2.2 \times 10^{-16}, r = .95$; fear = 34.19%, $t(73) = 22.21, p < 2.2 \times 10^{-16}, r = .93$; happiness = 86.17%, $t(73) = 57.42, p < 2.2 \times 10^{-16}, r = .99$; neutral = 93.47%, $t(73) = 108.32, p < 2.2 \times 10^{-16}, r = 1$; sadness = 53.20%, $t(73) = 19.47, p < 2.2 \times 10^{-16}, r = .92$; and surprise = 90.05%, $t(73) = 94.83, p < 2.2 \times 10^{-16}, r = 1$; chance = 14.29%).

Moreover, to examine whether observers could categorize emotions that each face expressed correctly, a 7 (facial expression of image) × 7 (categorized expression) rmANOVA was conducted. It showed a significant main effect of categorized expression, $F(6, 438) = 49.39, p < .0001$, $\eta_p^2 = .40$, and a significant interaction between facial expression of image and categorized expression, $F(36, 2628) = 517.09, p < .0001$, $\eta_p^2 = .88$. For each facial expression of image, follow-up analyses using a 7 (categorized expression) rmANOVA were conducted. The results showed that main effects of categorized expression were significant in all facial expressions of image, $F(6, 438) > 81.05, p < .0001$, $\eta_p^2 > .53$. Multiple comparisons using Shaffer’s method indicated that
Table 2. Means, Standard Deviations, Maximums, and Minimums of Categorization Performance of Each Face in the Database

|      | Anger | Disgust | Fear | Hapiness | Neutral | Sadness | Surprise |
|------|-------|---------|------|----------|---------|---------|----------|
| Anger | Mean  | 0.38    | 0.40 | 0.01 | 0.01 | 0.08 | 0.11 | 0.01 |
|      | SD    | 0.18    | 0.15 | 0.01 | 0.01 | 0.16 | 0.13 | 0.02 |
|      | Max   | 0.80    | 0.70 | 0.03 | 0.07 | 0.83 | 0.63 | 0.07 |
|      | Min   | 0.03    | 0.07 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Disgust | Mean  | 0.40    | 0.41 | 0.01 | 0.00 | 0.02 | 0.15 | 0.01 |
|      | SD    | 0.23    | 0.14 | 0.02 | 0.01 | 0.04 | 0.20 | 0.02 |
|      | Max   | 0.90    | 0.80 | 0.07 | 0.03 | 0.23 | 0.80 | 0.07 |
|      | Min   | 0.00    | 0.03 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Fear | Mean  | 0.10    | 0.10 | 0.34 | 0.01 | 0.01 | 0.11 | 0.33 |
|      | SD    | 0.09    | 0.10 | 0.13 | 0.01 | 0.02 | 0.14 | 0.23 |
|      | Max   | 0.40    | 0.47 | 0.57 | 0.03 | 0.10 | 0.60 | 0.83 |
|      | Min   | 0.00    | 0.00 | 0.03 | 0.00 | 0.00 | 0.00 | 0.00 |
| Hapiness | Mean  | 0.01    | 0.01 | 0.00 | 0.00 | 0.86 | 0.12 | 0.00 |
|      | SD    | 0.02    | 0.02 | 0.01 | 0.01 | 0.13 | 0.11 | 0.01 |
|      | Max   | 0.10    | 0.10 | 0.03 | 1.00 | 0.57 | 0.03 | 0.03 |
|      | Min   | 0.00    | 0.00 | 0.00 | 0.33 | 0.00 | 0.00 | 0.00 |
| Neutral | Mean  | 0.02    | 0.02 | 0.00 | 0.01 | 0.93 | 0.01 | 0.01 |
|      | SD    | 0.03    | 0.03 | 0.01 | 0.02 | 0.07 | 0.04 | 0.02 |
|      | Max   | 0.23    | 0.13 | 0.03 | 0.07 | 1.00 | 0.33 | 0.10 |
|      | Min   | 0.00    | 0.00 | 0.00 | 0.57 | 0.00 | 0.00 | 0.00 |
| Sadness | Mean  | 0.06    | 0.24 | 0.05 | 0.01 | 0.08 | 0.53 | 0.04 |
|      | SD    | 0.08    | 0.18 | 0.05 | 0.02 | 0.11 | 0.23 | 0.07 |
|      | Max   | 0.40    | 0.73 | 0.23 | 0.07 | 0.47 | 0.97 | 0.37 |
|      | Min   | 0.00    | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Surprise | Mean  | 0.01    | 0.01 | 0.05 | 0.02 | 0.02 | 0.01 | 0.00 |
|      | SD    | 0.02    | 0.02 | 0.05 | 0.06 | 0.02 | 0.01 | 0.08 |
|      | Max   | 0.10    | 0.10 | 0.20 | 0.53 | 0.13 | 0.03 | 1.00 |
|      | Min   | 0.00    | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.43 |

Note. Cells with bold numbers indicate the same categories as the images shown.
participants could categorize neutral, happy, sad, and surprised expressions as they were (adjusted $ps < .0001$). On the other hand, fearful expressions were frequently confused with surprised expressions (adjusted $p = 1.00$). Furthermore, angry expressions were frequently confused with disgusted expressions (adjusted $p = .78$) and vice versa (adjusted $p = 1.00$).

Attractiveness, which was measured in the discrimination experiment, is shown in Fig. 3. A one-way (facial expression of image) ANOVA showed that there was a significant main effect of facial expression, $F(6, 438) = 1805.30$, $p < .0001$, $\eta^2_p = .96$. Multiple comparisons with Shaffer’s method showed that all pairs except for anger and fear were significantly different (adjusted $ps < .0001$, but anger and fear, adjusted $p = .47$), indicating that happy expressions were rated as the most attractive as compared to other expressions.

**Discussion**

Validation experiments revealed that some facial expressions were constantly recognized as specific facial expressions (i.e., neutral, happy, and surprised). However, other facial expressions in the database were often confused even though they were made based on Ekman and Friesen’s (1978) manipulation. One possible reason for confusing some expressions is cultural differences. Previous studies suggest that East Asians, including Japanese, often confuse anger and disgust more than Westerners because of differences in how they see faces (Jack, Blais, Scheepers, Schyns, & Caldara, 2009; Jack, Garrod, Yu, Caldara, & Schyns, 2012). However, note that discrimination performance was much higher than chance in all facial expression categories. There might be another problem in the presentation method of the validity tests, specifically in the rating experiment. That is, faces were presented alone, which made facial expression recognition more difficult. As compared to this, in a previous study conducted by the authors (Ueda & Yoshikawa, 2018), participants could discriminate between angry and disgusted
expressions from this database when they were presented simultaneously.

Since the results of the validity test were based on averages of all images, some images could be recognized more precisely while others were not. The database includes all results of the validity tests, and researchers who want to use it should choose images based on them, specifically when they use angry, disgusted, and fearful expressions.

Attractiveness rating showed that happy expressions conveyed the most attractive impression as compared to other expressions. Impressions of people are often overgeneralized using a facial structure that resembles a particular emotional expression (Oosterhof & Todorov, 2008; Zebrowitz, 2017), indicating that emotional expression can also modulate impressions of them. The result in attractive rating was consistent with the findings of previous studies: O’Doherty et al. (2003) showed that brain activation (i.e., medial orbitofrontal cortex) produced by attractive faces was enhanced with smiling facial expressions.

**FACE EVALUATION EXPERIMENT**

Finally, to describe the characteristics of models in the database explicitly, we asked participants to evaluate the models’ personality traits. This is useful not only to grasp broad trends of Japanese faces in the database and compare them with other previous findings (e.g., Oosterhof & Todorov, 2008), but also to treat them as control variables when researchers want to use the database in new experiments.

In this study, we measured attractiveness, compassion, competence, distinctiveness, dominance, extroversion, maturity, and trustworthiness as the models’ personality traits. Oosterhof and Todorov (2008) demonstrated that trustworthiness and dominance are two main orthogonal dimensions of face evaluation using 2D computer modeling. Facial maturity can convey physical strength and is predicted to have a relationship with the dominance dimension (Oosterhof & Todorov, 2008). Facial attractiveness and distinctiveness relate to behavioral ratings of a face’s physical shape: the former is based on averageness and symmetry of a face, while the latter is a cue that represents how much a face deviates from the average. These evaluations would be helpful when a researcher wants to control for the memorability of a face. In addition, we asked participants to judge compassion (evaluation of empathic motivation), competence (evaluation in public office), and extroversion (evaluation of prosociality), which are evaluated in daily life as basic personality traits.

**Participants**

Thirty-one undergraduate and graduate students from Kyoto University participated in the experiment. All participants did not participate in any previous validation experiments. Written informed consent (purpose, methodology, risks, the right to withdraw, handling of individual information, and the voluntary nature of participation) was obtained from all participants prior to the test. Participants were paid 500 JPY for participating in a half-hour experiment.
Stimuli

Seventy-four images of each model’s frontal view with a neutral expression with direct gaze were used. In this experiment, facial expressions other than neutral were not presented.

Procedure

At first, all images used in this study were presented to participants for one second per image. This enabled participants to know what images were presented and to rate them in a larger range of scales. After completing the presentation, participants were instructed what personality traits they had to judge and presented with each image again. Participants kept judging the same personality traits for all images, and then the judgment was changed after an arbitrary short break.

Participants judged seven personality traits (attractiveness, compassion, competence, distinctiveness, dominance, extroversion, maturity, and trustworthiness) on a 9-point Likert scale (1 = very weak, 9 = very strong). Hence, each image was presented eight times in total (look through in the first block and seven types of judgments). The order of judgment and the order of image presentation in each judgment was randomly determined across participants.

The experiment was divided into two sessions: one for male images and the other for female images. The order of sessions was counter-balanced across participants.

Results

Table 3 shows the average, standard deviation, maximum, and minimum of mean ratings of each model. The results shown in Table 3 indicated that average ratings were close to the middle point of the scale (= 5), except for attractiveness judgments, in which the average rating was lower than the middle point. Moreover, Table 4 shows the correlation matrix among evaluations. The matrix indicated that personality traits estimated from faces were roughly divided into two groups: one consisted of attractiveness, compassion, competence, extroversion, and trustworthiness; the other consisted of distinctiveness, dominance, and maturity. Correlations between estimated personality traits within each group were relatively higher as compared to correlations across groups.

To identify dimensions of face evaluation, we conducted a principal components analysis (PCA) using these ratings. The results of PCA are shown in Fig. 4. The first principal component (PC1) accounted for 52.6% of the variance in the evaluations. PC1 loaded strongly on attraction, compassion, competence, extroversion, and trustworthiness, suggesting that it can be interpreted as being the same as valence evaluation, which was PC1 proposed in Oosterhof and Todorov (2008; which accounted for 63.3% of the variance in evaluations in their study). The second principal component (PC2) accounted for 23.1% of the variance in the evaluations. It loaded strongly on distinctiveness, dominance, and maturity, suggesting that it can also be interpreted as being the same as dominance evaluation, which was PC2 proposed in Oosterhof and Todorov (2008; which accounted for 18.3% of the variance in evaluations in their study).
|                      | Attractiveness | Compassion | Competence | Distinctiveness | Dominance | Extroversion | Maturity | Trustworthiness |
|----------------------|----------------|------------|------------|----------------|-----------|--------------|----------|----------------|
| **Male Images**      | **Mean**       | 4.22       | 4.78       | 4.80           | 4.74      | 4.95         | 5.18     | 5.39           | 4.58 |
|                      | **SD**         | 0.78       | 0.82       | 0.80           | 0.85      | 0.61         | 0.78     | 0.74           | 0.74 |
|                      | **Max**        | 6.00       | 6.42       | 5.90           | 6.97      | 5.97         | 6.52     | 7.26           | 5.90 |
|                      | **Min**        | 2.87       | 3.23       | 2.84           | 3.68      | 3.65         | 3.06     | 4.45           | 3.13 |
| **Female Images**    | **Mean**       | 4.00       | 4.79       | 4.96           | 4.84      | 5.05         | 4.87     | 5.31           | 4.77 |
|                      | **SD**         | 0.82       | 0.77       | 0.86           | 0.83      | 0.77         | 1.02     | 0.96           | 0.70 |
|                      | **Max**        | 6.06       | 6.23       | 6.77           | 7.35      | 6.61         | 7.16     | 7.26           | 6.45 |
|                      | **Min**        | 2.58       | 3.35       | 3.23           | 3.45      | 3.74         | 2.94     | 3.26           | 3.29 |
| **All Images**       | **Mean**       | 4.07       | 4.78       | 4.91           | 4.81      | 5.02         | 4.97     | 5.33           | 4.70 |
|                      | **SD**         | 0.81       | 0.78       | 0.84           | 0.83      | 0.72         | 0.96     | 0.89           | 0.71 |

Table 4. Correlation Matrix among Face Evaluations

|                      | Attractiveness | Compassion | Competence | Distinctiveness | Dominance | Extroversion | Maturity | Trustworthiness |
|----------------------|----------------|------------|------------|----------------|-----------|--------------|----------|----------------|
| Attractiveness       | –              |            |            |                |           |              |          |                |
| Compassion           | .71            | –          |            |                |           |              |          |                |
| Competence           | .85            | .72        | –          |                |           |              |          |                |
| Distinctiveness      | .05            | –.06       | .14        | –              |           |              |          |                |
| Dominance            | .30            | –.06       | .45        | .42            | –         |              |          |                |
| Extroversion         | .74            | .53        | .75        | .32            | .65       | –            |          |                |
| Maturity             | .03            | .05        | .33        | .19            | .44       | .26          | –        |                |
| Trustworthiness      | .80            | .91        | .84        | –.09           | .11       | .57          | .12      | –              |
Fig. 4. The result of principal component analysis (PCA) with the first principal component (PC1) and the second principal component (PC2). Loading of each evaluation is described in the figure using “arrows.”

Moreover, to examine the effect of gender, we conducted a PCA using ratings for female and male faces separately. The results with only female faces showed that PC1 loaded strongly on attraction, compassion, competence, extroversion, and trustworthiness, and accounted for 55.1% of the variance. PC2 loaded strongly on distinctiveness, dominance, and maturity, and accounted for 24.2% of the variance. The results with only male faces showed that PC1 loaded strongly on attraction, compassion, competence, and trustworthiness, and accounted for 49.5% of the variance. PC2 loaded strongly on distinctiveness, dominance, and maturity, and accounted for 20.9% of the variance. Although there were subtle differences between female and male (e.g., extroversion contributed to PC1 for female faces while it contributed to both PC1 and PC2 for male faces), trends were the same across females and males, indicating that PC1 can be interpreted as valence evaluation, and PC2 can be interpreted as dominance evaluation. The results of PCA for female and male faces are shown in the Appendix.
Discussion

These evaluations revealed the characteristics of faces in the database, and average ratings can be used as control variables in future experiments. Moreover, PCA results indicated that dimensions of face evaluation using real Japanese face images were the same as that in a previous study generated using 2D computer modeling (Oosterhof & Todorov, 2008). Valence dimension relates to signaling whether to approach or avoid a person, whereas dominance dimension relates to signaling the physical strength. This result suggests that dimensions of face evaluation might be consistent across races, cultures, and real images versus elaborated computer modeling. Recently, Sutherland et al. (2018) also showed consistency of facial impressions across cultures (i.e., China and the United Kingdom) and races (Asian and Caucasian). The result in this study supports the findings of those studies.

Note that Sutherland et al. (2018) proposed capability as a dimension instead of dominance, because dominance or physical power were not frequently mentioned in unconstrained facial impressions. The authors have proposed that the concept of dominance evaluation has been confused and should be divided into two (Ueda & Yoshikawa, 2018): trait dominance, which refers to an individual’s competence and tendency to engage in dominant behavior; and relative dominance, which refers to having social power or competence to command obedience in a social group. The later concept is close to capability, which reflects social ability, social attributes, or wisdom. Dominance in this experiment was measured with a single face without assuming any social groups (which was the same situation for measuring trait dominance in Experiment 1 of Ueda & Yoshikawa, 2018). Therefore, the present results concerning dominance might be related to a trait dominance or physical power evaluation, but not a relative dominance or capability evaluation.

General Discussion

In this paper, we developed a new facial expression database consisting of Japanese faces, and conducted two experiments to examine its validity and one experiment to grasp rough trends of models in database. As a result of the experiments, we can confirm that faces in the database showed appropriate facial expressions depending on the categories of anger, disgust, fear, happiness, neutral, sadness, and surprise. Moreover, the results of face evaluation demonstrated how to control for differences in models’ perceived personality traits.

Although face expressions in the database were formed in accordance with the explanations provided by Ekman and Friesen (1978) and discrimination performance was much higher than chance, angry, fearful, and disgusted expressions were often confused. This has often been observed in previous experiments with East Asian observers (Jack et al., 2009; Jack et al., 2012). Facial expressions in these previous studies and the present database were made based on Ekman and Friesen’s (1978) manipulation and might be different from natural expressions for Japanese (East Asian) customs. The primary
objective in this study is to make a facial expression database, in which facial expressions were formed by the same definitions, and natural (or specific) facial expressions in each race (or culture) should be further investigated as another interesting problem to be solved. Since discrimination performance seemed to be satisfactory, the database has a certain amount of validity.

A by-product of face evaluation revealed an interesting finding, which is that dimensions of face evaluation were consistent with a previous study (Oosterhof & Todorov, 2008) regardless of differences in race, culture, and use of actual individual versus computer modeling images. These results suggest that findings using manipulation of a computer modeling face, which has been often used recently, can be adopted for real person faces.

The database, including rating intensity, discrimination performance, and face evaluation, is available as the Kokoro Research Center (KRC) facial expression database for academic use (i.e., materials in psychological experiments). Notable features of this database are as follows: i) quite high quality and large enough for usage in psychological experiments; ii) illuminated with powerful photographic lights and full-colored; and iii) variety of images for all seven facial expressions, including multiple face angles (front, left, and right views) and gaze directions (direct gaze and left- and right-averted gaze). In addition, we plotted points of face features to morph between neutral and emotion expression faces. Researchers who want to use it should contact the first author (YU) of this paper. The database is forbidden for commercial use because the informed consent provided did not specify such purposes. Since each model declared whether their images can be shown in journal publications and conference/lecture documents, people who want to show the images must follow the database guidelines. (Please contact the author directly for details.)

AUTHOR CONTRIBUTIONS

Y.U., M.N., and S.Y. designed the recording and rating methods. M.N. made video clips, and Y.U. and M.N. made images from the video clips and conducted the rating experiment. Y.U. first wrote the manuscript, and M.N. and S.Y. revised it. All authors confirmed the final version of the manuscript for submission.

CONFlict of interest

The authors declare no conflicts of interest.

REFERENCES

Bayliss, A. P., Paul, M. A., Cannon, P. R., & Tipper, S. P. (2006). Gaze cuing and affective judgments of objects: I like what you look at. Psychonomic Bulletin & Review, 13, 1061–1066. doi: 10.3758/
Biehl, M., Matsumoto, D., Ekman, P., Hearn, V., Heider, K., Kudoh, T., & Ton, V. (1997). Matsumoto and Ekman’s Japanese and Caucasian facial expressions of emotion (JACFEE): Reliability data and cross-national differences. *Journal of Nonverbal Behavior, 21*, 3–21. doi: 10.1023/A:1024902500935

Ekman, P., & Friesen, W. V. (1978). *Manual for the facial action coding system*. Palo Alto, CA: Consulting Psychologists Press.

Frischen, A., Bayliss, A. P., & Tipper, S. P. (2007). Gaze cueing of attention: Visual attention, social cognition, and individual differences. *Psychological Bulletin, 133*, 694–724. doi: 10.1037/0033-2909.133.4.694

Fujimura, T., & Umemura, H. (2018). Development and validation of a facial expression database based on the dimensional and categorical model of emotions. *Cognition and Emotion, 32*, 1663–1670. doi: 10.1080/02699931.2017.1419936

Gross, R. (2005). Face databases. In S. Z. Li & A. K. Jain (Eds.), *Handbook of face recognition* (pp. 301–327). New York, NY: Springer-Verlag.

Gross, R., Matthews, I., Cohn, J., Kanade, T., & Baker, S. (2008). Multi-PIE. *Proceedings of the 8th IEEE International Conference on Automatic Face and Gesture Recognition*, 1–8. doi: 10.1109/AFGR.2008.4813399

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences, 4*, 223–233. doi: 10.1016/S1364-6613(00)01482-0

Jack, R. E., Blais, C., Scheepers, C., Schyns, P. G., & Caldara, R. (2009). Cultural confusions show that facial expressions are not universal. *Current Biology, 19*, 1543–1548. doi: 10.1016/j.cub.2009.07.051

Jack, R. E., Garrod, O. G. B., Yu, H., Caldara, R., & Schyns, P. G. (2012). Facial expressions of emotion are not culturally universal. *Proceedings of the National Academy of Sciences of the United States of America, 109*, 7241–7244. doi: 10.1073/pnas.1200155109

Jones, B. C., DeBruine, L. M., Little, A. C., Conway, C. A., & Feinberg, D. R. (2006). Integrating gaze direction and expression in preferences for attractive faces. *Psychological Science, 17*, 588–591. doi: 10.1111/j.1467-9280.2006.01749.x

Kampe, K. K. W., Frith, C. D., Dolan, R. J., & Frith, U. (2001). Reward value of attractiveness and gaze. *Nature, 413*, 589. doi: 10.1038/35098149

Lyons, M., Akamatsu, S., Kamachi, M., & Gyoba, J. (1998). Coding facial expressions with Gabor wavelets. *Proceedings of the 3rd IEEE International Conference on Automatic Face and Gesture Recognition*, 200–205. doi: 10.1109/AFGR.1998.670949

Mason, M. F., Tatikow, E. P., & Macrae, C. N. (2005). The look of love: Gaze shifts and person perception. *Psychological Science, 16*, 236–239. doi: 10.1111/j.0956-7976.2005.00809.x

Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law, 7*, 3–35. doi: 10.1037/1076-8971.7.1.3

O’Doherty, J., Winston, J., Critchley, H., Perrett, D., Burt, D. M., & Dolan, R. J. (2003). Beauty in a smile: The role of medial orbitofrontal cortex in facial attractiveness. *Neuropsychologia, 41*, 147–155. doi: 10.1016/S0028-3932(02)00145-8

Okubo, K., & Takahashi, M. (2015). Overview of facial stimuli databases and that contain Asian faces. *Bulletin of the Graduate School of Education, the University of Tokyo, 55*, 193–205.

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences of the United States of America, 105*, 11087–11092. doi: 10.1073/pnas.0805664105

Strohminger, N., Gray, K., Chituc, V., Heffner, J., Schein, C., & Heagins, T. B. (2016). The MR2: A multi-racial, mega-resolution database of facial stimuli. *Behavior Research Methods, 48*, 1197–1204. doi: 10.3758/s13428-015-0641-9

Sutherland, C. A. M., Liu, X., Zhang, L., Chu, Y., Oldmeadow, J. A., & Young, A. W. (2018). Facial first impressions across culture: Data-driven modeling of Chinese and British perceivers’ unconstrained facial impressions. *Personality and Social Psychology Bulletin, 44*, 521–537. doi: 10.1177/0146167217744194

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *The Quarterly Journal of Experimental Psychology Section A, 46*, 225–245. doi: 10.1080/14640749308401045

KRC FACIAL EXPRESSION DATABASE
Todorov, A., Said, C. P., Engell, A. D., & Oosterhof, N. N. (2008). Understanding evaluation of faces on social dimensions. *Trends in Cognitive Sciences, 12*, 455–460. doi: 10.1016/j.tics.2008.10.001

Ueda, Y., & Yoshikawa, S. (2018). Beyond personality traits: Which facial expressions imply dominance in two-person interaction scenes? *Emotion, 18*, 872–885. doi: 10.1037/emo0000286

van der Schalk, J., Hawk, S. T., Fischer, A. H., & Doosje, B. (2011). Moving faces, looking places: Validation of the Amsterdam Dynamic Facial Expression Set (ADFES). *Emotion, 11*, 907–920. doi: 10.1037/a0023853

Watanabe, N., Suzuki, R., Yoshida, H., Tsuzuki, D., Bamba, A., Chandrasiri, N. P., . . . Yamada, H. (2007). Facial Information Norm Database (FIND): Constructing a database of Japanese facial images. *Japanese Journal of Research on Emotions, 14*, 39–53. doi: 10.4092/jsre.14.39

Yin, L., Wei, X., Sun, Y., Wang, J., & Rosato, M. J. (2006). A 3D facial expression database for facial behavior research. *Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition*, 211–216. doi: 10.1109/FGR.2006.6

Zebrowitz, L. A. (2017). First impressions from faces. *Current Directions of Psychological Science, 26*, 237–242. doi: 10.1177/0963721416683996

(Manuscript received 19 March, 2019; Revision accepted 17 September, 2019; Released online in J-STAGE as advance publication 28 May, 2020)
Fig. A1. The result of principal component analysis (PCA) for female faces. Loading of each evaluation is described in the figure using “arrows.”
Fig. A2. The result of principal component analysis (PCA) for male faces. Loading of each evaluation is described in the figure using “arrows.”