International Journal of Smart Computing and Artificial Intelligence
International Institute of Applied Informatics
2024, Vol. 8, No. 1, IJSCAI820

Human Brain Judgment and Automated Classification of Masked Facial Expressions

Koji Kashihara [*], Mizuki Shinguu [*]

Abstract

We investigated brain activity in response to facial expressions wearing masks. N170 responses at the T5 and T6 sites were synchronized with the vertex positive potential (VPP) response at the Cz site. The N170 responses were increased under masked face conditions, which may be associated with amodal completion. We then tested the facial emotion recognizer (FER) as a general classifier and the specifically created classifiers based on convolutional neural networks (CNNs) for predicting masked facial expressions. Although the accuracies in the FER were greatly lower for Japanese faces with masks than without masks, the specific CNN classifier improved the accuracies under the masked conditions.

Keywords: facial expressions, face masks, brain activity, classifiers

1 Introduction

For people to realize smooth communication, it is crucial that they discern emotional states from facial expressions, and face-to-face communication usually provides rich emotional information. With the ongoing spread of infectious diseases (e.g., COVID-19), there are more opportunities for people to wear face masks. Interacting with someone wearing a mask makes it difficult to grasp their facial expressions, presumably causing poor communication [1]. However, even when a face mask hides the nose and mouth, one can still recognize the person's face and estimate the facial parts and expressions. Such completion, predicting hidden parts based on empirical rules, is called "amodal completion" [2].

Emotional prediction of masked faces necessitates advanced judgment underlain by various brain activities. Event-related potentials (ERPs) in electroencephalogram (EEG) signals are an efficient index to estimate brain activity during perceptual or cognitive tasks. N170, vertex positive potential (VPP), and late positive potential (LPP) are known as the representative components of ERPs for facial recognition. In particular, the N170 waveform is observed in the posterior temporal lobes (T5 and T6 sites) during facial perception and recognition. Brain source analysis of N170 has revealed the facial area of the brain—the fusiform gyrus (FG) and superior temporal sulcus (STS) [3][4]—and its latency reflects the process of constructing faces.

Furthermore, the N170 amplitude is related to affective facial expressions [5] and increased by

[*] Ritsumeikan University, Shiga, Japan

emotional expressions [6]. The stimulus only in terms of eye information increases the amplitude of the N170 response and prolongs its latency [7]. However, it is unclear whether any differences exist in the N170 response to the faces or expressions of a person wearing a mask. The VPP response is a specific ERP component of facial expressions [8], indicating a similar latency to N170. Brain source analysis of VPP has identified the same area as N170 [4]. The LPP can be measured in the central and parietal areas, with a relatively long latency. The LPP is associated with high emotional arousal levels, and its amplitude is higher for pleasant or unpleasant facial expressions than for neutral faces [9].

If facial expressions are automatically and accurately identified from real-time camera images even under masks, it could help us predict emotions and facilitate nonverbal communication in our daily lives (e.g., patient care in hospitals, collaborate meetings, and school classes and seminars). Specially, convolutional neural networks (CNNs) can automatically create target filter kernels from input-output data, to divide biological images into categories [10]. Because CNNs are also suitable for analyzing facial emotion recognition [11], they could be used to efficiently assess the features of masked facial images, showing high accuracy in the classification of facial expressions. The facial emotion recognizer (FER) [11][12] has been developed as a general technique for the automatic detection of expressions; however, its effectiveness is unclear with masked faces. The optimal settings for the network structure and parameters in the CNN should be considered to increase the generalization to the specific case of masks, modifying the traditional FER.

The first purpose of this study was to assess the brain processing of faces with and without masks by analyzing the features of N170 and VPP responses. We also compared the effects of facial expressions on behavioral characteristics (i.e., accuracy and reaction times). Finally, the automatic classifiers were tested to recognize facial expressions with and without masks, and the results were compared to human judgment. Specific CNN classifiers were constructed to modify the performance of the traditional classifier [11][12] and identify Japanese facial expressions even in masked cases. The abstracted EEG features for this study could be applicable for automatically classifying facial expressions in future works.

2 EEG Study

We evaluated brain activity when recognizing emotional faces (neutral, smiling, fearful, and sad) with and without masks, and we measured EEG signals to analyze the response features of N170 (T5 and T6) and VPP (Cz).

2.1 Participants

The participants in this study were eight healthy male students (mean age: 21.4 ± 0.9 years) at Ritsumeikan University (Shiga, Japan). The experimenter checked the health condition of each participant, and they all had normal vision (including correction). After sufficiently explaining the experimental procedure, we obtained a signature on the consent form from each participant.

2.2 Stimuli and Environments

(1) Visual stimuli: Four types of images (4 expressions × 6 persons = 24 images) were selected from the ATR Facial Expression Image Database (DB99) (ATR-Promotions, Kyoto, Japan): fearful, neutral, smiling, and sad faces (two Japanese males and four Japanese females). For each image, we prepared an image with a face mask (24 images in total). The images were converted to grayscale (8-bit) and adjusted to the same luminance. The images, with a size of 380 × 380 pixels, were presented on a display screen (resolution: 1,024 × 768 pixels; size: 23.8 inches) in a blackout room. The viewing distance between the participant and the display was set at 60 cm, and the viewing angle to the presented object was set within 5°. The room temperature was kept at 24°C.

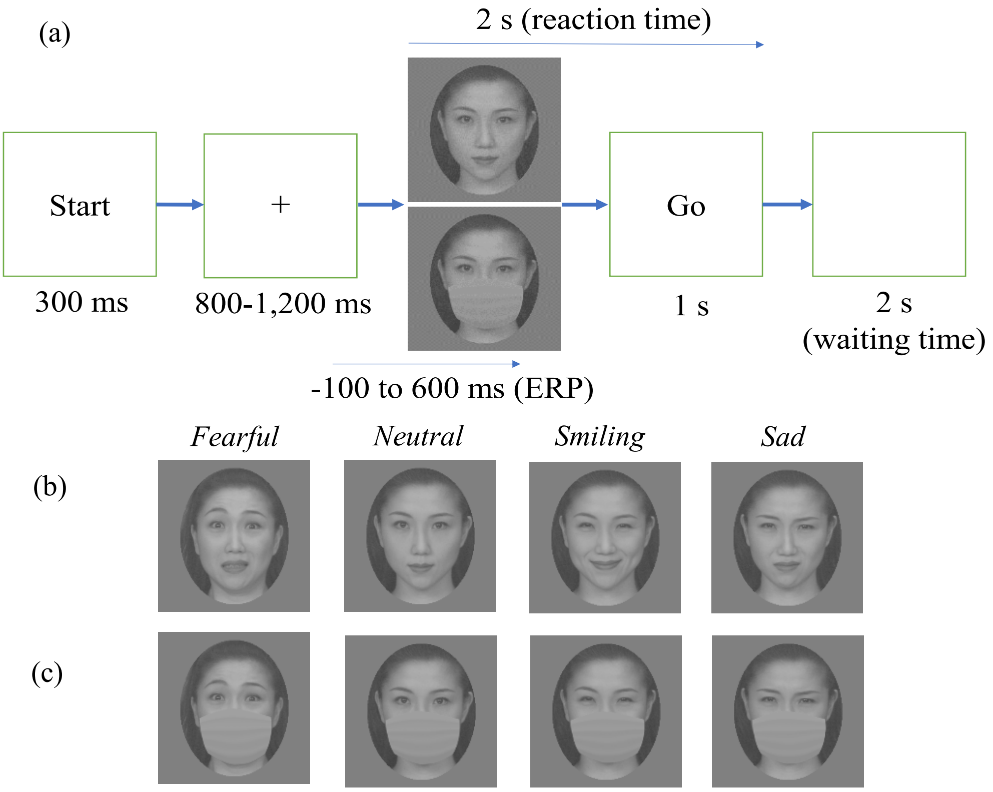

Figure 1: (a) The experimental protocol for the EEG study. Four types of expressions (i.e., fearful, neutral, smiling, and sad faces) (b) without or (c) with masks.

(2) EEG recordings: We continuously recorded EEG signals using BIOPAC MP150 and EEG100C amplifiers (sampling rate: 500 Hz). The EEG signals were measured at three locations, in accordance with the International 10-20 method, as follows: Cz related to the VPP; T5 and T6 related to the N170. The electrode for the body ground was located on the left ear, and that of the reference was on the right ear.

2.3 Experimental Procedure

After an initial practice task was sufficiently performed, we conducted the EEG experiment. In each trial (Figure 1), the "Start" signal was presented (300 ms) at the center of the presentation position of the stimulus image. After the "+" sign as the gazing point was presented for 800 to 1,200 ms, one of four different facial expressions was presented for 1 s. The "+" sign was then presented again (1 s). Participants were instructed to respond quickly and accurately in the period (i.e., 2 s) between the appearance of the face image and the disappearance of the last "+" sign. They were also instructed to concentrate on the next trial without worrying about making mistakes. Four keys on the keyboard (i.e., "D," "F," "J," and "K"), corresponding to the index and middle fingers of both hands, were used for responding to the four expressions. The correspondence to the keys was randomly changed for each participant.

One testing block consisted of 48 trials. In each trial of each block, one of 48 face images (each without or with a mask) was randomly displayed. We conducted two blocks during the practice task and four blocks during the EEG experiment. The break between each block lasted for a few minutes. To reduce noise during EEG recording, the experimenter instructed each participant to avoid blinking as much as possible between the "Start" signal and the end of the facial image presentation and not to move their head or entire body.

2.4 Data Analysis

(1) ERPs: The analyzed interval was set between -100 and 600 ms before and after presenting a face image (baseline: 100 ms immediately before presenting the face image). In preliminary experiments with three participants (i.e., the same protocol as the main experiment), the reaction times with a button press took at least 300 ms (i.e., the fastest reaction time), meaning there were no significant effects of the button press on the earlier ERP components such as N170 and VPP. If the EEG amplitude exceeded ±80 μV due to the influence of blinking or body movement, the trial was excluded from the data analysis.

(2) VPP and N170: After computing an addition average for each experimental condition and for each participant, the grand average was calculated across all participants. At the T5 and T6 sites, the minimum value between 150 and 250 ms after the stimulus onset of a face image was defined as the N170. At the Cz site, the maximum value between 150 and 250 ms after stimulus onset was defined as the VPP. After the N170 and VPP values (i.e., amplitude and latency) were averaged for each experimental condition, the grand averages were computed across all participants.

(3) Behavioral performance: For each subject and each experimental condition, we calculated the accuracy percentage and reaction times. Trials without a response were excluded from the data analysis. Grand averages were computed across all subjects.

(4) Statistics: Statistical treatment indicated a within-subjects analysis of variance (ANOVA) for two factors (facial expressions × presence of a mask). For significant results (p

< .05), multiple comparisons (Holm's method) were performed.

2.5 Experimental Results

(1) ERPs: Figure 2 shows the ERPs (VPP at Cz and N170 at T5 and T6) in facial expressions without (*left*) and with (*right*) masks. Overall, the VPP latencies were synchronized with the N170 responses. The N170 amplitudes were higher in expressions with masks than in those without masks, they were deeper at T6 (the *right* electrode) than at T5 (the *left* electrode).

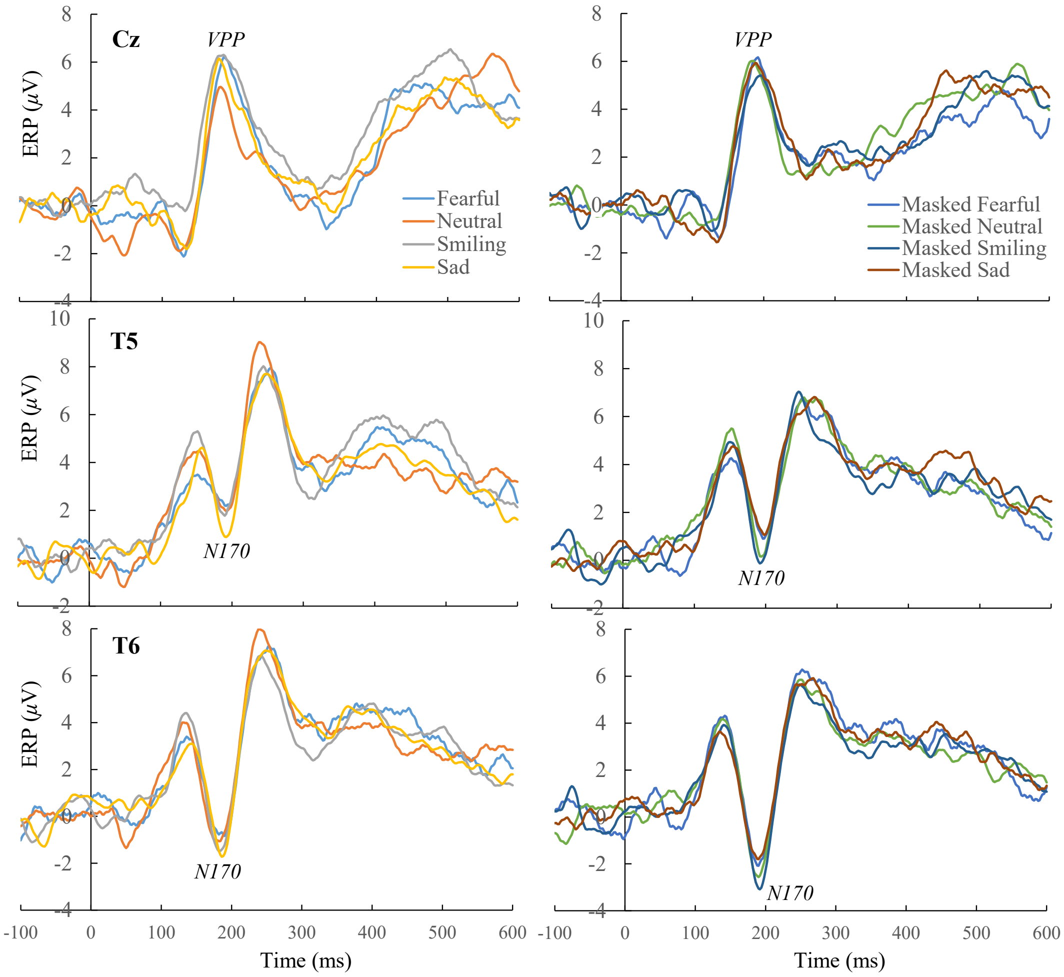

Figure 2: ERPs in facial expressions without (*left*) and with (*right*) masks: VPP at Cz (*top*) and N170 at T5 (*middle*) and T6 (*bottom*).

Figure 3 (*left*) represents the peak amplitudes of the N170 and VPP responses. The ANOVA revealed a significant tendency for the interaction at T5 [$F_{(3,21)} = 2.65$, $p < .1$] and a significant main effect of the mask at T6 [$F_{(1,7)} = 8.47$, $p < .05$]. In particular, the N170 response tended to be greater for smiling faces with masks. We suggest that predicting the invisible facial parts (especially around the mouth) because of the mask resulted in a larger

N170 amplitude.

Figure 3 (*right*) indicates the latencies of the N170 and VPP responses. In N170 at T5, the ANOVA showed significant main effects of the factors for the mask [$F_{(1,7)} = 10.75$, $p < .05$] and facial expression [$F_{(3,21)} = 4.49$, $p < .05$]. For the facial expression, the multiple comparisons showed a significant difference between the fearful and neutral conditions ($p < .05$). In N170 at T6, the ANOVA revealed a significant main effect of the mask factor [$F_{(1,7)} = 5.95$, $p < .05$].

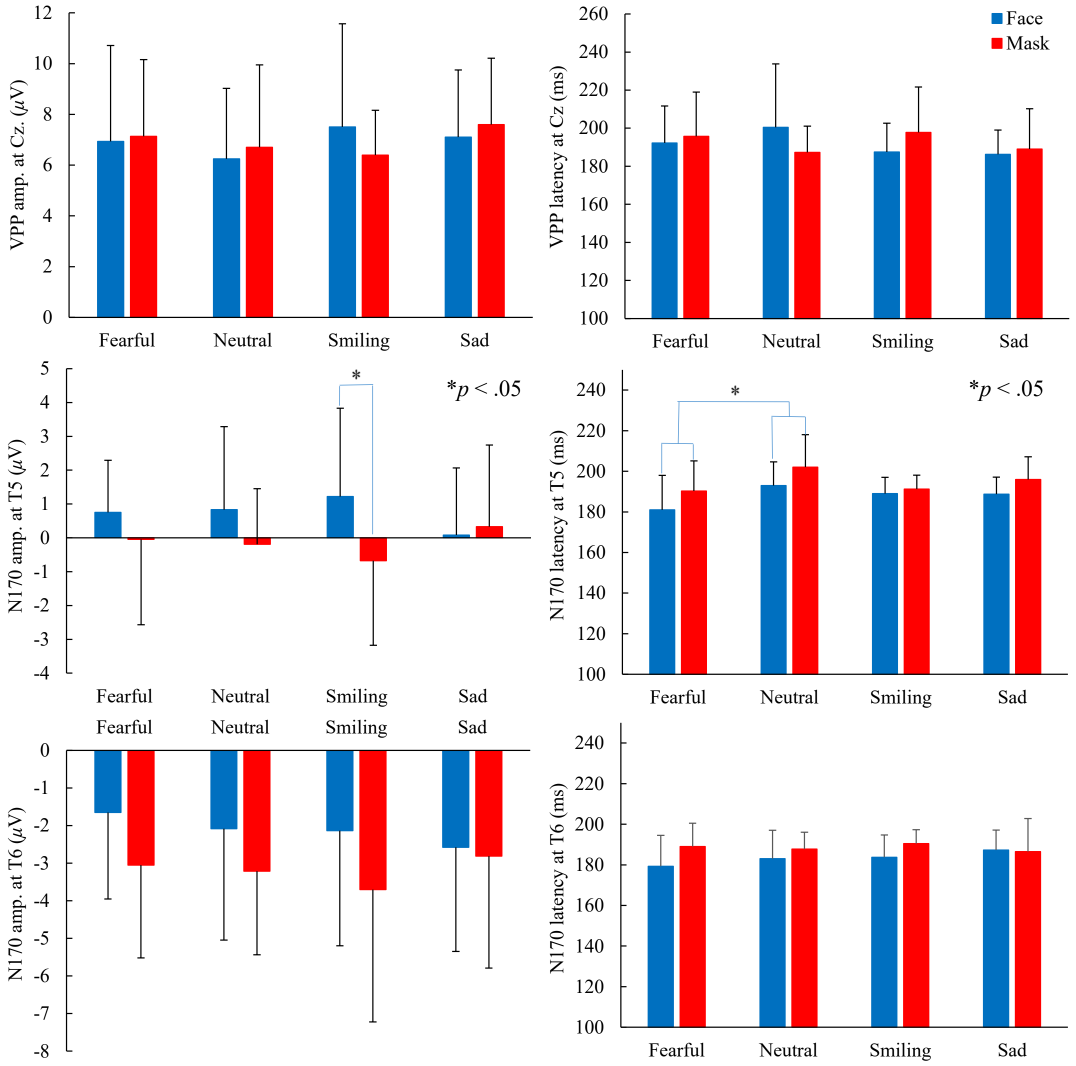

Figure 3: The peak amplitude values (*left*) and latencies (*right*) of the VPP at Cz (*top*) and the N170 at T5 (*middle*) and T6 (*bottom*): mean ± S.D. Facial expressions (i.e., fearful, neutral, smiling, and sad) without (*blue*) and with (*red*) masks.

(2) Behavioral performance: Accuracy and reaction times were calculated for each condition. Overall, for masked faces, the accuracy was worse for negative emotions, and the reaction times were delayed. Figure 4 shows the accuracy results (*left*) and reaction times (*right*). The ANOVA indicated that main effects existed for the mask [$F_{(1,7)} = 30{,}89$, $p < .01$]

and facial expression [$F_{(3,21)} = 6.57, p < .01$] in terms of accuracy, and for the mask [$F_{(1,7)} = 31.60, p < .01$] and facial expressions [$F_{(3,21)} = 11.04, p < .01$] in terms of reaction times. In particular, accuracy was significantly decreased for sad expressions. Regarding the reaction times, the interaction was also significant [$F_{(3,21)} = 3.48, p < .05$]. The detailed results of the multiple comparisons are shown in Figure 4 (*right*).

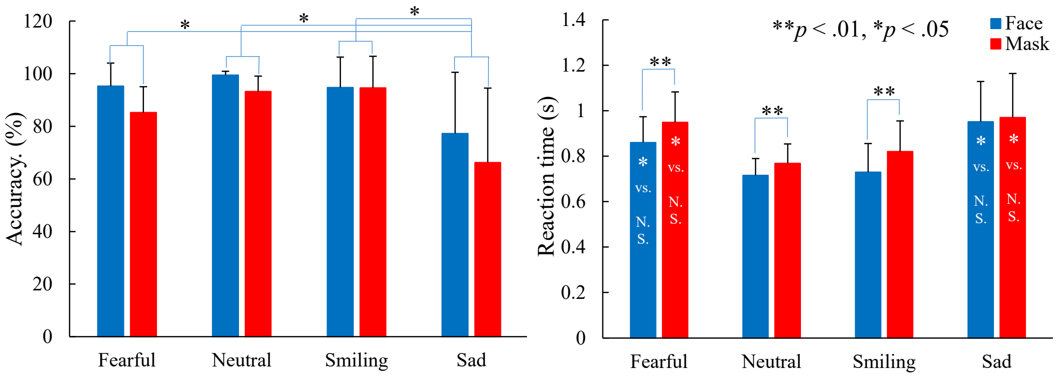

Figure 4: Behavioral performance (mean ± S.D.): accuracies (%) (*left*) and reaction times (s) (*right*) in facial expressions without (*blue*) and with (*red*) masks. N., neutral; S., smiling.

3 Automatic Classifiers for Masked Facial Expressions

The FER as a general classifier was assessed by comparing it with human accuracy for this study. The similar face images used in the EEG experiment were applied to this classification. The specific CNN classifiers were then evaluated to modify the traditional FER. Finally, the accuracies of the tested classifiers were compared with the human results.

3.1 Traditional Classifier for Face Expressions

(1) FER: This study used the FER library provided by Justin Shenk, which is based on hybrid methods [11,12]. OpenCV adopts the Haar cascade algorithm to capture a face, and eye cascades can acquire stable face detection. The detected faces were cropped for further evaluation, and regions other than the faces were rendered in gray scale.

The accuracy (%) was averaged among the tested results of six people (with or without masks) in the image dataset. Here, the FER calculated the probability in each of the facial expression categories. The FER resulted in six emotions (anger, fear, happiness, disgust, surprise, and sadness) with a neutral case, and one showing the highest probability was selected as a final output. Because the EEG study focused on four cases (i.e., fearful, neutral, smiling, and sad), the highest probability of the four emotions was determined as a final output for this study. In accordance with this probability, the FER determined the final facial expression; the highest value was selected to judge it. Therefore, the mean values of the probabilities in the tested cases (i.e., six images of unknown faces) were also computed for each experimental condition.

(2) Validation results: Figure 5 indicates the validation results of the FER under masked and unmasked face conditions. Accuracies were higher for faces without masks than for those with masks, and there was a tendency for the probability to be higher for faces without masks than for those with masks. In particular, compared with the probability of fearful and smiling faces with masks, the classification accuracies were greatly decreased for the masked faces.

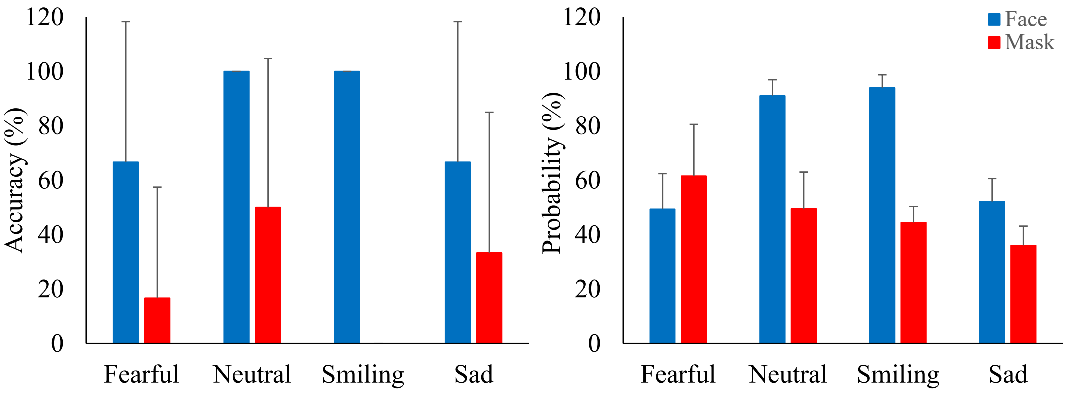

Figure 5: The accuracy (*left*) of the FER classifier in identifying facial expressions without (*blue*) and with (*red*) masks. The probability (*right*) of the FER classifier indicating the maximum value. Mean ± S.D.

3.2 Specific CNN Classifier for Masked Facial Expressions

To improve the general FER, specific CNN classifiers for masked faces were constructed for this study, and accuracy was evaluated by using similar images as the EEG study.

(1) Image augmentation: As a pre-preparation step, image augmentation was applied to increase the training and validation data for the CNN classifiers (Python ver. 3.9 with Keras ver. 2.10). OpenCV (ver. 4.5) was used to perform the image processing of the random rotation (0 ± 5 degrees) and brightness change (0.8 to 1.2 times at random). The face area was then detected and cropped as a region of interest, which was resized to 64 × 64 pixels for the CNNs. Finally, we prepared 100 images for each facial expression (four emotions in total) of the dataset (six persons in total). The CNN classifiers learned those images, and the accuracy and probability were validated.

(2) Specific CNN classifiers: To modify the low accuracy of the FER, especially regarding masked facial expressions, we explored the optimal construction and hyperparameters for the CNNs by trial and error. In the present work, the CNN structure consisted of two convolutional layers with a fully connected layer. Facial images were first applied to the input layer of the CNN. The convolutional layers then learned multiple filter kernels (i.e., feature maps). The size of the filter kernels ($c \times f_i \times f_i$) was set at $1 \times 3 \times 3$, $f_1 = f_2 = 3$ with $n_1 = n_2 = 64$; c represents the number of channels in an input image (i.e., a grayscale image in $c = 1$); f_i represents the size of the filter kernels; n_i denotes the number of filter kernels in the i-th convolutional layer. The padding process was followed to prevent a reduction in the image size by filtering.

The batch-normalization process and ReLU as the activation function were applied to the output in each convolutional layer, followed by the max-pooling layer to provide a form of translation invariance and the dropout process to increase and maintain generalization. The output of the final convolutional layer was passed to the fully connected layer, followed by the softmax function output to identify four classes of facial expressions. The CNNs were trained using the Adam optimization algorithm to optimize cross-entropy loss for multi-class classification.

(3) Training and validation methods: The CNNs were trained and validated to identify facial expressions without and with masks. A *K*-fold cross-validation test (the facial data of six persons) among the four levels of expressions (i.e., fearful, neutral, smiling, and sad) was performed to assess the accuracy of each classifier. Each classifier was trained using *K*-1 data (i.e., faces except for a target person) and validated by the remaining data (unknown faces of a target person). This procedure was repeated to compute the classification accuracy for *K* rounds.

The data number was set at 2,000 images (i.e., 100 images of the augmentation × 4 emotions × 5 persons' faces except for a target) for each training phase and 400 images (i.e., 100 images × 4 emotions in a target face) for each validation phase. In the hyperparameters of the CNN, the learning rate, dropout rate, mini-batch size, and number of epochs were set at 10^{-5}, 0.2, 10, and 100, respectively. This trial (i.e., training and test of the CNN) was repeated four times for every evaluation condition because of initially randomized weights. The heat maps based on weights in the CNN layers were also visualized to clarify the facial parts on which the trained and tested classifier focused [13].

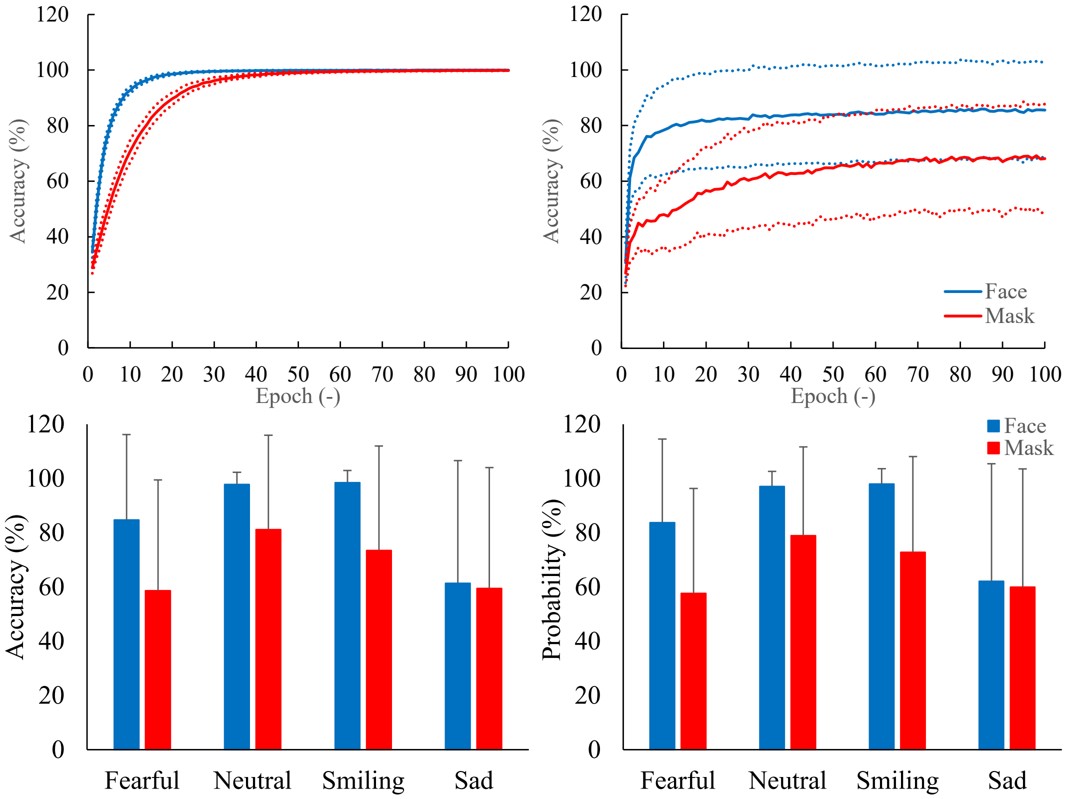

Figure 6: Learning curves in the accuracy (*upper left*) and the probability (*upper right*) of the specific CNN classifiers for facial expressions without (*blue*) and with (*red*) masks (mean ± S.D.). The accuracy (*lower left*) and probability (*lower right*) of each facial expression at the final epoch of test data (mean ± S.D.).

(4) Results of training and validation: Figure 6 (*upper*) shows the learning curves in the accuracy and the probability of each specific CNN classifier for the facial expressions without and with masks: the mean values for the faces of 6 persons (i.e., the results of 100 images) × 4 trials in each condition = 24 times. The accuracy values in the training data (*left*) converged to 100% around 50 epochs under the conditions both without and with masks. In the validation data (*right*), the accuracy values were distributed in both conditions as compared to those in the training data. The mean values of accuracy reached 85% for no masks and 65% for masks. Figure 6 (*lower*) indicates the accuracy and probability values of each facial expression in the final epoch of the test data. The accuracies and probabilities were increased for no mask conditions compared with those for masked faces. Although it was difficult to identify the negative (fearful and sad faces) conditions, the accuracies were greatly improved, especially for masked faces. The results of the probability were similar to those of accuracy.

Figure 7 shows typical examples of heat maps (i.e., gradation of weight in each convolution layer in the final epoch of the test data). In all expressions without and with masks, the roles of the convolutional layers were divided into two parts—detecting the whole area and shape in a target face (i.e., eyes, eyebrows, and mouth areas are hollowed out) in the first layer and characterizing facial parts (i.e., eyes, eyebrows, mouth, bottom of the nose, etc.) in the second layer. There was an inverted relationship between the first and second layers. Under masked conditions, the information on facial parts was limited to the upper half of the face, such as the eyes and eyebrows except for the masked region. Similar tendencies appeared even under the results with low accuracy (e.g., fearful and sad faces without and with masks). Especially in smiling faces without masks, rasing the corner of the mouth was mentioned as a feature of the second convolutional layer in the CNN classifier, although it was hidden in masked faces.

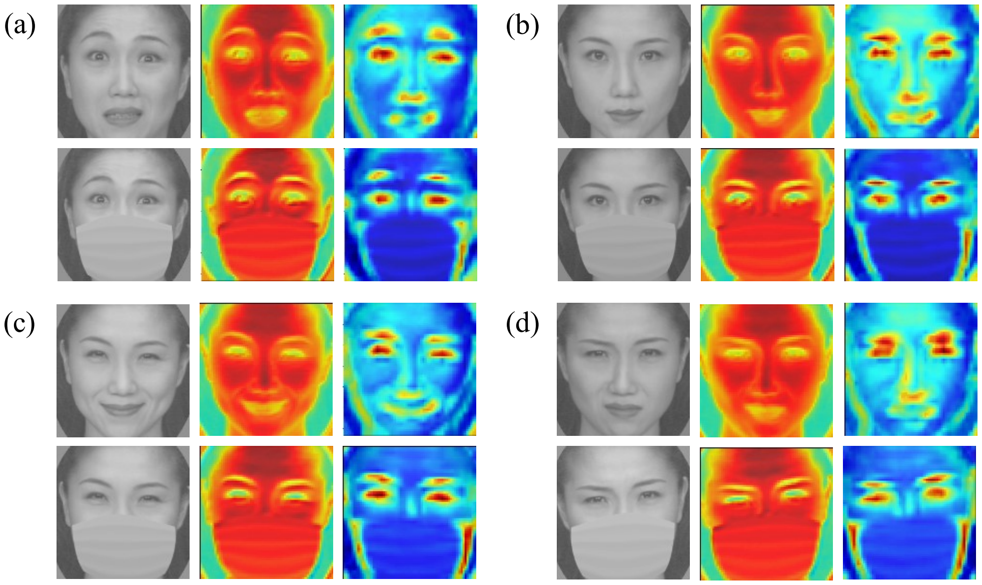

Figure 7: Typical examples of the heat maps (*red*, high weights; *blue*, low weights) of faces in each layer of the CNN classifier in the final epoch of the test data. (a) fearful, (b) neutral, (c) smiling, and (d) sad faces without (*upper*) and with (*lower*) masks.

3.3 Human Judgment versus Automated Classifiers

By comparing the accuracy between human recognition and classifiers, the automated classifiers can clarify the future points of modification to obtain a high recognition rate even under mask faces. Table 1 summarizes the accuracy of the human and classifier results in judging facial expressions with and without masks. For no masks (i.e., usual cases), the FER showed the highest accuracies (100%) under the neutral and smiling conditions. However, the accuracies for fearful and sad (i.e., negative) faces were lower for the classifiers than for human judgment. For both humans and classifiers, the accuracies in identifying facial expressions were lower for masked faces than for unmasked faces. In particular, the FER had worse results in judging emotional faces with masks compared to human judgment, which had higher accuracy. However, even in humans, the accuracies for the fearful and sad expressions were lower than those for the other expressions with and without masks.

Compared to the traditional FER, each specific CNN classifier showed the possibility to greatly improve accuracy (25.0% in the FER to 68.1% in the CNN) when it learned the specific features for masked facial expressions, regardless of various cases such as rotations and brightness of faces when assessing the CNN classifier. Moreover, the CNN classifiers tended toward lower accuracy regarding negative (i.e., fearful and sad) expressions compared to neutral and smiling faces, without and with masks. However, the recognition rates were higher in participants than in the specific classifiers for facial expressions, even in masked conditions.

Table 1: Accuracy of Human Judgment and Automatic Classifiers

Facial Expressions		**Accuracy (%)**		
		Human Judgment	***FER***	***CNN Classifier***
No Mask	***Fearful***	95.8	66.7	84.7
	Neutral	99.5	100.0[a]	97.8
	Smiling	94.9	100.0[a]	98.5
	Sad	77.9	66.7	61.4
	Mean (± S.D.)	92.0 (± 9.6)	83.4 (± 19.2)	85.6 (± 17.3)
Mask	***Fearful***	85.5	16.7	58.6
	Neutral	94.0	50.0	81.1
	Smiling	94.3	0.0[b]	73.4
	Sad	66.7	33.3	59.4
	Mean (± S.D.)	85.1 (± 12.9)	25.0 (± 21.5)	68.1 (± 11.0)

[a]Max. and [b]min. values in this table.

4 Discussion

We performed the EEG study first and found N170 modulation for masked facial expressions. Next, the same dataset used in the EEG experiment was applied to the automated classifier. The reaction times were delayed for masked facial expressions, and accuracies were worse for negative faces with masks. Changes in the orbicularis oculi muscle may influence the judgment of smiling.

4.1 EEG Study

We found that the masked facial expressions increased N170 amplitudes and delayed latencies, compared to the cases of unmasked faces that were also used in the previous reports [5][6][7]. One factor of N170 changes may be brain processing related to predicting the invisible parts of a masked face (e.g., the mouth and nose), suggesting the effects of amodal completion [2] in face recognition. These effects were remarkable for smiling faces in this study, perhaps because it is easier to judge a smile by changes in the eye area.

It is becoming increasingly normal to wear face masks, although some facial parts (i.e., the mouth, nose, and cheek muscles) are hidden. When looking at cropped images of facial parts, such as human eyes or mouths alone, people often feel uncomfortable or unnatural. In contrast, masked faces can be seen naturally, even when missing information about facial parts. Under such conditions, we revealed that face-specific ERPs were modulated by masked facial expressions. In fact, the N170 response was larger with masks than without masks, suggesting that facial composition and completion in the brain may be unconsciously performed (i.e., amodal completion).

ERPs used to judge masked facial emotions may vary according to gender, age, race, culture, and individual differences among participants. For example, regarding photographs of faces, research has shown that compared to Japanese, Americans weighted cues displayed in the mouth more when judging emotions, whereas Japanese tended to weight cues in the eyes more than Americans [14]. The differences between the stimulus methods—movies (i.e., dynamic change) and static photographs—also affect the recognition of facial expressions. Furthermore, although we used unknown faces to evaluate the ERPs, familiar faces may enhance N170 amplitudes [15][16]. A neutral face associated with memory (e.g., observed behavior and personal characteristics, familiar situations, etc.) may also induce various emotions and enhance brain activities [17][18]. Therefore, the presence of familiar and known faces should be considered when assessing the ERPs.

4.2 Automated Classifiers

The FER had a classification accuracy around 70% on the FER-2013 emotion dataset [11]. For this study, compared to the human recognition of facial expressions, the FER as a general classifier had lower accuracies, indicating that the basic classifier's emotional recognition should be modified to identify masked facial expressions. The constructed CNN classifiers improved the FER method, especially for masked facial expressions. This classifier could be applicable for the appropriate and quick detection of emotional change during nonverbal communication in various social situations, by judging facial expressions from camera images. If sufficiently trained CNNs are used to classify facial expressions, a similar speed as human brain processing (around 800 ms, the reaction time for this study) will be needed for the judgment, meaning the possibility of real-time processing, although the processing speed of the CNN classifier depends on the input image size and computer's specs (e.g., CPU, GPU, motherboard, and RAM [12]). Moreover, passive or affective brain-computer interfaces (BCIs) [17] enable the nonverbal communication of specific emotions, and such technologies should facilitate the real-time perception of facial expressions, even in masked cases. In approaches to the use of BCIs with EEG signals, an essential requirement abstracts brain activity at functional frequencies under reduced artificial noise [19], and support vector machines can efficiently extract meaningful brain changes [20].

The roles of the CNN layers constructed for this study are divided into two parts: grasping the whole face except for small parts in the first layer and recognizing the shapes of facial parts in the second layer. This judgment process of the CNN classifier may be similar to human brain processing to recognize faces and expressions. The neural system for face perception is generally classified into a core system (inferior occipital gyrus, lateral FG, and STS) for visual analysis and an extended system (intraparietal sulcus, amygdala, insula, etc.) for cognitive functioning in attention, mouth movement, facial expression or identity, and emotion [21]. This suggests that common and efficient processing for recognizing facial expressions is present in both the human brain and artificial neural networks. Interestingly, even if the information on faces is hidden by wearing masks, the effects of the recognition process for expressions would be retained in the CNN layers. Although the human brain can complement the hidden face area, the CNN classifier will judge facial expressions only from the remaining information on facial parts under masked conditions (e.g., facial parts such as eyes and eyebrows). However, the brain mechanisms while looking at masked faces remain unknown and should be clarified in future studies.

We searched the optimal parameters for the specific CNN classifier, by trial and error. However, we should explore the optimal structure (e.g., the number of convolutional and fully connected layers; number and size of filter kernels; combination of batch normalization, dropout, activation functions, and learning methods) and hyperparameters (e.g., learning rates, mini-batch size, and epoch number) of the CNNs. The weights of the CNN were also changed in every trial because of the small number of facial models. However, to acquire further generalization, it is necessary to converge to general values of weights, learning them by various faces. Furthermore, transfer learning or fine-tuning (e.g., VGG16 and VGG19 [22]) has the potential to increase the accuracy of facial expression recognition under masked conditions. As with human judgment, there are easy or difficult cases for the CNN to train and recognize facial features because of the individual differences among target faces and expressions in the dataset. The CNNs should also be expanded to identify basic emotions [23], such as faces of anger, disgust, and surprise faces, which were not tested in this study, and to show the generalization to various target faces.

5 Conclusion

Our purposes were to evaluate the brain mechanisms of masked facial expressions and explore the automatic classification method for facial expressions while wearing masks. Regarding the first objective, we found that the N170 responses were significantly changed by masked facial expressions. In particular, the increase of N170 amplitude may be associated with the magnitude of amodal completion. Future studies are required to consider gender, age, race, culture, and individual differences among participants and to assess other basic emotions, except for the tested facial expressions in this study.

Regarding the second objective, the specific CNN classifier for masked facial expressions improved the general FER. However, the accuracy was still higher in the human recognition of facial expressions than in the automated classifiers, indicating the requirement of further modifications. In future research, these results can be useful as machine learning and artificial intelligence features for predicting masked facial expressions. As possible features to identify masked facial expressions and increase accuracy, the components of ERPs such as N170 and VPP while looking at a target face may be applied to the hybrid automatic classifiers with facial images. To realize the automated classifiers for human facial expressions and predict emotions would be linked to facilitating nonverbal communication in various situations of daily life and might be useful for the screening of mental diseases based on the analysis of emotional changes.

Acknowledgments

We would like to acknowledge the Scientific Research (C) program of the Japan Society for the Promotion of Science (KAKENHI, 21K12113). This paper is an extended version of the conference paper [24].

References

[1] N. Mheidly, M. Y. Fares, H. Zalzale, and J. Fares, "Effect of face masks on interpersonal communication during the COVID-19 pandemic," Frontiers in Public Health, 898, 2020.

[2] J. Thielen, S. E. Bosch, T. M. van Leeuwen, M. A. van Gerven, and R. van Lier, "Neuroimaging findings on amodal completion: A review," i-Perception, vol. 10(2), 2041669519840047, 2019.

[3] R. J. Itier and M. J. Taylor, "Source analysis of the N170 to faces and objects," Neuroreport, vol. 15(8), pp. 1261–1265, 2004.

[4] W. Luo, W. Feng, W. He, N. Y. Wang, and Y. J. Luo, "Three stages of facial expression processing: ERP study with rapid serial visual presentation," Neuroimage, vol. 49(2), pp. 1857–1867, 2010.

[5] J. K. Hietanen and P. Astikainen, "N170 response to facial expressions is modulated by the affective congruency between the emotional expression and preceding affective picture," Biological Psychology, vol. 92(2), pp. 114–124, 2013.

[6] J. A. Hinojosa, F. Mercado, and L. Carretié, "N170 sensitivity to facial expression: A meta-analysis," Neuroscience & Biobehavioral Reviews, vol. 55, pp. 498–509, 2015.

[7] S. Bentin, T. Allison, A. Puce, E. Perez, and G. McCarthy, "Electrophysiological studies of face perception in humans," Journal of Cognitive Neuroscience, vol. 8(6), pp. 551–565, 1996.

[8] J. Zhao, Q. Meng, L. An, and Y. Wang, "An event-related potential comparison of facial

expression processing between cartoon and real faces," PLoS One, vol. 14(1), e0198868, 2019.

[9] H. T. Schupp, B. N. Cuthbert, M. M. Bradley, J. T. Cacioppo, T. Ito, and P. J. Lang, "Affective picture processing: the late positive potential is modulated by motivational relevance," Psychophysiology, vol. 37(2), pp. 257–261, 2000.

[10] K. Kashihara, "Iris recognition for biometrics based on CNN with super-resolution GAN," In 2020 IEEE Conference on Evolving and Adaptive Intelligent Systems (EAIS), pp. 1–6, May 2020.

[11] O. Arriaga, M. Valdenegro-Toro, and P. Plöger, "Real-time convolutional neural networks for emotion and gender classification," arXiv preprint arXiv:1710.07557, 2017.

[12] I. de Paz Centeno, "MTCNN," GitHub, Jan. 2018. https://github.com/ipazc/mtcnn/

[13] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," In Proceedings of the IEEE International Conference on Computer Vision, pp. 618–626, 2017.

[14] M. Yuki, W.W. Maddux, and T. Masuda, "Are the windows to the soul the same in the East and West? Cultural differences in using the eyes and mouth as cues to recognize emotions in Japan and the United States," Journal of Experimental Social Psychology, vol. 43(2), pp. 303–311, 2007.

[15] S. Caharel, S. Poiroux, C. Bernard, F. Thibaut, R. Lalonde, and M. Rebai, "ERPs associated with familiarity and degree of familiarity during face recognition," Int. J. Neurosci., vol. 112, pp. 1499–1512, 2002.

[16] B. Rossion, S. Campanella, C.M. Gomez, A. Delinte, D. Debatisse, L. Liard et al., "Task modulation of brain activity related to familiar and unfamiliar face processing: an ERP study," Clin. Neurophysiol., vol. 110, pp. 449–462, 1999.

[17] K. Kashihara, "A brain-computer interface for potential non-verbal facial communication based on EEG signals related to specific emotions," Frontiers in Neuroscience, vol. 8, 244, 2014.

[18] A. Todorov, M.I. Gobbini, K.K. Evans, and J.V. Haxby, "Spontaneous retrieval of affective person knowledge in face perception," Neuropsychologia, vol. 45, pp. 163–173, 2007.

[19] C. Tallon-Baudry, O. Bertrand, C. Delpuech, and J. Pernier, "Stimulus-specificity of phase-locked and non phase-locked 40-Hz visual responses in human," J. Neurosci., vol. 16, pp. 4240–4249, 1996.

[20] K. Kashihara, "Automatic discrimination of task difficulty predicted by frontal EEG activity during working memory tasks in young and elderly drivers," International Journal of Information Technology & Decision Making, pp. 1–43, 2022.

[21] J.V. Haxby, E.A. Hoffman, and M.I. Gobbini, "The distributed human neural system for

face perception," Trends Cogn. Sci., vol. 4, pp. 223–233, 2000.

[22] A. Sajjanhar, Z. Wu, and Q. Wen, "Deep learning models for facial expression recognition," In 2018 Digital Image Computing: Techniques and Applications (dicta), pp. 1–6, IEEE, December 2018.

[23] P. Ekman, "Basic emotions," In T. Dalgleish & M. Power (Eds.), Handbook of Cognition and Emotion, New York: Wiley, 1999.

[24] K. Kashihara and M. Shinguu, "The judgment of masked facial expressions by humans and classifiers," In 2022 12th International Congress on Advanced Applied Informatics (IIAI-AAI), pp. 283–286, July 2022.