



# Role of temporal processing stages by inferior temporal neurons in facial recognition

Yasuko Sugase-Miyamoto<sup>1\*</sup>, Narihisa Matsumoto<sup>1</sup> and Kenji Kawano<sup>2</sup>

<sup>1</sup> Human Technology Research Institute, The National Institute of Advanced Industrial Science and Technology, Tsukuba, Japan

<sup>2</sup> Department of Integrative Brain Science, Graduate School of Medicine, Kyoto University, Kyoto, Japan

## Edited by:

Gabriel Kreiman, Harvard University, USA

## Reviewed by:

Hans P. Op de Beeck, Catholic University of Leuven, Belgium

Charles Gross, Princeton University, USA

Scott L. Brincat, Massachusetts Institute of Technology, USA

Chou Hung, National Yang-Ming University, Taiwan

## \*Correspondence:

Yasuko Sugase-Miyamoto, Systems Neuroscience Group, Human Technology Research Institute, The National Institute of Advanced Industrial Science and Technology, 1-1-1 Umezono, Tsukuba, Ibaraki 305-8568, Japan.  
e-mail: y-sugase@aist.go.jp

In this review, we focus on the role of temporal stages of encoded facial information in the visual system, which might enable the efficient determination of species, identity, and expression. Facial recognition is an important function of our brain and is known to be processed in the ventral visual pathway, where visual signals are processed through areas V1, V2, V4, and the inferior temporal (IT) cortex. In the IT cortex, neurons show selective responses to complex visual images such as faces, and at each stage along the pathway the stimulus selectivity of the neural responses becomes sharper, particularly in the later portion of the responses. In the IT cortex of the monkey, facial information is represented by different temporal stages of neural responses, as shown in our previous study: the initial transient response of face-responsive neurons represents information about global categories, i.e., human vs. monkey vs. simple shapes, whilst the later portion of these responses represents information about detailed facial categories, i.e., expression and/or identity. This suggests that the temporal stages of the neuronal firing pattern play an important role in the coding of visual stimuli, including faces. This type of coding may be a plausible mechanism underlying the temporal dynamics of recognition, including the process of detection/categorization followed by the identification of objects. Recent single-unit studies in monkeys have also provided evidence consistent with the important role of the temporal stages of encoded facial information. For example, view-invariant facial identity information is represented in the response at a later period within a region of face-selective neurons. Consistent with these findings, temporally modulated neural activity has also been observed in human studies. These results suggest a close correlation between the temporal processing stages of facial information by IT neurons and the temporal dynamics of face recognition.

**Keywords:** inferior temporal cortex, categorization, facial identity, temporal dynamics

## INTRODUCTION

The ability to recognize meaningful information from faces is a crucial aspect of our social life. When we are exposed to an environment that is rich in visual features and stimuli, we can immediately detect a face and easily recognize its individual identity and emotional state. This suggests the presence of an efficient system in the primate brain for the visual analysis of faces. Recent functional magnetic resonance imaging (fMRI) studies have revealed the existence of several brain areas that participate in the analysis of visual characteristics of faces in both humans and macaque monkeys, thus pointing to multiple functionally specialized “face patches” in the face-processing system (Tsao et al., 2008).

Prior to these studies, face-responsive neurons were recorded in the macaque temporal cortex (Bruce et al., 1981; Perrett et al., 1985), which includes some of these face patches, and their functional properties have been studied from various aspects (Desimone et al., 1984; Yamane et al., 1988). In our previous studies (Sugase et al., 1999; Matsumoto et al., 2005b), we recorded neuronal activity in the anterior temporal cortex of monkeys while presenting visual stimuli that included monkey and human faces, and studied the temporal modulation of neuronal responses to the faces. We found that the initial transient responses of face-responsive neurons represented information about global categories, namely human vs.

monkey vs. simple shapes, and the later portion of the responses represented information about more detailed facial categories, i.e., facial expression or identity (Sugase et al., 1999; Matsumoto et al., 2005b). We identified an important role for temporal firing patterns in the information coding of visual facial features and suggested that the initial transient response and the later response encode different aspects of facial characteristics. Recent single-unit studies in non-human primates have provided further evidence supporting the important role of temporal neuronal modulation. In addition, human studies reveal a close correlation between neuronal events and the timing of face recognition. In this review, we will first describe studies using macaque monkeys and then summarize studies using human subjects. The accumulated body of evidence reveals a close correlation between the temporal encoding of facial information by inferior temporal (IT) cortex neurons and the temporal dynamics of face recognition.

## MONKEY STUDIES

### REPRESENTATION OF FACIAL IMAGES BY NEURONS IN THE INFERIOR TEMPORAL CORTEX

In the ventral visual cortical pathway, visual signals are processed sequentially through areas V1, V2, V4, and the IT cortex (areas TEO and TE; Mishkin et al., 1983). In the macaque IT cortex, neurons

show selective responses to complex visual images (Baylis et al., 1987; Richmond et al., 1987; Kobatake and Tanaka, 1994; Tsunoda et al., 2001; Brincat and Connor, 2004; Yamane et al., 2006, 2008), such as faces or animals (Gross et al., 1972; Bruce et al., 1981; Desimone et al., 1984; Perrett et al., 1985; Hasselmo et al., 1989; Nakamura et al., 1994; Tanaka, 1996; Tsao et al., 2006). Face-responsive/selective neurons give stronger responses, often twice as strong, to face images as compared with other images such as objects, geometric shapes, or scrambled face images. Information about faces is sparsely represented by a population of face-responsive neurons (Young and Yamane, 1992). The responses of these neurons are sensitive to the configuration of facial parts, such as the height of the forehead from the left eye to the hairline, the distance between the eyes and the mouth, or a combination of these parameters (Yamane et al., 1988; Freiwald et al., 2009). A possible role for IT face-responsive neurons in face recognition has been reported. The responses of the majority of face-responsive neurons are correlated with the animal's perceptual experiences (Sheinberg and Logothetis, 1997, 2001). IT neuronal responses have been shown to represent the configuration of facial parts that are useful for a categorization task when using monkeys that are well experienced with the categorization task (Sigala and Logothetis, 2002). In the anterior IT cortex, the response latency of neurons encoding view-invariant face-identity information correlates with the monkey's behavioral response latency during identification of a face (Eifuku et al., 2004). Performance in face discrimination tasks based on the configuration of facial parts is affected by cooling of the temporal cortex (Horel, 1993).

#### TIME COURSE OF NEURONAL INFORMATION PROCESSING OF FACES

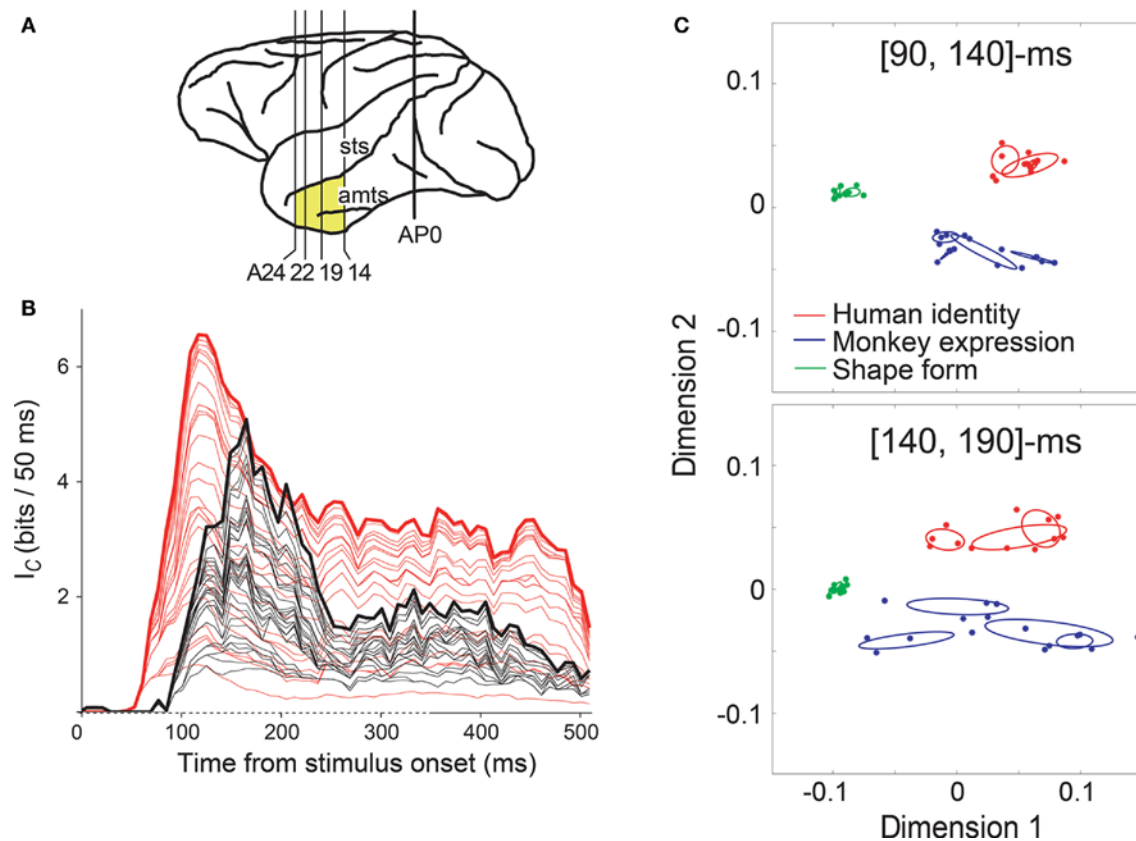
At each stage along the ventral pathway, namely areas V2, V4, and the IT cortex, stimulus selectivity of neural responses to complex images becomes sharper during the later portion of the response as compared with selectivity during the initial transient response. Hegde and Van Essen (2004, 2006) studied neuronal responses to gratings (sinusoidal, hyperbolic, and polar gratings) and contour stimuli (bars, crosses, and angles) in areas V2 and V4. They found that the population response of the neurons was better able to categorize the stimuli into broad groups, e.g., gratings vs. contour, during the initial phase of the response (area V2, 40–80 ms after stimulus onset; area V4, 80–100 ms), and was able to distinguish between individual stimuli within the stimulus groups after the initial phase (Hegde and Van Essen, 2004, 2006). Brincat and Connor (2006) examined neuronal responses in the posterior part of the IT cortex to simple shape stimuli combining convex, concave, and straight contours. They found that information about individual contour fragments is carried in the initial transient response (90% maximum at 122 ms after stimulus onset) and information about the specific multipart contour configurations emerges gradually (90% maximum at 184 ms after stimulus onset). Tamura and Tanaka (2001) showed that the stimulus selectivity of neuronal responses to photographs of natural objects and geometric shapes in the anterior part of the IT cortex is more selective in the later portion of the response (after 240 ms from stimulus onset) compared to the initial transient response (130 ms from stimulus onset).

The importance of temporal firing patterns during information coding of visual stimuli by IT neurons was first described by Optican, Richmond, and colleagues (Optican and Richmond,

1987; Richmond and Optican, 1987; Richmond et al., 1987) and has been confirmed since then. Tovee et al. (1993) showed that IT neuronal responses represent the greatest information about the test stimuli identity during the first 100–200 ms after stimulus onset (Tovee et al., 1993). The robustness of the temporal firing pattern across days and weeks of visual experience was also shown (Bondar et al., 2009).

We examined neuronal response in the anterior IT gyrus and the anterior part of the superior temporal sulcus in the IT cortex to visual stimuli, including geometric shapes and monkey/human faces with various expressions, using a fixation task (recording location shown in **Figure 1A**; Sugase et al., 1999). Some neurons showed initial transient responses to faces, but not to shapes, and maintained a sustained response to only a particular facial expression or identity, indicating sharper stimulus selectivity in the later portion of the response (e.g., Figure 1 of Sugase et al., 1999, and Figure 2 of Matsumoto et al., 2005b). Using a moving time window, we calculated the time course of the transmitted facial stimuli information with respect to both the global category (human faces vs. monkey faces vs. shapes) and sub- or fine categories within each category member (monkey facial expression, monkey identity, human facial expression, or human identity). We found that the latency and peak of information for the global categories were earlier than those of the fine categories (measured as the middle of the 50-ms analysis window; latency, 91 and 142 ms after stimulus onset; peak, 152 and 179 ms, for global and fine categories, respectively; **Figure 1B**). This result indicates that the temporal firing pattern is important for information processing of faces and suggests that the initial transient response and the later response encode different aspects of facial characteristics, i.e., information about global categories and about fine categories. The result that both the peak of the information for the global categories and the peak of the information for the fine categories were observed within 100–200 ms after stimulus onset is consistent with the finding by Tovee et al. (1993) and supports the notion that neuronal responses within this time period represent the greatest information about the stimuli identity. Because each neuron displays a different temporal firing pattern (as shown in Figure 2 of Matsumoto et al., 2005b), we further analyzed the responses of a population of neurons and found that the separation of human vs. monkey vs. shapes was maximized at the [90, 140]-ms period after the stimulus onset, and that the separation of monkey facial expressions and human facial identities was maximized at the [140, 190]-ms period (**Figure 1C**; Matsumoto et al., 2005b). This result indicates that global categorization occurs earlier than finer categorization at the population level.

Recent single-unit studies in non-human primates provide evidence that is consistent with our findings. Neurons in the anterior IT cortex respond to human faces on average 15 ms earlier (103 ms mean latency) than to non-primate animal faces (118 ms; Kiani et al., 2005), showing that the initial transient response is modulated with respect to the global category, i.e., human vs. animal. Tsao, Freiwald, and colleagues made remarkable advances in identifying face-processing systems along the ventral visual pathway, identifying six interconnected cortical regions that consist of face-selective neurons, i.e., face patches (**Figure 2A**; Tsao et al., 2003, 2006, 2008; Moeller et al., 2008; Freiwald et al., 2009; Freiwald and Tsao, 2010). In the middle face patch (ML and MF in **Figure 2A**), which is located within the superior



**FIGURE 1 | Time course of two different scales of facial information in the IT cortex. (A)** Areas of brain examined in Sugase et al. (1999) (yellow). AP0 represents the position of the external auditory meatus; A14, A19, A22, and A24 represent anterior 14, 19, 22, and 24 mm, respectively. amts, anterior middle temporal sulcus; sts, superior temporal sulcus. **(B)** Information transmission rate ( $I_c$ ) of face-responsive neurons. Summation of  $I_c$  curves for global (red, human faces vs. monkey faces vs. shapes) and fine categories (black; monkey facial expression, monkey identity, human facial expression, or human identity) from 32 neurons aligned at the stimulus onset.  $I_c$  was calculated over a 50-ms sliding analysis window, and plotted at the middle of the time window. Non-significant information was excluded from the analysis. The 32 neurons represent information about both global and fine categories. In the time axis (abscissa), the

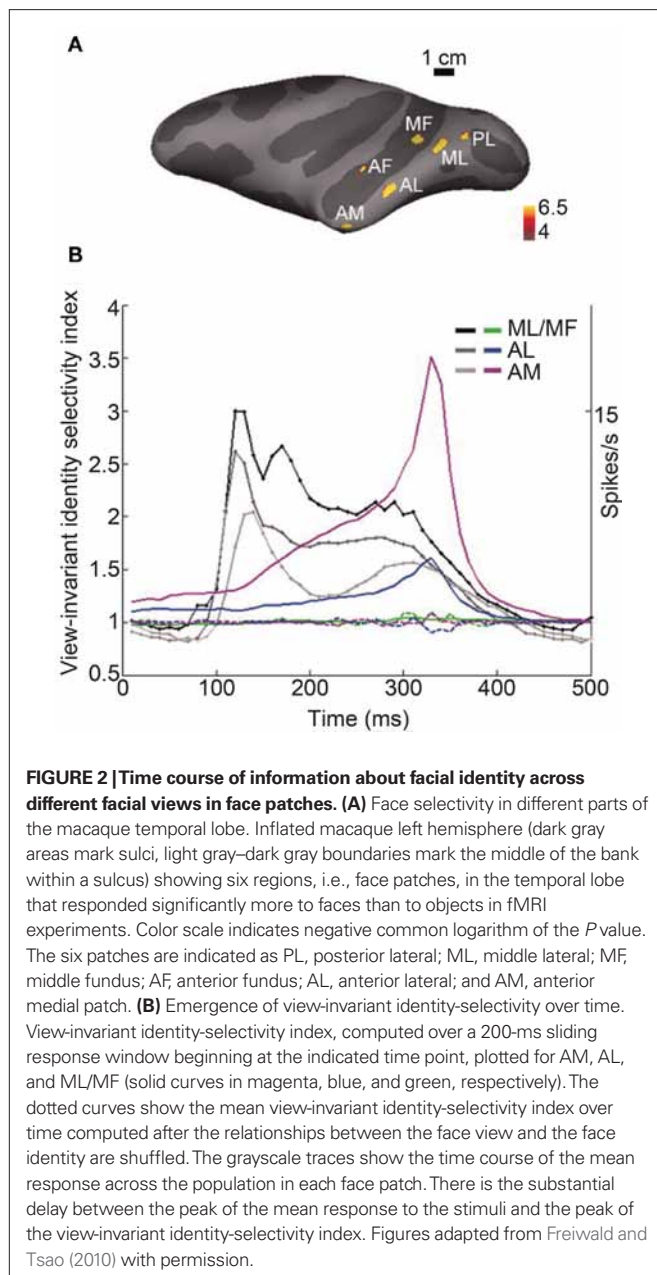
time of stimulus onset is 0 ms and the duration of stimulus presentation is indicated as a dashed line. **(C)** Population activity vectors in two-dimensional space rearranged using PCA in the [90, 140] and [140, 190]-ms windows. The horizontal axis represents the first principal component, while the vertical axis represents the second principal component. The points indicate the population activity vectors of 45 face-responsive neurons for the individual stimuli. The colors of the points represent the global category: the vectors for human faces, monkey faces, and shapes are shown in red, blue, and green, respectively. The red, blue, and green ellipses indicate the distributions of the population activity vectors for the fine categories human identity, monkey expression, and shape-form, respectively. Figures adapted from Sugase et al. (1999) (A,B) and Matsumoto et al. (2005b) (C) with permission.

temporal sulcus, categorization (faces, bodies, fruits, hands, gadgets, or scrambled images) performance by the neuronal response (133 ms) precedes identification performance (192 ms; Tsao et al., 2006), showing an earlier representation of the global category information than of the subcategory image identity information. Freiwald and Tsao (2010) found that in the anterior medial patch (AM in Figure 2A), which is located in the anterior part of the IT gyrus, information about facial identity across different facial views emerges gradually and peaks at around 300 ms after stimulus onset (Figure 2B). The fact that view-invariant facial identity information is represented in the response at a later period (~300 ms) indicates that additional information processing is necessary to compute facial identity across different facial views, which may take a considerable amount of time.

This type of coding seems to be a plausible mechanism underlying the temporal dynamics of face recognition, e.g., the process of face detection/categorization followed by facial identification

of an individual. The causal relationship between signals from face-responsive neurons in the IT cortex and face perception was addressed by Afraz et al. (2006). Microstimulation within face-selective sites around the time of the initial transient response, i.e., [50, 100] ms after stimulus onset, biased the monkey's decision toward faces during a face/non-face categorization task (Afrac et al., 2006). This result indicates that the initial transient response of face-selective neurons has a causal relationship with face perception, i.e., whether an image contains a face, and the perception of a global category. The causal relationship between the later portion of the response and face recognition, e.g., recognition of facial identity, remains to be elucidated.

There are at least three different possibilities for the underlying mechanism that governs how initial transient responses and later responses encode different aspects of facial characteristics. One possibility is that the initial transient response encodes achromatic



information through the magnocellular pathway, which arrives at the IT cortex earlier than chromatic information through the parvocellular pathway. However, this is unlikely. Edwards et al. (2003) found that colored images evoked larger responses from IT face-responsive neurons than did achromatic images during the earliest part of the response. However, the possibility of the magnocellular and parvocellular pathways playing different roles in encoding fine/coarse facial information remains open for future studies. The initial transient response may encode information about the low spatial frequency components of face images through the magnocellular pathway, and the later portion of the response may encode information about the high spatial frequency components through the parvocellular pathway. This idea is supported by human fMRI studies (Goffaux et al., 2010, see later section). The relationship

between responses at different temporal domains and the high and low spatial frequency components of images remains to be elucidated at the single-unit level in non-human primate studies.

The second possibility is that the representation of each facial part is followed by complete representation of multiple facial parts. The third possibility is that a visual stimulus is processed at different scales, from global-to-local, with the passage of time. Several studies support the second and third possibilities. With respect to the second possibility, individual object parts of abstract shapes are represented during the initial transient responses (peak, 122 ms after stimulus onset), and multipart configurations are represented in the later portion of the responses (peak, 184 ms; Brincat and Connor, 2006), in the posterior part of the IT cortex (area TEO and posterior TE), which sends efferent projections to the anterior temporal cortex (Saleem et al., 2000; Borra et al., 2010). Consistent with the suggestion by Brincat and Connor (2006), neurons in the anterior IT cortex may represent a diagnostic area that is useful in detecting facial features, such as the eyes (Lewis and Edmonds, 2003), during the initial transient responses and represents information about the configuration of facial parts that is useful for recognition of facial identity (Tanaka and Farah, 1993) during the later portion of the response. The third possibility has been investigated using Navon (1977) figures, e.g., a large letter N consisting of small letter H's (Tanaka and Fujita, 2000; Tanaka et al., 2001; Sripathi and Olson, 2009, 2010a,b). The behavioral response time to discriminate the global form is 20–30 ms earlier than that to discriminate the local form, for both monkey and human subjects (Tanaka and Fujita, 2000; Sripathi and Olson, 2009). The global form is represented during the initial transient response, and the local form is represented approximately 30 ms later in the IT cortex (Sripathi and Olson, 2009). Differences in the global–local latency of the neuronal response are related to the large–small latency difference. These results indicate that the global signal emerges earlier, because shapes at a larger scale elicit discriminative neuronal activity earlier. Therefore, characteristics of faces at a larger scale, e.g., the outline of the face, may elicit differential neuronal responses earlier than do facial characteristics at a smaller scale, e.g., face parts.

With respect to the cellular mechanisms underlying the representation of global-to-fine category information, we speculated that signals related to fine categories emerge through intra-areal contribution or feedback from other areas (Sugase et al., 1999). The amygdala is one of the candidate areas that send feedback signals and affect the later portion of the responses related to fine category information, since the IT cortex receives projections from the amygdala (Amaral and Price, 1984; Amaral et al., 1992) and neurons in the amygdala respond to specific facial identity or expression (Leonard et al., 1985; Nakamura et al., 1992; Kuraoka and Nakamura, 2006; Gothard et al., 2007). With respect to the contribution of intra-areal connections, the fine category information might emerge through their recurrent connections. We tested this possibility using an attractor network model that consisted of excitatory and inhibitory neurons (Matsumoto et al., 2005a). Since the model reproduced the temporal dynamics of the responses of the face-responsive neurons, we speculated that the recurrent processing within the IT cortex might be sufficient to give rise to the later portion of the response. We proposed two physiological experiments to investigate whether the attractor network model



is a plausible mechanism: one using noisy, fine feature degraded images, and the other using weakening connections between excitatory neurons within the anterior IT cortex. Although inhibition by  $\gamma$ -aminobutyric acid within the IT cortex contributes to neuronal stimulus selectivity (Wang et al., 2000), the relationship between the intra-areal inhibition (contribution of both the recurrent and lateral inhibition) and the temporal firing patterns of IT neurons still remains indefinite. In addition, the correlation between the responses of neuronal pairs in the IT cortex is higher during the presentation of face-like drawings than during the presentation of non-face-like drawings as early as [100, 300] ms after the stimulus onset (Hirabayashi and Miyashita, 2005). The result indicates that connections between neurons in a local circuit may be strengthened during information processing of face images.

There is another line of research, which suggests that the processing of visual stimuli occurs in a rapid feed-forward pass during the early portion of the response, with no role in the basic performance of recurrent/feedback processing during the later portion of the response (although it likely plays a role in top-down effects, such as attentional biases), i.e., the feed-forward hypothesis (Riesenhuber and Poggio, 1999; Serre et al., 2007). As supporting evidence for this hypothesis, Hung et al. (2005) showed that information about object categorization and identification can be read out from an unbiased sample of the IT neuronal site within a narrow time window (12.5 ms) at the earliest part of the response (starting from 125 ms after stimulus onset) using a support vector machine classifier. Since this study did not focus on face-responsive neurons, the time course to the readout of detailed information specific to human/monkey faces, facial identity and expression, is not reported.

## HUMAN STUDIES

At the behavioral level, visual recognition speed has been examined using large sets of images consisting of faces, animals, or objects. Thorpe and colleagues have shown that human observers are able to categorize briefly presented natural scenes with a reaction time of less than 400 ms and approximate accuracy of 94% using an animal vs. no-animal go/no-go task with manual responses (Thorpe et al., 1996; VanRullen and Thorpe, 2001; Rousselet et al., 2002). The reaction time of human subjects during the task was reported to be 100–180 ms (median reaction time) slower than monkey subjects (Fabre-Thorpe et al., 1998). Recently, these authors used saccadic eye movement for a choice task, and showed that subjects were especially rapid at detecting human faces. In the task, two images were simultaneously presented and the subjects made a saccade to an image of a particular target category, e.g., human faces or motor vehicles. The reaction time (start time) of the saccade was substantially shorter toward human faces than toward the other category images, e.g., motor vehicles (in their third experiment, human faces, 159 ms mean, 100 ms minimum; vehicles, 183 ms mean, 170 ms minimum; Crouzet et al., 2010). The reaction time was longer during the choice task than during a simple saccadic detection task, with only one stimulus being presented, both for the human face targets (at least 20 ms longer) and for the vehicle targets (80 ms longer). These results suggest that a lower number of processing steps is required to select faces than to select other objects. It remains to be clarified whether signals in the IT cortex, e.g., early

global category information (face vs. non-face), are required for the saccadic choice. The face bias during the saccadic choice task is at least partially related to the low-level physical characteristics of the images, as scrambling the orientation contents of the images but not scrambling their relative positions disrupts the face bias (Honey et al., 2008).

The detection of an object or categorization of an object at a basic categorical level (bird, car, dog, etc.) is approximately 65 ms faster than within-category identification (Grill-Spector and Kanwisher, 2005). Face detection using synthetic faces occurs 24 and 31 ms earlier than viewpoint and face identification, respectively (Or and Wilson, 2010). This result suggests that face recognition is a process of face detection/categorization followed by viewpoint/face identification and that additional processing is required to obtain the latter information. On the other hand, face identification becomes as fast as face categorization depending on the familiarity of the face. For example, identifying a face such as that of Bill Clinton is as fast as categorizing the face as human or monkey (Tanaka, 2001; Anaki and Bentin, 2009). Studies examining whether information about each facial part is analyzed, accumulated, and integrated over time during face recognition remain controversial (Singer and Sheinberg, 2006; Anaki et al., 2007; Cheung et al., 2011).

Humans and macaque monkeys have similar brain systems for face-processing (Tsao et al., 2008; Bell et al., 2009; Pinsk et al., 2009). The temporal dynamics of the neural correlates of facial information processing have also been studied in human subjects or patients using non-invasive brain-imaging techniques or invasive techniques. The results for these studies are consistent with studies on non-human primates and human behavioral studies. Studies using magnetoencephalography (MEG) showed that a face-selective response 100 ms after the stimulus onset (M100) was related to categorization of the stimuli as a face (Liu et al., 2002; Meeren et al., 2008) and that the later response at 170 ms after the stimulus onset (M170) was related to both facial categorization and recognition of individual faces (Liu et al., 2002). Both M100 and M170 were observed over the occipito-temporal cortex, though M100 was distributed slightly posteriorly. Since the existence of facial parts (eyes, nose, and mouth) is important for M100 regardless of their configuration, and the part-configuration is important for M170, Liu et al. (2002) speculated that M100 and M170 responses reflect processing related to the detection of diagnostic facial parts and to processing related to the analysis of the part-configuration, respectively, instead of global-to-local processing. The familiarity of faces affected the M170 response (Kloth et al., 2006) and other later responses (latency around 250–400 ms, M400) from the occipito-temporal sensors (Harris and Aguirre, 2008).

Event-related potential (ERP) studies have shown that the representation of facial categories and face-inversion effects occurs during an early negative component that peaks at around 170 ms after stimulus onset (N170, starting at 130–150 ms after stimulus onset; Rousselet et al., 2008). The strength of face selectivity of the N170 response (calculated from the peak amplitude between 140 and 200 ms) has been shown to be highly correlated across subjects with those of the hemodynamic response in the temporal lobe, i.e., the fusiform face area (FFA) and the posterior part of the right superior temporal sulcus, by a simultaneous ERP–fMRI study (Sadeh et al., 2010). A later response (e.g., a negativity between 300 and 500 ms

from a parietocentral electrode) was shown to be associated with the familiarity of faces, suggesting that the late signals are related to familiar facial identity (Eimer, 2000). Electrocorticography (ECoG) recordings indicate that an enhanced activation in gamma power (30–70 Hz) and evoked responses that occur 150–200 ms after stimulus onset (similar timing as N170) are associated with successful recognition of the stimulus category, i.e., face, house, or object (Fisch et al., 2009). Studies using intracranial or subdural electroencephalography (EEG) recordings showed that face-specific ERPs (face-N200) within the fusiform and IT gyri were affected neither by semantic priming and face-name learning/identification nor by selective attention, but that subsequent slow evoked potential was affected (P290 and N700, or ~240 ms after stimulus onset, respectively; Puce et al., 1999; Engell and McCarthy, 2010). These results indicate that the face-N200 is largely insensitive to cognitive task manipulations or attention, unlike the later responses, suggesting that the face-N200 reflects an initial and obligatory neural response that is related to the visual analysis of faces and that the later responses are susceptible to cognitive demands and the top-down control of attention. Quiroga et al. (2005) suggested an important role for the human medial temporal lobe (MTL; the hippocampus, amygdala, entorhinal cortex, and parahippocampal cortex) in recognizing facial identity. Some neurons in the MTL respond exclusively to images of a particular famous person regardless of different views, e.g., the actress Jennifer Aniston. The latency of the responses to a famous person are primarily observed during the later period, i.e., between 300 and 600 ms after stimulus onset (Quiroga et al., 2005; Mormann et al., 2008). The results indicate that the higher selectivity of the MTL neurons was found with longer latencies, in other words, in the later responses. Furthermore, these data suggest that interactions between the IT cortex and MTL may play an important role in the recognition of facial identity regardless of view variations.

A recent fMRI study revealed a role for temporal dynamics in face information processing, i.e., coarse-to-fine processing (Goffaux et al., 2010). By presenting face images that preserve either low spatial frequency or high spatial frequency for a 75, 150, or 300-ms duration, Goffaux et al. (2010) found that the coarse structure of a face, which is carried by low spatial frequency, is processed prior to the fine details transmitted by high spatial frequency in the FFA. At a 75-ms exposure, the responses to low spatial frequency face images are greater than the responses to high spatial frequency face images. In contrast, at 150 and 300-ms exposure, the responses to high spatial frequency face images, which contain local fine information useful for identification, are stronger than the responses to low spatial frequency face images. Another study suggests that coarse signals about visual stimuli, i.e., achromatic but with luminance-contrast stimuli (magnocellular biased stimuli), are derived rapidly through the dorsal visual pathway via the occipital visual cortex and

orbitofrontal cortex to the IT cortex (Kveraga et al., 2007). These results support the hypothesis that the initial and later portions of neuronal responses in the IT cortex play important roles in the coarse and fine processing of face images, respectively.

## COMPARISON OF MONKEY AND HUMAN STUDIES

Monkey and human studies suggest at least two temporal processing stages underlying the recognition of faces. The correlation between the two processing stages and neuronal activity has been shown by human MEG studies, indicating that M100 is related to face detection/categorization and M170 is related to face identification. The timing of the M100 response (averaged response peak time, 105 ms after stimulus onset, range 84.5–130.5 ms) and that of the M170 response (160 ms) overlaps with the peak time of global information (mean  $\pm$  SD, 152  $\pm$  57 ms) and that of fine information (179  $\pm$  49 ms) reported by Sugase et al. (1999), respectively. The M100 response appears slightly earlier compared to the peak time of the global information, most likely due to M100 being distributed slightly more posteriorly than M170, thus including signals from posterior parts of the brain, i.e., the occipital.

The N170 response of ERP studies and the face-N200 of intracranial or subdural EEG recordings seem to reflect at least processing of the detection/categorization of faces. It is not apparent whether the later response (M400 for MEG, 300–500 ms for ERP, P290, and N700 for EEG) that is correlated with the familiarity of faces parallels the fine information reported by Sugase et al. (1999), or a slowly emerging signal reflecting the view-invariant identity information reported by Freiwald and Tsao (2010). Future studies at the single-unit level using non-human primates should address which temporal domains are influenced by the familiarity of faces.

## CONCLUSION

The accumulated body of evidence on non-human primates shows that temporal firing patterns in IT cortex neurons are important for information coding of visual features of faces. It has been shown that the initial transient response represents the face category, and that the later response represents facial identity and facial expression. Future studies should investigate the role of signals in the later response of face recognition to reveal the causal relationship between the later response and face detection/categorization, recognition of facial identity, or even face familiarization. It should also be examined whether the temporal processing stages are specific to face images or are general properties of the visual system.

## ACKNOWLEDGMENT

This work was supported by Grant-in-Aid for Scientific Research on Innovative Areas, “Face perception and recognition” from MEXT KAKENHI (21119528, 23119732), and MEXT KAKENHI (22700161).

## REFERENCES

- Afraz, S. R., Kiani, R., and Esteky, H. (2006). Microstimulation of inferotemporal cortex influences face categorization. *Nature* 442, 692–695.
- Amaral, D. G., and Price, J. L. (1984). Amygdalo-cortical projections in the monkey (*Macaca fascicularis*). *J. Comp. Neurol.* 230, 465–496.
- Amaral, D. G., Price, J. L., Pitkanen, A., and Carmichael, S. T. (1992). *Anatomical Organization of the Primate Amygdaloid Complex*. New York: Wiley-Liss, Inc.
- Anaki, D., and Bentin, S. (2009). Familiarity effects on categorization levels of faces and objects. *Cognition* 111, 144–149.
- Anaki, D., Boyd, J., and Moscovitch, M. (2007). Temporal integration in face perception: evidence of configural processing of temporally separated face parts. *J. Exp. Psychol. Hum. Percept. Perform.* 33, 1–19.
- Baylis, G. C., Rolls, E. T., and Leonard, C. M. (1987). Functional subdivisions of the temporal lobe neocortex. *J. Neurosci.* 7, 330–342.
- Bell, A. H., Hadj-Bouziane, F., Frihauf, J. B., Tootell, R. B., and Ungerleider, L. G. (2009). Object representations in the temporal cortex of monkeys

- and humans as revealed by functional magnetic resonance imaging. *J. Neurophysiol.* 101, 688–700.
- Bondar, I. V., Leopold, D. A., Richmond, B. J., Victor, J. D., and Logothetis, N. K. (2009). Long-term stability of visual pattern selective responses of monkey temporal lobe neurons. *PLoS ONE* 4, e8222. doi: 10.1371/journal.pone.0008222
- Borra, E., Ichinohe, N., Sato, T., Tanifuji, M., and Rockland, K. S. (2010). Cortical connections to area TE in monkey: hybrid modular and distributed organization. *Cereb. Cortex* 20, 257–270.
- Brincat, S. L., and Connor, C. E. (2004). Underlying principles of visual shape selectivity in posterior inferotemporal cortex. *Nat. Neurosci.* 7, 880–886.
- Brincat, S. L., and Connor, C. E. (2006). Dynamic shape synthesis in posterior inferotemporal cortex. *Neuron* 49, 17–24.
- Bruce, C., Desimone, R., and Gross, C. G. (1981). Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. *J. Neurophysiol.* 46, 369–384.
- Cheung, O. S., Richler, J. J., Phillips, W. S., and Gauthier, I. (2011). Does temporal integration of face parts reflect holistic processing? *Psychon. Bull. Rev.* 18, 476–483.
- Crouzet, S. M., Kirchner, H., and Thorpe, S. J. (2010). Fast saccades toward faces: face detection in just 100 ms. *J. Vis.* 10, 16.1–17.
- Desimone, R., Albright, T. D., Gross, C. G., and Bruce, C. (1984). Stimulus-selective properties of inferior temporal neurons in the macaque. *J. Neurosci.* 4, 2051–2062.
- Edwards, R., Xiao, D., Keyser, C., Foldiak, P., and Perrett, D. (2003). Color sensitivity of cells responsive to complex stimuli in the temporal cortex. *J. Neurophysiol.* 90, 1245–1256.
- Eifuku, S., De Souza, W. C., Tamura, R., Nishijo, H., and Ono, T. (2004). Neuronal correlates of face identification in the monkey anterior temporal cortical areas. *J. Neurophysiol.* 91, 358–371.
- Eimer, M. (2000). Event-related brain potentials distinguish processing stages involved in face perception and recognition. *Clin. Neurophysiol.* 111, 694–705.
- Engell, A. D., and McCarthy, G. (2010). Selective attention modulates face-specific induced gamma oscillations recorded from ventral occipitotemporal cortex. *J. Neurosci.* 30, 8780–8786.
- Fabre-Thorpe, M., Richard, G., and Thorpe, S. J. (1998). Rapid categorization of natural images by rhesus monkeys. *Neuroreport* 9, 303–308.
- Fisch, L., Privman, E., Ramot, M., Harel, M., Nir, Y., Kipervasser, S., Andelman, F., Neufeld, M. Y., Kramer, U., Fried, I., and Malach, R. (2009). Neural “ignition”: enhanced activation linked to perceptual awareness in human ventral stream visual cortex. *Neuron* 64, 562–574.
- Freiwald, W. A., and Tsao, D. Y. (2010). Functional compartmentalization and viewpoint generalization within the macaque face-processing system. *Science* 330, 845–851.
- Freiwald, W. A., Tsao, D. Y., and Livingstone, M. S. (2009). A face feature space in the macaque temporal lobe. *Nat. Neurosci.* 12, 1187–1196.
- Goffaux, V., Peters, J., Haubrechts, J., Schiltz, C., Jansma, B., and Goebel, R. (2010). From coarse to fine? Spatial and temporal dynamics of cortical face processing. *Cereb. Cortex* 21, 467–476.
- Gothard, K. M., Battaglia, F. P., Erickson, C. A., Spitzer, K. M., and Amaral, D. G. (2007). Neural responses to facial expression and face identity in the monkey amygdala. *J. Neurophysiol.* 97, 1671–1683.
- Grill-Spector, K., and Kanwisher, N. (2005). Visual recognition: as soon as you know it is there, you know what it is. *Psychol. Sci.* 16, 152–160.
- Gross, C. G., Rocha-Miranda, C. E., and Bender, D. B. (1972). Visual properties of neurons in inferotemporal cortex of the macaque. *J. Neurophysiol.* 35, 96–111.
- Harris, A. M., and Aguirre, G. K. (2008). The effects of parts, wholes, and familiarity on face-selective responses in MEG. *J. Vis.* 8, 1–12.
- Hasselmo, M. E., Rolls, E. T., and Baylis, G. C. (1989). The role of expression and identity in the face-selective responses of neurons in the temporal visual cortex of the monkey. *Behav. Brain Res.* 32, 203–218.
- Hegde, J., and Van Essen, D. C. (2004). Temporal dynamics of shape analysis in macaque visual area V2. *J. Neurophysiol.* 92, 3030–3042.
- Hegde, J., and Van Essen, D. C. (2006). Temporal dynamics of 2D and 3D shape representation in macaque visual area V4. *Vis. Neurosci.* 23, 749–763.
- Hirabayashi, T., and Miyashita, Y. (2005). Dynamically modulated spike correlation in monkey inferior temporal cortex depending on the feature configuration within a whole object. *J. Neurosci.* 25, 10299–10307.
- Honey, C., Kirchner, H., and VanRullen, R. (2008). Faces in the cloud: Fourier power spectrum biases ultrarapid face detection. *J. Vis.* 8, 1–13.
- Horel, J. A. (1993). Retrieval of a face discrimination during suppression of monkey temporal cortex with cold. *Neuropsychologia* 31, 1067–1077.
- Hung, C. P., Kreiman, G., Poggio, T., and DiCarlo, J. J. (2005). Fast readout of object identity from macaque inferior temporal cortex. *Science* 310, 863–866.
- Kiani, R., Esteky, H., and Tanaka, K. (2005). Differences in onset latency of macaque inferotemporal neural responses to primate and non-primate faces. *J. Neurophysiol.* 94, 1587–1596.
- Kloth, N., Dobel, C., Schweinberger, S. R., Zwieterlood, P., Bolte, J., and Junghofer, M. (2006). Effects of personal familiarity on early neuromagnetic correlates of face perception. *Eur. J. Neurosci.* 24, 3317–3321.
- Kobatake, E., and Tanaka, K. (1994). Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *J. Neurophysiol.* 71, 856–867.
- Kuraoka, K., and Nakamura, K. (2006). Impacts of facial identity and type of emotion on responses of amygdala neurons. *Neuroreport* 17, 9–12.
- Kveraga, K., Boshyan, J., and Bar, M. (2007). Magnocellular projections as the trigger of top-down facilitation in recognition. *J. Neurosci.* 27, 13232–13240.
- Leonard, C. M., Rolls, E. T., Wilson, F. A., and Baylis, G. C. (1985). Neurons in the amygdala of the monkey with responses selective for faces. *Behav. Brain Res.* 15, 159–176.
- Lewis, M. B., and Edmonds, A. J. (2003). Face detection: mapping human performance. *Perception* 32, 903–920.
- Liu, J., Harris, A., and Kanwisher, N. (2002). Stages of processing in face perception: an MEG study. *Nat. Neurosci.* 5, 910–916.
- Matsumoto, N., Okada, M., Sugase-Miyamoto, Y., and Yamane, S. (2005a). Neuronal mechanisms encoding global-to-fine information in inferior-temporal cortex. *J. Comput. Neurosci.* 18, 85–103.
- Matsumoto, N., Okada, M., Sugase-Miyamoto, Y., Yamane, S., and Kawano, K. (2005b). Population dynamics of face-responsive neurons in the inferior temporal cortex. *Cereb. Cortex* 15, 1103–1112.
- Meeren, H. K., Hadjikhani, N., Ahlfors, S. P., Hamalainen, M. S., and de Gelder, B. (2008). Early category-specific cortical activation revealed by visual stimulus inversion. *PLoS ONE* 3, e3503. doi: 10.1371/journal.pone.0003503
- Mishkin, M., Ungerleider, L. G., and Macko, K. A. (1983). Object vision and spatial vision: two cortical pathways. *Trends Neurosci.* 6, 414–417.
- Moeller, S., Freiwald, W. A., and Tsao, D. Y. (2008). Patches with links: a unified system for processing faces in the macaque temporal lobe. *Science* 320, 1355–1359.
- Mormann, F., Kornblith, S., Quiroga, R. Q., Kraskov, A., Cerf, M., Fried, I., and Koch, C. (2008). Latency and selectivity of single neurons indicate hierarchical processing in the human medial temporal lobe. *J. Neurosci.* 28, 8865–8872.
- Nakamura, K., Matsumoto, K., Mikami, A., and Kubota, K. (1994). Visual response properties of single neurons in the temporal pole of behaving monkeys. *J. Neurophysiol.* 71, 1206–1221.
- Nakamura, K., Mikami, A., and Kubota, K. (1992). Activity of single neurons in the monkey amygdala during performance of a visual discrimination task. *J. Neurophysiol.* 67, 1447–1463.
- Navon, D. (1977). Forest before trees: the precedence of global features in visual perception. *Cogn. Psychol.* 9, 353–383.
- Optican, L. M., and Richmond, B. J. (1987). Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. III. Information theoretic analysis. *J. Neurophysiol.* 57, 162–178.
- Or, C. C., and Wilson, H. R. (2010). Face recognition: are viewpoint and identity processed after face detection? *Vision Res.* 50, 1581–1589.
- Perrett, D. I., Smith, P. A., Potter, D. D., Mistlin, A. J., Head, A. S., Milner, A. D., and Jeeves, M. A. (1985). Visual cells in the temporal cortex sensitive to face view and gaze direction. *Proc. R. Soc. Lond. B Biol. Sci.* 223, 293–317.
- Pinsk, M. A., Arcaro, M., Weiner, K. S., Kalkus, J. F., Inati, S. J., Gross, C. G., and Kastner, S. (2009). Neural representations of faces and body parts in macaque and human cortex: a comparative fMRI study. *J. Neurophysiol.* 101, 2581–2600.
- Puce, A., Allison, T., and McCarthy, G. (1999). Electrophysiological studies of human face perception. III: effects of top-down processing on face-specific potentials. *Cereb. Cortex* 9, 445–458.
- Quiroga, R. Q., Reddy, L., Kreiman, G., Koch, C., and Fried, I. (2005). Invariant visual representation by single neurons in the human brain. *Nature* 435, 1102–1107.
- Richmond, B. J., and Optican, L. M. (1987). Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. II. Quantification of response waveform. *J. Neurophysiol.* 57, 147–161.
- Richmond, B. J., Optican, L. M., Podell, M., and Spitzer, H. (1987). Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. I. Response characteristics. *J. Neurophysiol.* 57, 132–146.



- Riesenhuber, M., and Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nat. Neurosci.* 2, 1019–1025.
- Rousset, G. A., Fabre-Thorpe, M., and Thorpe, S. J. (2002). Parallel processing in high-level categorization of natural images. *Nat. Neurosci.* 5, 629–630.
- Rousset, G. A., Husk, J. S., Bennett, P. J., and Sekuler, A. B. (2008). Time course and robustness of ERP object and face differences. *J. Vis.* 8, 1–18.
- Sadeh, B., Podlipsky, I., Zhdanov, A., and Yovel, G. (2010). Event-related potential and functional MRI measures of face-selectivity are highly correlated: a simultaneous ERP-fMRI investigation. *Hum. Brain Mapp.* 31, 1490–1501.
- Saleem, K. S., Suzuki, W., Tanaka, K., and Hashikawa, T. (2000). Connections between anterior inferotemporal cortex and superior temporal sulcus regions in the macaque monkey. *J. Neurosci.* 20, 5083–5101.
- Serre, T., Oliva, A., and Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. *Proc. Natl. Acad. Sci. U.S.A.* 104, 6424–6429.
- Sheinberg, D. L., and Logothetis, N. K. (1997). The role of temporal cortical areas in perceptual organization. *Proc. Natl. Acad. Sci. U.S.A.* 94, 3408–3413.
- Sheinberg, D. L., and Logothetis, N. K. (2001). Noticing familiar objects in real world scenes: the role of temporal cortical neurons in natural vision. *J. Neurosci.* 21, 1340–1350.
- Sigala, N., and Logothetis, N. K. (2002). Visual categorization shapes feature selectivity in the primate temporal cortex. *Nature* 415, 318–320.
- Singer, J. M., and Sheinberg, D. L. (2006). Holistic processing unites face parts across time. *Vision Res.* 46, 1838–1847.
- Sripati, A. P., and Olson, C. R. (2009). Representing the forest before the trees: a global advantage effect in monkey inferotemporal cortex. *J. Neurosci.* 29, 7788–7796.
- Sripati, A. P., and Olson, C. R. (2010a). Global image dissimilarity in macaque inferotemporal cortex predicts human visual search efficiency. *J. Neurosci.* 30, 1258–1269.
- Sripati, A. P., and Olson, C. R. (2010b). Responses to compound objects in monkey inferotemporal cortex: the whole is equal to the sum of the discrete parts. *J. Neurosci.* 30, 7948–7960.
- Sugase, Y., Yamane, S., Ueno, S., and Kawano, K. (1999). Global and fine information coded by single neurons in the temporal visual cortex. *Nature* 400, 869–873.
- Tamura, H., and Tanaka, K. (2001). Visual response properties of cells in the ventral and dorsal parts of the macaque inferotemporal cortex. *Cereb. Cortex* 11, 384–399.
- Tanaka, H., and Fujita, I. (2000). Global and local processing of visual patterns in macaque monkeys. *Neuroreport* 11, 2881–2884.
- Tanaka, H., Onoe, H., Tsukada, H., and Fujita, I. (2001). Attentional modulation of neural activity in the macaque inferior temporal cortex during global and local processing. *Neurosci. Res.* 39, 469–472.
- Tanaka, J. W. (2001). The entry point of face recognition: evidence for face expertise. *J. Exp. Psychol. Gen.* 130, 534–543.
- Tanaka, J. W., and Farah, M. J. (1993). Parts and wholes in face recognition. *Q. J. Exp. Psychol. A* 46, 225–245.
- Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annu. Rev. Neurosci.* 19, 109–139.
- Thorpe, S., Fize, D., and Marlot, C. (1996). Speed of processing in the human visual system. *Nature* 381, 520–522.
- Tovee, M. J., Rolls, E. T., Treves, A., and Bellis, R. P. (1993). Information encoding and the responses of single neurons in the primate temporal visual cortex. *J. Neurophysiol.* 70, 640–654.
- Tsao, D. Y., Freiwald, W. A., Knutsen, T. A., Mandeville, J. B., and Tootell, R. B. (2003). Faces and objects in macaque cerebral cortex. *Nat. Neurosci.* 6, 989–995.
- Tsao, D. Y., Freiwald, W. A., Tootell, R. B., and Livingstone, M. S. (2006). A cortical region consisting entirely of face-selective cells. *Science* 311, 670–674.
- Tsao, D. Y., Moeller, S., and Freiwald, W. A. (2008). Comparing face patch systems in macaques and humans. *Proc. Natl. Acad. Sci. U.S.A.* 105, 19514–19519.
- Tsunoda, K., Yamane, Y., Nishizaki, M., and Tanifuji, M. (2001). Complex objects are represented in macaque inferotemporal cortex by the combination of feature columns. *Nat. Neurosci.* 4, 832–838.
- VanRullen, R., and Thorpe, S. J. (2001). The time course of visual processing: from early perception to decision-making. *J. Cogn. Neurosci.* 13, 454–461.
- Wang, Y., Fujita, I., and Murayama, Y. (2000). Neuronal mechanisms of selectivity for object features revealed by blocking inhibition in inferotemporal cortex. *Nat. Neurosci.* 3, 807–813.
- Yamane, S., Kaji, S., and Kawano, K. (1988). What facial features activate face neurons in the inferotemporal cortex of the monkey? *Exp. Brain Res.* 73, 209–214.
- Yamane, Y., Carlson, E. T., Bowman, K. C., Wang, Z., and Connor, C. E. (2008). A neural code for three-dimensional object shape in macaque inferotemporal cortex. *Nat. Neurosci.* 11, 1352–1360.
- Yamane, Y., Tsunoda, K., Matsumoto, M., Phillips, A. N., and Tanifuji, M. (2006). Representation of the spatial relationship among object parts by neurons in macaque inferotemporal cortex. *J. Neurophysiol.* 96, 3147–3156.
- Young, M. P., and Yamane, S. (1992). Sparse population coding of faces in the inferotemporal cortex. *Science* 256, 1327–1331.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 01 March 2011; accepted: 12 June 2011; published online: 23 June 2011.

Citation: Sugase-Miyamoto Y, Matsumoto N and Kawano K (2011) Role of temporal processing stages by inferior temporal neurons in facial recognition. *Front. Psychology* 2:141. doi: 10.3389/fpsyg.2011.00141

This article was submitted to *Frontiers in Perception Science*, a specialty of *Frontiers in Psychology*.

Copyright © 2011 Sugase-Miyamoto, Matsumoto and Kawano. This is an open-access article subject to a non-exclusive license between the authors and Frontiers Media SA, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and other Frontiers conditions are complied with.