Chinese Science Bulletin 2006 Vol. 51 No. 1 1—7

DOI:

The time course of visual categorization: Electrophysiological evidence from ERP

CHEN Antao, LI Hong, QIU Jiang & LUO Yuejia

School of Psychology, Southwest University; Key Laboratory of Cognition and Personality, MOE, Chongqing, 400715, China
Correspondence should be addressed to LI Hong (email: lihong1@swu.edu.cn) or Luo Yuejia (email: luoyj@bnu.edu.cn)

Received November 24, 2005; accepted April 29, 2006

Abstract Visual categorization can be derived from interaction between inputting features set (IFS) from outside stimuli and anticipating features set (AFS) from concept, and such interaction can take two forms: match extraction and conflict control. Using ERP recording, we investigated the temporal course of visual categorization. The results indicated that the larger the AFS, the higher the amplitude of the N1 was generated, which demonstrated the effect of the AFS on the attention. When the size of the AFS was larger than or equal to 2, prominent N2 component was elicited, which demonstrated the effect of conflict when the feature of IFS mismatched with the feature of the AFS. The judgment of category was manifested on the LPC component, and this component was also sensitive to conflict control. Based on the results, we proposed that the temporal course of visual categorization was as follows: selective attention, feature perception, feature match/extraction and judgment of category/conflict control. Among those processes, the judgment of category is the core processing; however the former four sub-processes form the base of categorization. The results are in support of the idea that LPC is responsible for high-level categorization process.

Keywords: visual categorization, event-related potentials, AFS (anticipating features set), IFS (inputting features set), conflict control, LPC.

Categorization is the process to ascertain the identity of an object by a specific standard[1,2]. Categorization is a fundamental cognitive process; in fact, the raw perception would be of little use without the effective classification of original information into distinct categories[3]. There have been numerous studies on this topic exploring the psychological and neural mechanisms underlying categorization[4,5].

The early researches supposed that the amplitude of LPC reflected the degree of the categorization processing involving context updating[6,7]. Ito and Cacioppo[8] discovered that LPC component of ERP was sensitive to the explicit categorization task. Azizian *et al.*[9] explored the ERP reflection of similarity changes between target and standard stimuli in categorization and discovered that the amplitude of P300 can reflect the categorization effect of similarity. Besides, Thorpe *et al.*[5] found that human brain can discriminate correctly the natural objects very quickly against complex backgrounds: the waveforms elicited by the target stimuli (photos including animals) and distraction stimuli (photos without animals) diverge at 150 ms after stimuli onset, the waveform elicited by the non-animal photos went negatively from 150 ms while that elicited by animal photos went on positively until 180 ms when it turned negative. Antal *et al.*[10] found that non-animal photos elicited more negative N1 and N2 in 150—250 and 350—500 ms time windows than animal photos while animal photos elicited more positive P2 component. They thought that N1 was relevant to high level categorization processing. Fabre-Thorpe *et al.*[11] discovered that the categorization speed of natural objects could not be improved through sufficient training of subjects, which implied that visual categorization underwent many a few processing stage and these stages could not be compressed.

As illustrated above, from the ERP study on categorization, two different conclusions can be reached. One group concluded that categorization is reflected by late components LPC or P300 and the other group concluded that categorization is reflected in component as early as N1.

In our opinion, the reason for this discrepancy lies in different understanding of the exact cognitive process underlying categorization. It is reasonable to assume that categorization is a process involving interaction between the inputting features set (IFS) and anticipating features set (AFS) and such an interaction can take three different forms: extracting of the inputting feature if it matches that of anticipation, controlling the inputting feature if it mismatches that of expectancy and ignoring the inputting feature if there is no corresponding feature in the anticipation.

We take Thorpe *et al.*'s research as an example[5].

When categorization responses between animal and non-animal photos were required, the subjects' prior knowledge of the feature of a typical "animal" formed the base for the top-down AFS (i.e. four limbs, eyes, wings, hair, furs, etc.). At the same time, the visual features characterizing the object formed IFS (i.e. red feathers, the beaks, etc.) in a bottom-up manner. As long as any one feature in the IFS falls in the AFS, the subject can make "yes" response.

The majority of research, which suggests that the early ERP components reflect categorization processing, was conducted in the animal vs. non-animal paradigm used in Thorpe *et al.*[5]. When the inputting features came from known animals, nearly all the inputting features would fall into the AFS, therefore not causing any features conflict. It is important to note that the brain cannot complete perceptual processing until 150 ms after the stimuli onset[12]. Therefore, the ERP components reported in these studies may only reflect the perception of the individual features, but may not be responsible for the higher level processes, such as categorization including conflict control. Categorization including conflict control is more valuable and universal for human beings' adaptive behaviors, after all the gross differentiation of animal VS non-animal is rather simple considering the complexity of the requirement for cognitive process in real world.

As described above, categorization occurs based on the interaction between IFS and AFS; therefore in order to reveal the detailed mechanism, it is necessary to systematically manipulate the features of stimuli. Standard geometric figures can be a suitable stimulus set. In addition, experimenter can also easily control the subjects' experience on the stimulus[13]. In the present study, the subjects were required to first compare two geometric figures and find the common features (forming AFS), and then followed with a categorization task which required the subjects to take the AFS as standard and judge whether the third geometric figure shared a category with the preceding two geometric figures. The categorization task requirement is to make "yes" judgment if the IFS just has one feature that matches the AFS in correspondent dimension, and making "no" judgment if no any feature in the IFS matches the AFS in correspondent dimensions.

There were thus three different interactions scenarios between IFS and AFS: the inputting feature falls in AFS and matches the feature in correspondent dimension (causing match-extraction processing); the inputting feature fall in AFS and mismatches with the feature in correspondent dimension (causing conflict-control processing); the inputting feature does not fall in AFS and are irrelevant to the task (causing irrelevance-ignorance processing). Three experimental conditions were implemented: condition 1, AFS has one common feature (one matching feature, 1MF); condition 2, AFS has two common features (two matching features condition, 2MF); condition 3, AFS has three common features (three matching features condition, 3MF). Through this novel experimental design, we expect to gain better understanding of the brain mechanisms underlying the temporal process of higher visual categorization including conflict control.

1 Method

1.1 Subjects

Fifteen undergraduate students served as subjects, of which 7 males and 8 females, ages 21—28, 24 on average. All subjects were physically and mentally healthy, right-handed, and had normal or corrected-to-normal vision. All subjects signed an informed consent form for the experiment.

1.2 Materials

The stimuli were all regular geometrics, each stimulus having a feature in each of three dimensions: color (yellow, blue, green, and red), shape (triangle, square, circle, and cross), and stripe orientation (0°, 45°, 90°, and 135°). RGB value of color is 225, 235 and 0 (for yellow), 0, 221 and 255 (for blue), 0, 255 and 50 (for green), 255, 51 and 0 (for red), respectively. The combination of the four feature levels and the three dimensions makes up 4×4×4=64 different stimuli (Fig. 1). The stimuli were all drawn in CorelDRAW11, and were individually exported and saved as a bmp file. The sizes of the figures are: 4.56 cm base and 5.92 cm high for triangles, 4.28 cm edge for squares, 4.28 cm diameter for circles, and 4.24 wheelbase for crosses (in experiment, the distance between subjects' eyes and the screen is larger than 1 m; therefore horizontal and vertical angles were both less than 3.5°). To avoid the impact of dazzling flashes on brain electrical activities, black screen background was adopted.

1.3 ERP Recording

The EEG was recorded from 64 scalp sites using tin electrodes mounted in an elastic cap (NeuroScan Inc.),

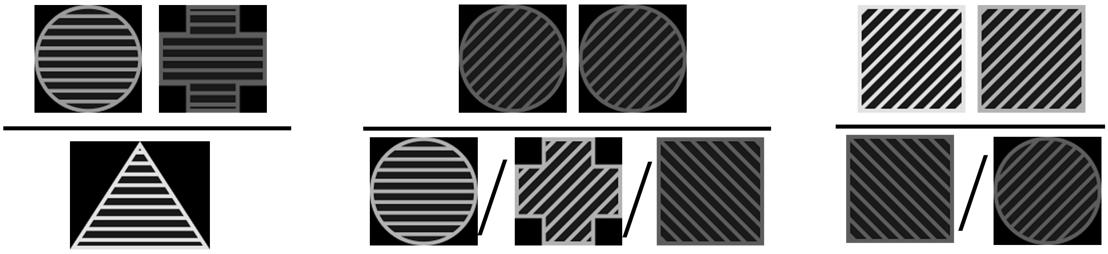

Fig. 1. Illustration of stimuli material, experimental tasks and correspondent answers. only shape and orientation were presented here.

with the linked reference on the left and right mastoids, and a ground electrode was placed on the medial aspect of the frontal. The horizontal and vertical electrooculograms (EOGs) were recorded. The EEG and EOG were amplified using a DC-100 Hz bandpass and continuously sampled at 500 Hz/channel. All inter-electrode impedance was maintained below 5 kΩ. Averaging of ERPs was computed off-line; the analysis course began 100 ms before onset of stimulus, and continued until 700 ms after the onset of the stimuli. Trials with EOG artifacts (mean EOG voltage exceeding ±80 μV) and those contaminated with artifacts due to amplifier clipping, peak-to-peak deflection exceeding ±80 μV were excluded from averaging.

1.4 Procedure and Task

The procedure of each trial was as follows: first, a central fixation cross ‘‘+’’ was presented in center of screen for 500 ms; second, S1 and S2 was presented simultaneously, one in each side of the center of screen for 1500 ms followed by a blank screen for 1000—1500 ms as the interval, the subjects were asked to compare the two stimuli and find out their common features (forming a new category C); third, target stimulus (S3) was presented, and the subjects had to judge by pressing one of two response keys (F or J) within 2000 ms whether S3 belonged to the category C. Completing the categorization, there was a blank screen for 1000 ms.

Category C was defined with features. There are three conditions (Fig. 1): (1) C was defined with just one feature. For example, “the stripe orientation was 0°”. If strip orientation of S3 was 0°, then the subject should give a positive response (Fig. 1(a)). (2) C was defined with two features. For example, “the shape is square and the stripe orientation was 45°”. If shape of S3 was square *or* stripe orientation of S3 was 45°, then the subject should give a positive response (Fig. 1(c)). (3) C was defined with three features. For example, “the shape was circle, the color was red, and the stripe orientation was 45°”. If shape of S3 was a circle, *or* the color of S3 was red *or* the stripe orientation of S3 was 45°, then the subject should give a positive response (Fig. 1(b)). Therefore, three categorization tasks were obtained: one matching feature condition (1MF, one match abstraction processing, two irrelevant ignorance processing), two matching features condition (2MF, one match abstraction processing, one conflict control processing, and one irrelevant ignorance processing), and three matching features condition (3MF, one match abstraction processing, two conflict control processing).

Only the subjects whose accuracy in the practice arrived at 90% would step into the formal experiment. Because the classification happened after S3 onset, the ERP waveforms were time-locked to the onset of S3.

1.5 ERP data analysis and statistics

The recorded EEGs in the three conditions were overlapped and averaged respectively. The ERP of the following electrode sides were selected for statistic analyses: F3, Fz, F4, Fc3, Fcz, Fc4, C3, Cz, C4, Fc5, Fc6, C5, C6 (13 sites for anterior); Cp3, Cpz, Cp4, P3, Pz, P4, Po3, Po4, Poz (9 sites for posterior). The N1 component may have separable anterior and posterior subcomponents. The amplitudes (baseline to peak) of anterior N1 and posterior N1 were thus measured separately in the 80—120 and 120—160 ms time windows, respectively. The amplitudes (baseline to peak) of P2 (190—250 ms) and N2 (250—310 ms) were measured at both anterior and posterior electrodes, respectively. Mean voltage of LPC in the time window of 330—480 ms was measured at both anterior and posterior electrodes.

Amplitudes (baseline to peak) of the early components (N1, P2, and N2) and mean amplitude in the time window of 330—450 ms (LPC) were analyzed using two-way repeated measures analysis of variance (ANOVA). For all analyses, *P* was corrected for deviations according to Greenhouse-Geisser.

2 Results

2.1 Behavioral data

The accuracies of classifying S3 were recorded. Averaged accuracy was 98.08% in 1MF, 96.63% in 2MF, and 95.65% in 3MF. According to the Signal Detection Theory, if a subject's *P*(A) value is larger than 0.5, his/her responses were reliable (*P*(A) = (hit number + correct rejection number)/total stimulus number). Analysis results indicated that all 15 subjects' responses were reliable. The ANOVA result of response time (RT) indicated that the condition main effect was significant ($F(2,28) = 43.129$, $P<0.001$), the response way (yes or no) main effect was significant ($F(1,14) = 14.721$, $P = 0.002$), but the condition by response way interaction was not significant ($F(2,28) = 3.511$, $P>0.05$).

Obviously, when the classification task needs a positive response, the involving cognitive processes consist of match abstraction, conflict control and irrelevant ignorance; but when the negative response is needed, the involving cognitive processes do not include match abstraction. Because the match abstraction is the key processing in the classification, only the processing under positive response could be identified as classification, and in the present paper only the correct positive response was analyzed. In the positive response way, RT of 1MF, 2MF and 3MF was 651 ms, 818 ms and 844 ms, respectively. The condition main effect of RT was significant ($F(2,28) = 82.29$, $P<0.001$), mean amplitudes pairwise comparison of RT showed that 1MF was significantly faster than 2MF ($F(1,14) = 129.68$, $P<0.001$), 2MF was significantly faster than 3MF ($F(1,14) = 6.609$, $P = 0.022$).

2.2 ERP Components

For the anterior N1 (N100), there was significant condition main effect ($F(2,28) = 3.575$, $P = 0.047$), electrode effect was, too, $F(12,168) = 4.947$, $P = 0.003$, but the condition by electrode interaction was not, $F(24,336) = 0.904$, $P = 0.105$. Mean amplitudes pairwise comparison of condition showed that there was not significant difference between 1MF and 2MF ($F(1,14) = 1.728$, $P = 0.210$), and there was not any significant difference between 2MF and 3MF, either, ($F(1,14) = 1.644$, $P = 0.221$), but there was significant difference 1MF and 3MF ($F(1,14) = 8.371$, $P = 0.012$). For the posterior N1 (N140), there was not any significant condition main effect ($F(2,28) = 0.493$, $P = 0.540$), electrode effect was significant ($F(8,112) = 7.547$, $P = 0.003$), and the condition by electrode interaction was not significant ($F(16,224) = 1.554$, $P = 0.218$).

For P2, there was significant condition main effect ($F(2,28) = 4.213$, $P = 0.036$), electrode effect was not significant ($F(21,294) = 1.046$, $P = 0.357$), nor was the condition by electrode interaction $F(42,588) = 2.447$, $P = 0.046$. Mean amplitudes pairwise comparison of condition showed that there was significant difference between 1MF and 2MF ($F(1,14) = 7.622$, $P = 0.015$), there was no significant difference between 2MF and 3MF ($F(1,14) = 0.001$, $P = 0.975$), there was significant difference between 1MF and 3MF ($F(1,14) = 9.442$, $P = 0.008$).

For N2, there was significant condition main effect ($F(2,28) = 15.636$, $P<0.001$), electrode effect was significant ($F(21,294) = 6.202$, $P = 0.01$), and the condition by electrode interaction was significant ($F(42,588) = 4.647$, $P = 0.001$). Mean amplitudes pairwise comparison of condition showed that there was significant difference between 1MF and 2MF ($F(1,14) = 12.051$, $P = 0.004$), there was no significant difference between 2MF and 3MF ($F(1,14) = 2.133$, $P = 0.166$), there was significant difference 1MF and 3MF ($F(1,14) = 45.820$, $P<0.001$).

For LPC, there was significant condition main effect ($F(2,28) = 38.508$, $P<0.001$), electrode effect was significant ($F(21,294) = 8.016$, $P = 0.001$), and the condition by electrode interaction was significant ($F(42,588) = 2.731$, $P = 0.47$). Mean amplitudes pairwise comparison of condition showed that there was significant difference between 1MF and 2MF ($F(1,14) = 43.303$, $P<0.001$), there was no significant difference between 2MF and 3MF ($F(1,14) = 0.674$, $P = 0.425$), there was significant difference 1MF and 3MF ($F(1,14) = 96.992$, $P<0.001$).

3 Discussion

The behavioral data of RT revealed that the response in 1MF was faster significantly than that in 2MF which reversely was significant faster than that in 3MF. As stated above, 1MF includes one match extraction processing and two irrelevant ignorance processings. 2MF includes one match extraction processing, one conflict control processing, and one irrelevant ignorance processing. And 3MF includes one match extraction processing and two conflict control processings. Obviously, the three conditions involve the same number of processing and each has one match extraction processing, and therefore, the difference among them lies in conflict control or irrelevant ignorance processing. The result indicates that the conflict control processing is more difficult than irrelevant ignorance processing, implying that conflict control processing is likely to happen in more complex thinking stage while irrelevant ignorance processing in a simpler stage (such as attention).

Three conditions elicited prominent N1 in central-parietal sites at 100 ms after S3 onset and the amplitudes of three conditions showed regular descending relations in negative direction: 3MF elicited the highest amplitudes, the amplitude of 2MF was medium and 1MF showed the lowest amplitudes. According to most ERP study results, N1 is related to attention[15]. The subjects should attend to the greatest number (three) of features when they keep three possible features in working memory, the number (two) of features to be attended was less than the former when the subjects keep two possible features in working memory and they need only to attend one when keeping one feature in working memory. Therefore, the negative amplitudes in N1 just reflected the number of features subjects should attend to, that is to say, the greater the number, the higher the amplitude of N1.

From 150 ms after S3 onset, the waveform of 1MF diverged from those of the other two conditions, and the waveforms of all three conditions reached their maximal amplitudes at about 220 ms, but the prominent P2 was found only in 2MF and 3MF. P2 also appeared in 1MF and its amplitude was larger than the other two conditions. According to the relevant researches[5], the divergence arising at 150 ms between 1MF and 2MF, 3MF signaled the beginning of feature perception, and this processing may end at 220 ms. Larger P2 amplitude in 1MF revealed that this condition caused stronger perceptual processing. Otherwise, the smaller P2 amplitudes of 2MF and 3MF indicated the decentralization of feature perception because greater number of target features brought about weaker perceptual match. Besides, significant condition and electrode interaction appeared in P2, greater difference happened at central-frontal sites while no difference occurred in occipital-parietal sites, which is consistent with the waveform showed in Fig. 2, indicating that perceptual processing is mainly related to the central-frontal sites[13].

N2 is closely related with mental conflict[16,17]. In the present research, 2MF and 3MF both elicited prominent N2 whereas almost no component appeared in 1MF, which indicated that 2MF and 3MF elicited stronger mental conflict. Significant condition and electrode interaction appeared in N2; this was probably caused by the fact (Fig. 2) that weaker differences happened at occipital-parietal sites while stronger difference occurred in central-frontal sites, suggesting that perceptual processing is mainly related to the central-frontal sites[17]. The interaction is likely related to conflict detection correlated with the activation in anterior cingulated cortex[17]. Analysis showed that conflict control happened in two dimensions in 3MF, in one dimension in 2MF and no conflict-inhibition appeared in 1MF. Obviously, the existence of conflict control would raise mental conflict, and the more conflict processes, the greater the mental conflict. The N2 amplitude relations among three conditions well reflected the point stated above: 1MF raised no obvious N2 component, while 2MF and 3MF elicited obvious N2. 3MF elicited slightly higher amplitude than 2MF, which stemmed from conflict control occurring in two dimensions for the former, while it happened only in one dimension for the latter.

After N2, 1MF elicited very obvious LPC component peaking at 370 ms while no obvious LPC component appeared in 2MF and 3MF (Fig. 2). The relation could be explained by the fact that the human brain began to control the conflicting features on conscious thinking level after conflict detection. It is generally accepted that LPC reflects the advanced and complex mental processing. Donchin thought that the latency of LPC represents the time needed to evaluate and categorize the stimuli[6]. Through analysis and taking into account the nature of the present task, we can conclude that LPC represents the end of categorization that in essence is the judgment of category based on perceptual analysis.

Besides, some researchers hold that LPC represents some inhibition on stimulus processing[18]. In the present research, 1MF includes no inhibition processing while 2MF and 3MF both include inhibition processing, as is clearly reflected on the amplitude relations among three conditions: the amplitude elicited by 1MF was significantly higher than those elicited by 2MF or 3MF. However, there was no significant difference in P3 amplitude between 2MF and 3MF, which is possibly re-

是否要对X,Y轴进行说明？

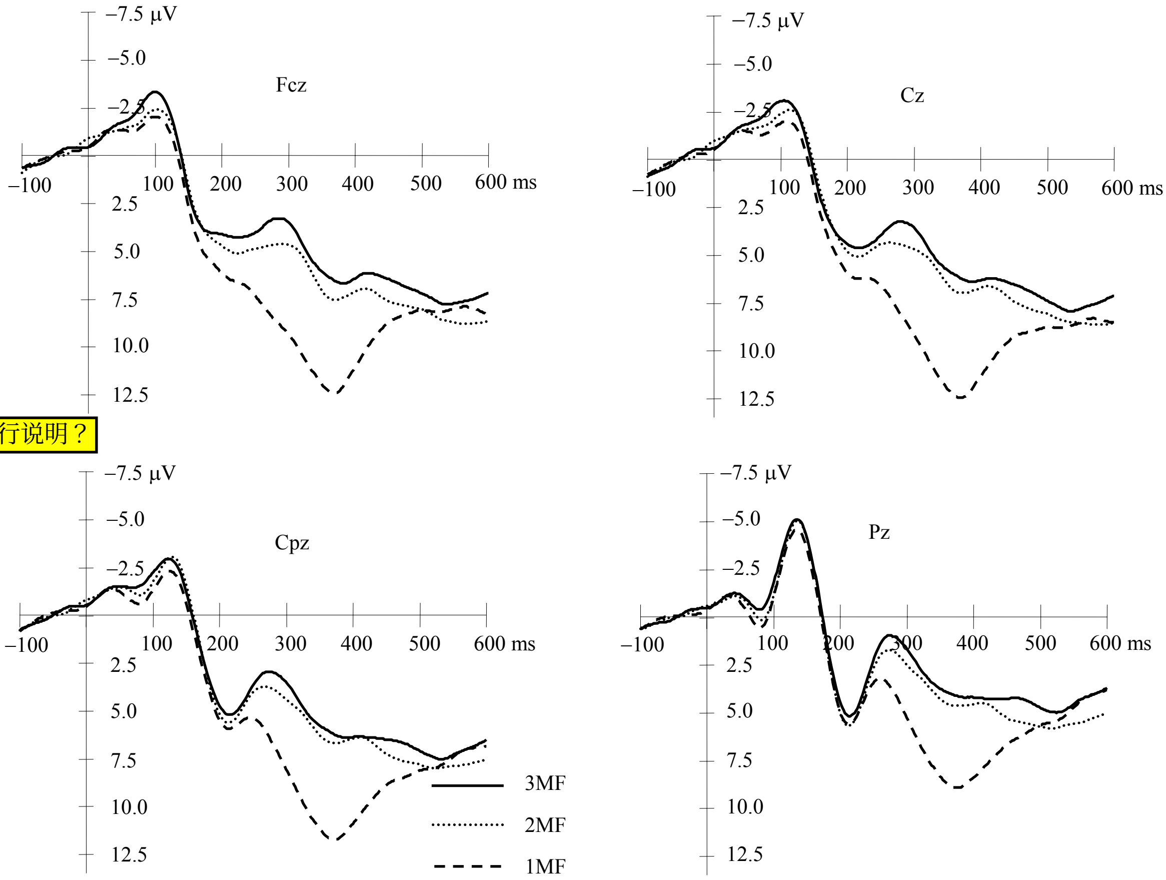

Fig. 2. Grand average ERPs at FCz, Cz, CPz and Pz for all conditions.

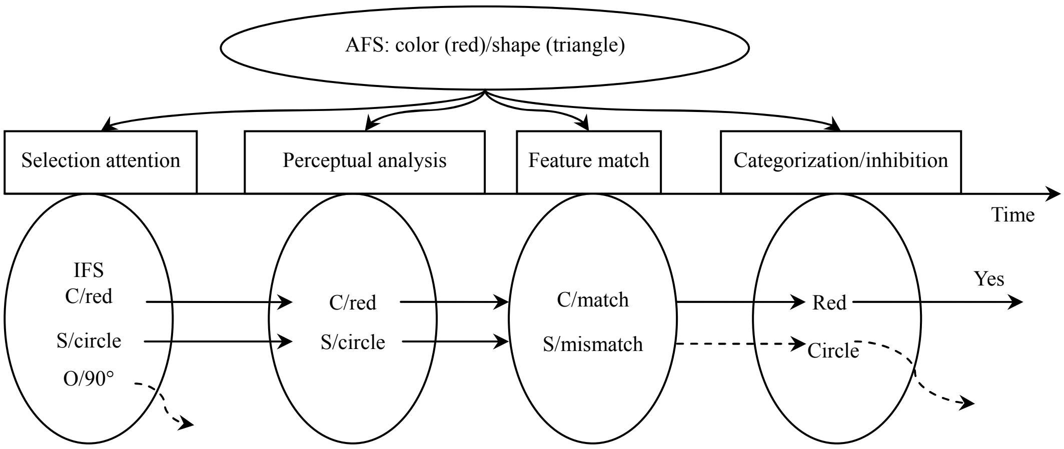

Fig. 3. The temporal course of visual categorization (take 2MF as the example).

lated to the parallel processing of thinking[19] but pending further research.

With the result of the present study, we can describe the temporal course of categorization (Fig. 3). Subjects would form two kinds of preparations according to the expectancy: dimension preparation and feature preparation. Dimension preparation was embodied in attention processing: attending to target dimension (i.e. color, shape) and ignoring irrelevant dimension (i.e. stripe). 100 ms after stimulus onset, subjects only paid attention to target dimension (solid line) and ignored irrelevant dimension (orientation, dashed line), that is, target dimension was further processed and the waveform manifestation of attention processing was N1 component. Further processing was perceptual analysis by which the subject could recognize what the features were in target dimension (red, circle) and the waveform manifestation of perceptual analysis is P2.

After attention and perceptual analysis, the categorization stepped into feature match stage in which the effect of feature preparation embodied. The inputting feature originating from perceptual analysis would be compared with the features in AFS. If the inputting feature matched with the correspondent feature in AFS, then the inputting feature would be extracted for next processing. If the inputting feature mismatched with the correspondent feature in AFS, then the inputting feature would cause conflict and elicit N2.

After the termination of perceptual analysis was the judgment of category and conflict control stage in which subjects further controlled the conflicting features and made the positive judgment according to the matched feature. And the waveform manifestation of it was typically LPC whose amplitude would decrease in 2MF and 3MF because of the control on the conflicting features. It is worthwhile to notice that the brain cannot eliminate easily the feature impression when the feature has been represented in the thought. Therefore, the inhibition on the conflicting features would last for a while, which could account for the smaller LPC amplitude of 2MF and 3MF than that of 1MF.

References

1 Zentall T R, Galizio M, Critchfield T S. Categorization, concept learning, and behavior Analysis: An introduction. J Exp Anal Beha, 2002, 78(3): 237—248

2 Mai X, Luo Y. ERP study on human visual categorization of natural objects. Adv Psychol Sci (in Chinese). 2004, 12(2): 168—175

3 Freedman D J, Riesenhuber M, Poggio T, et al. Categorical representation of visual stimuli in the primate prefrontal cortex. Science, 2001, 291(5502): 312—316

4 Anderson J R. The adaptive nature of human categorization. Psychol Rev, 1991, 98(3): 409—429

5 Thorpe S, Fize D, Marlot C. Speed of processing in the human visual system. Nature, 1996, 381(6582): 520—522

6 Donchin E. Surprise! …Surprise? Psychophysiology, 1981, 18(5): 493—513

7 Donchin E, Coles M G H. Is the P300 component a manifestation of context updating? Behav Brain Sci, 1988, 11(3): 357—374

8 Ito T A, Cacioppo J T. Electrophysiological evidence of implicit and explicit categorization processes. J Exp Social Psychol, 2000, 36(6): 660—676

9 Azizian A, Freitas A, Watson T D, et al. Electrophysiological correlates of categorization: P300 amplitude as index of target similarity. Biol Psychol, 2006, 71(3): 278—288

10 Antal A, Keri S, Kovacs G, et al. Early and late components of visual categorization: An event-related potential study. Cognitive Brain Res, 2000, 9(1): 117—119

11 Fabre-Thorpe M, Delorme A, Marlot C, et al. A limit to the speed of processing in ultra-rapid visual categorization of novel natural scenes. J Cognitive Neurosci, 2001, 13(2): 171—180

12 Hillyard S A, Anllo-Vento L. Event-related brain potentials in the study of visual selection attention. Proc Natl Acad Sci USA, 1998, 95(3): 781—787

13 Bigman Z, Pratt H. Time course and nature of stimulus evaluation in category induction as revealed by visual event-related potentials. Biol Psychol, 2004, 66(2): 99—127

14 Chen A, Li H, Feng T, et al. The diversity effect of inductive reasoning under segment manipulation of complex cognition. Sci China Ser C-Life Sci, 2005, 48(6): 658—668

15 Mangun G R. Neural mechanisms of visual selective attention. Psychophysiology, 1995, 32(1): 4—18

16 Wang Y, Kong J, Tang X, et al. Event-related potential N270 is elicited by mental conflict processing in human brain. Neurosci Lett, 2000, 293(1): 17—20

17 van Veen V, Carter C S. The timing of action monitoring processes in anterior cingulate cortex. J Cognitive Neurosci, 2002, 14(4): 593—602

18 Goldstein A, Spencer K M, Donchin E. The influence of stimulus deviance and novelty on the P300 and Novelty P3. Psychophysiology, 2002, 39(6): 781—790

19 Rousselet G A, Fabre-Thorpe M, Thorpe S. Parallel processing in high-level categorization of natural images. Nat Neurosci, 2002, 5(7): 629—630