Picture categorizing processing in a cross-modal interference task

ATSUKO ISHIO
Department of Educational Psychology, Faculty of Education, Kyoto University, Sakyo-ku, Kyoto 606

By using a cross-modal, auditory-visual picture-word interference task, two experiments were conducted to examine how picture categorization proceeds. To clarify the various conditions under which interference effects are expected to rise, several types of different degrees of picture-word semantic relation and the stimulus onset asynchrony (SOA) were varied, and the comparison between auditory-visual and visual-visual task was made. Two types of distracting words were used; names of pictures (Exp. 1) or categories (Exp. 2). Twenty-one (Exp. 1) and 24 (Exp. 2) students participated and they were instructed to categorize pictures as rapidly as they could, ignoring word components. In Exp. 1, there were no significant interference or facilitation effects at all SOAs for both tasks. In Exp. 2, significant interference or facilitation effects were shown around 0 ms of SOA for both tasks. These results were discussed in relation to the effect of task demand on picture processing, to the locus of interference and to the difference in interference pattern between cross-modal and unimodal tasks.

Key words: picture categorization, Stroop effects, cross-modality, interference, facilitation.

When a word is printed inside the line drawing of a concrete object and the word refers to an object different from that in the picture, the presence of the incongruent word significantly delays picture naming compared with the presence of a meaningless CVC (consonant-vowel-consonant) trigram (Rosinski, Golinkoff, & Kukish, 1975). The interference between a picture and a word has been considered as an analogy to the Stroop task (Stroop, 1935). Glaser and Dungelhoff (1984) concluded from their several experiments that there was an identical underlying cognitive process in both tasks. However, the question of the locus of interference has not yet solved. Three types of hypothesis have been proposed and these hypothesis have located the effects at the stages of perceptual encoding, semantic evaluation, or response production (Glaser & Dungelhoff, 1984).

Recently, Cowan and Barron (1987) examined the effects of auditory color-word interference on the visual Stroop task. When a color-word and a color-patch were presented by different modality, we could avoid the possibility that two visual stimuli interacted at the stage of early perceptual encoding. Thus, observed cross-modal interference effects suggested that the interference effects occurred at least after the stage of early perceptual encoding. Cowan and Barron (1987) located the cross-modal interference effects at the stage of response production.

In picture-word interference tasks, Schriefers and Meyer (1990) and Ishio (1990) demonstrated that cross-modal interference effects could be obtained. Schriefers and Meyer (1990), in the experiment of varying the stimulus onset asynchrony (SOA), considered the temporal relation of target and distractor presentation to be most important for the cross-modal interference effects. Ishio (1990), on the other hand, suggested that the interference effects depended upon not only the timing of stimulus presentation but also upon the semantic relations between pictures and words: There was more inhibition when a word and a picture belonged to the same category than when they each other belonged to different categories. This result pointed

---

1 Part of this research was presented at the 22nd International Congress of Applied Psychology, Kyoto, Japan, 1990.
2 I would like to express my appreciation to Dr. Noboru Sakano of Kyoto University for his valuable comments on this article. I also would like to thank Dr. Akira Yamazaki of Hiroshima University for his helpful advice.
out that the stage of semantic evaluation was related to the interference in a cross-modal task.

In addition, it should be noted that it was only in the cross-modal task that nonsense words resulted in much interference in naming pictures. Visual nonsense words caused no significant interference at all SOAs. Considering nonsense words did not require the access of semantic information, this interference was not caused by the conflict at the stage of semantic evaluation. Probably the auditory nonsense words passed through (or bypassed) the stage of the semantic evaluation and occurred the interference at the stage of response production. The interference effects which was caused by auditory nonsense words were located at the stage of response production.

The result that visual nonsense words caused no interference even when they were presented prior to pictures suggested the possibility that visual nonsense words didn’t reach at the stage of response production. If visual nonsense words had arrived at the output stage, they should have occurred some interference effects like auditory nonsense words had done. Visual nonsense words might have been interrupted to process before the stage of response production, as soon as they have been recognized that they had no meaning (Ishio, 1990). So far, the question whether interference was occurred at the stage of semantic evaluation, at the stage of response production or at both stages has remained unsolved.

Glaser and Düngelhoff (1984) examined the mechanism of the Stroop interference by using picture naming, word reading, and picture and word categorizing tasks. Since the deeper semantic processing is demanded in a categorizing task, the results from that task contributed to elaborate the processing model of the Stroop interference including the structure of semantic information. In earlier categorizing studies, however, nonsense words, especially auditory nonsense words, have not been used as word stimuli. In a cross-modal naming task, auditory nonsense words caused inhibitions at the stage of response production. Similarly, in a categorizing task, auditory nonsense words will reach at the stage of response production and come into conflict with the phonological information of the relevant response, for subjects should be required verbal responses like a naming task. If auditory nonsense words cause any interference effects in a cross-modal categorizing task, it will suggest not only the stage of semantic evaluation but also the stage of response production to be the locus of interference. Then we have to reconsider the whole mechanism of processing as there are two loci of interference.

In a picture-word task, the interference pattern is reversed when the task is changed from naming to categorization. Incongruent words cause strong inhibitions in a picture naming task, whereas picture categorization is relatively immune to interference from incongruent words. On the other hand, word categorization is easily disrupted by the presence of an incongruent picture, whereas word naming is hardly affected by the incongruent picture. Researchers have been interested in this phenomenon, because it suggests that word recognition and picture recognition have different semantic processing. Considering that a categorizing task requires the access of semantic information, the reversal of the interference pattern means that semantic and articulatory information become differently available between pictures and words: pictures access semantic information prior to articulatory information and words have the reverse order of information access (Smith & Magee, 1980). However, Glaser and Düngelhoff (1984) showed that words couldn’t disrupt picture categorizing when words were presented much prior to pictures and had much time to process, and threw doubt on the explanation which Smith and Magee (1980) had proposed. In addition, recent literatures demonstrated that pho-
nological recoding for the purpose of lexical access is not necessary in the cognitive process of reading (Besner & Hildebrandt, 1987). Some other reasons why the reversal of the interference pattern occurs should be examined to construct the processing model of interference.

The first purpose of the present study was aimed at examining how interference pattern can be obtained in a cross-modal picture categorizing task: for example, whether auditory nonsense words cause any interference in a categorizing task and whether the reversal of interference is shown. The expected results may be suggestive to the theories concerning the locus of interference. Few studies about cross-modal interference have conducted with refined methods. To clarify the features of a cross-modal categorizing task the comparison between auditory-visual and visual-visual tasks was introduced. The main task was almost the same as that of Ishio (1990). To clarify the factors that influence on the amount of interference, picture-word semantic relation and stimulus onset asynchrony (SOA) were varied.

The second purpose was to examine picture processing in categorization. Whether pictures always access semantic information prior to articulatory information was investigated in Experiment 2.

Experiment 1

Method

Subjects. Twenty-one undergraduate and graduate students participated, with 10 in the auditory-visual task (A-V Task) and 11 in the visual-visual task (V-V Task). All subjects had normal or corrected-to-normal vision.

Stimulus materials. Twenty-four line drawings each of which belong to one of six categories (animals, clothing, vehicles, stationery, fruit, furniture) were chosen from the list constructed by Yoshikawa and Inui (1986). All pictures were presented as black line drawings on a white background within the size of 6.5 × 6.5 cm at the center of the CRT screen connected to the personal computer (NEC PC9801M). The mean luminance of pictures was about 22.48 cd/m². Two types of word stimuli were varied by Task Type (A-V Task or V-V Task). Auditory word stimuli spoken by a female speaker were 24 objects’ names of the pictures, 24 nonsense words and a control stimulus, the sound produced by knocking on a desk. The mean intensity of auditory stimuli was around 63 dB. The visual word stimuli were written words of auditory stimuli except control stimulus. In V-V Task, the control stimulus was a row of four “×”s, “××××”. Visual word stimuli were written in Japanese orthography, Katakana and each letter was 1 cm high and 1 cm long on the CRT screen. Visual words were presented at the center of the CRT screen and consequently a part of each picture was covered by a word when the picture and the word existed simultaneously.

Each picture was combined with five types of word stimuli: same stimulus (SS; both stimuli refer to the same objects), same category (SC; both stimuli belong to the same category), incongruent (IC; stimuli each other belong to different categories), neutral (N; pictures are combined with nonsense words) and control (C; “××××” or the sound knocking on a desk is presented with the picture, as mentioned above) conditions.

Procedure. Task Type was manipulated between subjects. Six types of SOA (−500, −200, −100, 0, +100, +500 ms) and five types of Congruency (SS, SC, IC, N, C) were varied within subjects.

Before the experiment, the subjects were shown the 24 pictures’ list and instructed to name each picture and its category to make sure of their names. At the experimental session, the subjects sat about 100 cm apart from the CRT screen and were asked to name the category that belonged to the presented picture as rapidly and accurately as they could. Each trial began by presenting “+” on the CRT screen with a warning sig-
nal (buzzer). After 500 ms interval, a picture stimulus (target) was presented during SOA, which was followed by a word stimulus (distractor). Thus, a minus sign with a SOA indicates preexposure of a word stimulus. The timer started to measure the reaction time (RT) at the onset of the picture and stopped simultaneously when subjects produced their verbal responses. The intertrial interval was approximately 3 s.

Subjects were given 48 practice trials followed by 120 experimental presentations. A short rest occurred after 60 trials. Each picture-word combination was presented once with each SOA. The stimuli were administered to each subject in six constant SOA blocks. Within each SOA pictures were presented in randomized order avoiding repetitions of identical category in direct succession. The order of SOA conditions was counterbalanced across subjects. Subjects participated over two days, with three SOA blocks in a day.

Results

For each subject, mean RTs for each within-subject cell were calculated, the RTs made by error response and deviated by more than two standard deviations from the respective subjects’ mean RTs being excluded. In order to illustrate the effects of Congruency conditions, the individual mean RT of C condition in each SOA was subtracted from the mean RTs of SS, SC, IC and N conditions. Figures 1 and 2 show mean facilitation and inhibition scores as a function of SOAs in four types of picture-word combinations differing in degree of congruency for A-V and V-V tasks.

A three way 2 (Task Type) × 6 (SOA) × 5 (Congruency) ANOVA was performed on the mean RTs, which revealed significant main effect of SOA [F(5, 550) = 6.91, p < .001] and significant interaction of Task Type × SOA [F(5, 550) = 10.93, p < .001]. The main effect of Congruency was marginally significant [F(4, 550) = 2.01, p < .1]. The main effect of Task Type and other interactions showed no significant differences.

To examine the effects of Congruency or Task Type within each SOA, an ANOVA (Task Type × Congruency) was conducted on the mean RTs. There were significant or marginally significant main effects of Congruency at -500, -200, -100, 0, +100 and +500 ms of SOA [F(4, 76) = 3.06, p < .03; F(4, 76) = 2.93, p < .03; F(4, 76) = 2.26, 

Figure 1. Mean facilitation and inhibition scores as a function of stimulus onset asynchronies (SOAs) in four types of picture-word combinations differing in degree of congruency (auditory-visual task).

Figure 2. Mean facilitation and inhibition scores as a function of stimulus onset asynchronies (SOAs) in four types of picture-word combinations differing in degree of congruency (visual-visual task).
Interaction of Task Type by Congruency were significant at -500, -200 and -100 ms of SOA \( F(4,76) = 4.04, p < .005; F(4,76) = 4.82, p < .002; F(4,76) = 2.82, p < .04 \). However, further analysis for Congruency within each SOA in each task showed that there were no differences among Congruency conditions at all SOAs. Similarly, further analysis for Task Type in each Congruency condition within each SOA revealed no significant differences between in A-V and in V-V Tasks at all SOAs.

**Discussion**

The purpose of Experiment 1 was to investigate what interference pattern there might be in a cross-modal picture categorizing task. In particular, it should be noticed whether auditory nonsense words took any effects on picture categorization.

When the RT of each SS, SC, IC and N condition within each SOA for each task significantly longer (or shorter) than that of C condition within the same cell, it is regarded as a sign of significant interference (or facilitation) effect. However, the results indicated no significant interference or facilitation effects at all SOAs in both tasks. Although these results, for V-V Task, replicated those by the previous researchers (Glaser & Düngelhoff, 1984; Smith & Magee, 1980), almost the same results could be obtained also for A-V Task. Incongruent words, irrespective of their modalities, couldn't affect picture categorizing. Even auditory nonsense words could not cause any interference. Does the present fact that picture categorizing is immune to interference necessarily mean that pictures have preferential access to semantic information?

Yoshikawa (1987), in her investigation of priming effects on picture naming and categorizing, designed several types of prime-target combination with the SOA of 700 ms. From the theory of spreading activation model (Collins & Loftus, 1975), she expected for both naming and categorizing tasks the facilitation effects in both ID (prime and target are identical pairs) and RE (related pairs; both items belong to the same semantic category) conditions, and additionally, much more amount of facilitation in ID condition than that in RE condition. However, for the naming task, facilitation was obtained only in ID condition but not in RE condition, and approximately the same amount of facilitation was observed in both ID and RE conditions for the categorizing task. She discussed that priming effects were dependent upon the identification levels (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976) of target pictures. According to Rosch et al. (1976), in taxonomies of concrete objects, there are several levels of abstraction, superordinate (e.g., vehicle), basic level (e.g., car) and subordinate (e.g., sports car). For the picture naming task, the basic level of pictures might be more activated as compared with superordinate, and thus the activation of superordinate by a prime word in RE condition might not affect picture naming. On the other hand, for the categorizing task subject might activate higher identification level, and the activation of basic level by a prime word in ID condition could not affect picture categorizing more than in RE conditions. Consequently, no difference in the amount of facilitation effects had occurred between ID and RE conditions.

This theory might explain that the absence of interference effects was obtained in the conditions of auditory nonsense words. In the categorizing task, subjects might have activated higher identification level, and nonsense words, which were regarded as the words of basic level, might have been interrupted to process, even if they should have reached at the stage of response production.

In the following Experiment 2, such a prediction proposed by Yoshikawa (1987) was examined by presenting the names of catego-
ories as distracting words instead of the names of pictures. If pictures always access semantic information prior to articulatory information, there will be little interference in picture categorization even if distracting words are changed to the names of categories. However, if subjects in accordance with the task demand change the identification level to activate, presentation of the names of categories as distracting words would result in any interference in a categorizing task. To construct the processing model in which more results from different tasks should be explained and to compare a cross-modal task with a unimodal task, both auditory-visual and visual-visual tasks were conducted.

Experiment 2

Method

Subjects. Twenty-four graduate and undergraduate students were served, with 12 in each task. None had participated in Experiment 1 and each had normal or corrected-to-normal vision.

Stimulus materials. Picture stimuli were identical to those of Experiment 1. Word stimuli were changed from the names of pictures to six names of categories (e.g., animals, clothing), and control stimulus. Nonsense words were excluded. Three types of picture-word combination were used: S (Same; word stimulus within a picture is the name of the category to which the picture belong), I (Incongruent; word is the name of the incongruent category to which the picture belong), and C (control) conditions.

Procedure. Task Type (A-V or V-V) was manipulated between subjects. Five types of SOA (−500, −100, 0, +100, +500 ms) and three types of Congruency (S, I, C) were varied within subjects. Other procedures were the same as those of Experiment 1.

Results

Mean RTs of individual subject were calculated as in Experiment 1. Figures 3 and 4 show mean facilitation and inhibition scores as a function of SOAs in two types of picture-word combinations for A-V and V-V tasks.

A three way 2 (Task Type)×5 (SOA)×3 (Congruency) ANOVA yielded significant
main effects of Congruency $[F(2,308) = 176.22, p < .001]$ and SOA $[F(4,308) = 7.40, p < .001]$. Interactions of Task Type $\times$ SOA and Congruency $\times$ SOA were significant at the 0.1 per cent level $[F(4,308) = 5.64; F(8,308) = 13.07]$. The main effect of Task Type and other interactions were not significant.

An ANOVA (Task Type $\times$ Congruency) was conducted on the mean RTs within each SOA. The main effects of Congruency were shown at -500, -100, 0 and +100 ms of SOAs $[F(2,44) = 69.03; F(2,44) = 44.42; F(2,44) = 73.25; F(2,44) = 87.39, p < .001]$. Significant main effect of Task Type was obtained at 0 ms of SOA $[F(1,44) = 6.51, p < .02]$. At +100 ms of SOA, the main effect of Task Type was marginally significant $[F(1,44) = 2.95, p < .1]$. Further analysis were performed for Congruency within each SOA in each task. For A-V Task, there were significant facilitation effects in S conditions at -500, -100 and 0 ms of SOAs and a significant interference effect in I condition at 0 ms of SOA. For V-V Task, a significant facilitation effect was found in S condition at -500 ms of SOA and significant interference effects were observed in I conditions at 0 and +100 ms of SOA.

Further analysis for Task Type revealed that the RTs of I conditions in V-V Task were longer than those of I conditions in A-V Task at 0 and +100 ms of SOA $[t(12) = 4.90; t(12) = 1.67, p < .1]$ and the RTs of S conditions in A-V Task were significantly shorter than those of S conditions in V-V Task at 0 and +100 ms of SOA $[t(12) = 3.71; t(12) = 2.49, p < .05]$. Discussion

The purpose of Experiment 2 was to investigate picture processing in a categorizing task. When the names of categories were presented as distracting words, incongruent words could disrupt picture categorizing. These results can not be explained completely by the theory concerning the different order of information access. As Yoshikawa (1987) proposed, subjects might have changed the identification level to activate in accordance with the task demand. Considering that subjects has already understood the task demand, it is possible that subjects may intend to process pictures at certain specific level (Yoshikawa, 1987).

Alternative explanation, however, may be possible. Dunbar and MacLeod (1984), following the theory derived from Interactive Models, discussed that the members of response set primed certain response more than the nonmembers throughout the experiment. The threshold for activation of primed responses should be lower than that for unprimed responses and it is easier for the members of response set to be activated and to interfere with targets. According to this theory, the task of categorizing pictures activated not the names of pictures but the names of categories, because the names of pictures were not the members of response set. Thus interference effects should be shown in Experiment 2, not in Experiment 1.

Another interesting results were obtained in Experiment 2. When pictures and words were presented under SOA = 0 and +100 ms, the RTs of I conditions in V-V Task were longer than those of the same Congruency conditions in A-V Task. On the other hand, the RTs of S conditions in A-V Task were significantly shorter than those in V-V Task. The result suggests that visual words are more likely to interfere with the processing of visual pictures than auditory words, that is, the interference effects between the same modalities is stronger than those between different modalities. For the facilitation effects, the cross-modal task may produce them much more than the unimodal task. However, more detailed experiments should be necessary before this explanation can be accepted, because consistent results about this issue weren't obtained for the cross-modal, picture naming task (Ishio, 1990).

In addition, some suggestions for the locus
of interference were obtained. Although the semantic evaluation hypothesis seems more favorable to the explanation of the interference, it fails to explain the results of the present study, that is, the reason why the interference effects reverse in two different (naming or categorization) tasks. According to the semantic evaluation hypothesis, the degrees of which words semantically overlapped with pictures should not be influenced by the difference in tasks (Young, Ellis, Flude, McWeeny, & Hay, 1986). Furthermore the response competition hypothesis at a production stage can not explain why the interference effects emerged in the categorizing task as observed in the present study by changing the contents of word distractors. If one follows the hypothesis, all words should have been processed more rapidly than pictures and should have arrived at the response stage prior to pictures in the naming task and the reverse should be available in the categorizing task. Finally, it may be insufficient to explain the mechanism of interference in relation only to the questions at which stage the interference occurs.

Cohen, Dunbar, and McClelland (1990) presented a model of attention using the Stroop effect as an example. They assumed that the information of task demand had a role of allocation of attention and had to be provided as input to the model. When we examine the processing mechanisms of various tasks, we should at least consider not only the factors associated with stimulus materials but also those associated with task demand. As discussed by Toma and Toshima (1989), a normal organism was not a passive information processing machine but an active entity who performed the task in a goal-oriented manner following the task demand. In the future, further research derived from the wide view point in which includes the information of task demand will be needed to investigate the mechanism of interference.

References
Young, A.W., Ellis, A.W., Flude, B.M., McWeeny, K.H.,

(Received May 7, 1991; accepted Nov. 14, 1992)