EFFECTS OF CONCEPTUAL CATEGORIZATION ON EARLY VISUAL PROCESSING

SIWEI LIU

NATIONAL UNIVERSITY OF SINGAPORE

2013

EFFECTS OF CONCEPTUAL CATEGORIZATION ON EARLY VISUAL PROCESSING

SIWEI LIU
(M.Sc., University of York, UK)

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN PSYCHOLOGY

DEPARTMENT OF PSYCHOLOGY

NATIONAL UNIVERSITY OF SINGAPORE

2013

DECLARATION

I hereby declare that this thesis is my original work and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in the thesis.

This thesis has also not been submitted for any degree in any university previously.

Siwei Liu

13 March, 2014

Acknowledgements

I would like to thank the following people:

Trevor Penney, for his guidance, his humor, his support in my difficult times, and his patience with my mistakes. For the freedom of exploring, and the timely advice in the midst of confusion.

Annett Schirmer, for her instructions in my learning, her help, and her offer in spite of inconvenience.

Lonce Wyse, for his encouragement and his optimism.

Angela and Seetha, for their help in the data recording phases of several experiments. Hui Jun for her involvement in my research projects.

Nicolas Escoffier, my companion on the path of doing a PhD. For his answers to my questions at various stages of this research. For his calm supports and insights when I ran into problems of both research and life. For the coffee breaks, the music, the trips, and the beer we shared. And for the friendship that we are proud of.

Eric Ng and Ranjith, my weekend and late-night neighbours. Eric, for his answers to my statistics-related questions and for our common interests in Hongkong. Ranjith, for the philosophical, political, and all other intellectual debates we had, and the long lists of movie and book recommendations.

All the BBL labmates. For the good times we spent together, their willingness to participate in my pilots and help during the EEG setup. Adeline, Darshini, Latha, Pearlene, Darren, Ivy, Attilio, Yong Hao, Shi Min, Yng Miin, Ann, Shawn, Suet Chian, Tania, April, Karen, Ling, Brahim, Christy, Claris, Maria,

Shan, Antarika, Stella and Steffi. Bounce, for extinguishing my fear of dogs.

All other friends. For the good times we had, and the help when needed. Lidia, Joe, Saw Han, Smita, Pek Har, Mei Shan, Yu, and Hui.

Uncle 9, for accommodating me for more than five years since my arrival.

Michael, for the love, and for the joy and the hardship we shared. Especially since the second half of last year, for his emotional and financial support, and for taking care of me during my illness.

My mother and father, and the rest of my family for the unconditional love.

Contents

Summary

The effects of conceptual categorization on early visual processing were examined in six experiments by measuring how familiar and individually-identifiable auditory stimuli influenced event-related potential (ERP) responses to subsequently presented visual stimuli. Early responses to the visual stimuli, as indicated by the P1 component, were modulated by whether the auditory and the visual stimuli belonged to the same basic-level category (e.g., dogs) and whether, in cases where they were not from the same basic-level category, the categorization levels were congruent (i.e., both stimuli from basic level categories versus one from the basic level and the other from the subordinate level). The current study points to the importance of the interplay between categorization level and basic-level category congruency in cross-modal object processing.

List of Figures and Tables

List of Abbreviations

ANOVA	Analysis of variance
EEG	Electroencephalography
ERP	Event-related Potential
fMRI	Functional magnetic resonance imaging
SOA	Stimulus onset asynchrony
SD	Semantic dementia

Chapter 1 Introduction

The human brain categorizes information from the world, both to understand the information and to generate predictions. We learn to slot objects into different conceptual categories, such as that a poodle belongs to the Dog category and a dog belongs to the Animal category. Knowing the category of an object helps us to infer features that may not be perceivable immediately or directly. For example, knowing that the object is a dog allows us to infer that it can bark, even though we have not heard it do so. Given that a dog is an animal, we can readily apply the features that belong to the Animal category to the dog. Moreover, if we also know that the dog is a poodle, we can further apply the features that specifically belong to the Poodle category to the dog.

Research suggests that there is a hierarchical structure comprising different levels of category inclusiveness and specificity for the knowledge space in humans (Medin, 1989; Cohen & Lefebvre, 2005). Whereas the concept animal belongs to the superordinate level of abstraction, poodle belongs to the subordinate level. Rosch, Mervis, Gray, Johnson, and Boyes-Braem (1976) pointed out that these abstraction levels are not equally easy to access. They argued that we tend to recognize objects at the basic level, where we find a balance between inclusiveness and specificity. Categorizing objects at more specific levels allows us to predict their characteristics more precisely. For instance, a poodle behaves differently from a porcelaine, but the trade-off is that categorization may take more processing time and effort at the subordinate level (Murphy & Smith, 1982).

In an experiment designed to examine the features participants added when moving from a less specific to a more specific level, Rosch et al. (1976) asked participants to describe objects at the subordinate, basic, and superordinate levels. They found that the number of additional features was larger when moving from the superordinate level to the basic level than when moving from the basic level to the subordinate level. In subsequent experiments, they found that the basic level objects shared similar visual images; people handled them with similar motor programs; and concepts at the basic level were also easiest to access and were learned earlier by children (Rosch et al., 1976). Hence, objects from the same category share the largest number of common features at the basic level. Tanaka and Taylor (1991) pointed out that most of the added features comprise perceptual information, rather than functions. They quantified four types of additional information, including additional object parts, modified object parts, different object dimensions (e.g., size and color), and behaviors (or functions). They found that the first three were more critical than the last one. Moreover, for domain experts, who were known for their rich knowledge about the object categories at the subordinate level, the type of perceptual information that became more important with more specific categorization depended on the domain of expertise. For example, dog experts tended to add in more object parts as additional features at the subordinate level, whereas bird experts tended to modify attributes listed at the basic level. Interestingly, the bird experts were more likely to use the subordinate category names to describe birds, but the dog experts did not show a preference for particular category level names. Both expert groups added more features to the subordinate level than did novices.

Of course, relatively few humans are bird experts, but most are experts in perceiving human faces. We understand that a face comprises two eyes, one nose and one mouth and that the two eyes must be next to each other and above the nose, which in turn is above the mouth. Both the nose and the mouth must be roughly aligned to the center of the eyes. These are the feature configurations of a face, but we are usually not satisfied with merely knowing that the object is a face. We prefer to individually recognize the face. Feature configuration, such as the distance between the eyes and the eye height relative to the mouth, plays an important role in face recognition (Tanaka & Sengco, 1997; Goffaux & Rossion, 2007; Young, Hellawell, & Hay, 1987; Sigala & Logothetis, 2002). Changing the eye height in an image can lead people to believe that the face belongs to a different person (Haig, 1984). Feature configuration is important when moving from the basic level of classifying an image as a face to the more specific subordinate level of determining face identity (Barton, Press, Keenan, & O’Connor, 2002).

Although most studies of conceptual categorization have used visual stimuli, Adams and Janata (2002) showed that auditory stimuli are subject to the same categorization level effects as visual stimuli. They asked participants to match visually presented words with either pictures or sounds. They manipulated the category levels of the words to form four conditions for each modality. For example, for a sound or picture of a crow, the subordinate-level match was crow; the subordinate-level mismatch was sparrow; the basic-level match was bird; and the basic-level mismatch was cat. Note that in the subordinate-level mismatch condition, words still matched with the visual or auditory stimuli at the basic level (e.g., both crows and sparrows are birds). The results indicated that participants were faster and more accurate for basic level stimuli than subordinate level stimuli for both pictures and sounds. Moreover, although participants were more accurate and faster matching pictures than sounds, there was no significant interaction between stimulus modality and categorization level.

1.1 Conceptual Categorization in the Brain

Object categorization appears to occur both within and beyond the primary sensory cortices. For example, Adams and Janata (2002) showed that the inferior frontal regions in both hemispheres responded to object categorization regardless of stimulus modality. Responses in the fusiform gyri also corresponded to the categorization level, though it was restricted to the left hemisphere for auditory stimuli.

Lee (2010) instructed participants to remember sounds and pictures from different categories without instructing them about the category level at which they should discriminate the sounds. Hence, most participants were expected to use the basic level by default. Lee contrasted brain responses to the different basic-level categories and found that the regions recruited to discriminate animate basic level categories were more lateral on the STG than the inanimate discriminative regions. In a second experiment, Lee extended the comparison to the visual modality. He compared the discriminative areas for animate versus inanimate categories when the stimuli were sounds or pictures. Modality-specific early sensory areas were found for both sounds and pictures (e.g., the middle portion of the middle temporal sulcus for sounds, and the lingual gyrus for pictures) and there were also areas activated by both sounds and pictures (e.g., the right supplementary motor area).

Studies of the structure of the perceptual system support a multimodal processing model. Information is processed within each modality and later bound into a unitized perception either by direct connections or synchronizations between different modalities or by convergence zone(s) (Calvert & Thesen, 2004). The conceptual system is argued to be multimodal in that object representations are also distributed among modalities (Barsalou, 1999; Rogers & Patterson, 2007).

Rogers and Patterson (2007) adopted the Rosch et al. (1976) paradigm to compare normal participants and patients with semantic dementia (SD). While they replicated Rosch et al.'s findings with the normal participants, the SD patients showed better performance at the superordinate level than at the basic and subordinate levels. These findings contradict the idea that processing first occurs at the basic level before spreading to other levels. The superordinate categorization was preserved among the SD patients and only the basic or subordinate levels of categorization were impacted. Rogers & Patterson (2007) proposed a parallel distributed model. Following the view that object representations are distributed across different regions in the brain, they argued that the representations are not only stored in the modality-specific regions. Though regions in sensory cortex store modality-specific features, the concepts that link the features are stored in the anterior temporal cortex and are represented in distributed patterns according to the similarities among features. The basic level advantage is explained based on the distinctive-and-informative principle. The inputs from the sensory areas share more similarity at the basic level than at the superordinate level. For example, pears share similar shapes with lightbulbs. However, pears and lightbulbs differ greatly in color, texture, function, and so on. The representations of pears and lightbulbs may be similar in the “shape”-processing regions. But representations in other regions will be very different. Their representation in the anterior temporal cortex will be distinctive. The subordinate level representations in the anterior temporal cortex share more structural similarity with each other and require more precise input from the sensory areas. Therefore, objects are identified faster at the basic level in the category verification task. They further suggested that when receiving input from the sensory areas, the superordinate level concepts begin to be activated earlier than the basic or subordinate levels, but are slower to reach the threshold for output production. However, if forced to respond before any threshold is reached, the participants recognize the objects more accurately at the superordinate level.

1.2 EEG and ERPs

In light of the behavioral evidence for different categorization levels, potential brain electrophysiological signatures that reflect this distinction have also been examined. Three ERP components that have been used as tools to examine object categorization, the P1, N170, and P2, are described in this section.

1.2.1 P1

The P1 is a positive ERP deflection that occurs between 50 and 150ms after visual stimulus onset. The scalp distribution is bilateral over occipito-temporal electrode sites with sources believed to be in the lateral occipital cortex, extrastriate visual cortex, ventral visual cortex, and around the posterior fusiform area (Mangun, Buonocore, Girelli, & Jha, 1998; Mangun et al., 2001).

The P1 is sensitive to the physical features of visual stimuli, such as contrast, spatial frequency, and luminance. For example, Ellemberg and colleagues (2001) tested responses to stimuli with different contrasts and spatial frequencies and the P1 component was elicited by all spatial frequencies in the experiment except the highest (i.e., 16 c per degree). At each spatial frequency, the P1 component appeared at low contrast, its amplitude rapidly increased as the contrast increased to a medium level, but did not increase further as contrast was increased. Indeed, at the highest tested spatial frequency, the P1 amplitudes experienced a sharp drop. The authors suggested that this P1 amplitude profile resembled features one may expect from magnocellular responses. Osaka and Yamamoto (1978) reported that as stimulus luminance increased, the P1 latency decreased. However, Johannes, Münte, Heinze, and Mangun (1995) failed to find an effect of luminance on P1 latency, but did find that the higher the stimulus luminance the larger the P1 amplitude.

Stimulus feature complexity also increases P1 amplitude. Martinovic, Gruber, and Müller (2008) examined the effects of different features on the P1 component, which included surface details, visual complexity, and color typicality. They found that adding surface details and increasing visual complexity enhanced the P1 amplitude, but the P1 latency was shorter for more complex stimuli than less complex ones. However, presenting an object with an atypical color did not affect the P1 amplitude or latency.

Most important for the present work is evidence that the P1 component reflects object recognition processes. For example, inverted faces elicited larger P1 amplitudes and/or P1 latencies compared to upright faces (e.g., Itier and Taylor, 2004b; Allison et al., 1999; Taylor, 2002) even though inverted and upright faces shared the same low-level physical features, which suggests that a higher level of visual processing was involved. Furthermore, Freunberger and colleagues (2008a) presented images with different levels of distortion. By reducing the distortion, half of the images resolved into pictures of objects while the others remained meaningless patterns. Participants were instructed to respond as quickly and accurately as possible when they recognized the objects in the pictures. Comparison of the ERP responses to pictures of living objects, non-living objects, and distorted images revealed that the P1 amplitudes elicited by the distorted images were significantly larger than the responses to the living and the non-living object images.

The P1 amplitude is also affected by attention. Early research tended to focus on the effect of spatial attention effect on the P1 amplitude. For example, in a spatial cueing paradigm (e.g., Van Voorhis & Hillyard, 1977), participants were instructed to fix their eyes on the screen center while attending to either the left or the right visual field. Across trials, visual stimuli appeared in the attended or the unattended visual field. The participant’s task was to detect visual targets within the attended visual field, ignoring stimuli appearing in the other visual field. Each stimulus appeared briefly (e.g., 200ms) to avoid saccades. ERP responses to the stimuli appearing in the same visual field were compared between the attended and the unattended conditions. The P1 component elicited by the stimuli was larger over the contralateral hemisphere when the visual field was attended compared to when it was not attended.

Allocation of attention during an experiment can be manipulated in different ways. For example, attention can be allocated to different visual fields in different blocks (e.g., Mangun et al., 2001) or allocated on a trial-by-trial basis (e.g., Eimer, 1998; Mangun & Hillyard, 1991). In the trial-by-trial allocation experiments, attention was directed by a symbol (e.g., an arrow) placed at screen center. Most of the time, the symbol correctly predicted where the next visual stimulus would appear, but occasionally the prediction was incorrect. Comparison of the P1 responses between the validly and the invalidly cued trials when the visual stimuli were presented in the same visual field revealed that valid trials elicited a larger P1 over the contralateral hemisphere. Given attention was directed to the relevant visual field in the valid trials, but not in the invalid trials, attention accounts for the difference in P1 amplitudes.

Attending to different visual fields from block to block and directing attention using symbols involves voluntary allocation of attention. However, attention can also be directed automatically to a visual region by briefly presenting a stimulus in the peripheral visual field. The effects of attention in this case are slightly different from those of voluntary attention allocation. In a study by Hillyard, Luck, and Mangun (1994), four dots at the four corners of the screen marked the possible target locations. One dot disappeared 50ms before the target was presented. This acted as a valid cue most of the time (75%). Although the behavioral results showed that RT was facilitated for the valid trials compared to the invalid trials, no cue validity effect was observed for P1 amplitudes. As noted by Briand and Klein (1987), peripheral cueing versus symbol cueing may involve different attention systems.

The invalid cue trials in the spatial cueing paradigm also reveal the P1 response elicited by the visual field that does not contain a stimulus. The P1 response on these trials can be compared to valid trials in which attention was correctly directed away from the visual field without a stimulus. P1 amplitudes over the hemisphere ipsilateral to the target stimulus were larger on invalid than on valid trials. This means that the attention effect on P1 amplitude does not require a stimulus input (Mangun et al., 2001).

Klimesch (2011) proposed that the P1 component reflects inhibition such that when more inhibition is needed, the P1 amplitude is larger. For the hemisphere contralateral to the stimulus presentation field, the P1 reflects inhibition processes that enhance the signal to noise ratio (SNR). For the ipsilateral hemisphere, the P1 component reflects reduction of task-irrelevant activations. He (Klimesch, 2011) also pointed to the link between the P1 amplitudes and stimulus complexity. Using inhibition theory, he argued that longer word length, inverted or scrambled faces as compared to upright faces, and distorted images, all increased the complexity of the stimuli and therefore required more inhibition efforts in stimulus processing. Therefore, the P1 amplitudes were larger for long word length, inverted faces, scrambled faces, and distorted images.

To summarize, the P1 component is related to object recognition, which is linked to conceptual categorization. The P1 amplitude differentiates between meaningless patterns and meaningful objects. It reflects efforts to slot objects into conceptual categories. Therefore, unfamiliar, distorted, and/or complex objects, all of which require more categorization effort, increase P1 amplitude. The P1 component is also subject to selective attention modulation. Increasing attention allows better stimulus processing and the effect of selective attention on the P1 component may reflect facilitation of object categorization.

1.2.2 N170

The N170 is a negative-polarity ERP component that typically peaks approximately 170ms after the onset of a visual stimulus. It is distributed over the occipital temporal area with sources in the fusiform gyrus and the superior temporal areas (e.g., Itier & Taylor, 2004a; Herrmann, Ehlis, Muehlberger, & Fallgatter, 2005). The N170 is affected by physical features such as spatial frequency, contrast, size, and visual noise (e.g., Rossion & Jacques, 2008; Eimer, 2011). Eimer (2000b) found that the viewing angle of a face also affects the N170; full front upright view (0 degree) and profile view (90 degree) elicited similar N170 amplitudes, whereas the back side-view (135 degree) and back view (180 degree) elicited smaller N170s.

In addition, the N170 component is more sensitive to faces than other objects (Bentin, Allison, Puce, Perez, & McCarthy, 1996; Rossion & Jacques, 2008). Specifically, the N170 component elicited by human faces is larger than those elicited by non-face objects, such as houses, cars, furniture, and human hands (Bentin et. al., 2007; Bentin et. al., 1996; Carmel & Bentin, 2002; Kovacs, et. al., 2006). See Thierry, Martin, Downing, & Pegna (2007) and Bentin et al. (2007) for discussions of whether Interstimulus Perceptual Variance accounts for the N170 face effect.

Comparing the N170 elicited by human faces and by faces of other species (e.g., monkeys, dogs) reveals that the N170 differs in amplitude between humans and other animals (e.g., Eimer, 2000b; Itier & Taylor, 2004a; Carmel & Bentin, 2002; Rousselet, Macé, & Fabre-Thorpe, 2004; de Haan, Pascalis, & Johnson, 2002; Itier, Van Roon, & Alain, 2011; Bentin et al., 1996; Gajewski & Stoerig, 2011). For example, de Haan, Pascalis, and Johnson (2002) compared the N170 responses to human faces versus monkey faces and found a smaller N170 for human faces. Itier et al. (2011) also found that ape, cat, and dog faces elicited larger N170s than human faces. Putting faces in natural scenes, Rousselet et al. (2004) compared human faces from different races with faces from various species (e.g., mammals, birds, fish, and reptiles) and still observed a tendency of smaller N170 to human faces. However, Gajewski and Stoerig (2011) reported a larger N170 to human than to monkey faces and a larger N170 to monkey faces than to dog faces. Similar results were also found in the earlier study (Bentin et al. 1996). The inconsistency across experiments might be due to the diversity of the face stimuli (Gajewski & Stoerig, 2011). Gajewski and Stoerig (2011) used human faces from African, Asian, and Caucasian, dog faces from a wide range of breeds (e.g., Fox Terrier, Bull Terrier, German Shepard, Husky, Cocker Spaniel, and Maltese), and monkey faces ranging from Chimpanzee to Loris. Bentin et al. (1996) had excluded non-human primate faces in the comparison "because of their similarity to human faces". A larger N170 for human faces tended to come from studies using a wide range of animal faces, though diversity of the faces did not necessarily lead to a larger N170 for the human faces (e.g., Rousselet et al., 2004; Itier et al., 2011). The N170 differences between human faces and other animal faces were more reliable among studies examining the face inversion effect (as discussed below)..

Distortion of faces also modulates the N170 response. For example, violating the up-down feature arrangement of a face (i.e., face inversion) increases the N170 amplitude and latency (Rossion et al., 2000; Itier, Latinus, & Taylor, 2006; Itier et al., 2011; Macchi Cassia, Kuefner, Westerlund, & Nelson, 2006). As face inversion preserves the physical features of a face but disrupts its configuration, the inversion effect shows that the N170 is sensitive to face configuration. Interestingly, de Haan et al. (2002) observed the typical N170 inversion effect for human faces, but inversion of monkey faces did not modulate the N170 amplitude. Itier, Latinus, and Talor (2006) compared inversion effect of human faces, cars, houses, ape faces, chairs and human eyes. While inversion delayed the N170 for all stimulus categories, it only increased the N170 amplitude for human faces. Itier et al. (2011) also found that inversion enhanced the N170 for human faces, but did not affect the N170 for ape faces. The N170 amplitudes for upright ape faces were similar to those for the human upright faces. The study further revealed that inversion decreased the N170 amplitude for dog faces and cat faces. These results suggest that not only the N170 is sensitive to face configuration, but it is more sensitive to the configuration of human faces.

Although studies indicate that infants are better in recognizing upright than inverted human faces (see Nelson, 2001 for review), inverted faces do not elicit a different infant-N170 than upright faces even though human faces do elicit larger infant-N170 than the non-human faces (Haan, Pascalis, & Johnson, 2002). Taylor, Batty, and Itier (2004) reported that the onset of the inversion effect on the N170 can be as late as mid childhood. Hence, the sensitivity of the N170 to human face configuration might require experience.

This relatively late development of the sensitivity to human face configuration is in line with research on how experience modulates the N170. Repeated presentation of human faces from the same view can decrease the N170 amplitudes, but associating the faces with the person's information can reduce the repetition effect (Heisz & Sheddon, 2008). Presenting faces of the same person, but from different views can also eliminate the repetition effect (Ewbank, Smith, Hancock, & Andrews, 2008). Moreover, for people with perceptual expertise in certain object categories, the N170 is larger for objects of their expertise (Tanaka & Curran, 2001).

The N170 amplitude increase with expertise in an object category is related to conceptual categorization levels because subordinate-level categorization elicits a larger N170 than basic-level categorization (Tanaka, Luu, Weisbrod, & Kiefer, 1999). Experts automatically categorize objects of their expertise at the subordinate level, while novices by default categorize the same objects at the basic level (Tarr & Gauthier, 2000). Hence, the larger N170 amplitudes for human faces than for other objects have been interpreted as a consequence of expertise in human face recognition because human faces are categorized by default at the subordinate level. Rossion, Gauthier, Goffaux, Tarr, and Crommelinck (2002) examined the N170 response elicited by novel objects (Greebles) in participants trained in the lab to expertly recognize the Greebles and untrained control participants. Human faces and Greebles elicited the N170 and the human face inversion effect (delayed and enhanced N170) was observed for both groups. However, an inversion effect for Greebles was found only in the Greeble experts. Rossion, Kung, and Tarr (2004) presented human faces together with Greebles or non-trained objects. Their results revealed that, for the lab-trained Greeble experts, there was a competition effect on the N170 when human faces were presented with Greebles, but not when human faces were presented with non-trained objects. This result indicates that both human faces and objects of expertise recruit similar processes that are not shared with non-trained objects.

Unlike the P1 component, the N170 seems to be less sensitive to attention. For example, Cauquil, Edmonds and Taylor (2000) compared the N170 responses to four stimulus categories (full human faces, human eyes only, human faces with closed eyes, and flowers) in two tasks where participants had to detect either eyes only or faces with eyes closed. The N170 responses were not affected by whether or not the stimuli were targets. Carmel and Bentin (2002) presented four categories of stimuli (human faces, cars, furniture, and birds) to participants in two tasks (i.e., Animacy Decision and Car Monitoring). In the Animacy Decision Task, all stimuli were judged as animal or non-animal and the N170 to human faces was larger than every other stimulus category. In the Car Monitoring Task, cars were the targets and the N170 response to human faces was larger than those to birds and furniture, but there was no significant N170 difference between human faces and cars. However, cars did not elicit a larger N170 amplitude than furniture either, though cars did elicit larger a N170 than birds. The authors of both papers (Cauquil, Edmonds & Taylor, 2000; Carmel & Bentin, 2002) questioned the influence of attention on the N170. They argued that in the target detection task, targets should attract more attention than non-targets, so the absence of the N170 difference between targets and non-targets implies attention does not affect on the N170 to human faces. Moreover, Carmel and Betin (2002) argued that attention modulates the N170 response to other objects (e.g., cars), but not the response to human faces.

However, other researchers have reported effects of attention on the

N170 to faces (Eimer, 2000a; Sreenivasan, Goldstein, Lustig, Rivas, & Jha, 2009; Mohamed, Neumann, & Schweinberger, 2009). For example, Eimer (2000a) manipulated attention to human faces when they were presented either at the central or the peripheral visual field in a target detection task and found an effect of attention on the N170 only when faces were presented centrally. When human faces were superimposed on scenes (Sreenivasan et al., 2009) the attention effect was observed only when human faces were not highly discriminable. Finally, Mohamed, Neumann, and Schweinberger (2009) argued that the N170 can be modulated by attention load. They replicated the N170 difference between human faces and houses in a low attention load task, but found that the N170 amplitude was reduced for human faces and enhanced for houses in a high attention load task.

Interestingly, the N170 can be larger when people expect to see human faces rather than words, even though neither faces nor words are actually presented. Wild and Busey (2004) embedded human faces or words in visual noise and compared the responses in high contrast, low contrast and noise only conditions. Human faces elicited a larger N170 in both high and low contrast conditions. In the noise only condition, the N170 amplitude was larger if participants expected to see human faces rather than words. For those participants who reported seeing faces or words in the noise only condition, the N170 amplitudes were larger if they reported seeing faces than seeing words. These results support the notion that top-down processes modulate the N170.

To summarize, the N170 component is an occipito-temporal distributed component that peaks around 170ms after stimulus onset. It is modulated by categorization level because human faces and other objects of expertise, which are categorized at the subordinate level by default, elicit larger N170s than non-expertise objects (e.g., cars, flowers, furniture), which are categorized at the basic level by default. Moreover, it is sensitive to configural processing because stimulus inversion enhances N170 amplitude and delays its latency. Finally, it is less sensitive to attention, but not completely free from attentional influence.

1.2.3 P2

The P2 component follows the P1 and the N170 components. The P2 literature is not as established as that for the P1 or the N170. In general, there are anterior and posterior P2 components (Luck & Hillyard, 1994). The anterior P2 has been related to working memory (e.g., Lefebvre, Marchand, Eskes, & Connolly, 2005), while the posterior P2 has been linked to semantic processing (e.g., Gruber & Muller, 2005).

The posterior P2 is sensitive to the semantic aspect of the stimuli. For example, using an N-back task, Rose, Chmid, Winzen, Sommer, and Büchel (2005) embedded task-relevant letters in background pictures with different degrees of degration. The posterior P2 amplitudes increased as the background pictures became more and more visible. Moreover, there was no significant difference in the posterior P2 between the 1-back and the 2-back tasks, which suggests that the posterior P2 is not very sensitive to working memory load.

Sensitivity of the posterior P2 to semantic information can also be observed in the semantic priming paradigm. For example, Rossell, Price and Nobre (2003) presented words or nonwords as primes and targets in a lexical decision task. The primes and the targets were either related or unrelated. For the related pairs, the words were exemplars of a given category. The stimulus onset asynchrony (SOA) was either 200ms or 1000ms. The posterior P2 was larger for related prime-target pairs than the unrelated pairs, but only when the SOA was 200ms.

For longer SOAs, semantic processing can be probed using a repetition priming paradigm. Gruber and Muller (2005) presented a sequence of line drawings that were either meaningful or meaningless. The SOAs ranged from 2.7s to 3 s. Each drawing was displayed for 700ms and repeated 3 times non-consecutively. The participant’s task was to judge the meaningfulness of the drawings. Results showed that at the first presentation, meaningless drawings elicited a larger P2 than meaningful ones. A reduction in P2 amplitude was observed when the repetitions were compared to the first presentations. Interestingly, neither the P1 nor the N1 differed between the meaningful and the meaningless conditions. The P2 difference was not due to the low-level physical differences of the stimuli, but more likely reflecting semantic processing.

The P2 responses to different attributes of a conceptual category may vary. For example, Hoenig, Sim, Bochev, Herrnberger, and Kiefer (2008) required participants to verify visual or action-related attributes of either natural or man-made object concepts. They found that the posterior P2 was larger for visual attribute verification of man-made object concepts, followed by action-related attribute verification of man-made object concepts. Visual attribute verification of natural object concepts elicited the smallest posterior P2. The authors argued that the dominant attributes are action-related attributes for man-made objects and visual attributes for natural objects (also see Kiefer, 2005). The results indicated that the P2 amplitudes were smaller for the dominant attributes of a concept category.

Therefore, the posterior P2 is sensitive to semantic information of visual stimuli, regardless of task relevance and working memory (Gruber & Muller, 2005; Rose et al., 2005). Processing the dominant attributes of concept elicits smaller P2 amplitudes than processing the nondominant ones (Hoenig et al., 2008).

The posterior P2 component can be affected by attention. For example, Luck and Hillyard (1994) reported two EEG experiments using visual search tasks. In Experiment 1, they presented four types of visual stimuli that each contained eight small colour bars. For one type of stimuli, all eight bars were small, blue and vertical. In the other three types, one of the eight bars was different from the rest in colour (i.e., green), orientation (i.e., horizontal), or size (i.e. large). In different stimulus blocks, deviant bars of one dimension became the targets while those of the other dimensions remained non-targets. Participants were instructed to detect the target bars. Analysis of the posterior P2 component showed that visual stimuli with target bars elicited larger P2 amplitudes than those without a target. Hence, the posterior P2 can be affected by attention. Moreover, the visual stimuli with a nontarget deviant and those without deviant bars did not show significant P2 difference. This suggests that the P2 effect was due to task-directed, voluntary attention, instead of an automatic attention shift to the deviants.

In Experiment 2, Luck and Hillyard (1994) further compared responses to targets that were either deviant bars in a single dimension or in any dimension. There were again four types of visual stimuli. But in different conditions, participants responded to the stimuli 1) when one of the eight bars was blue (colour deviant condition); 2) when one of the eight bars was horizontal (orientation deviant condition); 3) when one of the eight bars was larger; or 4) when one of the eight bars was different from the rest regardless of the dimension of the deviation. The proportion of the targets was the same across the four conditions. Results showed again that the P2 amplitudes were larger for the targets than for the non-targets. Furthermore, the P2 difference between targets and nontargets was larger when targets were deviant bars in any dimension (condition 4) than in a single dimension (any of the other three conditions). This suggests that when task relevant, an automatic attention shift can affect the P2 component.

An attention effect on the posterior P2 can be observed not only when attention is allocated to the stimulus, but also when attention is absent from the stimulus despite task relevance. Vogel and Luck (2002) presented sequences of 20 stimuli where each stimulus was displayed for 33ms with a 50ms inter-stimulus interval. Two targets, a digit and a letter “E”, were embedded in a sequence consisting of the other letters. The second target was either at the 3rd (inside attentional blink period) or the 7th (outside attentional blink) position after the first target. Moreover, the second target was either in the middle (masked by the following nontargets) or at the end (not masked) of the sequence. It was found that the posterior P2 responses to the second target were suppressed when they fell within the attentional blink. This occurred regardless of whether the second target was masked or not. Therefore, the posterior P2 not only increased when attention was directed to the stimulus (as in Luck & Hillyard, 1994), it also decreased when attention was absent from the stimulus (Vogel & Luck, 2002). The latter effect occurred even when the stimulus was task relevant.

Therefore, the posterior P2 component is related to semantic processing. Processing different attributes of a concept and attention can modulate the posterior P2 amplitudes, with task relevance affecting the attention effect.

1.3 Audio - Visual Processing

The multi-modal distribution of conceptual categorization in the brain was discussed briefly in Section 1.1, but the interaction between processing in different modalities requires further clarification. When the auditory stimulus is simple and the stimulus onset asynchrony (SOA) between the auditory and the visual stimuli is brief, attention can spread between the auditory and visual modalities. For example, Busse and colleagues (2005) instructed participants to attend to only one side of a computer screen and to respond whenever a checkerboard with dots appeared on the attended side, but not to respond to checkerboard stimuli presented on the unattended side. Half of the checkerboards were presented simultaneously with a task-irrelevant tone. Overall, attention affected the P1 regardless of tone presentation. To examine the audiovisual interaction, the authors obtained two difference waves. They subtracted the visual-only condition ERP waveform from the audiovisual condition ERP waveform when the checkerboard appeared in the attended hemi-field and they subtracted the visual-only condition ERP waveform from the audiovisual condition ERP waveform when the checkerboard appeared in the unattended hemi-field. Comparing the difference waves from the attended checkerboard conditions to the difference waves from the unattended checkerboard conditions, they found a negativity starting at 220ms after tone onset in the attended condition difference waveform, which they interpreted as indicating a spread of attention between the auditory and the visual modalities.

Donohue, Roberts, Grent-‘t-Jong, and Woldorff (2011) used a similar paradigm. The centrally located tones in the audio-visual conditions were played with a SOA of 0ms (i.e., simultaneously), 100ms, or 300ms. Difference waveforms were created as described above for the Busse et al. (2005) study and compared for the three SOA conditions. There was negative-polarity activity at frontal central electrode sites starting 200ms after onset of the auditory stimulus in both the simultaneous presentation and the 100ms SOA condition. The findings indicate that spread of attention can occur from the visual modality to the auditory modality both when the auditory and the visual stimuli are presented simultaneously and when there is a brief SOA. Although attention spread was not observed when the SOA was 300ms, this does not necessarily mean that there is no cross-modality effect at longer SOAs.

Facilitation effects from the visual modality to the auditory modality have been reported in priming studies (e.g., Noppeney, Josephs, Hocking, Price, & Friston, 2008; McKone & Dennis, 2000; Schneider, Engel, & Debener, 2008). Even with an SOA of 1400ms, auditory object recognition is faster and more accurate when the category of the preceding visual stimulus is congruent rather than incongruent with the category of the auditory stimulus (Schneider et. al., 2008).

When the SOA is longer, the brain may rely on an amodal mechanism to manage information from different modalities. In the Donohue et al. (2011) study, the scalp distributions of the negative-polarity difference waves 200 ms after auditory stimulus onset were not identical for the simultaneous and the 100ms-delay conditions. The authors suggested that the sources contributing to the effect may differ between the simultaneous condition and the 100ms-delay condition. Separating responses to the left and the right side of the screen, the authors found that the early negativity in the simultaneous condition was biased towards the hemisphere contralateral to the visual stimulus presentation hemi-field. Given that sounds were presented centrally, this suggests that within-modality processes were involved in the simultaneous condition, but when there was a delay between the auditory and the visual stimuli, attention spread relied on amodal processes.

The audiovisual studies described above focused on the effect of a visual stimulus on an auditory stimulus. The influence of an auditory stimulus on a visual stimulus is less consistent. For example, Greene, Easton, and LaShell (2001) failed to find a facilitation effect from the auditory modality to the visual modality, whereas Schneider and colleagues (2008) reported a priming effect of auditory stimuli on visual object recognition. The inconsistency may be due partially to the picture quality difference between the two studies. Schneider et al. (2008), presented degraded pictures. Hence, category congruency from the auditory modality could help with recognition of the visual object. Chen and Spence (2011), who presented an auditory stimulus followed 346ms later by a 13ms visual stimulus that was then masked, found a similar effect. Visual sensitivity was significantly increased when the category of the auditory stimulus was congruent with the category of the visual object. Together with the Schneider et al. (2008) study, these results suggest that an auditory effect on visual object recognition is more likely to be observed when visual recognition is difficult.

Moreover, the delay of visual stimulus presentation in both studies (i.e., Schneider et al., 2008 and Chen & Spence, 2011) was necessary for the auditory effect to be observed. When the SOA between the auditory stimulus and the visual stimulus was shortened the facilitation effect vanished (Chen & Spence, 2011). This is due to the intrinsic difference in visual versus auditory object processing. Visual recognition of environmental objects occurs within 150ms of stimulus onset (Thorpe, Fize, & Marlot, 1996), but environmental sound recognition takes longer because the information in the sounds is revealed over time. A short or 0ms SOA does not provide enough time for extraction of auditory information that can benefit visual object recognition.

1.4 Categorization Level Manipulations

In the studies by Chen and Spence (2011) and by Schneider and colleagues (2008), the auditory and visual stimulus categories could be congruent or incongruent while the categorization levels were held constant at the basic level. The visual object recognition difficulty obtained by degrading the visual stimuli was also at the basic level of categorization. Therefore, hearing a dog bark at the basic level should activate the representation of the basic-level concept dog. This representation includes features not only within the auditory modality, but also in the visual modality. Given spreading activation of representations is associative, the visual representation of the basic-level concept dog is primed. Therefore, when the visual input of a dog is later received, recognition at the basic level becomes easier. It leaves open, however, the effect of the categorization level in cross-modality priming. Does the congruency between the categorization level of the auditory stimulus (and therefore, the activated concept) and the visual stimulus affect visual stimulus processing?

Repetition priming studies where categories were held constant but categorization levels were manipulated showed priming effect both within and across categorization levels (Martinovic, Gruber, & Müller, 2009; Francis, Corral, Jones, & Saenz, 2008). Behavioural responses were facilitated for both priming within the superordinate level and priming within the basic level. For priming across categorization levels, Francis and colleagues (2008) observed a facilitation effect of priming across categorization levels in both directions. Martinovic and colleagues (2009) found a facilitation effect of superordinate-to-basic level priming. The priming effect was stronger for within than between categorization level priming. As additional perceptual information is needed for a more specific level of categorization (Tanaka & Taylor, 1991), more feature processing regions are primed after a concept at a specific categorization level is activated. Processing of subsequent visual stimulus from the same category and at the same categorization level is easier, compared to a more general categorization level. Martinovic and colleagues (2009) also reported a tendency that the P1 amplitudes were enhanced in the basic-level categorization compared to superordinate categorization. But whether categorization levels affect visual processing in cross-modality priming remains unclear.

Based on the evidence, we hypothesized that both category and categorization level congruency should play a role in sensory processing of subsequent visual stimuli. As categories can be congruent or incongruent depending on categorization levels, basic level congruence should be particularly important because most objects are by default categorized at this level. Basic level categorization also has a balance between inclusiveness and specificity of feature processing. It provides a good starting point for investigating how specificity of categorization may affect visual processing.

1.4.1 Present Experiments

The current study examined whether the level of conceptual categorization of an auditory stimulus affects the processing of a subsequently presented visual stimulus. More specifically, we are interested in effect of categorization levels on the P1 and the N170 components. The P1 component has been an indicator of early visual processing. The N170 component is related to configural processing, sensitive to face-like stimuli and changes as default categorization level shifts. We expected that conceptual representation activated by auditory stimuli would affect visual sensory processing given concepts are represented as multi-modal distributed pattern. A subsequent visual stimulus that is congruent with the activated representation may consequently be processed more efficiently and attract more sensory processing resources. Therefore, the P1 component should be affected by the categorization level of the auditory stimulus. We were also interested to see whether the N170 amplitudes increase when the preceding auditory stimulus could be categorized at a more specific level. The N170 component should have a larger amplitude when the visual stimulus is to be categorized at a more specific level of categorization. We used faces (dog faces or human faces) or face-like stimuli (the front view of a car) to ensure an N170 component was elicited. Given that conceptual categorization of an auditory stimulus requires time, we adopted a priming paradigm in which the auditory stimulus preceded the visual stimulus. In addition, across experiments the category congruency between the auditory and the visual stimuli was varied to determine whether categorization level and stimulus category interact to affect visual processing.

In the first experiment (Dog-Dog experiment), we used dog barks as the auditory stimuli and novel dog faces as the visual stimuli. The auditory and the visual stimuli were congruent at the basic level (i.e., dog). By training our participants, we created two categorization level conditions for the auditory stimuli. Half of the auditory stimuli were untrained and were therefore expected to be categorized at the basic level whereas the other half were trained and therefore were individually identifiable and expected to be categorized at the subordinate level (i.e., associated with a dog whose name was known). Participants were not trained to recognize the dog faces. Thus, all the dog faces were assumed to be categorized at the basic level.

In the second experiment (Dog-Car experiment), the training procedure for the auditory stimuli was the same as in the Dog-Dog experiment. However, auditory dog bark stimuli were followed by front views of cars as the visual stimuli. The car pictures were all novel and therefore were assumed to be categorized at the basic level, like the dog faces. Hence, there was category incongruency between the barks and the cars.

In the third experiment (Dog-Human experiment) the training procedure for the auditory stimuli was the same as in the Dog-Dog experiment. The auditory dog bark stimuli were followed by human face visual stimuli. Hence, the categories were different at the basic level, but the same at the superordinate level (i.e., animals). More importantly, unlike dog faces and cars, human faces are believed to be individually categorized (i.e., at the subordinate level). Hence, the visual stimuli were congruent with the trained auditory stimuli in categorization level as both could be individually categorized.

In the fourth experiment (Human-Dog experiment), the auditory stimuli were human voices saying “dada”. As in Experiments 1-3, participants learned to associate a specific person’s name with the voice for half of the auditory stimuli. The auditory stimuli were followed by novel dog faces. Hence, the categories of the auditory and the visual stimuli were different at the basic level, but the same at the superordinate level (i.e., animals), as in the Dog-Human experiment. Given that all the dog faces were untrained and categorized at the basic level, the dog faces were more congruent with the untrained voices than with the trained voices in terms of categorization level. Together, the third and the fourth experiments (i.e., the Dog-Human experiment and the Human-Dog experiment) examine whether categorization level congruency between the auditory and the visual stimulus affects visual stimulus processing.

In the fifth experiment (Human-Human experiment), the auditory stimuli and the training were the same as in the Human-Dog experiment, but the visual stimuli were the human faces used in the Dog-Human experiment. Both voices and human faces belong to the same basic-level category, human. This is similar to the Dog-Dog experiment. However, while the dog faces and the untrained dog barks were both categorized at the basic level (i.e., category-level congruent), the human faces were more congruent in categorization level with the trained voices than the untrained voices because both the human faces and the trained voices should be individually categorized.

The sixth and final experiment (Dog-Mix experiment) was slightly different from the previous five. In the first five experiments, the participants knew what basic level stimulus category would follow the auditory stimulus. For example, in Experiment 3 after hearing a dog bark the participants expected to see a human face. Although they had never seen the individual faces previously, they knew that it would be a human face. It is possible that in addition to activation of the representation of the auditory stimulus, the representation of the visual category was also activated before seeing the visual stimulus. The last experiment reduced this expectation by adding uncertainty to the category of the visual stimuli. Dog barks were the auditory stimuli and the same training procedure was applied. However, the visual stimuli following the dog barks could be dog faces, cars, or human faces, so the category congruency between the visual and the auditory stimuli was uncertain. The role of categorization levels in visual processing could then be explored without the basic-level category expectation effect.

Chapter 2 General Method

2.1 Data Recording and Processing

The data acquisition was done using a Biosemi Active Two System. The EEG was recorded from 64 Ag/AgCl electrodes mounted to an elastic cap based on the extended 10-20 system. Two electrodes (EX1 and EX2) were placed above and below the left eye to measure eyeblinks. In the Dog-Dog experiment, another two electrodes (EX3 and EX4) were attached to the outer canthus of both eyes to measure horizontal eyemovements. In the other five experiments, one electrode (EX4) was attached to the outer canthus of the right eye. The horizontal eyemovements can be observed by subtracting data of the EX4 from the average of the EX1 and the EX2. In addition, one electrode (EX5) was placed on the nose tip. The EEG and EOG were recorded continuously from 0Hz to 52Hz at a sampling rate of 256 Hz. Data was referenced to the common mode sense active electrode during recording (Metting van Rijn, Peper, & Grimbergen, 1990).

The raw EEG data were imported using the nose as the reference. A low-pass filter of 30Hz was applied before a high-pass filter of 0.1Hz. Artefacts introduced by muscle movements were removed through inspection. ICA was used to remove eye movements. The continuous data were epoched with a 100ms pre-stimulus baseline. Data were then re-referenced to average values from 64 channels on the scalp.

2.2 ERP Components

The processed data were averaged across epochs within conditions, channels and participants. There are three ERPs elicited by the visual stimuli that are of interest: the P1, the N170, and the P2. The half area latency of each component was calculated for each participant by condition (Luck, 2005). We used a time window from 55-150ms for the visual P1, from 130-190ms for the N170, and 155-250ms for the P2. By definition, the half area latency for a given component is the time point where the area under the EEG waveform within the time window was divided in half.

The reported half-area latencies were determined across a subset of channels. The visual P1 component has a bi-lateral posterior distribution and studies of the visual P1 typically report effects in channels such as O1/O2, PO7/PO8, P7/P8, and PO3/PO4 (e.g., Sausung, et. al., 2005; Freunberger, et. al., 2008a, 2008b; Mangun, et. al., 2001). Here, six regions, comprising six channels each, were used for the statistical analyses of the P1 (left anterior: F3, F5, F7, FC3, FC5, and FT7; left middle: C3, C5, T7, CP3, CP5, and TP7; left posterior: P3, P5, P7, PO3, PO7, and O1; right anterior: F4, F6, F8, FC4, FC6, and FT8; right middle: C4, C6, T8, CP4, CP6, and TP8; and right posterior: P4, P6, P8, PO4, PO8, and O2). Channels in the left posterior and right posterior regions were used for calculating the half area latencies. For the N170, the half-area latencies were averaged across six occipito-temporal channels (P7, P8, PO7, PO8, P9, and P10). These six channels are commonly used in N170 studies (see Rossion & Jacques, 2008 for review). The same six channels were used for the visual P2 half –area latency analysis.

The mean amplitudes of the ERP components were defined as the averaged values within a 20ms time window surrounding the ERP peaks of the grand average waves. They were calculated using the following steps. First, data were averaged across participants, but within channels and conditions, to obtain the grand average waveforms. Then, the ERP peaks for these average waves were calculated and averaged across channels and conditions. Third, a 20ms time-window was centered at the ERP peak mean. Fourth, the average amplitudes within this time window were determined for each channel, condition, and participant.

2.3 Statistical Analyses

For the analysis of the visual P1, there were three factors: Old/New condition (Old: the preceding barks/voices had been learned in training; New: the preceding barks/voices were new), Hemisphere (the left/right hemisphere) and AMP region (the anterior, the middle, and posterior regions). For the N170 and the visual P2, three factors were used, including Old/New condition (Old: the preceding barks/voices had been learned in training; New: the preceding barks/voices were new), Hemisphere (the left/right occipito-temporal area) and Channel Group (P9/P10; P7/P8; PO7/PO8).

2.4 Stimuli

The auditory stimuli (16 bit mono, 44.1 kHz sampling rate, 10 ms fade-in and fade-out) were sixty audio recordings of barks from 60 dogs or voices from 30 males and 30 females. The barks were taken from the online sound databases, such as Freesound (www.freesound.org) and Findsounds (www.findsounds.com). The voices were recorded by the experimenters in our recording booth. The speakers were thirty male and thirty female Singaporean Chinese undergraduates from the National University of Singapore. Each speaker was paid 2 SGD for the recording. None of the speakers was a participant in any of the six EEG experiments.

The visual stimuli (690x397 pixels, 72 dpi, grayscale) were sixty face pictures of 60 dogs, sixty front-views of 60 cars, and/or sixty faces pictures of 60 Chinese (30 males and 30 females). The dog faces and the front-views of cars were taken from the internet (e.g., dog faces from http://www.dailypuppy.com/). The human faces were taken from the face database created by W. G. Hayward in the Visual Cognition Lab at the University of Hong Kong.

Chapter 3 Dog-Dog Experiment

3.1 Methods

Each participant came to the lab on two consecutive days. On Day 1, participants learned to identify each of 30 dog barks with a unique dog name (See Figure 1). The 30 dog barks were divided into four groups, with seven or eight barks per group. Participants listened to each bark while the dog name was presented on the screen. At the end of each group of barks, there was a multiple-choice test to ensure that participants could remember the name associated with each bark. Participants were encouraged to repeat the training

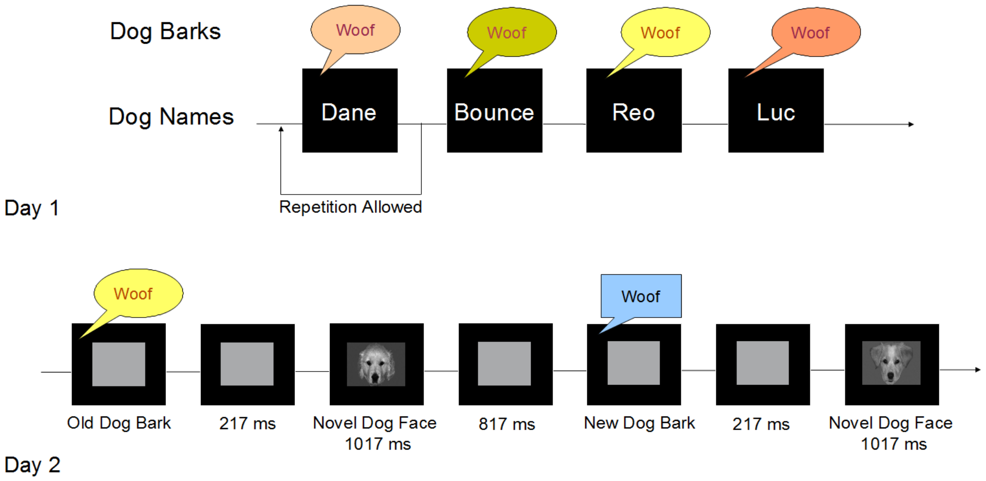


Figure 3.1: Procedure. On Day 1, participants learned to associate each of 30 different dog barks with a unique dog name. On Day 2, all participants were tested for recognition of the individual dog barks. If bark recognition accuracy was 80% or greater, they proceeded to the EEG experiment. In the EEG experiment, participants saw a sequence of 60 photographs of upright dog faces. Each dog face was preceded by either a dog bark for which the participant had learned a name (old bark, 50%) or a new bark (50%). In addition, 6 inverted dog faces, also preceded by a bark, were presented. The participant's task was to detect and respond to the inverted dog faces. Upright faces did not require a response.

session if more than one bark in the group was misnamed. After learning all barks, participants were given an identification test in which they listened to the barks presented in random order and they were required to type in the name corresponding to each bark. On Day 2, participants learned the barks and the names again. At the end of the revision, they were given the identification test again. Participants proceeded to the EEG experiment only if they correctly identified at least 80% of the barks.

In the EEG experiment, participants performed a target detection task in which they heard a sequence of dog barks presented centrally and in random order, each of which was followed by a picture of a dog face, presented for 1017ms. Half of the barks had been learned in the training phase and the other half were new. The interval between the offset of the bark and the onset of the dog face was 217ms. The interval between the offset of the dog face and the onset of the next bark was 817ms. Each dog face was presented once only. The participant's task was to press a key whenever an inverted dog face was presented. Upright dog faces did not require a response. There were 60 upright dog faces and 6 inverted ones in total. Bark assignment to old and new conditions and assignment of dog faces to follow an old or new bark was counterbalanced across participants.

Twenty-eight participants (14 females, average age of 22) from the National University of Singapore were included in the experiment. Informed written consent was obtained from each participant before the experiment. They were reimbursed at a rate of 9 SGD an hour for their time.

3.2 Results

3.2.1 Behaviour

The average hit rate was 94.05%, the average reaction time was 503ms, and the average false alarm rate was 0.48%. As the number of target trials was too small, no further analyses were conducted on the behavioral data.

3.2.2 P1

For the P1 half area latency, the main effect of Old/New condition (*F*(1, 27) = 1.27, *p* = .27) and the Old/New x AMP (*F*(2, 54) = 1.36, *p* = .27), Old/New x Hemisphere (*F*(1, 27) = 0.41, *p* = .53), and Old/New x AMP x

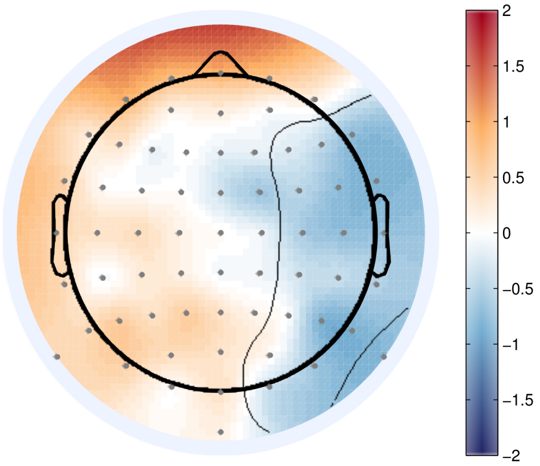


Figure 3.2: Topographic distribution of the ERP response difference between the Old and the New conditions at the P1 latency in the Dog-Dog experiment. The ERP amplitudes were averaged within the 20ms time window surrounding the P1 peak. The averaged amplitudes in the New condition were subtracted by the averaged amplitudes in the Old condition. The amplitudes were significantly different over the right hemisphere. The grey dots indicate the location of the 64 electrodes. The contour shows the significant area.

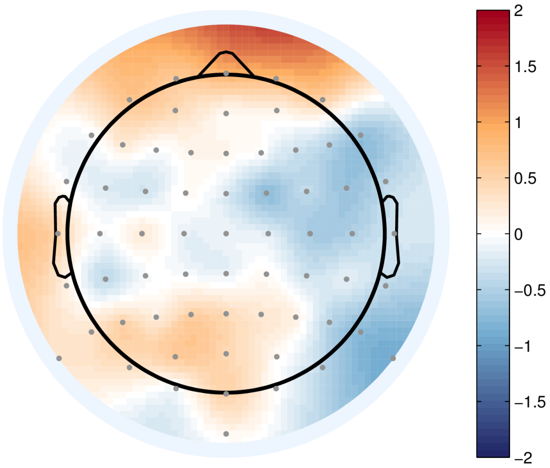


Figure 3.3: Topographic distribution of the ERP response difference between the Old and the New conditions at the N170 latency in the Dog-Dog experiment. The ERP amplitudes were averaged within the 20ms time window surrounding the N170 peak. The averaged amplitudes in the New condition were subtracted by the averaged amplitudes in the Old condition. The amplitudes were not significantly different between the Old and the New conditions. The grey dots indicate the location of the 64 electrodes.

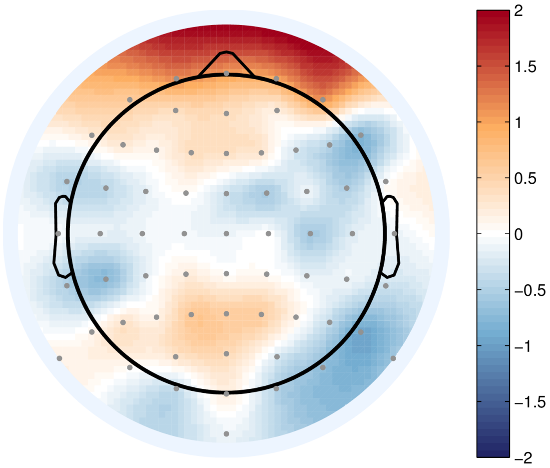


Figure 3.4: Topographic distribution of the ERP response difference between the Old and the New conditions at the P2 latency in the Dog-Dog experiment. The ERP amplitudes were averaged within the 20ms time window surrounding the P2 peak. The averaged amplitudes in the New condition were subtracted by the averaged amplitudes in the Old condition. The amplitudes were not significantly different between the Old and the New conditions. The grey dots indicate the location of the 64 electrodes.

Hemisphere (*F*(2, 54) = 0.13, *p* = .88) interactions were not significant.

The peak latency of the P1 derived from the grand average waveforms in the left and right posterior regions (see General Method) was 113ms. As noted in the Methods section, the grand average peak latency was used as the center of the 20ms time window used for analyzing the mean P1 amplitudes obtained from individual participants. The main effect of Old/New condition (*F*(1, 27) = 1.77, *p* = .19) was not significant, but the Old/New x Hemisphere interaction was significant (*F*(1, 27) = 6.00, *p* = .02). A follow-up analysis revealed that for right hemisphere electrodes the P1 amplitude was significantly larger in the Old condition than in the New condition (*F*(1, 27) = 7.21, *p* = .01), whereas for left hemisphere electrodes there was no such effect (*F*(1, 27) = 2.75, *p* = .11). There was no significant Old/New x AMP (*F*(2, 54) = 0.15, *p* = .86) or Old/New x AMP x Hemisphere (*F*(2, 54) = 0.07, *p* = .93) interaction.

3.2.3 N170

For the N170 half area latency, the main effect of Old/New (*F*(1, 27) = 0.79, *p* = .38) and the Old/New x Hemisphere (*F*(1, 27) = 1.31, *p* = .26), Old/New x Channel Group (*F*(2, 54) = 1.46, *p* = .24), and Old/New x Hemisphere x Channel Group (*F*(2, 54) = 0.35, *p* = .71) interactions were not significant.

The peak latency of the N170 derived from the grand average waveforms at the six occipito-temporal electrodes (see General Method) was 159ms. The analysis of the mean amplitudes within the 20ms time window

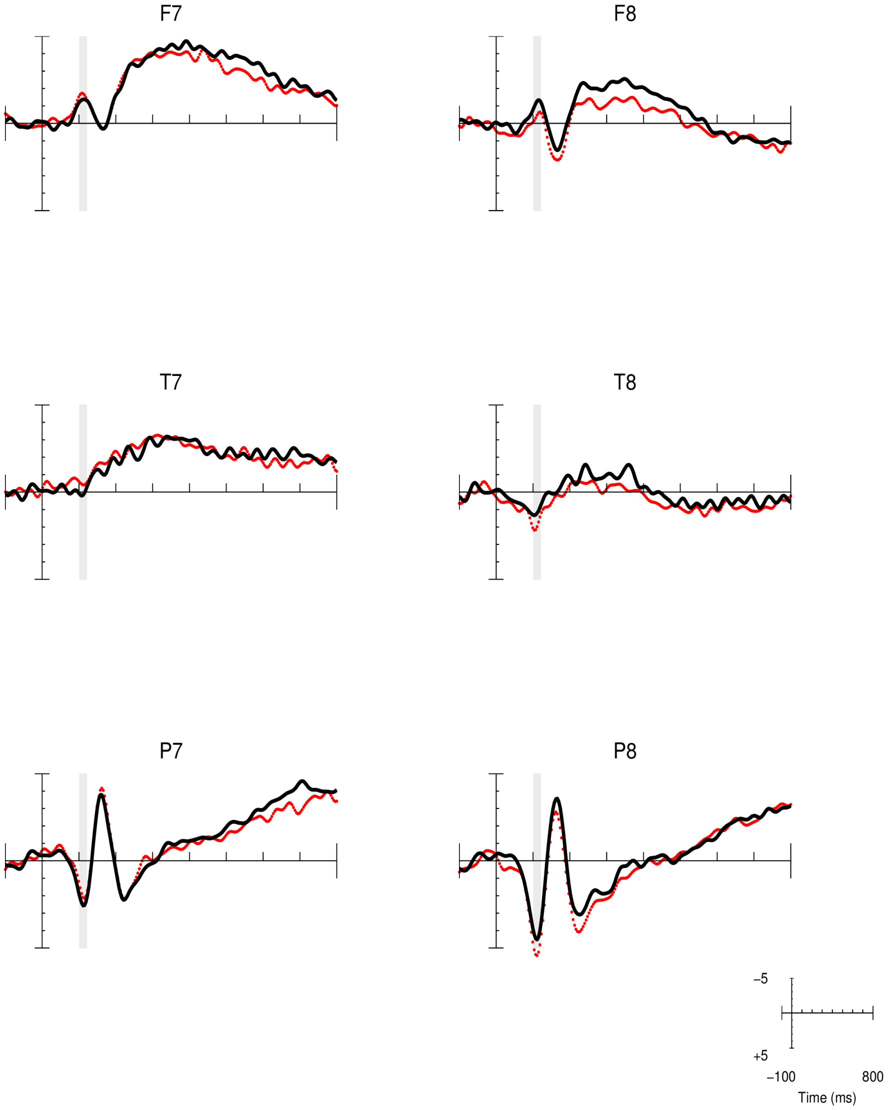


Figure 3.5: The ERP responses to dog faces in the Dog-Dog experiment. The red dotted lines show the ERP in the Old condition. The black solid lines show the ERP in the New condition. The light grey bars indicate the time windows for computing the average P1 amplitues.

surrounding this N170 peak revealed that the main effect of Old/New (*F*(1, 27) = 0.72, *p* = .40), the Old/New x Channel Group (*F*(2, 54) = 0.28, *p* = .76), and Old/New x Hemisphere x Channel Group (*F*(2, 54) = 0.26, *p* = .78) interactions were not significant. However, the Old/New x Hemisphere (*F*(1, 27) = 8.39, *p* = .01) interaction was significant. Over the right occipito-temporal area, the N170 in the Old condition was marginally smaller than that in the New condition (*F*(1, 27) = 3.31, *p* = .08), whereas there was no such effect over the left occipito-temporal area (*F*(1, 27) = 0.40, *p* = .53).

3.2.4 P2

For the P2 half area latency, the main effect of Old/New (*F*(1, 27) = 2.39, *p* = .13), the Old/New x Hemisphere (*F*(1, 27) = 0.12, *p* = .73), Old/New x Channel Group (*F*(2, 54) = 0.91, *p* = .41), and Old/New x Hemisphere x Channel Group (*F*(2, 54) = 0.95, *p* = .39) interactions were not significant.

The peak latency of the P2 derived from the grand average waveforms at the six occipito-temporal electrodes (see General Method) was 220ms. Analysis of the mean amplitudes within the 20ms time window surrounding this peak indicated that the main effect of Old/New (*F*(1, 27) = 0.54, *p* = .47), the Old/New x Hemisphere (*F*(1, 27) = 3.65, *p* = .07), Old/New x Channel Group (*F*(2, 54) = 0.11, *p* = .90), and Old/New x Hemisphere x Channel Group (*F*(2, 54) = 0.17, *p* = .84) interactions were not significant.

3.3 Discussion

The Dog-Dog experiment compared responses to the dog faces preceded by individually identifiable barks (i.e., the old condition) with those preceded by new barks (i.e., the new condition). The P1 amplitude was significantly larger following Old than New barks. It indicated that the preceding barks did affect the early processing stage of subsequent dog faces which share the same basic-level category with the barks. It suggested that categorization levels of concepts that were activated by preceding auditory stimuli can affect early visual processing stage of subsequent stimuli. However, the scalp distribution of the difference waveform varied from the expected P1 distribution. Instead of having an occipito-temporal distribution, the response difference spread over the temporal region and only occurred over the right hemisphere. On one hand, the P1 component is subject to attention modulation (e.g., Taylor, 2002; Klimesch, 2011). The P1 amplitude was larger for attended visual stimuli compared to the unattended ones. It is possible that the old barks brought more attention to the following dog faces, and hence increased the P1 amplitude. On the other hand, differences in scalp distribution might reflect the audiovisual interaction between barks and faces (Molholm et al., 2002). Molholm and colleagues (2002) compared responses to visual-only and auditory-only stimuli with responses to audiovisual stimuli. They found a multi-modal interaction in various time windows from 46ms up to 200ms after stimulus onset. For example, the earliest audiovisual interaction they found started 46ms after stimulus onset with a parieto-occipital distribution. They also found a frontal-distributed audiovisual interaction around 100ms. In both time windows, audiovisual stimuli elicited more positive responses than the sum of the visual-only and the auditory-only responses. Our results may reflect both the auditory stimulus and the visual stimulus from the same basic-level category forming a unitary representation by further processing the visual stimulus at a less-specific categorization level. The authors also linked this right lateral positivity to the attentional orienting systems, suggesting the functional significance of attention in audiovisual stimulus processing. Therefore, the response difference between the old and the new conditions might reflect modulation of the categorization levels of the preceding barks on early processing of the dog faces.

There was a marginally significant effect of the barks on the N170 amplitude. The N170 following the old barks tended to be smaller than the N170 following the new barks. However, the scalp distribution of the effect was similar to the effect in the P1 time window, which was different from the typical N170 scalp distribution. The difference occurred over the right hemisphere covering the frontal, the temporal and the occipital regions. The difference was also in the same direction as the effect in the P1 time window. Similar positivities were observed in a study by Molholm et al. (2002) who found that from 160ms to 180ms, responses to the audiovisual stimuli were less negative than the sum of responses to the audio-only and the visual-only stimuli, with a right occipito-temporal scalp distribution. The N170 is related to configural processing of visual stimuli. The N170 differences between the dog faces following the old barks and the new barks suggests the modulation of the categorization levels of the preceding barks on the configural processing of the dog faces.

There was no significant P2 difference between the faces following the old barks and the new barks. The posterior P2 component has been linked to semantic processing. Visual stimuli elicited larger P2 when preceded by related versus unrelated stimuli (Rossell et al., 2003; Federmeier & Kutas, 2002). By changing the categorization level of the barks, we potentially altered the categorical relatedness between the barks and the dog faces. For example, both the barks and the dog faces belong to the same basic-level category – Dog, but they might not be from the same dog (i.e., categorically unrelated). The new barks and the dog faces were categorically related, while the old barks and the dog faces might be categorically unrelated. Therefore, it appears that the P2 should be larger for the dog faces following the new barks than the old barks. However, the SOA between the two stimuli may have constrained the effect of categorical relatedness. Earlier studies revealed an effect of categorical relatedness for both 200ms (Rossell et al., 2003) and 500ms (Federmeier & Kutas, 2002) SOAs, it was not observed for 1000ms SOA (Rossell et al., 2003). The minimum SOA between the barks and the faces was 617ms, which might be too long for the categorical relatedness effect to occur.

The posterior P2 component can also be affected by attention, but only when it is task relevant. As observed in the P1 time window, attention increased on the dog faces following the old barks. The current task was to detect the inverted dog faces and neither the dog barks nor the relationship between the barks and the dog faces were task relevant, so the P2 component was not affected by the preceding barks.

The early effects in this experiment raise the question of the circumstances under which familiar sounds modulate visual processing. Both barks and dog faces belong to the same basic-level category, Dog. Even for a non-expert, a given stimulus pair might be recognized as from the same dog, which is a more specific category than the basic level one. If the effect were due to object binding, then pairing the barks with visual stimuli from a different basic-level category should eliminate it. In Experiment 2, we used front views of cars as the visual stimuli. Not only are cars from a different basic-level category, but they are also from a different superordinate-level category, Vehicle. Thus, the auditory and the visual stimuli should not be recognized as from the same object.

Chapter 4 Dog-Car Experiment

4.1 Methods

The Dog-Car experiment was the same as the Dog-Dog experiment, with the exception that in the visual target detection task the pictures were front views of cars. There were 60 upright car pictures and 6 inverted car pictures in total.

Twenty participants (12 females, mean age of 22.9) from the National University of Singapore were included in the experiment. Informed written consent was obtained from each participant before the experiment. They were reimbursed at a rate of 9 SGD an hour for their time.

4.2 Results

4.2.1 Behavioral Results

The average hit rate was 100%. The average reaction time was 528ms, and the average false alarm rate was 0.33%. As the number of the target trials was too small, no further analysis was conducted on the behavioral data.

4.2.2 P1

For the P1 half area latency, the main effect of Old/New (Old: 111ms; New: 108ms; $F(1, 19) = 4.87$, $p = .04$) was significant, but the Old/New x Hemisphere ($F(1, 19) = 0.04$, $p = .84$) and Old/New x AMP ($F(2, 38) = 0.05$, p

= .95) interactions were not. Although the Old/New x AMP x Hemisphere interaction (*F*(2,38) = 4.19, *p* = .02) was also significant, follow-up analyses did not reveal any significant effects that included the Old/New factor.

The peak latency of the P1 derived from the grand average waveforms in the two occipito-temporal regions was 119ms. The analysis of the mean amplitudes within the 20ms time window surrounding this peak revealed that the main effect of Old/New (F(1, 19) = 1.18, p = .29), the Old/New x AMP (F(2, 38) = 0.27, p = .76), and Old/New x AMP x Hemisphere (F(2, 38) = 0.04, p = .96) interactions were not significant. However, the Old/New x Hemisphere interaction was significant (F(1, 19) = 4.81, p = .04). For right hemisphere electrodes, the P1 amplitude in the New condition was marginally

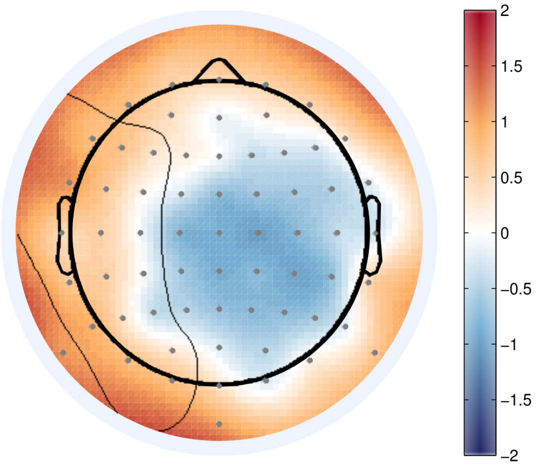


Figure 4.1: Topographic distribution of the ERP response difference between the Old and the New conditions at the P1 latency in the Dog-Car experiment. The ERP amplitudes were averaged within the 20ms time window surrounding the P1 peak. The averaged amplitudes in the New condition were subtracted by the averaged amplitudes in the Old condition. The amplitudes were significantly different over the left hemisphere. The grey dots indicate the location of the 64 electrodes. The contour indicates the significant area.

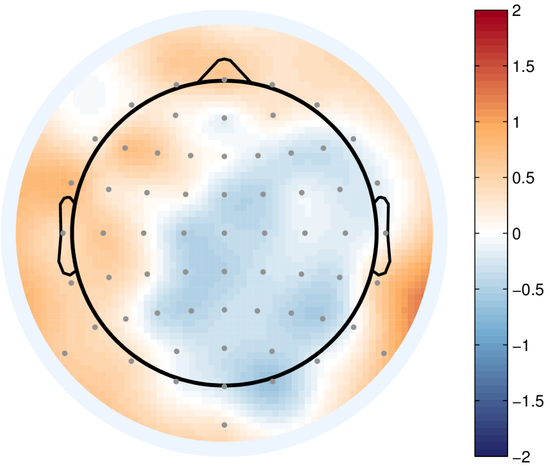


Figure 4.2: Topographic distribution of the ERP response difference between the Old and the New conditions at the N170 latency in the Dog-Car experiment. The ERP amplitudes were averaged within the 20ms time window surrounding the N170 peak. The averaged amplitudes in the New condition were subtracted by the averaged amplitudes in the Old condition. The amplitudes were not significantly different between the Old and the New conditions. The grey dots indicate the location of the 64 electrodes.

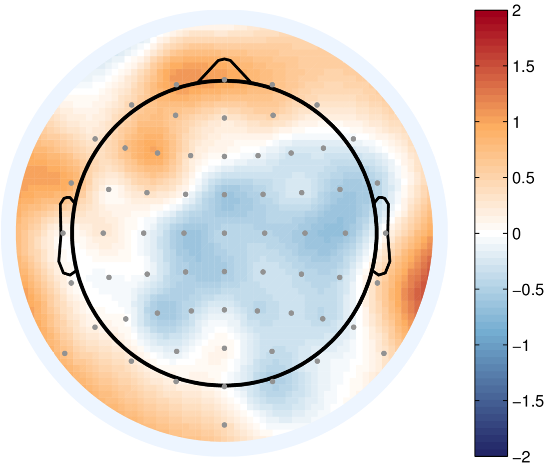


Figure 4.3: Topographic distribution of the ERP response difference between the Old and the New conditions at the P2 latency in the Dog-Car experiment. The ERP amplitudes were averaged within the 20ms time window surrounding the P2 peak. The averaged amplitudes in the New condition were subtracted by the averaged amplitudes in the Old condition. The amplitudes were not significantly different between the Old and the New conditions. The grey dots indicate the location of the 64 electrodes.

larger than that in the Old condition (F(1, 19) = 4.07, p = .06), whereas for left hemisphere electrodes the P1 amplitude in the New condition was significantly larger than that in the Old condition (F(1, 19) = 4.63, p = .04).

4.2.3 N170

For the N170 half area latency, the main effect of Old/New (*F*(1, 19) = 0.17, *p* = .69), the Old/New x Hemisphere (*F*(1, 19) = 2.08, *p* = .17), Old/New x Channel Group (*F*(2, 38) = 0.35, *p* = .71), and Old/New x Hemisphere x Channel Group *F*(2, 38) = 0.87, *p* = .43) interactions were not significant.

The peak latency of the N170 derived from the grand average waveforms was 168ms. The analysis of the mean amplitudes within the 20ms time window surrounding this N170 peak revealed that the main effect of Old/New (*F*(1, 19) = 0.44, *p* = .51), the Old/New x Hemisphere (*F*(1, 19) = 0.42, *p* = .52), Old/New x Channel Group (*F*(2, 38) = 1.17, *p* = .32), and Old/New x Hemisphere x Channel Group (*F*(2, 38) = 1.18, *p* = .32) interactions were not significant.

4.2.4 P2

For the P2 half area latency, the main effect of Old/New (*F*(1, 19) = 0.01, *p* = .92), the Old/New x Hemisphere (*F*(1, 19) = 0.56, *p* = .46), Old/New x Channel Group (*F*(2, 38) = 1.68, *p* = .20), and Old/New x Hemisphere x Channel Group (*F*(2, 38) = 1.77, *p* = .18) interactions were not significant.

The peak latency of the P2 derived from the grand average waveforms was 228ms. Analysis of the mean amplitudes within the 20ms time window

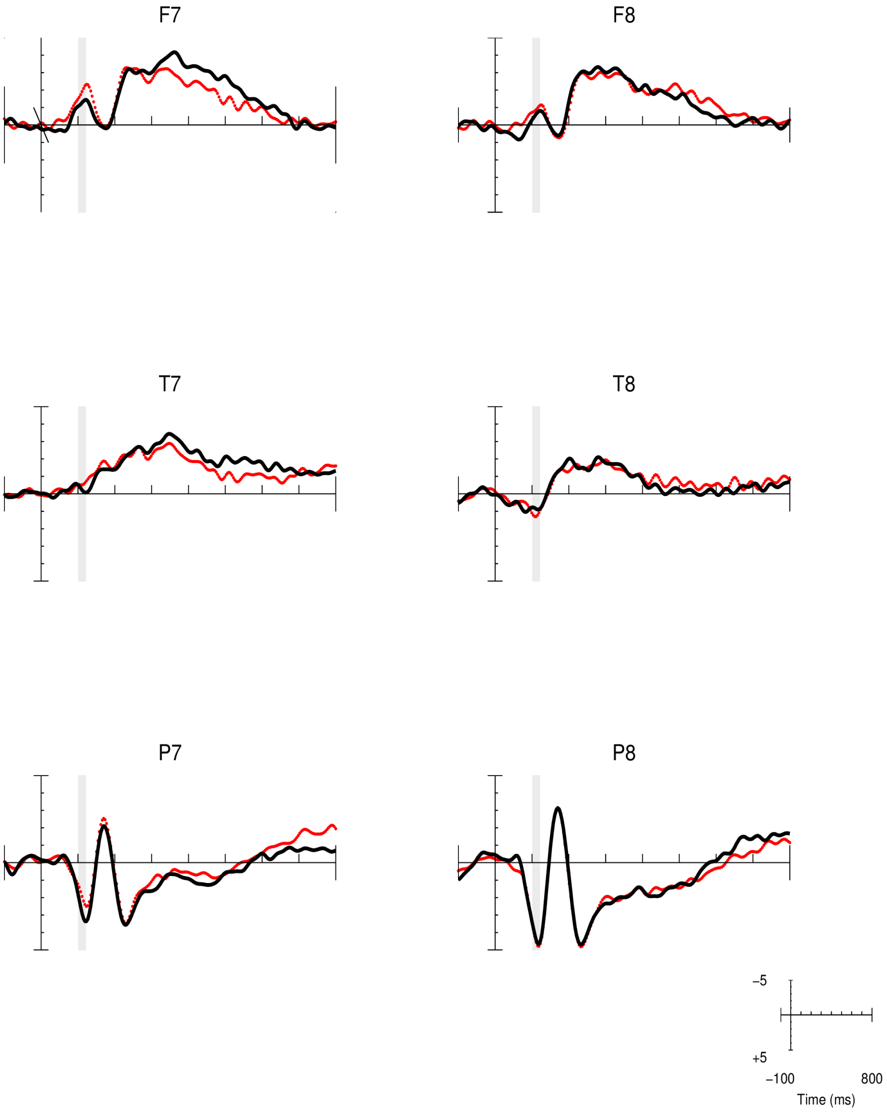


Figure 4.4: The ERP responses to cars in the Dog-Car experiment. The red dotted lines show the ERP in the Old condition. The black solid lines show the ERP in the New condition. The grey bars indicate the time windows to compute the average P1 amplitudes.

surrounding the peak revealed that the main effect of Old/New (*F*(1, 19) = 0.17, *p* = .68), the Old/New x Hemisphere (*F*(1, 19) = 1.05, *p* = .32), Old/New x Channel Group (*F*(2, 38) = 1.49, *p* = .24), and Old/New x Hemisphere x Channel Group (*F*(2, 38) = 0.08, *p* = .92) interactions were not significant.

4.3 Discussion

The Dog-Car experiment presented front views of cars following either individually identifiable barks (the old condition) or unknown barks (the new condition). The P1 amplitude elicited by car stimuli following old dog barks was significantly smaller than that elicited by car stimuli following new barks, particularly over the left hemisphere. This was opposite to the results at the P1 time window in the Dog-Dog experiment, in which responses in the P1 time window were more positive over the right hemisphere to the dog faces following the old barks than to those following the new barks. If the effects in the Dog-Dog experiment reflect binding of the visual and the auditory stimuli from the same basic-level category, the barks and the cars are not from the same basic-level category. However, the categorization level of the New barks were congruent with the cars. It was consistent with the hypothesis that the conceptual representation activated by the auditory stimulus can affect early visual processing of the subsequent stimulus.

As P1 amplitude is related to attention (e.g., Mangun et. al., 2001; Taylor, 2002; Klimesch, 2011), it is also possible that preceding the cars with individually identifiable barks decreased attention to the car stimuli. These findings are consistent with previous research that demonstrated the interplay between attention and cross-modal integration (see Talsma, Senkowski, Soto-Faraco, & Woldorff, 2010 for review). Although attention is not always necessary for cross-modal integration, when semantic processing is involved, such as detecting semantic congruency, attention to the sensory input from at least one of the modalities of the multimodal stimulus is required (van de Burg, Brederoo, Nieuwenstein, Theeuwes, & Olivers 2010). Moreover, multi-modal stimuli can capture attention and lead to spread of attention across modalities (Busse et al., 2005). Semantic congruency can facilitate processing of audiovisual stimuli, as reflected in faster reaction times and higher accuracy (e.g., Calvert, 2005; Laurienti, Kraft, Maldjian, Burdette, & Wallace, 2004; Calvert & Thesen, 2004), and affect attention to stimuli that briefly follow (van de Burg et al., 2010). Attention and cross-modal integration are thus closely related. Here, we paired stimuli that were at the basic level of categorization, but were from different conceptual categories. Half of the auditory stimuli were associated with individual dog names and were therefore more informative than the new barks. When these known barks were paired with unknown cars, they modulated attention to the car stimuli.

Taken together, the Dog-Dog and the Dog-Car experiments indicate that individually identifiable sounds can modulate responses to visual stimuli that follow shortly thereafter. The Dog-Dog experiment showed that individually identifiable auditory stimuli affect responses to visual stimuli when both the auditory and the visual stimuli are from the same basic-level category. The Dog-Car experiment showed that auditory stimuli also affect responses to visual stimuli when the visual stimuli belong to a different basic-level category. But the results from the two experiments were opposite both in the direction of amplitude changes and the scalp distributions. Moreover, both dogs and cars have default categorization at the basic level. This raises the question of whether the effect of auditory stimuli would still be present if the default level of the visual stimuli is at the subordinate level.

We can use human faces to address this question because most people categorize human faces at the subordinate level (Tanaka, 2001). Using human faces has another two advantages. First, the N170 is related to the categorization level of visual stimuli in that categorization at a more specific level elicits a larger N170 amplitude (Tanaka & Curran, 2001). Hence, the absence of a significant effect on the N170 in Experiments 1 and 2 may mean that the preceding sounds did not change the categorization level of the dog and car stimuli. However, we need to exclude the possibility that the N170 responses to objects of no expertise, such as dog and car images, are too small to reveal a detectable modulation of the N170 response by a preceding known auditory stimulus (i.e., a bark). The N170 component has a larger amplitude for human faces than for other objects (e.g., Eimer, 2000b), and some authors have argued that it is face specific (see Earp & Everett, 2013 for a review). Therefore, using human faces allows us to test this explanation.

The second advantage lies in the categorization-level congruency between human faces and individually identifiable barks. Although the first two experiments showed significant modulation of the preceding sounds on the visual responses, the details of this modulation demand further investigation. For example, the Dog-Car experiment showed that the preceding sounds can affect attention to cars, but did not explain why attention to cars decreased. Given the close relationship between attention and semantic congruency, one possible explanation is that the effect on attention may not rely solely on the categorization level of the auditory stimuli. Congruency between the categorization levels of the auditory and the visual stimuli may also play a role. Cars have a default categorization at the basic level, as do the unknown dog barks. Hence, cars are more congruent with the unknown barks than with the individually identifiable barks and this might allow cars following the unknown barks to attract more attention. If the hypothesis stands, using human faces as the visual stimuli should give the opposite result. That is, a human face should attract more attention when following an individually identifiable bark because both stimuli are at the subordinate level.

Chapter 5 Dog-Human Experiment

5.1 Methods

The Dog-Human experiment was the same as the Dog-Dog experiment, with the exception that the pictures in the visual target detection task were male and female Chinese faces. There were 60 upright Chinese faces and 6 inverted ones in total. Sixteen participants (8 females, mean age of 22.1) from the National University of Singapore were included in the experiment.

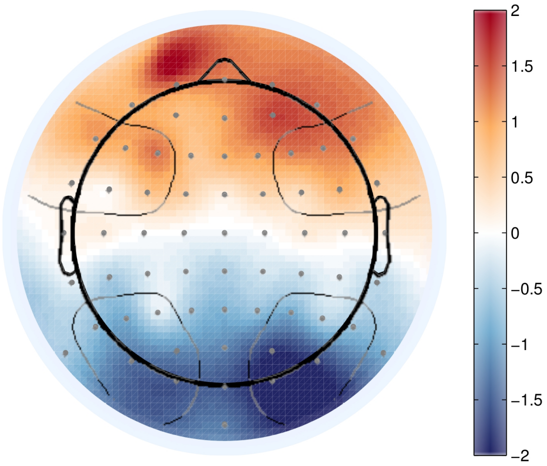


Figure 5.1: Topographic distribution of the ERP response difference between the Old and the New conditions at the P1 latency in the Dog-Human experiment. The ERP amplitudes were averaged within the 20ms time window surrounding the P1 peak. The averaged amplitudes in the New condition were subtracted by the averaged amplitudes in the Old condition. There was significant interaction between the Old/New conditions and the Anterior/Middle/Posterior regions. The human faces following the new barks elicited larger amplitudes than the human faces following the old barks over the anterior regions. There was no significant difference over the middle regions. The human faces following the new barks elicited smaller amplitudes than the human faces following the old barks over the posterior regions. The grey dots indicate the location of the 64 electrodes. The contours indicate the significant areas.

The participants were ethnically Chinese, so the pictures were own-race faces. Informed written consent was obtained from each participant before the experiment. They were reimbursed at a rate of 9 SGD an hour for their time.

5.2 Results

5.2.1 Behavioral Results

The average hit rate was 100%. The average reaction time was 487ms. The average false alarm rate was 0.31%. As the number of the target trials was too small, no further analysis was done on the behavioral data.

5.2.2 P1

For the P1 half area latency, the main effect of Old/New ($F(1, 15) = 0.03$, $p = .86$), the Old/New x AMP ($F(2, 30) = 0.17$, $p = .84$), Old/New x Hemisphere ($F(1, 15) = 0.49$, $p = .50$), and the Old/New x AMP x Hemisphere ($F(2, 30) = 0.06$, $p = .94$) interactions were not significant. The peak latency of the P1 derived from the grand average waveforms 114ms. The analysis of the mean amplitudes within the 20ms time window surrounding this peak revealed that the main effect of Old/New ($F(1, 15) = 1.22$, $p = .29$), the Old/New x Hemisphere ($F(1, 15) = 0.00$, $p = .95$), and Old/New x AMP x Hemisphere ($F(2, 30) = 2.12$, $p = .14$) interactions were not significant. However, the Old/New x AMP interaction was significant ($F(2, 30) = 5.10$, $p = .01$). At anterior electrode sites, the human faces following the new barks elicited larger P1 amplitudes than the human faces following the old barks ($F(1, 15) =$

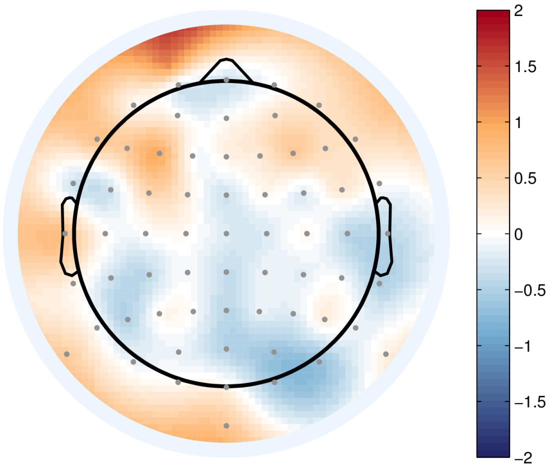


Figure 5.2: Topographic distribution of the ERP response difference between the Old and the New conditions at the N170 latency in the Dog-Human experiment. The ERP amplitudes were averaged within the 20ms time window surrounding the N170 peak. The averaged amplitudes in the New condition were subtracted by the averaged amplitudes in the Old condition. The amplitudes were not significantly different between the Old and the New conditions. The grey dots indicate the location of the 64 electrodes.

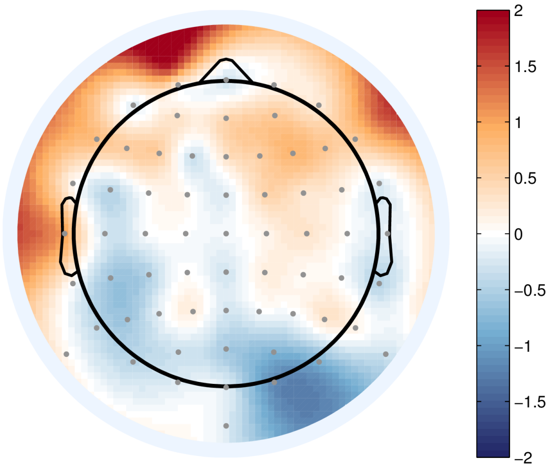


Figure 5.3: Topographic distribution of the ERP response difference between the Old and the New conditions at the P2 latency in the Dog-Human experiment. The ERP amplitudes were averaged within the 20ms time window surrounding the P2 peak. The averaged amplitudes in the New condition were subtracted by the averaged amplitudes in the Old condition. The amplitudes were not significantly different between the Old and the New conditions. The grey dots indicate the location of the 64 electrodes.

5.22, *p* = .04). At middle electrode sites, there was no significant effect of Old/New condition (*F*(1, 15) = 0.03, *p* = .87). At posterior electrode sites, the human faces following the new barks elicited smaller amplitudes than the human faces following the old barks (*F*(1, 15) = 4.94, *p* = .04).

5.2.3 N170

For the N170 half area latency, the main effect of Old/New (*F*(1, 15) = 0.00, *p* = .95), the Old/New x Hemisphere (*F*(1, 15) = 1.72, *p* = .21), and Old/New x Hemisphere x Channel Group (*F*(2, 30) = 0.83, *p* = .45) interactions were not significant. Although the Old/New x Channel Group interaction was significant (*F*(2, 30) = 3.62, *p* = .04), follow-up analyses failed to reveal a significant effect of Old/New for any Channel Group (P9/P10: *F*(1, 15) = 2.16, *p* = .16; P8/P7: *F*(1, 15) = 0.70, *p* = .42; PO8/PO7: *F*(1, 15) = 1.68, *p* = .21).

The peak latency of the N170 derived from the grand average waveforms was 150ms. The analysis of the mean amplitudes within the 20ms time window surrounding this peak revealed that the main effect of Old/New (*F*(1, 15) = 0.00, *p* = .96), the Old/New x Hemisphere (*F*(1, 15) = 0.50, *p* = .49), Old/New x Channel Group (*F*(2, 30) = 1.70, *p* = .20), and Old/New x Hemisphere x Channel Group (*F*(2, 30) = 0.24, *p* = .79) interactions were not significant.

5.2.4 P2

For the P2 half area latency, the main effect of Old/New (*F*(1, 15) =

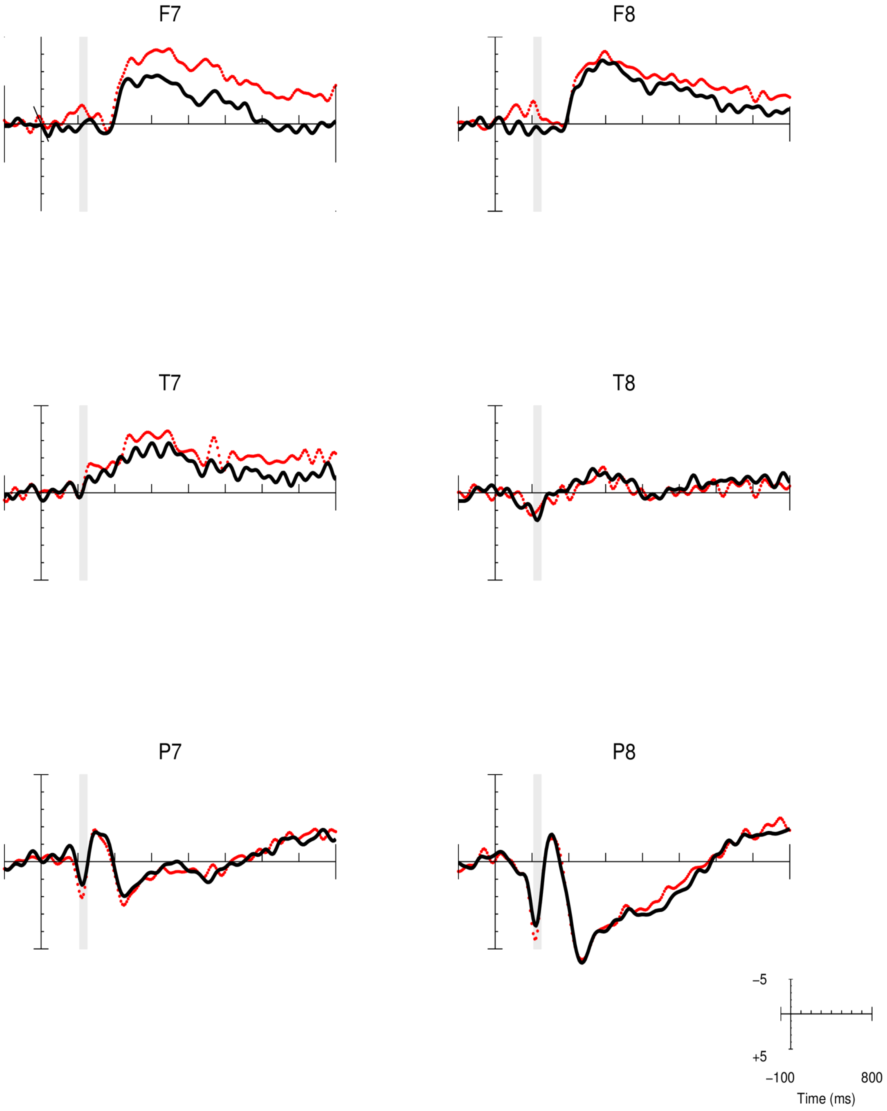


Figure 5.4: The ERP responses to human faces in the Dog-Human experiment. The red dotted lines show the ERP in the Old condition. The black solid lines show the ERP in the New condition. The grey bars indicate the time windows to compute the average P1 amplitudes.

1.69, *p* = .76), the Old/New x Hemisphere (*F*(1, 15) = 0.09, *p* = .76), Old/New x Channel Group (*F*(2, 30) = 2.92, *p* = .07), and Old/New x Hemisphere x Channel Group (*F*(2, 30) = 0.19, *p* = .83) interactions were not significant.

The peak latency of the P2 derived from the grand average waveforms was 227ms. Analysis of the mean amplitudes within the 20ms time window surrounding this peak indicated that the main effect of Old/New ($F(1, 15) = 0.39$, $p = .54$), the Old/New x Hemisphere ($F(1, 15) = 0.02$, $p = .88$), Old/New x Channel Group ($F(2, 30) = 1.02$, $p = .37$), and Old/New x Hemisphere x Channel Group ($F(2, 30) = 1.00$, $p = .38$) interactions were not significant.

5.3 Discussion

As in the Dog-Dog and the Dog-Car experiments, the barks preceding the human faces modulated the P1 response to the faces. However, unlike in the Dog-Car experiment, but similar to the Dog-Dog experiment, the P1 amplitudes over posterior electrode sites were larger in the old condition than in the new condition. Taken together, these results suggest that the preceding sounds alone did not determine the direction of the P1 amplitude change to the visual stimulus. Rather, the category congruency between the bark and the subsequent visual stimulus may also be important. Cars, by default, are categorized at the basic level, whereas human faces, by default, are categorized at the subordinate level. Hence, in the Car experiment the car stimuli were categorically congruent with the unknown barks, whereas in the Human experiment the face stimuli were categorically congruent with the individually-identifiable barks. In both experiments, the P1 amplitudes were larger in the condition for which the categorization levels of the auditory and the visual stimuli were congruent. That is to say, cars in the new condition elicited larger P1 amplitudes than those in the old condition, while human faces in the old condition elicited larger P1 amplitudes than those in the new condition. This suggests that it is the categorization-level congruency, or lack thereof, between the auditory and the visual stimuli that determines the direction of the effect on the P1 amplitude between the old and the new conditions.

There was a significant frontal difference between the Old and the New conditions in the P1 time window. The responses to the human faces following the old barks versus the new barks were more negative. This may reflect modulation of the anterior subcomponent of the visual N1 (Vogel & Luck, 2000) during the P1 time window. The visual N1 can be divided into two subcomponents with different latencies and scalp distributions. The anterior visual N1 occupied the frontal region and peaks around 100ms, while the posterior visual N1 distributed over the occipito regions and peaks around 160ms (Vogel & Luck, 2000). The anterior visual N1 can be affected by attention (Talsma et al., 2010). Attention increases the anterior N1 amplitudes. Particularly, Talma and colleagues (2009) observed that the attention effect increased when the auditory and the visual stimuli were not at the optimal SOA for audiovisual integration (e.g., SOA longer than 50ms). Moreover, Töllner, Gramann, Müller, and Eimer (2008) reported that the anterior visual N1 was related to modality shifting. Shifting modality enhanced the N1 amplitudes. Thus, the increase negativity over the anterior region for the human faces following the old barks may reflect enhanced attention for better audiovisual integration.

By using human faces as the visual stimuli, the experiment also permits elimination of a possible account of the lack of an Old/New effect on the N170 in the first two experiments. There was no significant effect on the N170 in the Dog-Human experiment, suggesting that the categorization level of the human faces was not affected. As human faces typically elicit larger N170 amplitudes than other objects, the absence of an effect of the preceding sounds would not be due to the small amplitude of the N170. Instead, it is more likely that the N170 is not sensitive to the modulation by the preceding sounds. Preceding visual stimuli with individually identifiable sounds cannot immediately alter the categorization level of the visual stimuli. Note that the visual stimuli in all three experiments were presented only once and the participants' task was to indicate whether the visual stimuli were inverted, which required only basic-level categorization. Modulation of the N170 may require presenting the audiovisual pairs repeatedly (e.g., Heisz, Watter, & Shedden, 2006).

Experiments 1 to 3 used barks as the auditory stimuli, but there is no reason to believe that barks are unique in causing the effects on the responses to visual stimuli. We argued above that associating auditory stimuli with names, which allows them to be individually identified, modulates responses to visual stimuli that follow shortly thereafter. Specifically, when the auditory and the visual stimuli belong to different basic-level categories (as in the Dog-Car and the Dog-Human experiments), categorization-level congruency between the auditory and the visual stimuli will increase allocation of attention to the visual stimuli. Auditory stimuli from other categories should have the same effect on the visual responses, so in the fourth experiment we replaced the barks with human voices.

Human voices were chosen because compared to barks, unknown human voices, by default, are categorized at a more specific level (Guastavino, 2007; Belin, Fecteau, & Bédard, 2004). If the P1 effects are due to congruency between stimulus categorization levels, then changing the default categorization level of the auditory stimulus to the subordinate level should also modulate the P1 effect. Voices in the old condition would still be more specific than the unknown voices, as the former could be individually identified following association with human names. Voices in the new condition would still be more congruent in categorization level with dog faces than those in the old condition. Therefore, the dog faces in the new condition should attract more attention than the dog faces in the old condition and the P1 in the new condition should be larger than that in the old condition.

Chapter 6 Human-Dog Experiment

6.1 Methods

The Human-Dog experiment was the same as the Dog-Dog experiment with the exception that the sounds were human voices, recorded from 60 Chinese students. There were 30 male voices and 30 female voices. All of the speakers said “dada”. The names the participants learned to associate with the voices were human names that matched the sex of the voice. Sixteen participants (8 females, mean age of 22.2) from the National University of Singapore were included in the experiment. The participants were ethnically Chinese, so the voices were own-race voices. Informed written consent was obtained from each participant before the experiment. They were reimbursed at a rate of 9 SGD an hour for their time.

6.2 Results

6.2.1 Behavioral Results

The average hit rate was 100%. The average reaction time was 498ms, and the average false alarm rate was 0.73%. As the number of the target trials was too small, no further analysis was conducted on the behavioral data.

6.2.2 P1

For the P1 half area latency, the main effect of Old/New ($F(1, 15)$ =

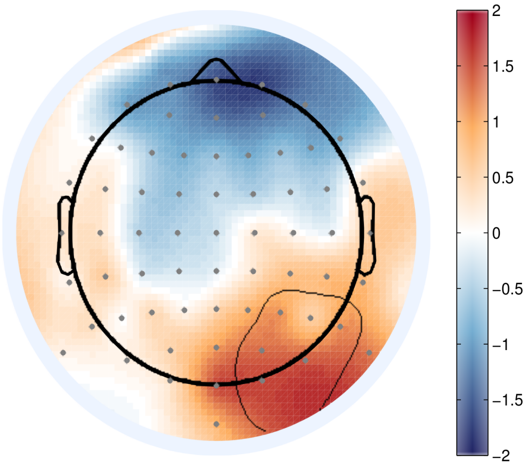


Figure 6.1: Topographic distribution of the ERP response difference between the Old and the New conditions at the P1 latency in the Human-Dog experiment. The ERP amplitudes were averaged within the 20ms time window surrounding the P1 peak. The averaged amplitudes in the New condition were subtracted by the averaged amplitudes in the Old condition. There was significant 3-way interaction (Old/New conditions X Hemispheres X Anterior/Middle/Posterior regions). Over the anterior regions, the Old condition did not significantly differ from the New condition. Over the middle regions, the old condition did not significantly differ from the new condition. Over the posterior regions, there was significant interaction between the Old/New conditions and the Hemispheres. The dog faces following the new barks elicited larger amplitudes than the dog faces following the old barks over the right posterior regions. The dog faces did not elicit different amplitudes between those following the new barks and those following the old barks over the left posterior regions. The grey dots indicate the location of the 64 electrodes. The contour indicates the significant area.

3.07, p = .10), the Old/New x AMP (F(2, 30) = 0.38, p = .69), Old/New x Hemisphere (F(1, 15) = 3.46, p = .08), and tOld/New x AMP x Hemisphere (F(2, 30) = 0.88, p = .43) interactions were not significant.

The P1 peak latency derived from the grand average waveforms was 119ms. The analysis of the mean amplitudes within the 20ms time window surrounding this peak revealed a marginally significant main effect of Old/New (F(1, 15) = 4.01, p = .06). The Old/New x AMP (F(2, 30) = 2.53, p =

.10) and Old/New x Hemisphere (*F*(1, 15) = 1.37, *p* = .26) interactions were not significant. However, the Old/New x AMP x Hemisphere interaction (*F*(2, 30) = 7.44, *p* = .00) was significant. Over the anterior, middle, and posterior regions the simple main effect of Old/New was not significant, *Fs*(1, 15) = 1.67, 3.26, and 3.05, respectively, all *p*s > .05. The Old/New x Hemisphere interaction was not significant at anterior, (*F*(1, 15) = 1.20, *p* = .29) and middle (*F*(1, 15) = 1.66, *p* = .22) electrode sites, but was significant at posterior electrode sites (*F*(1, 15) = 6.05, *p* = .03). Over the right posterior regions, the dog faces following the new voices elicited larger amplitudes than the dog faces following the old voices (*F*(1, 15) = 6.91, *p* = .02), whereas there was no such effect over the left posterior regions (*F*(1, 15) = 0.42, *p* = .53).

6.2.3 N170

For the N170 half area latency, the main effect of Old/New (*F*(1, 15) = 0.54, *p* = .48), the Old/New x Hemisphere (*F*(1, 15) = 0.74, *p* = .40), Old/New x Channel Group (*F*(2, 30) = 0.14, *p* = .87), and Old/New x Hemisphere x Channel Group (*F*(2, 30) = 0.31, *p* = .74) interactions were not significant.

The peak latency of the N170 derived from the grand average waveforms was 163ms. The analysis of the mean amplitudes within the 20ms time window surrounding the peak revealed that the main effect of Old/New (*F*(1, 15) = 1.07, *p* = .32), the Old/New x Hemisphere (*F*(1, 15) = 0.38, *p* = .55), Old/New x Channel Group (*F*(2, 30) = 0.11, *p* = .90), and Old/New x Hemisphere x Channel Group (*F*(2, 30) = 0.37, *p* = .69) interactions were not significant.

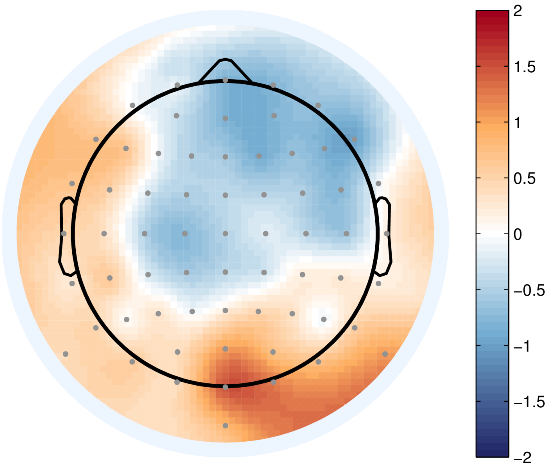


Figure 6.2: Topographic distribution of the ERP response difference between the Old and the New conditions at the N170 latency in the Human-Dog experiment. The ERP amplitudes were averaged within the 20ms time window surrounding the N170 peak. The averaged amplitudes in the New condition were subtracted by the averaged amplitudes in the Old condition. The amplitudes were not significantly different between the Old and the New conditions. The grey dots indicate the location of the 64 electrodes.

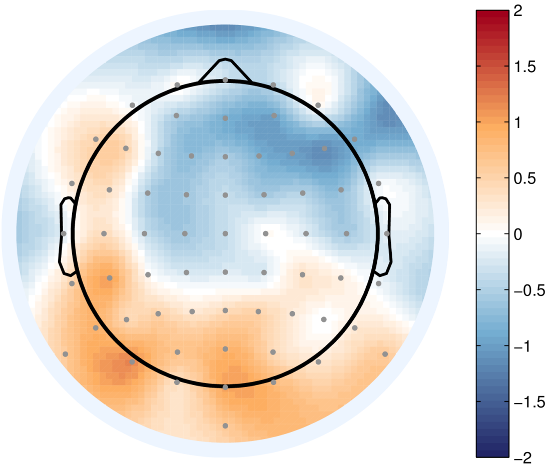


Figure 6.3: Topographic distribution of the ERP response difference between the Old and the New conditions at the P2 latency in the Human-Dog experiment. The ERP amplitudes were averaged within the 20ms time window surrounding the P2 peak. The averaged amplitudes in the New condition were subtracted by the averaged amplitudes in the Old condition. The amplitudes were not significantly different between the Old and the New conditions. The grey dots indicate the location of the 64 electrodes.

6.2.4 P2

For the P2 half area latency, the main effect of Old/New (*F*(1, 15) = 1.09, *p* = .31), the Old/New x Hemisphere (*F*(1, 15) = 0.08, *p* = .78), Old/New x Channel Group (*F*(2, 30) = 0.65, *p* = .53), and Old/New x Hemisphere x

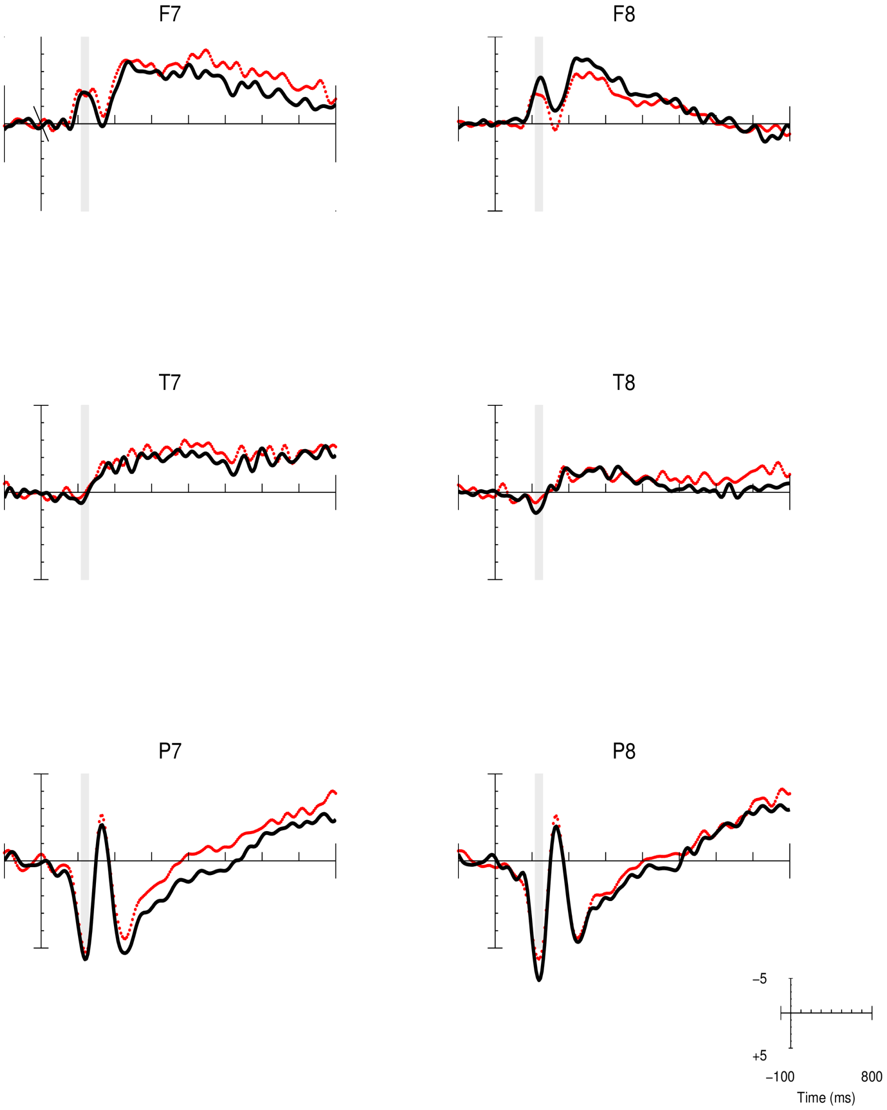


Figure 6.4: The ERP responses to dog faces in the Human-Dog experiment. The red broken lines show the ERP in the Old condition. The black solid lines show the ERP in the New condition. The grey bars indicate the time window to compute the average P1 amplitudes.

Channel Group (*F*(2, 30) = 1.32, *p* = .28) interactions were not significant.

The peak latency of the P2 derived from the grand average waveforms was 223ms. Analysis of the mean amplitudes within the 20ms time window surrounding this peak indicated that the main effect of Old/New (*F*(1, 15) = 1.30, *p* = .27), the Old/New x Hemisphere (*F*(1, 15) = 1.82, *p* = .20), Old/New x Channel Group (*F*(2, 30) = 0.07, *p* = .94), and Old/New x Hemisphere x Channel Group (*F*(2, 30) = 0.63, *p* = .54) interactions were not significant.

6.3 Discussion

The Human-Dog experiment compared responses to the dog faces following the individually-identifiable voices (the Old condition) with those following the unknown voices (the New condition). The dog faces following the individually-identifiable voices elicited smaller P1 over the right occipito-temporal region than the dog faces following the unknown voices.

As expected, the P1 amplitudes were larger in the New condition when the categorization levels of the auditory and the visual stimuli were more congruent and the results were consistent with those of the Dog-Car and the Dog-Human experiments. Dog faces have a default categorization at the basic level and the unknown human voices at the subordinate level, but the human voices in the Old condition can be individually identified. Hence, the Dog faces were more congruent in categorization level with the unknown voices than with the individually-identifiable voices. Consistent with the Dog-Car and the Dog-Human experiments, the P1 amplitudes were larger when the categorization levels are more congruent (i.e., the New condition). It demonstrated again that the direction of the P1 amplitude change does not depends solely on the categorization level of the auditory stimuli. The human voices were by default categorized at the subordinate level, while the barks were by default at the basic level, but both auditory stimuli increased the visual P1 amplitude when they were more congruent with the categorization level of the visual stimuli.

Categorization-level congruency accounts for the P1 effects in the Dog-Car, the Dog-Human, and the Human-Dog experiments. However, in the Dog-Dog experiment, no P1 effect was found. Responses to dog faces in the P1 time window were modulated by barks. But the topographic distribution of this effect was mainly over the right temporal region, which is different from the typical P1 scalp distribution. This effect was not replicated in any of the subsequent experiments. One important difference between the Dog-Dog experiment from the other three is that both the auditory and the visual stimuli were from the same basic-level category. Participants might recognize them as from the same objects and this might give rise to the unique results in the Dog-Dog experiment. To test this hypothesis, we conducted the fifth experiment, where both the auditory and the visual stimuli were also from the same basic-level category. Instead of Dogs, however, the fifth experiment used the basic category of Human. Since the participants did not know any of the persons in the visual stimuli, they might conclude that the voices and the faces belong to the same person. This was the same as in the Dog-Dog experiment, but the Human-Human experiment differed from the Dog-Dog experiment in the default categorization level. Human faces are by default categorized at the subordinate level, while dog faces are at the basic level.

Chapter 7 Human-Human Experiment

7.1 Methods

The Human-Human experiment was the same as the Human-Dog experiment with the exception that the pictures in the visual target detection task were Chinese faces. The faces were the same as those in the Dog-Human experiment. Sixteen participants (8 females, mean age of 21.9) from the National University of Singapore were included in the experiment. The participants were ethnically Chinese, so both the voices and the pictures were of own race. Informed written consent was obtained from each participant before the experiment. They were reimbursed at a rate of 9 SGD an hour for their time.

7.2 Results

7.2.1 Behavioral Results

The average hit rate was 100%. The average reaction time was 484ms, and the average false alarm rate was 0%. As the number of the target trials was too small, no further analyses were conducted on the behavioral data.

7.2.2 P1

For the P1 half area latency, the main effect of Old/New (*F*(1, 15) = 0.04, *p* = .84), the Old/New x AMP (*F*(2, 30) = 1.18, *p* = .32), Old/New x

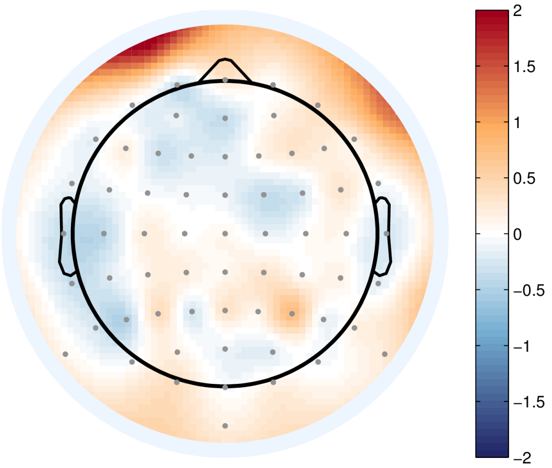


Figure 7.1: Topographic distribution of the ERP response difference between the Old and the New conditions at the P1 latency in the Human-Human experiment. The ERP amplitudes were averaged within the 20ms time window surrounding the P1 peak. The averaged amplitudes in the New condition were subtracted by the averaged amplitudes in the Old condition. The amplitudes were not significantly different between the Old and the New conditions. The grey dots indicate the location of the 64 electrodes.

Hemisphere (*F*(1, 15) = 1.03, *p* = .33), and Old/New x AMP x Hemisphere (*F*(2, 30) = 0.02, *p* = .98) interactions were not significant.

The peak latency of the P1 derived from the grand average waveforms was 117ms. The analysis of the mean amplitudes within the 20ms time window surrounding this peak revealed that the main effect of Old/New (*F*(1, 15) = 0.22, *p* = .65), the Old/New x AMP (*F*(2, 30) = 1.18, *p* = .32), Old/New x Hemisphere (*F*(1, 15) = 0.22, *p* = .65), and Old/New x AMP x Hemisphere (*F*(2, 30) = 0.02, *p* = .98) interactions were not significant.

7.2.3 N170

For the N170 half area latency, the main effect of Old/New, the Old/New x Hemisphere (*F*(1, 15) = 1.27, *p* = .28), Old/New x Channel Group

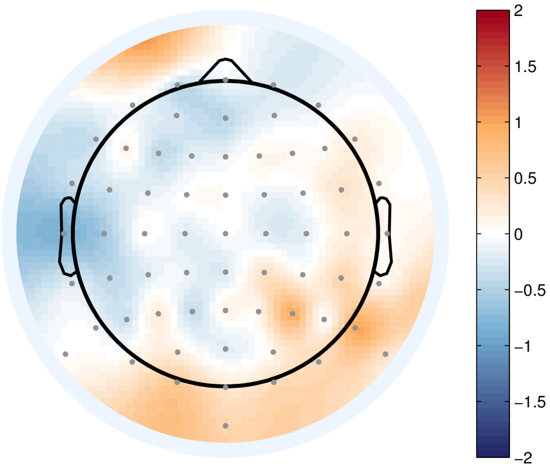


Figure 7.2: Topographic distribution of the ERP response difference between the Old and the New conditions at the N170 latency in the Human-Human experiment. The ERP amplitudes were averaged within the 20ms time window surrounding the N170 peak. The averaged amplitudes in the New condition were subtracted by the averaged amplitudes in the Old condition. The amplitudes were not significantly different between the Old and the New conditions. The grey dots indicate the location of the 64 electrodes.

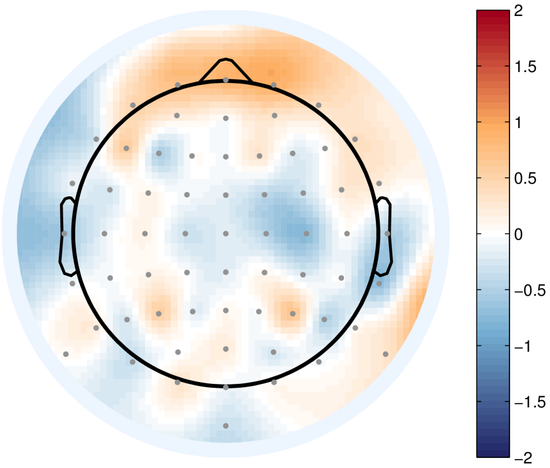


Figure 7.3: Topographic distribution of the ERP response difference between the Old and the New conditions at the P2 latency in the Human-Human experiment. The ERP amplitudes were averaged within the 20ms time window surrounding the P2 peak. The averaged amplitudes in the New condition were subtracted by the averaged amplitudes in the Old condition. The amplitudes were not significantly different between the Old and the New conditions. The grey dots indicate the location of the 64 electrodes.

(*F*(2, 30) = 0.43, *p* = .66), and Old/New x Hemisphere Channel Group (*F*(2, 30) = 1.15, *p* = .33) interactions were not significant.

The peak latency of the N170 derived from the grand average waveforms was 154ms. The analysis of the mean amplitudes within the 20ms time window surrounding this peak revealed that the main effect of Old/New (*F*(1, 15) = 0.67, *p* = .43), the Old/New x Hemisphere (*F*(1, 15) = 1.54, *p* = .27), Old/New x Channel Group (*F*(2, 30) = 0.29, *p* = .75), and Old/New x Hemisphere x Channel Group (*F*(2, 30) = 1.05, *p* = .36) interactions were not significant.

7.2.4 P2

For the P2 half area latency, the main effect of Old/New (*F*(1, 15) = 3.89, *p* = .07), the Old/New x Hemisphere (*F*(1, 15) = 3.09, *p* = .10), Old/New x Channel Group (*F*(2, 30) = 2.15, *p* = .13), and Old/New x Hemisphere x Channel Group (*F*(2, 30) = 1.03, *p* = .37) interactions were not significant.

The peak latency of the P2 derived from the grand average waveforms was 217ms. The analysis of the mean amplitudes within the 20ms time window surrounding the P2 peak revealed that the main effect of Old/New (*F*(1, 15) = 0.01, *p* = .94), the Old/New x Hemisphere (*F*(1, 15) = 0.09, *p* = .77), Old/New x Channel Group (*F*(2, 30) = 0.67, *p* = .52), and Old/New x Hemisphere x Channel Group (*F*(2, 30) = 0.09, *p* = .77) interactions were not significant.

7.3 Discussion

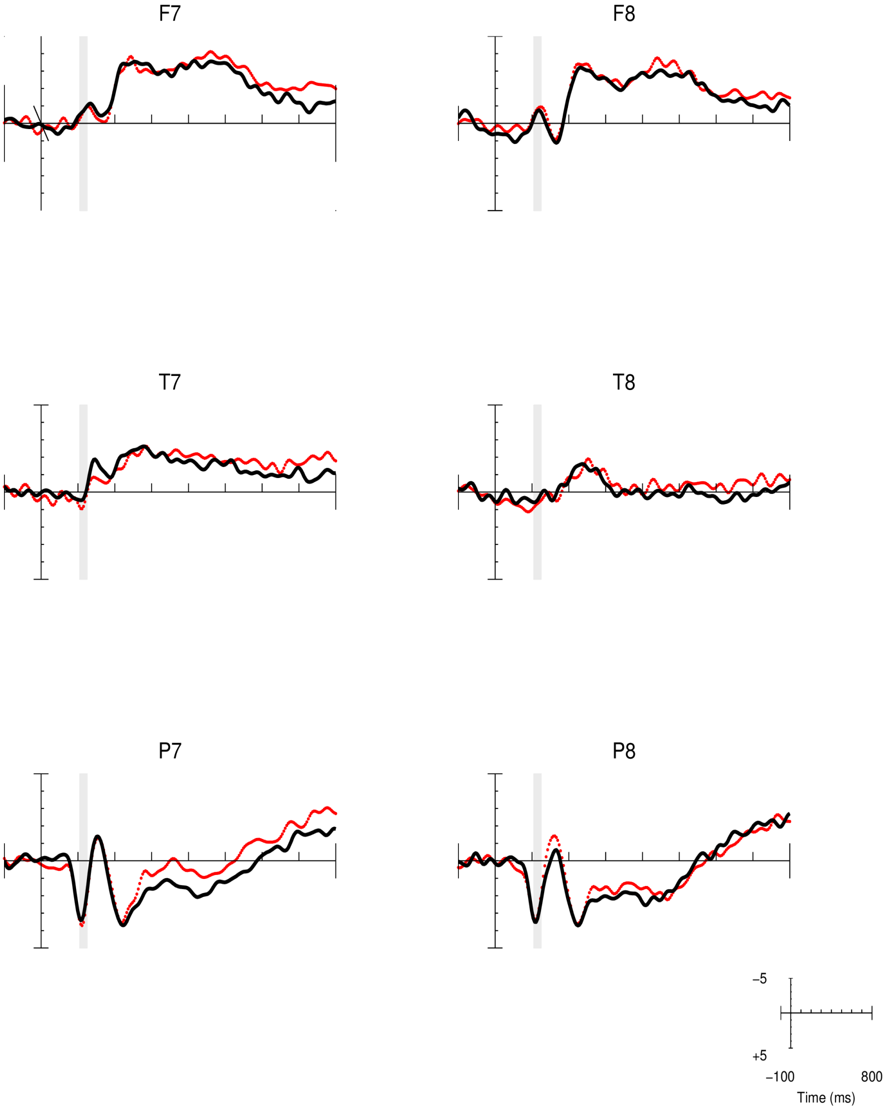


Figure 7.4: The ERP responses to human faces in the Human-Human experiment. The red dotted lines show the ERP in the Old condition. The black solid lines show the ERP in the New condition. The grey bars indicate the time window to compute the average P1 amplitudes.

The Human-Human experiment compared responses to the human faces following the individually-identifiable voices (the Old condition) with those following the unknown voices (the New condition). The P1 amplitudes were not significantly different between the Old and the New conditions.

The results in this experiment partly replicated the findings of the previous experiments. As in the Dog-Dog experiment, there was no effect on the P1 component. Unknown human voices and human faces were categorized at the subordinate level, while the learned voices could be individually identified. Hence, the human faces were more congruent with the unknown voices than with the known voices, but the human faces in the congruent condition did not elicit larger P1 amplitudes. This suggests that early responses to visual stimuli did not differ between the Old and the New condition and indicates that the categorization-level congruency effect observed in the Dog-Car, the Dog-Human, and the Human-Dog experiments occurs only when the auditory and the visual stimuli belong to different basic-level categories. Early visual responses at the posterior regions do not further increase when the auditory and the visual stimuli belong to the same basic-level category and this is true regardless of the default categorization level of visual stimuli. Dog faces were by default categorized at the basic level, while human faces were by default categorized at the subordinate level, but the early visual responses at the posterior regions to neither dog faces in the Dog-Dog experiment nor human faces in the Human-Human experiment increased when the auditory and the visual stimuli were congruent in categorization level.

However, not all the results in the Dog-Dog experiment were replicated. In the Dog-Dog experiment, there was an early response difference over the right temporal electrode sites. This effect was not found in the

Human-Human experiment. In fact, early responses over the right temporal electrode sites was not observed in any of the experiments following the Dog-Dog experiment.

The early response difference might be related multisensory integration (Stevenson & James, 2009; Barraclough, Xiao, Baker, Oram, & Perrett, 2005). While the auditory and the visual stimuli were obviously from different objects, participants may process the dog faces and the barks as from the same dogs. Preceding dog faces with barks could lead to integration of the barks and the dog faces to form multi-sensory object representations of dogs. The early response difference suggests that familiarity with the barks might be facilitating this integration. Moreover, the facilitation might be more effective for dog faces than human faces. Dog faces are by default categorized at the basic level, while human faces are categorized at the subordinate level. People are so sophisticated in human face-voice integration (Campanella & Belin, 2007) that familiarity with the voices would not further benefit the process. Thus, there was significant early response difference in the Dog-Dog experiment, but not in the Human-Human experiment.

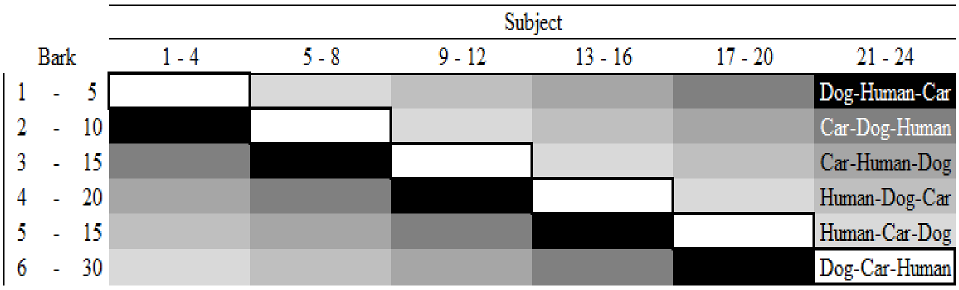
We argued that two factors were particularly important for the preceding auditory stimuli to modulate the processing of the visual stimuli. One was whether the auditory and the visual stimuli belong to the same basic-level category. The other was the categorization-level congruency between the auditory and the visual stimuli. It follows that the categorization level of the visual stimulus needs to be processed before the auditory and the visual stimuli can be compared. However, the modulation of the preceding sounds appeared as early as on the P1 component. Though categorization at the superordinate level (i.e., living or non-living objects) could occur as early as 70ms after stimuli onset, differentiation between processing at different categorization levels was reported at a much later stage (Large et. al., 2004). One explanation is that the visual stimulus categories were known even before the visual stimulus onset. This allowed the modulation effects of the preceding sounds to occur early. In all five experiments above, the category of the visual stimuli was the same throughout the experiments. Participants could anticipate the basic-level category of the visual stimuli. This allowed comparison of the basic-level categories and categorization level congruency between the auditory and the visual stimuli before visual stimulus onset. Therefore, the modulation effects could be effective at an early stage of visual processing.

To understand the effect of expectation, a final experiment was conducted. An auditory stimulus was followed by a visual stimulus, but instead of from only one stimulus category, the visual stimulus could be from any of the three basic-level categories: Dog, Car, or Human. If anticipation of the basic-level category of the visual stimulus does not matter, the results in the Dog-Mix experiment should replicate the findings in the Dog-Dog, the Dog-Car, and the Dog-Human experiments. Attention, as reflected in the P1, should increase for the human faces, decrease for the cars, and not change for the dog faces. Otherwise, conceptual categorization of the visual stimuli is required after visual stimulus onset for preceding sounds to affect visual stimulus processing, in which case the early response differences should be eliminated.

Chapter 8 Dog-Mix Experiment

8.1 Methods

The Dog-Mix experiment used the same sound training procedure as in Experiments 1 - 5. During the EEG recording, participants performed a target detection task. Participants heard a sequence of dog barks. Half of the barks were learned in the training phase and the other half were new. Instead of presenting the barks only once, each dog bark was repeated three times non-consecutively. A picture from one of three categories (a dog face; a front view of a car; a human face) followed each bark presentation. The Dog-Mix experiment used the same pictures as Experiments 1-5. All the human faces were Chinese faces. There were in total 180 upright pictures and 18 inverted target pictures. The categories of the three pictures that followed the same bark were counterbalanced across participants and barks, using a Latin square design (see Table 1). Twenty-four participants (13 females, mean age of 21.5)



	Subject					
Bark	1 - 4	5 - 8	9 - 12	13 - 16	17 - 20	21 - 24
1 - 5	□	░	▒	▓	█	Dog-Human-Car
2 - 10	■	□	░	▒	▓	Car-Dog-Human
3 - 15	█	■	□	░	▒	Car-Human-Dog
4 - 20	▓	█	■	□	░	Human-Dog-Car
5 - 15	▒	▓	█	■	□	Human-Car-Dog
6 - 30	░	▒	▓	█	■	Dog-Car-Human

Table 8.1: The assignment of barks to picture categories across bark repetitions is illustrated. E.g., Dog-Car-Human indicates that when the bark was presented for the first time, it was followed by a dog face; when the bark was presented for the second time, it was followed by a car; and when the bark was presented for the third time, it was followed by a human face.

from the National University of Singapore were included in the experiment. The participants were ethnically Chinese, so the human-face pictures were own-race faces. Informed written consent was obtained from each participant before the experiment. They were reimbursed at a rate of 9 SGD an hour for their time.

8.2 Results

8.2.1 Behavioral Results

The average hit rate was 97.45%. The average reaction time was 558ms, and the average false alarm rate was 0.58%. As the number of the target trials was too small, no further analyses were conducted on the behavioral data.

8.2.2 P1

8.2.2.1 Dog Faces

For the P1 half area latency the main effect of Old/New ($F(1, 23) = 0.00$, $p = .95$), the Old/New x AMP ($F(2, 46) = 1.26$, $p = .29$), Old/New x Hemisphere ($F(1, 23) = 0.12$, $p = .73$), and Old/New x AMP x Hemisphere ($F(2, 46) = 0.19$, $p = .83$) interactions were not significant.

The peak latency of the P1 derived from the grand average waveforms was 113ms. The analysis of the mean amplitudes within the 20ms time window surrounding this peak revealed that the main effect of Old/New ($F(1, 23) = 1.30$, $p = .27$), the Old/New x AMP ($F(2, 46) = 2.41$, $p = .10$), Old/New

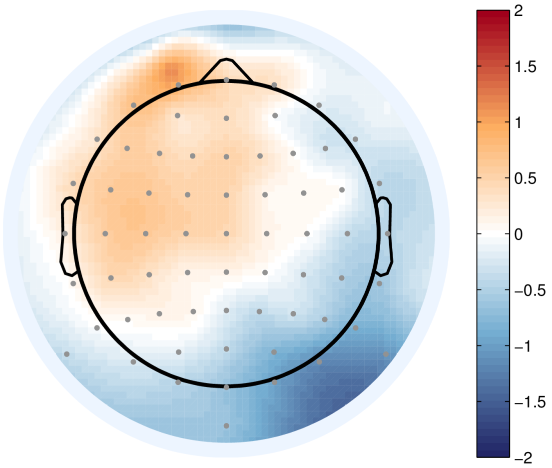


Figure 8.1: Topographic distribution of the ERP response difference to dog faces between the Old and the New conditions at the P1 latency in the Dog-Mix experiment. The ERP amplitudes were averaged within the 20ms time window surrounding the P1 peak. The averaged amplitudes in the New condition were subtracted by the averaged amplitudes in the Old condition. The amplitudes were not significantly different between the Old and the New conditions. The grey dots indicate the location of the 64 electrodes.

x Hemisphere (*F*(1, 23) = 4.04, *p* = .06), and Old/New x AMP x Hemisphere (*F*(2, 46) = 0.04, *p* = .96) interactions were not significant.

8.2.2.2 Cars

For the P1 half area latency, the main effect of Old/New (*F*(1, 23) = 1.16, *p* = .29), the Old/New x AMP (*F*(2, 46) = 0.22, *p* = .80), Old/New x Hemisphere (*F*(1, 23) = 0.10, *p* = .75), and Old/New x AMP x Hemisphere (*F*(2, 46) = 1.63, *p* = .21) interactions were not significant.

The peak latency of the P1 derived from the grand average waveforms was 113ms. The analysis of the mean amplitudes within the 20ms time window surrounding this peak revealed that the main effect of Old/New (*F*(1, 23) = 0.05, *p* = .82), the Old/New x AMP (*F*(2, 46) = 0.22, *p* = .80), Old/New

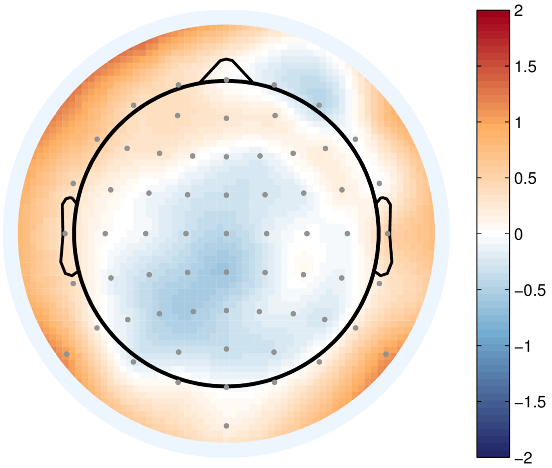


Figure 8.2: Topographic distribution of the ERP response difference to cars between the Old and the New conditions at the P1 latency in the Dog-Mix experiment. The ERP amplitudes were averaged within the 20ms time window surrounding the P1 peak. The averaged amplitudes in the New condition were subtracted by the averaged amplitudes in the Old condition. The amplitudes were not significantly different between the Old and the New conditions. The grey dots indicate the location of the 64 electrodes.

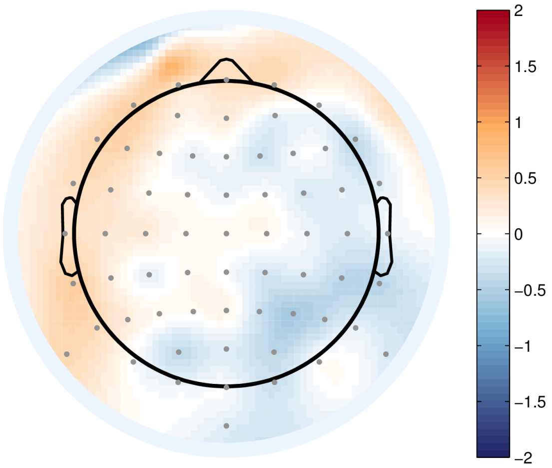


Figure 8.3: Topographic distribution of the ERP response difference to human faces between the Old and the New conditions at the P1 latency in the Dog-Mix experiment. The ERP amplitudes were averaged within the 20ms time window surrounding the P1 peak. The averaged amplitudes in the New condition were subtracted by the averaged amplitudes in the Old condition. The amplitudes were not significantly different between the Old and the New conditions. The grey dots indicate the location of the 64 electrodes.

x Hemisphere (*F*(1, 23) = 0.02, *p* = .89), and Old/New x AMP x Hemisphere (*F*(2, 46) = 0.75, *p* = .48) interactions were not significant.

8.2.2.3 Human Faces

For the P1 half area latency, the main effect of Old/New (*F*(1, 23) = 1.16, *p* = .29), the Old/New x AMP (*F*(2, 46) = 0.40, p = .67), Old/New x Hemisphere (*F*(1, 23) = 0.12, *p* = .73), and Old/New x AMP x Hemisphere (*F*(2, 46) = 0.19, *p* = .83) interactions were not significant.

The peak latency of the P1 derived from the grand average waveforms was 106ms. The analysis of the mean amplitudes within the 20ms time window surrounding this peak revealed that the main effect of Old/New (*F*(1, 23) = 0.16, *p* = .70), the Old/New x AMP (*F*(2, 46) = 0.06, *p* = .94), Old/New x Hemisphere (*F*(1, 23) = 1.18, *p* = .20), and Old/New x AMP x Hemisphere (*F*(2, 46) = 0.09, *p* = .91) interactions were not significant.

8.2.3 N170

8.2.3.1 Dog Faces

For the N170 half area latency, the main effect of Old/New *F*(1, 23) = 0.31, *p* = .58), the Old/New x Hemisphere (*F*(1, 23) = 3.76, *p* = .06), Old/New x Channel Group (*F*(2, 54) = 0.15, *p* = .86), and Old/New x Hemisphere x Channel Group (*F*(2, 54) = 0.14, *p* = .88) interactions were not significant.

The peak latency of the N170 derived from the grand average waveforms was 168ms. The analysis of the mean amplitudes within the 20ms time window surrounding this peak revealed that the main effect of Old/New

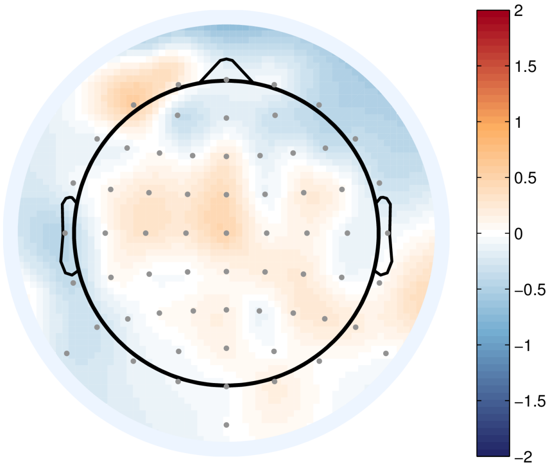


Figure 8.4: Topographic distribution of the ERP response difference to dog faces between the Old and the New conditions at the N170 latency in the Dog-Mix experiment. The ERP amplitudes were averaged within the 20ms time window surrounding the N170 peak. The averaged amplitudes in the New condition were subtracted by the averaged amplitudes in the Old condition. The amplitudes were not significantly different between the Old and the New conditions. The grey dots indicate the location of the 64 electrodes.

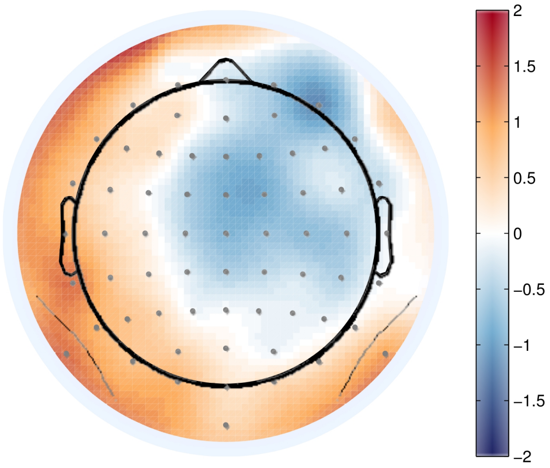


Figure 8.5: Topographic distribution of the ERP response difference to cars between the Old and the New conditions at the N170 latency in the Dog-Mix experiment. The ERP amplitudes were averaged within the 20ms time window surrounding the N170 peak. The averaged amplitudes in the New condition were subtracted by the averaged amplitudes in the Old condition. The amplitudes in P9/P10 channels were significantly different between the Old and the New conditions. The grey dots indicate the location of the 64 electrodes. The contours indicate the significant area.

(*F*(1, 23) = 0.04, *p* = .84), the Old/New x Hemisphere (*F*(1, 23) = 1.14, *p* = .30), Old/New x Channel Group (*F*(2, 54) = 0.07, *p* = .93), and Old/New x Hemisphere x Channel Group (*F*(2, 54) = 0.18, *p* = .84) interactions were not significant.

8.2.3.2 Cars

For the N170 half area latency, the main effect of Old/New (*F*(1, 23) = 0.65, *p* = .43), the Old/New x Hemisphere (*F*(1, 23) = 1.01, *p* = .33), Old/New x Channel Group (*F*(2, 54) = 2.14, *p* = .13), and Old/New x Hemisphere x Channel Group (*F*(2, 54) = 0.03, *p* = .97) interactions were not significant.

The peak latency of the N170 derived from the grand average waveforms was 167ms. The analysis of the mean amplitudes within the 20ms time window surrounding this peak revealed a marginally significant main effect of Old/New (*F*(1, 23) = 4.14, *p* = .05). Although the Old/New x Hemisphere (*F*(1, 23) = 2.04, *p* = .17) and Old/New x Hemisphere x Channel Group (*F*(2, 54) = 0.52, *p* = .60) interactions were not significant, the Old/New x Channel Group interaction (*F*(2, 46) = 3.63, *p* = .03) was significant. Cars in the Old condition elicited a significantly larger N170 than cars in the New condition over the lateral channel group (i.e., P9/P10: *F*(1, 23) = 7.32, *p* = .01), but not the other two channel groups (P8/P7: *F*(1, 23) = 3.06, *p* = .09; PO8/PO7: *F*(1, 23) = 1.23, *p* = .28).

8.2.3.3 Human Faces

For the N170 half area latency, the main effect of Old/New (*F*(1, 23)

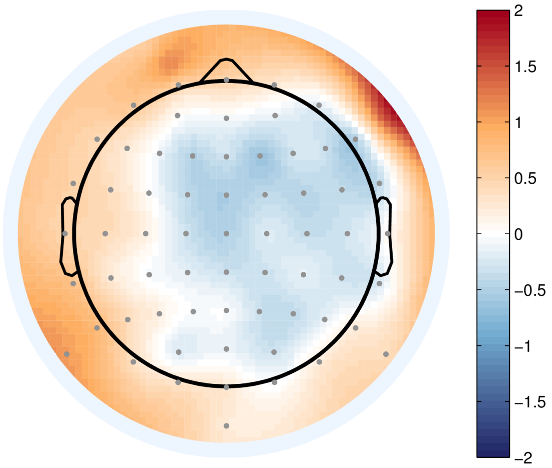


Figure 8.6: Topographic distribution of the ERP response difference to human faces between the Old and the New conditions at the N170 latency in the Dog-Mix experiment. The ERP amplitudes were averaged within the 20ms time window surrounding the N170 peak. The averaged amplitudes in the New condition were subtracted by the averaged amplitudes in the Old condition. The amplitudes were not significantly different between the Old and the New conditions. The grey dots indicate the location of the 64 electrodes.

= .04, *p* = .84), the Old/New x Hemisphere (*F*(1, 23) = 1.14, *p* = .30), Old/New x Channel Group (*F*(2, 54) = 0.48, *p* = .62), and Old/New x Hemisphere x Channel Group (*F*(2, 54) = 0.38, *p* = .69) interactions were not significant.

The peak latency of the N170 derived from the grand average waveforms was 153ms. Analysis of the mean amplitudes within the 20ms time window surrounding this peak revealed that the main effect of Old/New (*F*(1, 23) = 1.44, *p* = .24), the Old/New x Hemisphere (*F*(1, 23) = 0.60, *p* = .44), Old/New x Channel Group (*F*(2, 54) = 1.18, *p* = .32), and Old/New x Hemisphere x Channel Group (*F*(2, 54) = 1.12, *p* = .33) interactions were not significant.

8.2.4 P2

8.2.4.1 Dog Faces

For the P2 half area latency, the main effect of Old/New *F*(1, 23) = 1.46, *p* = .24), the Old/New x Hemisphere (*F*(1, 23) = 0.02, *p* = .89), and Old/New x Hemisphere x Channel Group (*F*(2, 54) = 0.12, *p* = .89) interactions were not significant. However, the Old/New x Channel Group interaction (*F*(2, 46) = 3.97, *p* = .03) was significant. The Old and the New conditions were different only for the PO8/PO7 channel group (PO8/PO7: *F*(1, 23) = 4.67, *p* = .04; P8/P7: *F*(1, 23) = 1.63, *p* = .21; P10/P9: *F*(1, 23) = 0.53, *p* = .47).

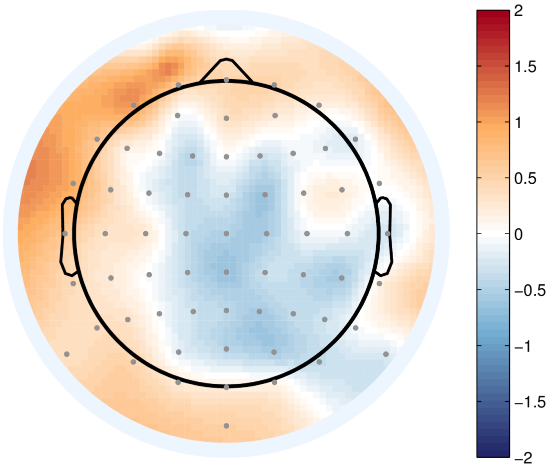


Figure 8.7: Topographic distribution of the ERP response difference to dog faces between the Old and the New conditions at the P2 latency in the Dog-Mix experiment. The ERP amplitudes were averaged within the 20ms time window surrounding the P2 peak. The averaged amplitudes in the New condition were subtracted by the averaged amplitudes in the Old condition. The amplitudes were not significantly different between the Old and the New conditions. The grey dots indicate the location of the 64 electrodes.

The peak latency of the P2 derived from the grand average waveforms was 237ms. Analysis of the mean amplitudes within the 20ms time window surrounding this peak indicated that the main effect of Old/New (*F*(1, 23) = 0.04, *p* = .84), the Old/New x Hemisphere (*F*(1, 23) = 2.22, *p* = .15), Old/New x Channel Group (*F*(2, 54) = 0.07, *p* = .93), and Old/New x Hemisphere x Channel Group (*F*(2, 54) = 0.53, *p* = .59) interactions were not significant.

8.2.4.2 Cars

For the P2 half area latency, the main effect of Old/New (*F*(1, 23) = 0.61, *p* = .44), the Old/New x Hemisphere (*F*(1, 23) = 0.10, *p* = .75), and Old/New x Hemisphere x Channel Group (*F*(2, 54) = 0.12, *p* = .89) interactions were not significant. However, the Old/New x Channel Group interaction (*F*(2, 46) = 3.97, *p* = .03) was significant. The Old and the New conditions were different only for the PO8/PO7 channel group (PO8/PO7: *F*(1, 23) = 4.67, *p* = .04; P8/P7: *F*(1, 23) = 1.63, *p* = .21; P10/P9: *F*(1, 23) = 0.53, *p* = .47).

The peak latency of the P2 derived from the grand average waveforms was 234ms. Analysis of the mean amplitudes within the 20ms time window surrounding this peak indicated that the main effect of Old/New (*F*(1, 23) = 3.36, *p* = .08), the Old/New x Hemisphere (*F*(1, 23) = 1.88, *p* = .18), and Old/New x Hemisphere x Channel Group (*F*(2, 54) = 0.20, *p* = .82) interactions were not significant. However, the Old/New x Channel Group interaction (*F*(2, 46) = 3.75, *p* = .03) was significant. Cars in the Old condition elicited a significantly smaller P2 than cars in the new condition over the

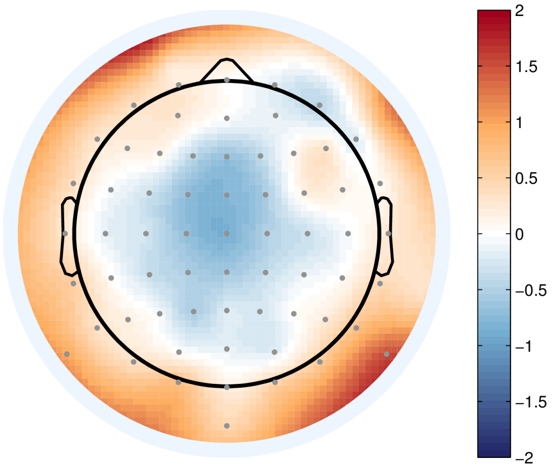


Figure 8.8: Topographic distribution of the ERP response difference to cars between the Old and the New conditions at the P2 latency in the Dog-Mix experiment. The ERP amplitudes were averaged within the 20ms time window surrounding the P2 peak. The averaged amplitudes in the New condition were subtracted by the averaged amplitudes in the Old condition. The amplitudes in P9/P10 channels were significantly different between the Old and the New conditions. The grey dots indicate the location of the 64 electrodes.

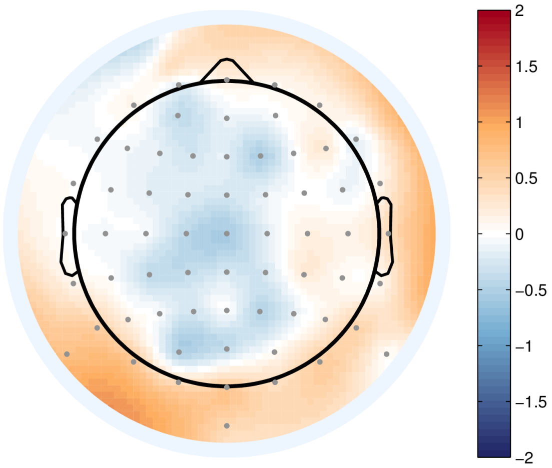


Figure 8.9: Topographic distribution of the ERP response difference to human faces between the Old and the New conditions at the P2 latency in the Dog-Mix experiment. The ERP amplitudes were averaged within the 20ms time window surrounding the P2 peak. The averaged amplitudes in the New condition were subtracted by the averaged amplitudes in the Old condition. The amplitudes were not significantly different between the Old and the New conditions. The grey dots indicate the location of the 64 electrodes.

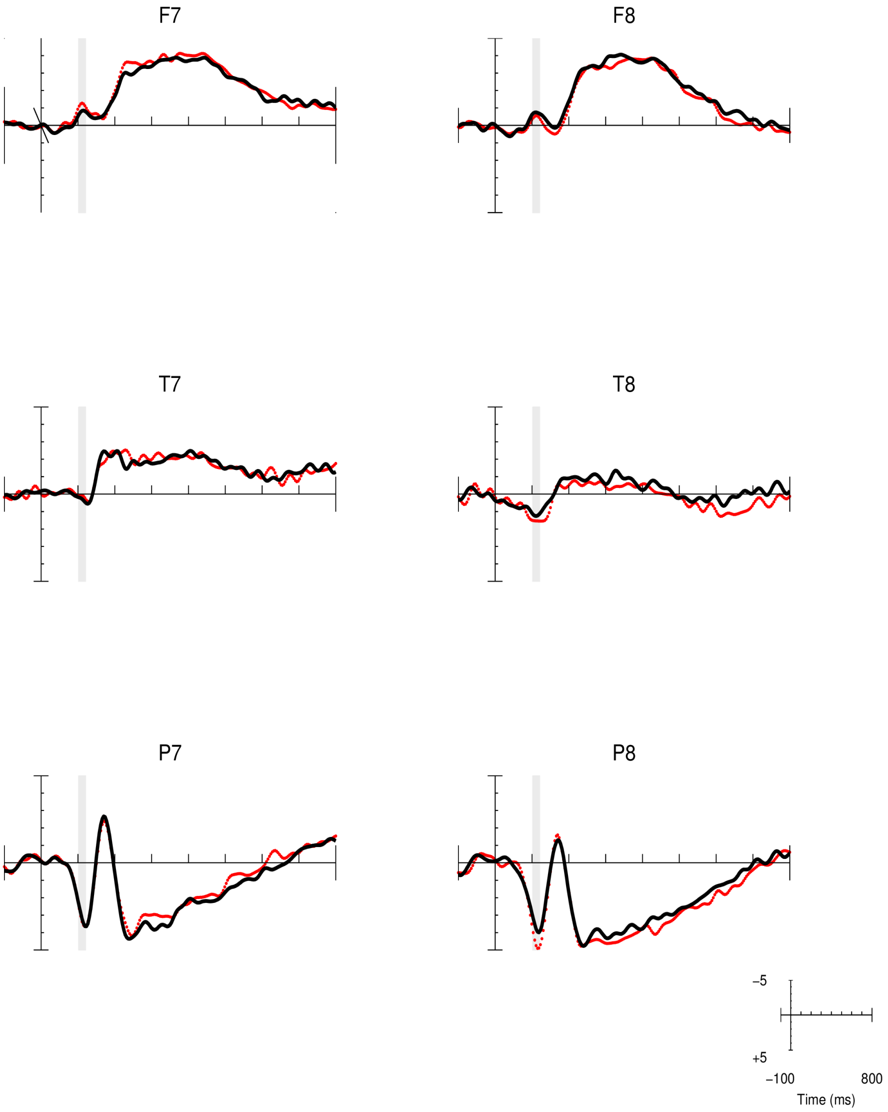


Figure 8.10: The ERP responses to dog faces in the Dog-Mix experiment. The red broken lines show the ERP in the Old condition. The black solid lines show the ERP in the New condition. The grey bars indicate the time window to compute the average P1 amplitudes.

lateral channel group (i.e., P9/P10: *F*(1, 23) = 6.32, *p* = .02), but not the other two channel groups (P8/P7: *F*(1, 23) = 1.98, *p* = .17; PO8/PO7: *F*(1, 23) = 1.05, *p* = .32).

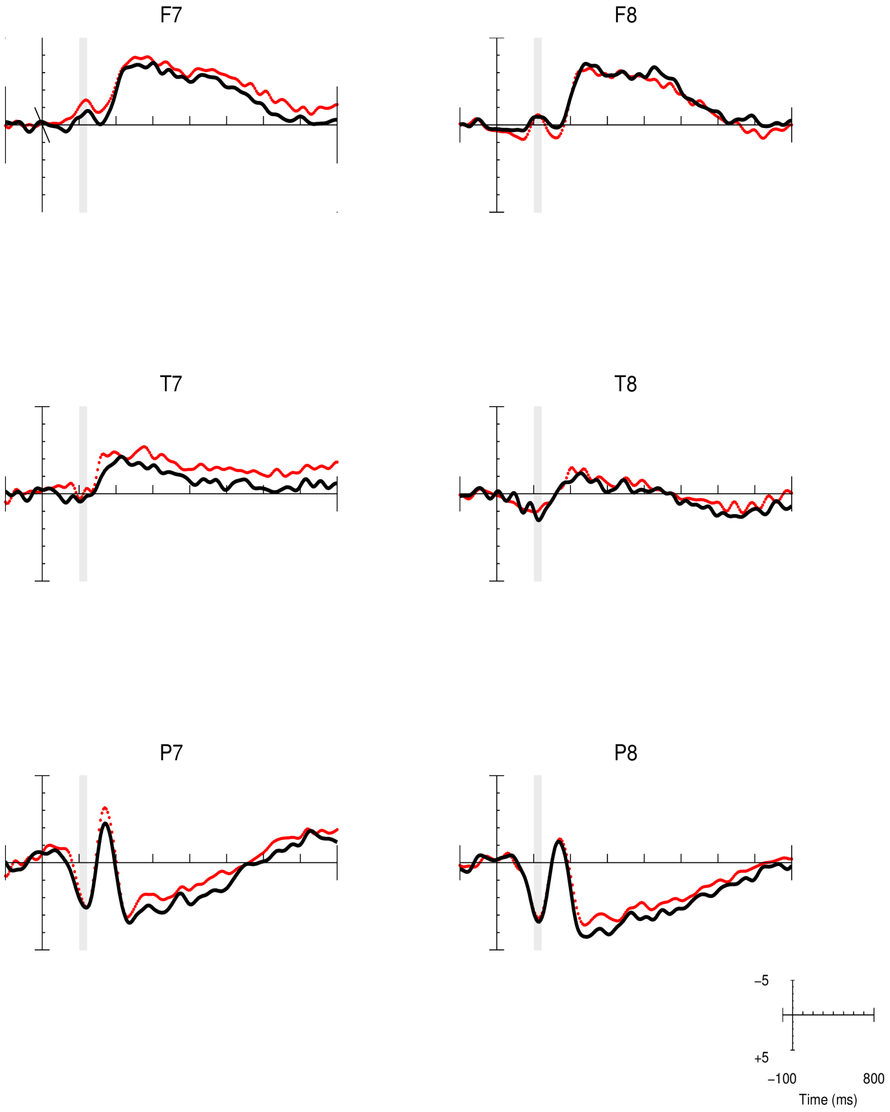


Figure 8.11: The ERP responses to cars in the Dog-Mix experiment. The red dotted lines show the ERP in the Old condition. The black solid lines show the ERP in the New condition. The grey bars indicate the time window to compute the average P1 amplitudes.

8.2.4.3 Human Faces

For the P2 half area latency, the main effect of Old/New (*F*(1, 23) = 0.03, *p* = .86), the Old/New x Hemisphere (*F*(1, 23) = 1.71, *p* = .20), Old/New x Channel Group (*F*(2, 46) = 1.08, *p* = .35), and Old/New x Hemisphere x

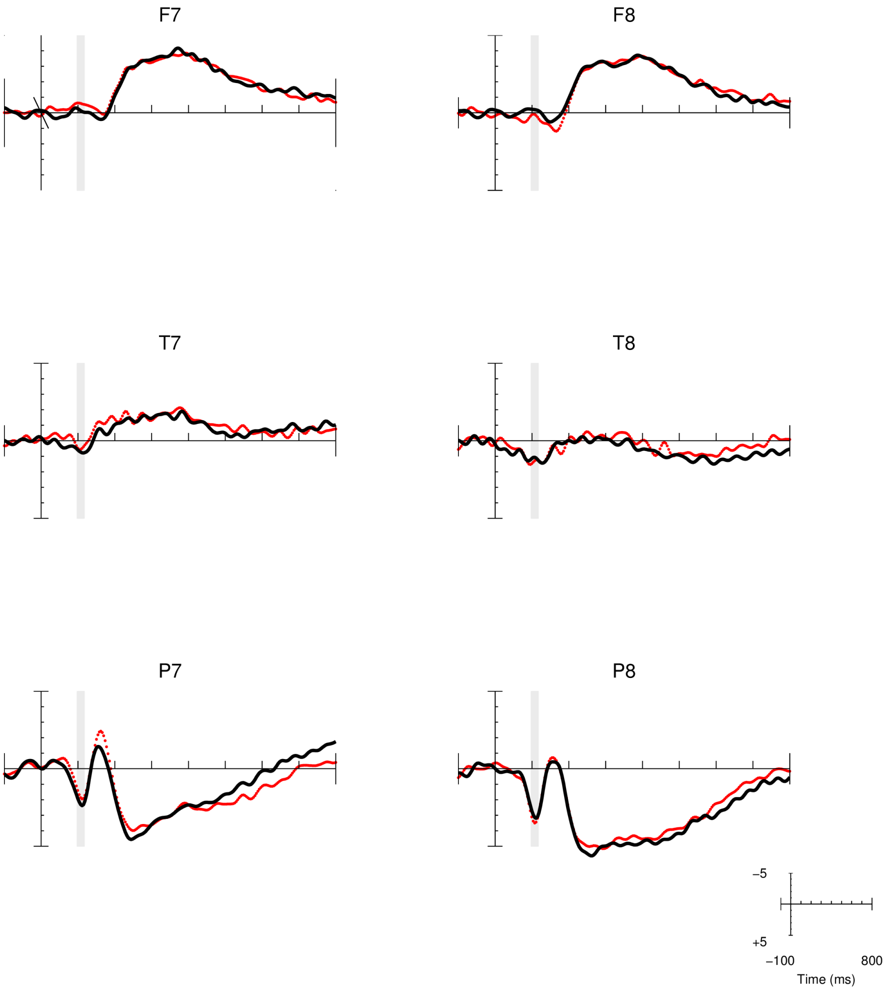


Figure 8.12: The ERP responses to human faces in the Dog-Mix experiment. The red dotted lines show the ERP in the Old condition. The black solid lines show the ERP in the New condition. The grey bars indicate the time window to compute the average P1 amplitudes.

Channel Group (*F*(2, 54) = 0.50, *p* = .61) interactions were not significant.

The peak latency of the P2 derived from the grand average waveforms was 246ms. Analysis of the mean amplitudes within the 20ms time window surrounding this peak indicated that the main effect of Old/New (*F*(1, 23) = 2.11, *p* = .20), Old/New x Hemisphere (*F*(1, 23)= 0.59, *p* = .45), Old/New x

Channel Group (*F*(2, 54) = 0.18, *p* = .83), and Old/New x Hemisphere x Channel Group (*F*(2, 54) = 1.26, *p* = .29) interactions were not significant.

8.3 Discussion

The Dog-Mix experiment compared responses to visual stimuli following individually identifiable barks with those following unknown barks. There was no significant effect from the preceding barks on the P1 component for any category of the visual stimuli. For the N170, responses to cars, but not dog faces or human faces, were modulated by the preceding sounds. Cars in the Old condition elicited a larger N170 than cars in the New condition.

This experiment failed to find similar effects on the P1 component as the Dog-Dog, the Dog-Car, and the Dog-Human experiments did. This outcome suggests that in the earlier experiments anticipation of the basic-level category of the visual stimuli contributed to the effect of preceding sounds on the P1 component. Given the visual stimuli in this experiment could be dog faces, cars, or human faces, participants were less likely to know the basic-level category of the visual stimulus before stimulus onset. Hence, to judge whether the auditory and the visual stimuli belonged to the same basic-level category, and/or shared a categorization level, processing of the visual stimuli had to reach the conceptual categorization stage. Though superordinate-level categorization can occur as early as 70ms, conceptual categorization of visual stimuli was more often observed in the N170 time window. While information about the basic-level categories of the visual stimuli might advance the effects from the preceding sounds to the P1 time range, difficulty in predicting the basic-level category of the visual stimulus could eliminate the early effects.

Although the N170 had not been affected in the previous five experiments, here larger N170 amplitudes were obtained in the Old condition than the New condition, but only for the car stimuli.. It is possible that the effect was related to anticipation of the visual stimulus category. Visual stimuli in this experiment could be from one of three basic-level categories, among which only dog faces were from the same basic-level category as the auditory stimuli (i.e., barks). It was more likely to encounter a visual stimulus from a different basic-level category (i.e., Car or Human) than one from the same basic-level category (i.e., Dog). Furthermore, by default, unknown barks, dog faces, and cars are categorized at the basic level, while human faces are categorized at the subordinate level. It was also more likely to see a visual stimulus at the basic level than at the subordinate level. Therefore, after hearing an unknown bark, participants were more likely to see a visual stimulus at a congruent categorization level (i.e., cars or dog faces) than at an incongruent level (i.e., human faces). Previous research showed that category expectation modulated the N170 amplitudes (Aranda, Madrid, Tudela, & Ruz, 2010), such that valid prediction of the visual category increased the N170 amplitude for words, but decreased the amplitude for human faces. Enhanced N170 amplitude for cars in our experiment is consistent with the category expectation effect for words in their study.

Experiment	**Visual P1**	**N170**	**Visual P2**
Dog-Dog	right, old > new	ns	ns
Dog-Car	left, old < new	ns	ns
Dog-Human	anterior, new > old; posterior, old > new	ns	ns
Human-Dog	posterior right, old < new	ns	ns
Human-Human	ns	ns	ns
Dog-Mix (Dog)	ns	ns	ns
Dog-Mix (Car)	ns	P9/P10, old>new	PO7/PO8, new>old
Dog-Mix (Human)	ns	ns	ns

Table 8.2: Summary of the results in all six experiments

Chapter 9 General Discussion

This study examined the effect of conceptual categorization on visual processing in a cross-modal context. Participants heard auditory stimuli that were either new or which they had previously learned to associate with the name of a dog or a person. Each sound was followed by a visual stimulus, with a constant interval between sound offset and visual stimulus onset. Potentially, the learned sounds could be categorized at a more individual categorization level than the new sounds. The default categorization level of the visual stimuli varied across experiments: Dog faces and cars are usually categorized at the basic level, while human faces are categorized at the subordinate level. We expected the categorization level of the sound to modulate the ERP components elicited by the subsequent visual stimulus (e.g., the P1 component).

9.1 Cross-modal Priming and Visual Processing

The main findings in this study were that the P1 responses to the visual stimuli were differently modulated in different experiments. In the Dog-Dog experiment and the Human-Human experiment, there was no effect of preceding sounds on the P1 component. In the Dog-Human experiment, the P1 amplitudes were larger in response to the human faces following learned dog barks than those following the new barks. In both the Dog-Car experiment and the Human-Dog experiment, the opposite pattern was observed. In the Dog-

Car experiment, the P1 amplitudes were smaller in response to the cars following the learned dog barks than those following the new barks. In the Human-Dog experiment, the P1 amplitudes were smaller in response to the dog faces following the learned human voices than those following the new voices. In the Dog-Mix experiment, no effect from the preceding sounds was found on the P1 components for any visual stimuli.

9.2 P1 Modulation as a Function of Categorization-Level Congruency and Basic-Level Category

The results can be understood in light of the categorization levels and the basic-level categories of the auditory and visual stimuli. Among the visual stimuli, dog faces and cars are categorized at the basic level by default, while human faces are categorized at the subordinate level by default. Regarding the auditory stimuli, dog barks that had not been associated with names would be categorized at the basic level by default, while human voices are categorized at the subordinate level. However, participants in different experiments learned to associate some dog barks and human voices with names, so those dog barks and the human voices could be identified individually leading to a change in the categorization level: the old auditory stimuli – whether dog barks or human voices – would now be categorized at a more individual level compared to the new sounds. This helps to account for the differential effects of the auditory stimuli on the P1 amplitudes in this study.

We argued that the P1 responses were modulated primarily by two factors: 1) the congruency between the basic levels of the auditory and the visual stimuli and 2) the congruency between the categorization level of the auditory stimulus and the default categorization level of the visual stimulus. The former determined whether the P1 components were modulated, while the latter influenced the direction of P1 amplitude change. This was particularly true when visual stimuli from only one basic-level category followed the auditory stimuli (e.g., in the Dog-Human experiment, all visual stimuli came from one basic-level category, 'Human'). These results pertain only to the first five experiments. Results from the last experiment, where the visual stimuli came from three basic-level categories, will be discussed later.

In the first five experiments, the learned auditory stimuli did not affect the P1 amplitude when both the auditory and the visual stimuli were from the same basic-level category. This was true regardless of the default categorization level of the visual stimuli. In the Dog-Dog experiment, dog faces were categorized at the basic level. Though the auditory and the visual stimuli were from different dogs, they both belonged to the same basic-level category, 'Dog'. The P1 responses to the dog faces were not affected by the categorization level of the preceding barks. In the Human-Human experiment, human faces were categorized at the subordinate level and the P1 responses to the human faces also were not affected by the categorization level of the preceding voices.

When the auditory and the visual stimuli were from different basic-level categories, the visual stimuli that were more congruent in categorization level with the auditory stimuli elicited larger P1 amplitudes. In the Dog-Car experiment, cars were categorized at the basic level. The new barks were also categorized at the basic level, but the old barks were individually identifiable and could be categorized at the subordinate level. Thus, the cars were more congruent in terms of categorization level with the new barks than the old barks. Therefore, the P1 responses to the cars preceded by the new barks were larger than those preceded by the old barks. Similarly, in the Dog-Human experiment, because human faces are normally categorized at the subordinate level, so the human faces were more congruent categorically with the old barks than with the new barks. Therefore, the P1 responses to the human faces preceded by the old barks were larger than those preceded by the new barks. Finally, in the Human-Dog experiment, the dog faces were categorized at the basic level, the old voices, like the old barks, were individually identifiable and categorized at the subordinate level, while the new voices were less individualized. Therefore, the dog faces were relatively more congruent with the new voices than with the old voices and the P1 responses to the dog faces preceded by the new voices were larger than those preceded by the old voices. To conclude, the results from the three experiments indicate that the P1 amplitudes were larger when the categorization levels of the auditory and the visual stimuli were congruent.

This effect depends only on the whether the auditory and the visual stimuli belong to different basic-level categories. It does not matter whether they belong to different superordinate-level categories. For instance, in the Dog-Car experiment, the dogs and the cars belonged to different categories at both the basic level (i.e., dog vs. car) and the superordinate level (i.e., animal vs. vehicle). In the Dog-Human experiment, the dogs and the humans belonged to the different basic-level categories (i.e., dog vs. human), but they were in the same superordinate-level category (i.e., animal). However, in both experiments, categorization level congruency elicited larger P1 amplitudes.

9.3 Sensory Processing Modulation as a Result of Cross-modal Semantic Congruency

But why were the P1 components modulated? Rogers and Patterson (2007) proposed a parallel distributed model for object representations. Following the view that object representations are distributed across different regions in the brain, they argued that the representations are stored in both the modality-specific regions and the amodal regions. While regions in sensory cortex store modality-specific features, the concepts that link the features are stored in the anterior temporal cortex and are represented in distributed patterns according to the similarities among features. Based on this view, we proceeded as follows to predict the effect of activated object representations on sensory processing of upcoming stimuli. Object representations are often multi-modal. A dog has four legs, can bark, and is also covered with fur. Representations activated by features from one modality can spread activity to other sensory modalities. Processing of the subsequent stimulus in those sensory modalities is expected to be affected. More specifically, we hypothesized that a subsequent stimulus that is congruent with the activated representation will attract more sensory processing resources. In the visual modality, it may lead to larger P1 amplitude.

But this only tells half of the story. The subsequent stimulus can be congruent with the activated representation in different ways, such as categories and abstract levels. In some experiments, we observed larger P1 amplitudes when the categorization levels were congruent than when they were incongruent. These were experiments when the visual stimuli were from a different basic-level category compared to the auditory stimuli. But we did not find significant P1 differences in the Human-Human experiment. In the Dog-Dog experiment, dog faces after learnt barks elicited larger activities over the right hemisphere, though the categorization level of the dog faces was more congruent with the new barks. In both experiments, the visual and the auditory stimuli were from the same basic-level category. It is possible that when both visual and auditory stimuli are from the same basic-level category, multisensory integration is involved to form a unified representation. It is commonly observed that multisensory stimuli elicit larger responses than the sum of unisensory responses. Evidence from intracranial recordings in human subjects (Calvert, 2001) suggests that auditory stimuli can modulate response phase in the visual cortex. During multisensory integration, higher phase concentration was observed compared to the expected sum of unisensory responses. It is possible that a more informative auditory stimulus enriches subsequent visual processing, especially for stimuli that were not at the individual processing level.

As reviewed in the Introduction, the P1 component has also been linked to attention. A second possibility is that the P1 component was modulated by attention. Though early research often used a spatial cueing paradigm, the P1 component is also affected by non-spatial attention (Taylor, 2002). Larger amplitudes are generally observed when the stimulus is attended to as compared to unattended. This is true even when no stimulus is actually presented (Mangun et al., 2001) in that simply attending to the visual field elicits a larger P1.

But why was attention to the visual stimuli modulated? In our study, the modulation of the P1 components was related to cross-modal congruency between the auditory and the visual stimuli. This cross-modal congruency was semantic rather than perceptual. Though cross-modality interaction does not have to rely on top-down regulation (e.g., spatial or temporal relationship of inputs from different modalities), semantic congruency between the auditory and the visual stimuli can affect cross-modal interaction (Hein et. al., 2007). Semantic congruency can facilitate processing of the audiovisual stimuli that leads to faster reaction times and higher accuracy (e.g., Calvert, 2005; Laurienti et. al., 2004). In addition, ERP responses to multimodal stimuli are larger than the summation of the responses to unimodal stimuli (Talsma, Doty, & Woldorff, 2007). However, this enhancement seems to require attention to the sensory input from at least one of the modalities of the multimodal stimuli (van de Burg, et. al., 2010). Van de Burg and colleagues (2010) presented multimodal stimuli followed by visual targets in an attentional blink paradigm. The multimodal stimuli were audiovisual letters that could either be congruent (e.g., hearing F and seeing F) or incongruent (e.g., hearing F and seeing Z). In a series of experiments, they required participants to either attend to the visual target alone, to one sensory channel of the audiovisual stimulus and the visual target, or to both sensory channels of the audiovisual stimulus and the target. Across all experiments, attentional blink was observed. However, the semantic congruency effect on the attentional blink at the target was observed when at least one sensory channel of the multimodal stimulus was attended. When the multisensory stimulus was completely task irrelevant, no semantic congruency effect was found.

In contrast, if no attention is devoted to the multimodal stimulus, the response can be smaller than the sum of the responses to the unimodal stimuli (Talsma et al., 2007). This is true even when attention is devoted to a sensory channel, but not the stimulus location. For example, Talsma and colleagues (2007) presented participants with two streams of visual stimuli at the upper and the lower visual field. The stimuli in the lower visual field were either accompanied by auditory stimuli or not and the auditory stimuli could also occur in the absence of a lower visual field stimulus. The participant's task was to detect the occasionally presented targets (i.e., a number), which could appear in the upper or the lower streams depending on the condition. When the lower stream was attended, responses to the multimodal stimuli were larger than the summation of the responses to the unimodal stimuli. However, when the upper stream was attended instead, responses to the multimodal stimuli were smaller than the summation of the unimodal stimuli in the lower visual field. Note that the reverse ERP response pattern between the attended and the unattended conditions occurred before the P1 component in their study. It could possibly be due to the simple multimodal stimuli used in their experiments (i.e., white square horizontal gratings in the visual modality and pips in the auditory modality). Moreover, the P1 amplitudes in the experiment were larger in the attended condition compared to the unattended one.

Considering the results from both studies (Talsma, et al., 2007; van de Burg, et al., 2010), it appears that semantic congruency across audiovisual modalities can facilitate performance, but to benefit from this interaction, one has to attend to the sensory input in at least one of the modalities of the multimodal stimulus (i.e., the auditory input, the visual input, or both). Attention is necessary to bring about the semantic congruency effect. These conclusions can apply to our study. First, the high accuracy of the behavioural results on the visual stimuli suggested that participants did pay attention to the task (and therefore to the visual stimuli). This allowed cross-modal interaction at the semantic level. Our experiments showed that congruency of categorization level can prompt increased attention to the stimuli. Although the auditory and the visual stimuli belong to different basic-level categories, congruent categorization levels may act to facilitate association between different stimuli by increasing attention to at least one of them. On the other hand, the effect of the category-level congruency does not seem to increase attention to the visual stimuli when the stimuli from both modalities belong to the same basic-level modality. This resulted in no category-level congruency effect on the P1 components in the Dog-Dog experiment and the Human-Human experiment. It is possible that belonging to the same basic-level category attracted enough attention to the stimuli, and category congruency did not further increase attention.

Compared to the cross-modal binding by synchronized presentation, semantic congruency was considered as a top-down effect. In our experiments, this top-down effect seemed to occur at quite an early stage of the visual stimulus processing (i.e., in the time range of the P1), which may seem surprising. However, it has been shown that the P1 component is sensitive to top-down modulation. Furthermore, it has been shown that, while a 100ms SOA (stimulus onset asynchrony) was too short, 300ms is a sufficient SOA for a semantic congruency effect to occur (van de Burg, et al., 2010). Compared to simultaneous presentation of the auditory and the visual stimuli, the priming paradigm used in our study allowed more than 100ms for the semantic processing to occur. The visual stimuli were presented 217ms after sound offset and the stimulus onset asynchrony (SOA) was more than 617ms in all six experiments. Moreover, in the first five experiments, the basic-level categories of the visual stimuli were known to the participants before visual stimulus presentation, which may have further facilitated the top-down process. As soon as the visual stimulus was presented, attention was readily brought to it.

Note that there was a significant difference between the Old and the New conditions over the right hemisphere in the Dog-Dog experiment. The amplitudes in the old condition were more positive than those in the new condition. The difference occurred within the P1 range, but the topographic distribution showed that the difference peaked over the temporal electrode sites. This is different from the P1 distribution, which spreads over the occipito-temporal electrode sites. It might reflect the process of binding barks with the dog faces, given that it was mostly above the temporal lobe, but the neural sources of EEG signals can be quite far away from the peak responses observed over the scalp due to volume conduction (Luck, 2005). The design of the current study does not allow verification of this hypothesis.

9.5 N170 Component

Linking a dog bark or a human voice to a dog or a person by name allows the dog or the person to be individually recognized. However, in the first five experiments, this individualization effect did not affect the N170 elicited by the visual stimuli. Previous studies demonstrated that stimuli that were categorized at a more individual abstract level elicited a larger N170 (e.g., Tanakan & Curran, 2001). Our results suggest that the effect of training participants to individually recognize the auditory stimuli does not immediately carry over to change the categorization level of the visual stimuli to a more individual level. In the Dog-Mix experiment, a difference between the old and the new conditions was observed within the N170 time window for car stimuli. These results may not be surprising, however, because all the visual stimuli were presented only once, and this means that there was only one opportunity for the association between the auditory and the visual stimuli to be established. Consider that participants had two-day learning sessions and far more exposure to the sounds before they could individually recognize most of the sounds in the old condition and that it requires far more learning before a person can become a domain expert, who can categorize objects of their expertise at the subordinate level.

However, these N170 results should not lead us to hastily conclude that auditory stimuli do not influence the categorization of the visual stimuli because N170 modulation during categorization might be influenced by attention level. Heisz, Watter and Shatten (2006) reported that attention controlled whether repeating same-identity faces reduced the N170 amplitude. Viewing a sequence of same-identity faces progressively habituated the N170 only when the faces appeared at the unattended location. Neither viewing a sequence of different-identity faces at the unattended location nor seeing the same-identity faces appear at the attended location led to attenuation of the N170. Attention to the faces, though face identity was task-irrelevant, seems able to reduce the N170 habituation from repetition. Heisz and Shadden (2008) further examined the repetition effects of the faces on the N170 responses when the faces were associated with personal information, compared with control faces that were presented repeatedly with unrelated stories (e.g., rocks and trees). They found that the face repetition effect on the N170 was attenuated when the faces were associated with personal information. Therefore, as the N170 habituation is linked to neuronal response specificity (Korinth, Sommer, & Brenznitz, 2013), both attention and associating with personal information can help the N170 resist habituation over repetition.

9.6 The Dog-Mix Experiment

The final experiment deserves further comment because it was the least similar to the previous experiments in terms of design and results. There was no effect of the Old/New bark condition on the P1 responses to any of the three categories visual stimuli. One possible explanation is that the link between the prime and the target changed in that the dog barks could be followed by visual stimuli from any of the three basic-level categories. The visual stimuli could be from the same (i.e., dog faces) or different (i.e., cars and human faces) basic-level category as the barks. Moreover, the categorization level of the visual stimuli and the preceding auditory stimuli could be congruent (e.g., cars following the new barks, with both stimuli categorized at the basic level) or incongruent (e.g., human faces, categorized at the subordinate level, following the new barks). The systematic relationships between stimuli in the previous experiment thus cannot be easily established in this experiment.

It is worth pointing out that the tasks in all experiments did not require participants to attend to the auditory stimuli. Research has shown that semantic congruency only affect cross-modal interaction when input from at least one of the modalities was attended (van de Burg, et. al., 2010). Though participants may still attend to the barks or the voices in our study, task instructions to attend to both modalities do not seem to be necessary. Neither did our tasks explicitly require conceptual categorization on the part of participants. Nevertheless, the results of our study suggest that both basic-level categories and abstract levels were still compared implicitly.

Chapter 10 Summary

We live in a world that constantly requires conceptual categorization. Concepts help us understand the perceived information from the environment and to generate predictions. Instead of dealing with novel objects all the time, we assign objects to pre-existing categories and assimilate them into a hierarchy of knowledge. Thus, we use our top-down knowledge to associate or differentiate objects, rather than solely rely on bottom-up information.

Concepts are represented both within and beyond sensory cortex (Rogers and Patterson, 2007). Neural activity specific to each modality spreads across brain regions and interacts to form a coherent multimodal representation of the new concept. By the same token, the existing representation is able to interact with inputs from other modalities. Moreover, conceptual representations are hierarchical such that members in the same hierarchy that share an essential set of intra- or inter-modal features are said to form a category. Research shows that conceptual categorization based on visual information can affect visual processing, which is an intramodal effect (e.g., Tanaka and Curran, 2001). The current study set out to examine the effect of conceptual categorization on visual processing in cross-modal context (i.e., by preceding visual stimuli with auditory stimuli that varied in their categorical similarity with the visual counterparts).

More specifically, we measured how familiar and individually-identifiable auditory stimuli (e.g., barks from known or unknown dogs) may affect responses to visual stimuli that briefly followed and were from various basic-level categories. We found that the early responses to the visual stimuli were modulated by the experimental manipulations. Two factors are particularly important for these effects to occur: 1) whether the auditory and the visual stimuli belong to the same basic-level categories, and if not, 2) the audiovisual congruency of the categorization levels. Early visual responses were not further increased if the auditory and the visual stimuli belonged to the same basic-level category, but they were enhanced if the categorization levels were congruent between the auditory and the visual stimuli from different basic-level categories. It is possible that attention was also modulated.

Studies examining semantic association often emphasize category congruency among different stimuli. The current study points to the importance of the interplay between the level of categorization and the congruency of the basic-level category in affecting visual object processing. The current study also provides knowledge about how multimodal conceptual categorization affects unimodal perceptual processing by putting it in the context of cross-modal processing. Previous research on categorization levels often involved only one modality, especially the visual modality. As concepts involve information from different modalities, we also need to understand how categorization levels function across modalities. Moreover, tasks in this study are perceptual and required no attention to the auditory modality. Results from this study contribute to our understanding of top-down processes in cross-modal interaction.

References

Adams, R. B., & Janata, P. (2002). A Comparison of Neural Circuits Underlying Auditory and Visual Object Categorization. *NeuroImage*, *16*(2), 36 1–377.

Aranda, C., Madrid, E., Tudela, P., & Ruz, M. (2010). Category expectations: A differential modulation of the N170 potential for faces and words. *Neuropsychologia*, *48*(14), 4038–4045.

Barraclough, N. E., Xiao, D., Baker, C. I., Oram, M. W., & Perrett, D. I. (200 5). Integration of Visual and Auditory Information by Superior Tempor al Sulcus Neurons Responsive to the Sight of Actions. *Journal of Cognitive Neuroscience*, *17*(3), 377–391.

Barsalou, L. W. (1999). Perceptual symbol systems. *The Behavioral and brain sciences*, *22*(4), 577–609; discussion 610–660.

Barton, J. J. S., Press, D. Z., Keenan, J. P., & O’Connor, M. (2002). Lesions of the fusiform face area impair perception of facial configuration in prosopagnosia. *Neurology*, *58*(1), 71–78.

Belin, P., Fecteau, S., & Bédard, C. (2004). Thinking the voice: neural correlates of voice perception. *Trends in Cognitive Sciences*, *8*(3), 129–135.

Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological Studies of Face Perception in Humans. *Journal of cognitive neuroscience*, *8*(6), 551–565.

Bentin, S., Taylor, M. J., Rousselet, G. A., Itier, R. J., Caldara, R., Schyns, P. G., ... Rossion, B. (2007). Controlling interstimulus perceptual varianc e does not abolish N170 face sensitivity. *Nature Neuroscience*, *10*(7), 801–802.

Busse, L., Roberts, K. C., Crist, R. E., Weissman, D. H., & Woldorff, M. G. (2005). The spread of attention across modalities and space in a multisensory object. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(51), 18751–18756.

Calvert, G. A. (2001). Crossmodal Processing in the Human Brain: Insights from Functional Neuroimaging Studies. *Cerebral Cortex*, *11*(12), 1110–1123. doi:10.1093/cercor/11.12.1110

Calvert, G. A., & Thesen, T. (2004). Multisensory integration: methodological approaches and emerging principles in the human brain. *Journal of Physiology-Paris*, *98*(1–3), 191–205.

Campanella, S., & Belin, P. (2007). Integrating face and voice in person perception. *Trends in Cognitive Sciences*, *11*(12), 535–543.

Carmel, D., & Bentin, S. (2002). Domain specificity versus expertise: factors influencing distinct processing of faces. *Cognition*, *83*(1), 1–29.

Cauquil, A. S., Edmonds, G. E., & Taylor, M. J. (2000). Is the face-sensitive N170 the only ERP not affected by selective attention? *Neuroreport*, *11*(10), 2167–2171.

Chen, Y.-C., & Spence, C. (2011). The crossmodal facilitation of visual object representations by sound: Evidence from the backward masking paradigm. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(6), 1784–1802.

de Haan, M., Pascalis, O., & Johnson, M. H. (2002). Specialization of Neural

Mechanisms Underlying Face Recognition in Human Infants. *Journal of Cognitive Neuroscience*, *14*(2), 199–209.

Donohue, S. E., Roberts, K. C., Grent-‘t-Jong, T., & Woldorff, M. G. (2011). The Cross-Modal Spread of Attention Reveals Differential Constraints for the Temporal and Spatial Linking of Visual and Auditory Stimulus Events. *The Journal of Neuroscience*, *31*(22), 7982–7990.

Eimer, M. (1998). Mechanisms of Visuospatial Attention: Evidence from Event-related Brain Potentials. *Visual Cognition*, *5*(1-2), 257–286.

Eimer, M. (2000a). Attentional Modulations of Event-Related Brain Potentials Sensitive to Faces. *Cognitive Neuropsychology*, *17*(1-3), 103–116.

Eimer, M. (2000b). The face-specific N170 component reflects late stages in the structural encoding of faces. *Neuroreport*, *11*(10), 2319–2324.

Eimer, M. (2011). The Face-Sensitivity of the N170 Component. *Frontiers in Human Neuroscience*, *5*.

Ewbank, M. P., Smith, W. A. P., Hancock, E. R., & Andrews, T. J. (2008). The M170 Reflects a Viewpoint-Dependent Representation for Both Famili ar and Unfamiliar Faces. *Cerebral Cortex*, *18*(2), 364–370.

Francis, W.S., Corral, N.I., Jones, M.L., & Sáenz, S.P. (2008). Decomposition of repetition priming components in picture naming. *Journal of experi mental psychology. General, 137*(3), 566-590. doi:10.1037/0096-3445. 137.3.566

Freunberger, R., Höller, Y., Griesmayr, B., Gruber, W., Sauseng, P., & Klimesch, W. (2008a). Functional similarities between the P1 component and alpha oscillations. *European Journal of Neuroscience*, *27*(9), 2330–2340.

Freunberger, R., Klimesch, W., Griesmayr, B., Sauseng, P., & Gruber, W. (2008b). Alpha phase coupling reflects object recognition. *NeuroImage*, *42*(2), 928–935.

Gajewski, P. D., & Stoerig, P. (2011). N170 – An Index of Categorical Face Perception? *Journal of Psychophysiology*, *25*(4), 174–179.

Goffaux, V., & Rossion, B. (2007). Face inversion disproportionately impairs the perception of vertical but not horizontal relations between features. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(4), 995–1002.

Greene, A. J., Easton, R. D., & LaShell, L. S. (2001). Visual-auditory events: c ross-modal perceptual priming and recognition memory. *Consciousness and cognition*, *10*(3), 425–435.

Gruber, T., & Müller, M. M. (2005). Oscillatory Brain Activity Dissociates bet ween Associative Stimulus Content in a Repetition Priming Task in the Human EEG. *Cerebral Cortex*, *15*(1), 109–116.

Guastavino, C. (2007). Categorization of environmental sounds. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *61*(1), 54–63.

Haig, N. D. (1984). The effect of feature displacement on face recognition. *Perception*, *13*(5), 505–512.

Hein, G., Doehrmann, O., Müller, N. G., Kaiser, J., Muckli, L., & Naumer, M. J. (2007). Object Familiarity and Semantic Congruency Modulate Responses in Cortical Audiovisual Integration Areas. *The Journal of*

Neuroscience, *27*(30), 7881–7887.

Heisz, J. J., Watter, S., & Shedden, J. M. (2006). Progressive N170 habituation to unattended repeated faces. *Vision Research*, *46*(1–2), 47–56.

Heisz, J. J., & Shedden, J. M. (2008). Semantic Learning Modifies Perceptual Face Processing. *Journal of Cognitive Neuroscience*, *21*(6), 1127–1134.

Herrmann, M. J., Ehlis, A.-C., Muehlberger, A., & Fallgatter, A. J. (2005). Source Localization of Early Stages of Face Processing. *Brain Topography*, *18*(2), 77–85.

Hillyard, S. A., Luck, S. J., & Mangun, G. R. (1994). The Cuing of Attention to Visual Field Locations: Analysis with ERP Recordings. In H.-J. Heinze, T. F. Münte, & G. R. Mangun (Eds.), *Cognitive Electrophysiol ogy* (pp. 1–25). Birkhäuser Boston.

Hoenig, K., Sim, E.-J., Bochev, V., Herrnberger, B., & Kiefer, M. (2008). Conceptual Flexibility in the Human Brain: Dynamic Recruitment of Semantic Maps from Visual, Motor, and Motion-related Areas. *Journal of Cognitive Neuroscience*, *20*(10), 1799–1814.

Itier, R. J., Latinus, M., & Taylor, M. J. (2006). Face, eye and object early proc essing: What is the face specificity? *NeuroImage*, *29*(2), 667–676. doi: 10.1016/j.neuroimage.2005.07.041

Itier, R. J., & Taylor, M. J. (2004a). N170 or N1? Spatiotemporal Differences between Object and Face Processing Using ERPs. *Cerebral Cortex*, *14* (2), 132–142.

Itier, R. J., & Taylor, M. J. (2004b). Effects of repetition learning on upright, inverted and contrast-reversed face processing using ERPs. *NeuroI-*

mage, *21*(4), 1518–1532.

Itier, R. J., Van Roon, P., & Alain, C. (2011). Species sensitivity of early face and eye processing. *NeuroImage*, *54*(1), 705–713.

Johannes, S., Münte, T. F., Heinze, H. J., & Mangun, G. R. (1995). Luminance and spatial attention effects on early visual processing. *Cognitive Brain Research*, *2*(3), 189–205.

Klimesch, W. (2011). Evoked alpha and early access to the knowledge system: The P1 inhibition timing hypothesis. *Brain Research*, *1408*, 52–71.

Korinth, S. P., Sommer, W., & Breznitz, Z. (2013). Neuronal response specificity as a marker of reading proficiency. *NeuroReport*, *24*(2), 96–100.

Laurienti, P., Kraft, R., Maldjian, J., Burdette, J., & Wallace, M. (2004). Sema ntic congruence is a critical factor in multisensory behavioral performa nce. *Experimental Brain Research*, *158*(4).

Lee, Y.-S. (2010). *Neural basis underlying auditory categorization in the human brain*. Dartmouth College, Hanover, New Hampshire.

Lefebvre, C. D., Marchand, Y., Eskes, G. A., & Connolly, J. F. (2005). Assessment of working memory abilities using an event-related brain potential (ERP)-compatible digit span backward task. *Clinical Neurophysiology*, *116*(7), 1665–1680.

Luck, S. J., & Hillyard, S. A. (1994). Electrophysiological correlates of feature analysis during visual search. *Psychophysiology*, *31*(3), 291–308.

Macchi Cassia, V., Kuefner, D., Westerlund, A., & Nelson, C. A. (2006). Modulation of Face-sensitive Event-related Potentials by Canonical and Distorted Human Faces: The Role of Vertical Symmetry and Up-Down

Featural Arrangement. *Journal of Cognitive Neuroscience*, *18*(8), 1343–1358.

Mangun, G. R., Buonocore, M. H., Girelli, M., & Jha, A. P. (1998). ERP and fMRI measures of visual spatial selective attention. *Human brain mapping*, *6*(5-6), 383–389.

Mangun, G. R., & Hillyard, S. A. (1991). Modulations of sensory-evoked brain potentials indicate changes in perceptual processing during visual-spatial priming. *Journal of experimental psychology. Human perception and performance*, *17*(4), 1057–1074.

Mangun, G. R., Hinrichs, H., Scholz, M., Mueller-Gaertner, H. W., Herzog, H., Krause, B. J., … Heinze, H. J. (2001). Integrating electrophysiology and neuroimaging of spatial selective attention to simple isolated visual stimuli. *Vision Research*, *41*(10–11), 1423–1435.

Martinovic, J., Gruber, T., & Müller, M. M. (2008). Coding of Visual Object Features and Feature Conjunctions in the Human Brain. *PLoS ONE*, *3*(11), e3781.

Martinovic, J., Gruber, T., & Müller, M. (2009). Priming of object categorization within and across levels of specificity. *Psihologija*, *42*(1), 27–46. doi:10.2298/PSI0901027M

McKone, E., & Dennis, C. (2000). Short-term implicit memory: visual, auditory, and cross-modality priming. *Psychonomic bulletin & review*, *7*(2), 341–346.

Mohamed, T. N., Neumann, M. F., & Schweinberger, S. R. (2009). Perceptual load manipulation reveals sensitivity of the face-selective N170 to attention. *NeuroReport*, *20*(8), 782–787.

Molholm, S., Ritter, W., Murray, M. M., Javitt, D. C., Schroeder, C. E., & Foxe, J. J. (2002). Multisensory auditory–visual interactions during early sensory processing in humans: a high-density electrical mapping study. *Cognitive Brain Research*, *14*(1), 115–128.

Murphy, G. L., & Smith, E. E. (1982). Basic-level superiority in picture catego rization. *Journal of Verbal Learning and Verbal Behavior*, *21*(1), 1–20.

Nelson, C. A. (2001). The development and neural bases of face recognition. *Infant and Child Development*, *10*(1-2), 3–18.

Noppeney, U., Josephs, O., Hocking, J., Price, C. J., & Friston, K. J. (2008). The Effect of Prior Visual Information on Recognition of Speech and Sounds. *Cerebral Cortex*, *18*(3), 598–609.

Osaka, N., & Yamamoto, M. (1978). VEP latency and RT as power functions of luminance in the peripheral visual field. *Electroencephalography and Clinical Neurophysiology*, *44*(6), 785–788.

Rogers, T. T., & Patterson, K. (2007). Object categorization: reversals and explanations of the basic-level advantage. *Journal of experimental psychology. General*, *136*(3), 451–469.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*(3), 382–439.

Rose, M., Schmid, C., Winzen, A., Sommer, T., & Büchel, C. (2005). The Functional and Temporal Characteristics of Top-down Modulation in Visual Selection. *Cerebral Cortex*, *15*(9), 1290–1298.

Rossell, S. L., Price, C. J., & Nobre, A. C. (2003). The anatomy and time course of semantic priming investigated by fMRI and ERPs. *Neuropsychologia*, *41*(5), 550–564.

Rossion, B., Gauthier, I., Goffaux, V., Tarr, M. J., & Crommelinck, M. (2002). Expertise training with novel objects leads to left-lateralized facelike electrophysiological responses. *Psychological science*, *13*(3), 250–257.

Rossion, B., Gauthier, I., Tarr, M. J., Despland, P., Bruyer, R., Linotte, S., & C rommelinck, M. (2000). The N170 occipito-temporal component is delayed and enhanced to inverted faces but not to inverted objects: an electrophysiological account of face-specific processes in the human brain. *Neuroreport*, *11*(1), 69–74.

Rossion, B., & Jacques, C. (2008). Does physical interstimulus variance account for early electrophysiological face sensitive responses in the human brain? Ten lessons on the N170. *NeuroImage*, *39*(4), 1959–1979.

Rossion, B., Kung, C.-C., & Tarr, M. J. (2004). Visual expertise with nonface objects leads to competition with the early perceptual processing of faces in the human occipitotemporal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(40), 14521–14526.

Rousselet, G. A., Macé, M. J.-M., & Fabre-Thorpe, M. (2004). Animal and human faces in natural scenes: How specific to human faces is the N170 ERP component? *Journal of vision*, *4*(1), 13–21.

Schneider, T. R., Engel, A. K., & Debener, S. (2008). Multisensory Identification of Natural Objects in a Two-Way Crossmodal Priming Paradigm.

Experimental Psychology (formerly "Zeitschrift für Experimentelle Psychologie"), *55*(2), 121–132.

Sigala, N., & Logothetis, N. K. (2002). Visual categorization shapes feature selectivity in the primate temporal cortex. *Nature*, *415*(6869), 318–320.

Sreenivasan, K. K., Goldstein, J. M., Lustig, A. G., Rivas, L. R., & Jha, A. P. (2009). Attention to faces modulates early face processing during low but not high face discriminability. *Attention, Perception, & Psychophysics*, *71*(4), 837–846.

Stevenson, R. A., & James, T. W. (2009). Audiovisual integration in human superior temporal sulcus: Inverse effectiveness and the neural processing of speech and object recognition. *NeuroImage*, *44*(3), 1210–1223.

Talsma, D., Doty, T. J., & Woldorff, M. G. (2007). Selective Attention and Audiovisual Integration: Is Attending to Both Modalities a Prerequisite for Early Integration? *Cerebral Cortex*, *17*(3), 679–690.

Talsma, D., Senkowski, D., Soto-Faraco, S., & Woldorff, M. G. (2010). The multifaceted interplay between attention and multisensory integration. *Trends in Cognitive Sciences*, *14*(9), 400–410.

Tanaka, J. W. (2001). The entry point of face recognition: evidence for face expertise. *Journal of experimental psychology. General*, *130*(3), 534–543.

Tanaka, J. W., & Curran, T. (2001). A neural basis for expert object recognition. *Psychological science*, *12*(1), 43–47.

Tanaka, J. W., & Sengco, J. A. (1997). Features and their configuration in face recognition. *Memory & cognition*, *25*(5), 583–592.

Tanaka, J. W., Luu, P., Weisbrod, M., & Kiefer, M. (1999). Tracking the time course of object categorization using event-related potentials. *Neuroreport*, *10*(4), 829–835.

Tanaka, J. W., & Taylor, M. (1991). Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychology*, *23*(3), 457–482.

Tarr, M. J., & Gauthier, I. (2000). FFA: a flexible fusiform area for subordinate-level visual processing automatized by expertise. *Nature Neuroscience*, *3*(8), 764–769.

Taylor, M. J. (2002). Non-spatial attentional effects on P1. *Clinical Neurophysiology*, *113*(12), 1903–1908.

Taylor, M. J., Batty, M., & Itier, R. J. (2004). The Faces of Development: A Review of Early Face Processing over Childhood. *Journal of Cognitive Neuroscience*, *16*(8), 1426–1442.

Thierry, G., Martin, C. D., Downing, P., & Pegna, A. J. (2007). Controlling for interstimulus perceptual variance abolishes N170 face selectivity. *Nature Neuroscience*, *10*(4), 505–511.

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

Töllner, T., Gramann, K., Müller, H. J., & Eimer, M. (2008). The Anterior N1 Component as an Index of Modality Shifting. *Journal of Cognitive Neuroscience*, *21*(9), 1653–1669.

van der Burg, E., Brederoo, S. G., Nieuwenstein, M. R., Theeuwes, J., & Olive rs, C. N. L. (2010). Audiovisual semantic interference and attention: Evidence from the attentional blink paradigm. *Acta Psychologica*, *134* (2), 198–205.

Vogel, E. K., & Luck, S. J. (2000). The visual N1 component as an index of a discrimination process. *Psychophysiology*, *37*(2), 190–203.

Vogel, E. K., & Luck, S. J. (2002). Delayed working memory consolidation during the attentional blink. *Psychonomic Bulletin & Review*, *9*(4), 739–743.

Voorhis, S. V., & Hillyard, S. A. (1977). Visual evoked potentials and selective attention to points in space. *Perception & Psychophysics*, *22*(1), 54–62.

Wild, H. A., & Busey, T. A. (2004). Seeing faces in the noise: Stochastic activity in perceptual regions of the brain may influence the perception of ambiguous stimuli. *Psychonomic Bulletin & Review*, *11*(3), 475–481.

Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configurational information in face perception. *Perception*, *16*(6), 747–759.