Stony Brook University

Behavioral and Neural Bases of the Effect of Pronunciation Length

on Visual Object Working Memory

A Dissertation Presented

By

Hwamee Oh

To

The Graduate School

In Partial Fulfillment of the Requirements

For the Degree of

Doctor of Philosophy

in

Biopsychology

Stony Brook University

May 2009

Stony Brook University

The Graduate School

Hwamee Oh

We, the dissertation committee for the above candidate for the Doctor of Philosophy

degree, hereby recommend acceptance of this dissertation.

Hoi-Chung Leung, PhD - Dissertation Advisor
Associate Professor of Biopsychology

Nancy Squires, PhD - Chairperson of Defense
Professor of Biopsychology

Gregory J. Zelinsky, PhD
Associate Professor of Cognitive and Experimental Psychology

Bradley R. Postle, PhD
Associate Professor of Psychology and Psychiatry
University of Wisconsin-Madison

This dissertation is accepted by the graduate school.

Lawrence Martin
Dean of the Graduate School

Abstract of the Dissertation

Behavioral and neural bases of the effect of pronunciation length

on visual object working memory

by

Hwamee Oh

Doctor of Philosophy

Biopsychology

Stony Brook University

2009

Visual object working memory refers to the maintenance of visual information for a short period of time. Although previous behavioral and recent neuroimaging studies have indicated a potential role of verbal processing in diverse human cognition, systematic investigation of the effect of verbal processing on visual cognition and neural activity is still lacking. In the present study, I investigated the effect of phonological coding, which was operationalized as the pronunciation length effect, on visual object working memory. In 3 behavioral experiments, the level of pronunciation length effect was examined as a function of presentation time, interstimulus interval (ISI), delay duration, and probe type. Using a delayed recognition task with namable common objects grouped as short (mostly 1 syllable) and long (2-4 syllables), Experiment 1 showed the effect of pronunciation length during visual object working memory when the objects were presented for 900 ms but not for 200 ms (Exp. 1A). With a 200-ms presentation time, however, the pronunciation length effect on visual object working memory emerged with a 800-ms ISI but not with 100-ms ISI (Exp. 1B) and with a 7800-ms delay but not with a 2800-ms delay (Exp. 1C). Experiment 2 replicated the pronunciation length effect during visual object working memory with the additional demand of maintaining visual representations. Results from Experiment 3 further showed that phonological coding of object names influences not only word-probe recognition but

also picture-probe recognition, with higher recognition accuracy for word-probe conditions compared to picture-probe conditions. In an fMRI experiment, the effect of pronunciation length on visual object working memory was significant in brain regions implicated in visual processing such as the fusiform gyrus and inferior occipital gyrus but did not reach significance in brain regions implicated in phonological processing such as the left inferior frontal gyrus. The present findings indicate that phonological processing during visual object working memory is likely automatic and influences brain areas traditionally considered to be involved in visual processing.

To my parents,

Oh, Yong-Tahk and Hong, Kwang-Su

Table of Contents

List of Figures

List of Tables

Acknowledgements

This dissertation would not have been possible without support and assistance from faculty, friends, and family. First, I would like to thank my advisor Dr. Hoi-Chung Leung for her guidance throughout my pursuit of this scientific journey during my graduate study at Stony Brook. Without her keen scientific mind and wholehearted support, I could have not completed this work. I would also like to thank my dissertation committee members Dr. Nancy Squires, Dr. Gregory Zelinsky, and Dr. Brad Postle for generously giving their time and expertise to better my work. I deeply thank them for their contribution to this dissertation and support throughout my graduate study.

My thanks also go to many other faculty members of this department, especially to Dr. Brenda Anderson, Dr. Patricia Whitaker, Dr. Turhan Canli, and Dr. John Robinson in the biopsychology area. They have provided me with a great graduate training through lectures, discussion, and personal feedback. I would also like to thank Dr. Arthur Samuel, Dr. Susan Brennan, and Dr. Harriet Waters for teaching me how to write and think through their invaluable courses. I also want to thank departmental staff members, Ms. Judy Thompson, Ms. Jean Bieselin, and Ms. Marilynn Wollmuth for their kind support throughout my graduate years in the department.

I am grateful to my colleagues in the department. Especially, I thank Yuji Yi, Weidong Cai, Vera Hua, John Borghi, and Hyejin Yang for their friendship and collaborative pursuit of knowledge. I also thank Dr. Eliza Congdon and Dr. Brian Haas for their friendship and help during my life at Stony Brook. My special thanks also go to Jadzia Jagiellowicz, Alexia Galati, Anna Kuhlen, Doreen Olvet, and Ilan McNamara for sharing their experiences and giving me their support and cheers when I needed them. They have consistently helped me keep perspective on what is important in life and shown me how to deal with reality.

I would also like to thank my parents and parents-in-law for their endless support and sacrifice for my education, for their wisdom throughout my life, and for believing in me all these years no matter what I have chosen to do with my life.

Last, but far from least, I want to express my deep appreciation and love for my husband, Sanghyun, who has guided me with his intellectual inspirations and unending

I. General Introduction

Working memory commonly refers to the temporary maintenance and manipulation of information to meet the moment-by-moment cognitive demands of everyday life. The involvement of working memory has been implicated in supporting higher cognitive functions in humans such as language, reading comprehension, and problem solving (Baddeley, 1986; Just & Carpenter, 1992). Therefore, research on the understanding of working memory has long been a focus of attention in the fields of cognitive psychology and neuroscience.

Several cognitive models have been proposed in order to explain the cognitive architecture of working memory. One of the most influential models of working memory is Baddeley's multi-component working memory model (Baddeley, 1986; Baddeley & Hitch, 1974). The model posits that verbal and visual information in working memory are coded and maintained in a separate storage buffer and coordinated by an attention-based controlling system called the central executive. The domain-specific organization of working memory proposed by this model has been supported by a substantial amount of evidence at both behavioral and neural levels (Courtney et al., 1998; Gruber & von Cramon, 2001; Leung et al., 2002; Levy & Goldman-Rakic, 1991; McCarthy et al., 1996; Mohr & Linden, 2005, 2006; Smith et al., 1996). Studies in cognitive psychology have suggested a role for the central executive in performing tasks requiring concurrent performance of both verbal and visuospatial processes (Rudkin et al., 2007); neuroimaging studies have further indicated that integration of information from different domains in working memory requires additional brain activity, especially in the prefrontal cortex (Collette & van der Linden, 2002; D'Esposito et al., 1995; Prabhakaran et al., 2000).

On the other hand, recent studies have suggested that the verbal processes are automatically and obligatorily engaged in processing visual information which is not readily verbalizable (Dent & Smyth, 2005; Postle et al., 2005, 2007; Winawer et al., 2007). However, the effect of concurrent verbal processing on visual cognition and neural activity remains largely unknown. In the present study, we aimed to investigate the cognitive and neural

mechanisms underlying the influence of verbal processing on visual information processing in working memory involving visual objects.

Visual object working memory refers to the cognitive process of maintaining visually presented information such as colors, shapes, and common objects for a short period of time. A theoretical basis of visual object working memory has been guided by Baddeley's multi-component working memory model (Baddeley, 1986). Accordingly, a significant amount of research has focused on studying visual object working memory based on the assumption that the perceived visual information is maintained in a visual form in working memory. Considering the inconsistent findings for the presence of verbal processing during visual object working memory (Brener, 1940; Dent & Smyth, 2005; Postle et al., 2005; Schiano & Watkins, 1981; Winawer et al., 2007), however, it is necessary to examine the extent to which verbal and visual information processing are involved in visual object working memory. In addition, it must be determined whether the inconsistent findings are brought about due to certain task factors influencing the level of verbal coding, and if so, what task factors might have caused the inconsistent findings in the literature.

Using behavioral and functional magnetic resonance imaging (fMRI) methods, we sought to answer the following questions. First, what task factors influence the level of verbal coding of visual information in working memory? Second, is verbal coding of visual objects automatic or does it involve the central executive? Third, what is the neural basis underlying the interaction between verbal and visual information processing in working memory? In a series of behavioral experiments using delayed visual recognition tasks, we examined the degree of verbal coding of visual information in working memory as a function of task factors: presentation time of study stimulus, interstimulus interval, duration of the delay, probe difficulty, and probe type. The degree of verbal coding of visual information was operationally defined as the degree of the pronunciation length effect, which is the behavioral benefit of remembering words with shorter pronunciation length than words with longer pronunciation length. In the fMRI experiment, we examined whether the brain regions implicated in verbal and visual information processing were modulated by pronunciation length of visual object names in working memory. Findings from the present study will help improve understanding of the precise functional architecture supporting visual object

2

working memory by considering both verbal and visual aspects of object information processing while teasing apart the relative contribution of each domain to visual object working memory.

II. Effect of Pronunciation Length on Visual Object Working Memory as a Function of Presentation Time, Interstimulus Interval, Duration of Delay, Probe Difficulty, and Probe Type

2.1. Background

Visual object working memory refers to the cognitive process of maintaining visually presented information (such as colors, shapes, and objects) for a short period of time. In the working memory literature, a theoretical basis of visual object working memory has been guided by Baddeley's multi-component working memory model, suggesting that different material domains are maintained by separate storage buffers: the phonological loop for verbal information, visual cache for visual object information, and inner scribe for spatial information (Baddeley, 1986; Logie, 1995). Neuropsychological and neuroimaging studies in humans have supported this material-specific proposal, showing separate neural substrates of working memory organized by stimulus type (Courtney et al., 1998; Gruber & von Cramon, 2001; Leung et al., 2002; McCarthy et al., 1996; Mohr & Linden, 2006; Smith et al., 1996). In parallel, research in nonhuman primates has delineated the functional organization of the prefrontal cortex, which is thought to be the neural basis of working memory, in a material-specific manner (Funahashi et al., 1989; Levy & Goldman-Rakic, 1991; Wilson et al., 1993), although some reported findings are inconsistent with the domain-specific organization of the prefrontal cortex (Rao et al., 1997). However, other investigators have found that dissociation between verbal and visual processing is not easy because the verbal processes are automatic and may even be obligatorily engaged for diverse visual information processing behaviors including perception, memory encoding, and maintenance of visual information (Dent & Smyth, 2005; Paivio, 1971; Postle et al., 2005; Potter, 1976; Winawer et al., 2007). Recent neuroimaging studies have shown that brain regions traditionally considered to be involved in phonological processing are active during the maintenance of unnamable visual information (Nystrom et al., 2000; Postle & Hamidi, 2007). The purpose of the present study is to determine the degree to which phonological processing is modulated by task parameters

during the maintenance of visual objects in working memory, and the effect of phonological processing on visual object working memory.

2.1.1. Behavioral framework for visual object working memory

Studies on visual object working memory have furthered our understanding of the cognitive mechanism for brief maintenance of visual information in humans. One of the main findings in visual object working memory studies relates to how much visual information can be held temporarily in working memory. Using a change detection paradigm, Vogel and colleagues (2001) reported that a maximum of 4 visual objects can be held in working memory regardless of the number of visual features comprising a given object. Alvarez and Cavanagh (2004), however, demonstrated that visual complexity of a given object influences search efficiency, indicating that the maximum number of visual objects held in working memory varies from 1 to 4 items as a function of visual complexity. More recently, Awh and colleagues (2007) showed that varying search efficiency observed in Alvarez and Cavanagh's (2004) study resulted from confounding factors of visual similarity between studied items and test items. They further indicated that the maximum number of visual objects in working memory remains 4 items, as previously suggested when the visual similarity factor was controlled.

In order to examine the mechanisms underlying the processing of pure visual information without verbal influences, three main methodological approaches were adopted. One was to use "unnamable" abstract shapes, such as Attneave shapes (Attneave & Arnoult, 1956), as study stimuli (e.g., Alvarez & Cavanagh, 2004; Linden et al., 2003). Another involved embedding a secondary verbal task into a main visual working memory task so that obligatory verbal processes during the main visual task would be suppressed (e.g., articulatory suppression; Gruber & von Cramon, 2001; Vogel et al., 2001; Xu & Chun, 2006). In a typical paradigm, a sequence of digits or letters is presented before the visual working memory task and those verbal items are tested at the end after the main visual working memory task is completed. In the third approach, study stimuli were presented briefly so that there was not sufficient time for visual stimuli to be subvocalized (e.g., Luck & Vogel, 1997; Vogel et al., 2001).

5

The validity of these three common methodologies to minimize or eliminate any verbal coding is still in question. Several studies reported findings that are inconsistent with the suggested effectiveness of these methodologies in preventing verbal coding during visual working memory (Baddeley et al., 1984; Intraub, 1981, 1984; Postle et al., 2005; Potter, 1993). Therefore, the question remains regarding the degree to which verbal processes may have influenced visual working memory performance.

2.1.2. Prior research on verbal coding of visual objects

The cognitive literature is relatively established for verbal and visual representation of visually presented objects. One of the early theories that framed the relationship between verbal and visual information during visual object processing is the dual coding theory (Paivio, 1971). The dual coding theory posits that namable visual objects can be processed both visually and verbally because verbal referents of visual objects are automatically activated when visual information is namable (Paivio, 1971). Supporting evidence has been provided showing that pictures were better recalled compared to words in a subsequent free recall task (Paivio & Csapo, 1973) because pictures were processed both verbally and pictorially while words were processed verbally only. Naming pictures was also shown to be more spontaneous than visualizing concrete words (Snodgrass et al., 1974).

Early behavioral studies have provided ample evidence on verbal coding of visual information. Brener (1940) asked subjects to verbally recall serially presented color strips and found the memory span for colors to be around 7, which was equivalent to that for letters and digits. Ternes and Yuille (1972) found that recall of both words and pictures of concrete objects was vulnerable to verbal interference during the delay period. Schiano and Watkins (1981) examined more comprehensively the verbal coding phenomenon in remembering visual objects by demonstrating the effects of phonological similarity, word length, and articulatory suppression on recall of namable visual objects. Together, these early behavioral studies suggested that visual short-term memory shares similar mental codes as verbal memory.

6

Further evidence comes from more recent studies on verbal coding of visual objects. Winawer et al. (2007) showed that color discrimination performance between Russians and native English speakers was differentially affected by verbal labeling of their native language and suggested an obligatory and automatic involvement of verbal processing in visual perception. Dent and Smyth (2005) showed that performance of a visual working memory task using Japanese Kanji characters was disrupted by concurrent articulatory suppression in native English speakers who did not know Japanese. Postle and colleagues (2005) demonstrated that visual object working memory involving nonverbalizable abstract shapes was vulnerable to a concurrent verbal task while spatial working memory performance was relatively unaffected. Neuroimaging studies have also shown that brain regions traditionally associated with verbal processing are activated during the maintenance of visual information that is not readily namable (Postle & Hamidi, 2007). These results together provide evidence suggesting potential involvement of verbal coding in memory for visual objects and visual perception, although the visual objects used in these recent studies were not readily namable.

Not only the nature of visual stimuli but also the task goal influences the way in which visual processing interacts with verbal processing. Zelinsky and Murphy (2000) trained subjects to associate pairs of faces and arbitrary surnames, which were either one or three syllables long. After training, they observed more fixations and longer gaze duration on faces with longer names compared to those with shorter names during encoding for a recognition task where subjects had to indicate whether a test probe matched one of the study items. Using the same face stimuli, however, eye movement related to the effect of face name length disappeared when subjects were required to find a studied face target among multiple face items in a search display. The researchers reasoned that this difference in eye movement between the two task goals occurs due to the difference in the level of verbal coding required of each task. In other words, verbal coding influences oculomotor behavior during encoding of visual objects in a recognition task, but not during search behavior. Using a delayed recognition task with line-drawing faces paired with names, Tversky (1969) tested whether pictorial and verbal coding of a visual object would be affected by task context and found that subjects tended to encode the initially presented visual items (line drawing of faces associated with specific names) in the modality of the testing item to which subjects made a

same/different decision. These results together suggest that the way visual information is encoded is not solely determined by the presentation format (i.e., visual percepts) but also by the task goal or anticipation of future use.

2.1.3. Questions addressed in the present study

Existing evidence on the verbal coding of visual objects suggests that visual information can be encoded into working memory both in verbal and visual representations and which domain will be the primary form of representation may be determined by task type. However, several questions remain to be answered. First, the cognitive mechanism underlying the verbal coding phenomenon of visual objects in visual object working memory is unclear. Studies investigating verbal coding of visual objects during visual object working memory seem to assume that verbal and visual representations are separate entities even though they are synchronized at a certain time point (i.e., visual processing occurs faster than verbal processing but waits for the verbal processing to complete). This view is also consistent with Baddeley's multicomponent working memory framework to some extent, suggesting that verbal and visual information are processed and stored by separate storage buffers, the phonological loop and visuospatial sketchpad, respectively, while coordinated by a central executive when information from both domains needs to be processed at the same time. Another possible mechanism by which verbal and visual information is processed, however, is that verbal and visual processing is one entity or two domains but simultaneously activated. Thus, processing information from one domain (e.g., verbal information) can automatically activate the processing of information from the other domain (e.g., visual information). As suggested by the behavioral studies described above, verbal labeling in the native language or verbal interference tasks inserted during the delay period of the main visual working memory task affect visual processing involved in visual perception, encoding, and maintenance of visual information in working memory.

Supporting evidence on automatic processing of both verbal and visual information also comes from the human neuroimaging literature. Visual association areas such as the inferior temporal cortex (IT) have been shown to be activated in verbal tasks where only words are presented as task stimuli in concrete/abstract word processing, mental imagery,

and repetition priming (Binder et al., 2005; D'Esposito et al., 1997; Wheatley et al., 2005). Findings from these studies cannot be the result of explicit verbalization or the coordination of verbal and visual information by the central executive, because, in these studies, it was not necessary to maintain information from different domains or word stimuli in pairs were presented very rapidly (150-ms presentation and 100-ms interstimulus interval).

If verbal coding is automatically involved in processing visual objects in working memory, the second remaining question is what task parameters influence the level of verbal coding of visual objects. As mentioned above, it is quite established that verbal coding occurs during the encoding and maintenance of visual objects in working memory. Although studies aiming to examine "pure visual" information processing in working memory utilized task stimuli and paradigms that were designed to minimize any verbal coding, there are inconsistent findings that do not support the validity of these approaches. It is unclear, however, whether inconsistent results are due to the different task stimuli or task designs that each study has adopted. Nevertheless, these approaches do not allow us to examine the effect of verbal coding on visual object working memory because they do not manipulate the verbal aspects associated with visual objects.

In the present study, we first aimed to examine the level of phonological coding of visual objects in working memory by manipulating task parameters using the same study materials. Specifically, we examined whether different findings on the effect of phonological coding of visual objects in working memory may have related to different time parameters assigned to encoding and maintenance of study items and probe type. Second, we aimed to examine one of the possible mechanisms underlying the influence of verbal coding of visual objects on visual object working memory by directly comparing results obtained from two task paradigms. One paradigm was designed to examine automatic verbal processing of visual objects in working memory and the other was designed to engage the central executive to maintain both verbal and visual information in working memory. Throughout the experiments, we examined changes in the level of verbal coding of visual objects as a function of task factors that we manipulated. The level of verbal coding of visual objects was operationally defined by the effect of name length of visual objects, which relates to the

9

word-length effect in verbal working memory studies. A theoretical basis of the word-length effect is discussed in the next section.

2.1.4. Theoretical basis of the word-length effect

The word-length effect refers to the behavioral benefit of remembering words with shorter pronunciation length than words with longer pronunciation length. It is worth mentioning that the phonological loop incorporated in Baddeley's working memory model is further divided into phonological storage and articulatory rehearsal components. Phonological storage is theorized as a memory store that retains acoustic-based information for a short period of time. It is sensitive to the phonological similarity effect, meaning that phonologically similar items such as "B" and "G" are less likely to be remembered than phonologically distinctive items such as "F" and "W" (Baddeley et al., 1975; Coltheart & Langdon, 1998; Lian et al., 2001; MacAndrew et al., 2002; Nimmo & Roodenrys, 2005). Articulatory researsal, on the other hand, is theorized as a maintenance mechanism for acoustic-based information. It has been shown to be subject to word length manipulation and concurrent articulation (or "articulatory suppression[1]"), meaning that overt or covert articulation of a different set of acoustic-based information interferes with remembering a set of verbal stimuli (Baddeley et al., 1975; Cowan et al., 1992; 2000: Mueller et al., 2003; Tehan et al., 2001). Although both phonological storage and articulatory rehearsal are involved in the maintenance of verbal information in working memory, the present study focuses on the mechanism of articulatory rehearsal by examining the effect of pronunciation length difference.

With respect to the word-length effect, there has been debate on the cause of this effect. One of fundamental issues concerns what stage of working memory this effect is attributed to. An original proposal of the word-length effect by Baddeley and colleagues

---

[1] Articulatory suppression refers to the method in which subjects are asked to repeat a word (e.g., "the"), letters, or digits either covertly or overtly.

(Baddeley et al., 1975) attributed this effect to a phonological rehearsal mechanism during encoding and delay stages, suggesting that the phonological information of longer items (e.g., polysyllabic words) compared to the shorter items (e.g., monosyllabic words) decays faster due to less frequent rehearsal of the former than the latter. Another account from the phonological explanation of this effect is an interference or decay during the output process (Cowan et al., 1992). By mixing short and long word items in a recall list and analyzing the time taken to recal items by their serial positions, Cowan et al. (1992) found that recall accuracy was related to the time taken to complete recalling the prior list of items. In line with the second view on the account of the word-length effect is an account of the involvement of the speech-production system during rehearsal and recall (Howard & Franklin, 1987; Martin et al., 1999; Monsell, 1987). According to this view, rehearsal consists of constant speech input/output processes. In order to test this conflicting account, Avons et al. (1994) used both serial recall and probed recall tasks, in which subjects did either recalling all heard items in a serial order or recalling one item at a serial position to which the experimenter pointed. Consistent with their hypotheses, the word length effect was reduced during probed recall compared to serial recall, indicating that the word length effect is partially due to the output process, not strictly due to the rehearsal process. Using a recognition task, a variant of the matching-span task, Coltheart et al. (2004) also showed that the word-length effect occurred even with the reduced output demands of recognition. Taken together, these results suggest that both rehearsal and output processes contribute to the word-length effect using word stimuli which are presented either auditorily or visually and that the word-length effect was observed even with a recognition task.

Another debate on the cause of the word length effect lies on whether or not the effects is caused by pronunciation duration or other phonetic characteristics such as phonological complexity or phonological dissimilarity. Results from earlier studies to test the effect of phonological duration are mixed. Some showed an effect of pronunciation duration by using word stimuli which differ only in spoken duration while being equal in number of syllables (Baddeley et al., 1975; Cowan et al., 1992). Others showed a reversed word length effect with the same manipulation but with the different word stimulus set (Caplan et al., 1992). This controversy was resolved in a recent study by Mueller et al. (2003), showing that

pronunciation length is responsible for this effect when other factors such as phonological complexity and phonological dissimilarity are carefully controlled.

For the purpose of the present study, we will use the term "pronunciation length effect" rather than the "word length effect," which has been commonly used in verbal working memory studies. Part of the reason is that, while visual words were used as study stimuli in verbal working memory studies, visual objects were used in the present study. Therefore, it would be more appropriate to use the term "pronunciation length effect" rather than word length effect, although we assume both generate the same behavioral effect.

2.1.5. Hypotheses of the present studies

Based on the previous literature, we hypothesize that the level of verbal coding will be influenced by multiple task factors. Specifically, we hypothesize that the degree of verbal coding will be greater (1) when visual information is presented for a longer time (Exp. 1A), (2) when time between visual items to be studied is longer (Exp. 1B), and (3) when retention duration is longer (Exp. 1C). In Exps. 2 and 3, we further examined effects of probe difficulty and probe type to eliminate alternative explanations for the results from Exp 1.

Overview of the Present Study

Given the foundation of the effect of phonological processing in verbal working memory, we applied this approach to examine the level of phonological coding in visual object working memory as a function of task parameters. Specifically, in Experiments 1A-C, we examined the effects of presentation time, interstimulus interval (ISI), and duration of the delay on the level of verbalization of visual objects in working memory. In Experiment 2, we tested the effect of pronunciation length on visual object working memory when the requirement of maintaining visual representations was increased. In Experiment 3, we tested the effect of probe type on the level of verbal coding of visual objects in order to ensure that the phonological coding observed in Experiment 2 reflects automatic processing of verbal aspects of visual information without an engagement of the central executive. We

hypothesized that the pronunciation length effect during visual object working memory would be greater with longer presentation time, longer ISI, and longer delay, and that a similar effect would be observed even when explicit verbal coding was not encouraged.

Experiment 1

Experiment 1A

In Experiment 1A, we sought to replicate the word-length effect that was shown in previous verbal working memory studies (Baddeley et al., 1975; Coltheart & Langdon, 1998; Mueller et al., 2003; Schiano & Watkins, 1981), using namable visual objects during a delayed-recognition task, and further hypothesized that the effect of pronunciation length would differ as a function of presentation time of visual objects. We also varied set size of study items between 1, 2, 4, and 6, in order to further examine any interaction between pronunciation length and set size. Participants were randomly assigned to either a long presentation condition task or a short presentation condition task. Therefore, presentation time condition (long presentation or short presentation) was tested as a between-subjects variable. In each presentation condition, pronunciation length of visual object names and set size were manipulated as within-subjects variables.

Method

Participants

Thirty-two right-handed young adults (19 females, mean age: 19.8 years) were recruited from the Stony Brook community. Sixteen participated in the long presentation condition task, and the other 16 participants did the short presentation condition task. All participants had no history of neurological and psychiatric disorders or drug abuse according to self-report, and had normal or corrected-to-normal vision. They were given course credits for their participation and prior to participation gave informed consent that was reviewed and approved by the Institutional Review Board of State University of New York at Stony Brook.

13

Apparatus and Materials

The experiment was programmed and executed with E-Prime software (Psychology Software Tools, Pittsburgh, PA). The task was presented on a 17-inch color monitor.

Between-category namable colored objects from the Hemera database (Hemera Technologies Inc., Gatineau, Quebec, Canada) and internet were used as stimuli. We selected between-category items for a study list because they were thought to reduce interference, which may influence performance beyond manipulations of interest, compared to within-category stimuli (Wickens, 1970). Objects were common objects and comprised 6 categories with 4 exemplars within each category. The final set of stimulus categories and characteristics is listed in Appendix. Objects were grouped into shorter name (mostly 1 syllable) or longer name (2-4 syllables) objects. Prior to the main experiment, we conducted two pilot studies. The first pilot study was conducted in order to control visual complexity of pictures between the two length conditions using the name-picture agreement task (Murphy & Brownell, 1985). The second pilot study was conducted in order to ensure pronunciation length differences between the two length conditions according to the method adopted in previous studies (Boutla et al., 2004). Ten subjects were recruited for each pilot study. Response times were the dependent measures for both pilot studies. Results showed equivalent visual complexity of finalized visual objects (M = 461 ms, SD= 138.33 ms, for short name; M = 462 ms, SD= 123.17 ms, for long name, p > 0.5) and a significant pronunciation length difference between the two length conditions (M = 352 ms, SD = 48 ms, for short name; M = 438 ms, SD = 74 ms, for long name, p = .006).

Besides these measures, written word frequency of object names were equated between length conditions based on values determined from the English Lexicon Database to control for the influence of long-term memory (Balota et al., 2007; See Appendix). In order to balance visual features of objects selected for each condition, colors and shapes of visual objects were chosen to maximize similarity between pictures of objects from each condition. Each picture was resized to be equivalent by area. Object pictures were subtended by a visual angle of 2° X 2° on average.

14

Design and Procedure

A 2 (pronunciation length: short vs. long name) X 4 (set size: 1, 2, 4, and 6) X 2 (presentation time: short vs. long) mixed design was used, where pronunciation length and set size were within-subjects factors and presentation time a between-subjects factor. The experimental paradigm was a delayed recognition task, as shown in Figure 1. After an initial fixation for 1 s followed by a 500 ms warning sign, a varying number of study stimuli (1, 2, 4, or 6) were presented consecutively at the center of the screen on a white background. In the long presentation condition, each study stimulus was presented for 900 ms with 100 ms interstimulus interval (ISI). In the short presentation condition, each stimulus was presented for 200 ms with the same ISI duration. At the end of the last item, a color-noise pattern mask was presented for 200 ms. Then, a delay period of 2800 ms followed until a test probe appeared. The test probe was presented for 2 s with an additional 1 s response window during which subjects were required to make a yes/no judgment to indicate whether the test probe was one of the study items. Responses were made by pressing one of two keys on the button box with either index finger. Hand-responses mapping was counterbalanced across subjects. Twenty-four object images for each pronunciation-length condition were used repeatedly throughout the experiment and drawn equally often as a study item and a test probe. In the study stimulus set, no category was repeated within a trial. A test probe for a "yes" response was selected from one of the studied items with the constraint that each serial position was equally selected as a "yes" probe. A test probe for a "no" response was selected with the constraint that half of the "no" probes were drawn from the same length group within the same category (e.g., a picture of "lime" for the study picture of "grapes") and the other half from the other length group of that category (e.g., a picture of "banana" for the study picture of "grapes").

The experimental task was conducted in 4 blocks with 48 trials in each block. There were 24 trials total for each condition. The number of "yes" and "no" trials was counterbalanced across blocks and conditions. Prior to the experimental task, subjects completed a naming test and practice sessions which took about 20 min together. The naming test was conducted in order to ensure the label subjects would associate with a given visual object. After the experimental task, the naming test was administered once again to record

what words were brought into the subjects' minds when they had to remember the study items during the experiment. A post-experiment questionnaire was administered to ask what strategy subjects used to perform the task. The experimental session lasted about an hour.

Results and Discussion

A separate analysis of variance (ANOVA) for A' scores and response time was conducted with length (short or long) and set size (1, 2, 4, or 6) as within-subjects factors and presentation time (200 ms or 900 ms) as a between-subjects factor. Response times which were 3 standard deviations away from the mean (less than 1% of all responses) were excluded from further data analysis. A' scores, corrected accuracy rates incorporating subjects' guessing, were calculated based on the formula from Stanislaw and Todorov (1999). Repeated-measures two-way ANOVAs were conducted for each presentation time condition. We first present the results from two-way ANOVAs with pronunciation length and set size as factors, followed by those from the overall 3-way ANOVA with pronunciation length, set size, and presentation time as factors.

Results are shown in Figure 2. In the long presentation condition, we found significant main effects of pronunciation length $[F(1,15) = 5.181, p < .05]$ and set size $[F(3, 45) = 22.115, p < .001]$ for A' scores and a significant main effect of set size only for response times. For A' scores, however, pairwise t-tests showed that the effect of pronunciation length was significant only in a set size of 6. An interaction between pronunciation length and set size approached significance $[F(1,15) = 2.711, p = .056]$, indicating that the pronunciation length effect became pronounced with larger set size. In the short presentation condition, however, the effect of pronunciation length disappeared $[F(1,15) = .031, p > .10]$, while the effect of set size, as expected, was significant for A' scores $[F(3, 45) = 36.836, p < .001]$. The three-way ANOVA showed a significant main effect of set size on A' scores $[F(3, 90) = 58.889, p < .001]$ and a significant interaction effect between set size and presentation time $[F(3, 90) = 3.788, p = .013]$. All other main effects and interactions were not significant.

16

To further examine whether response time for the short and long name objects differs by positive and negative probes, we calculated the mean response time for each pronunciation length condition separately for "yes" and "no" trials. With only correct trials in each trial type, similar results were yielded, showing only a significant main effect of set size for both long and short presentation time conditions (all ps < .01).

The present results are consistent with previous findings in verbal working memory studies showing the word-length effect during a serial recall task (Baddeley et al., 1975; Coltheart & Langdon, 1998; Mueller et al., 2003; Tehan et al., 2001). Different from previous studies, however, the present study used visual objects instead of words and a delayed recognition task instead of a recall task. Results from Experiment 1A suggest that the effect of pronunciation length on visual object working memory is observable when more time (900 ms presentation time with 100 ms ISI) is allowed to encode each visual object. Because verbal coding of visual objects is not necessary to perform the current task, in which only visual objects were used as stimuli, the present results are consistent with the view that verbal coding occurs automatically during visual object working memory as manifested by the effect of pronunciation length. As described above, the word length effect in verbal working memory studies has been considered to show an articulation-based maintenance mechanism for phonological information. Therefore, the current findings further suggest that encoding time of visual objects influences the level of phonological coding of visual objects, which in turn influences how much phonological representation of visual objects is maintained in visual object working memory.

Coltheart et al. (2004) showed that the word length effect occurred even with the reduced output demands of recognition. They further showed that orthographic length of words determined by the number of letters also influenced the word length effect when each word stimulus was presented visually at a very fast speed (i.e., 8 items/sec). When a list of 4-letter words and a list of 6-letter words are presented visually, recognition accuracy was higher for the former than the latter list at a rate of 8 items per second, but not at a rate of 1 item per second. Therefore, it is possible that the magnitude of the pronunciation length effect observed in the present study using visual objects might have been smaller compared to that observed in studies using verbal items.

17

When we examined response times separately to positive and negative probes, we did not find a difference between "yes" and "no" judgments as a function of pronunciation length of visual objects in working memory. Previous studies on recognition memory suggested that studied items are internally represented during the retention stage and compared with the newly encoded test stimuli in a serial and exhaustive manner, resulting in a setsize effect (Sternberg, 1966, 1967) and faster response time for "yes" trials compared to "no" trials (Hockley & Murdock, 1987; Ratcliff, 1978; Swanson, 1974). Therefore, the pronunciation length effect observed during the 900-ms presentation condition seems not to have interacted with a decision process during the probe stage.

Although insignificant (p = .056), there was a clear trend of pronunciation length interacting with set size. When we examined the effect of pronunciation length as a function of set size with the longer presentation time, performance between short and long name conditions differed only in a set size of 6, while no difference was observed in the other set sizes. Therefore, it may suggest that phonological coding occurs at a larger set size. On the other hand, it is possible that phonological coding may have occurred even with small sets of study stimuli, but that it is not observable at the behavioral level. However, whether or not phonological coding occurred at a smaller set size cannot be concluded based on the present experiment.

Experiment 1B

The overall results of Experiment 1A indicate that encoding time is critical for verbal coding of visual objects. However, it is unknown whether presentation time of visual objects or the interval duration between onset of study stimuli (i.e., interstimulus interval: ISI) facilitated phonological coding of visually presented items. In Experiment 1B, we examined this question by lengthening ISI to 800 ms while keeping presentation time of each study stimulus at 200 ms, allowing us to tease apart whether the pronunciation length effect observed in Exp. 1A was driven by presentation time or by ISI. We expected that if time, not necessarily with the presence of the visual stimuli, is responsible for the pronunciation length effect we observed in Exp. 1A, we would observe a similar pronunciation length effect with

lengthened ISI. However, if duration of the stimulus presentation was solely responsible for the pronunciation length effect, we will not observe a similar pronunciation length effect in Exp. 1B. To test this prediction, we compared the results from Exp. 1B with those from the short presentation condition of Exp. 1A by applying a 3-way ANOVA in the analysis.

Method

Participants

Sixteen new right-handed young adults (9 females, mean age: 20.6 years) were recruited from the Stony Brook community to participate in the long ISI condition. Other characteristics of participants were the same as in Exp. 1A.

Apparatus and Materials

Apparatus and materials for Exp. 1B were the same as in Exp. 1A.

Design and Procedure

A 2 (pronunciation length: short vs. long names) X 4 (set size: 1, 2, 4, and 6) X 2 (ISI: long vs. short) mixed design was used, where pronunciation length and set size were within-subjects factors and ISI was a between-subjects factor. The procedure of Exp. 1B was similar to that of the short presentation condition of Exp. 1A, except that study stimuli were presented for a short duration as of 200 ms while the ISI was lengthened to 800 ms.

Results and Discussion

As in Exp. 1A, dependent measures included A' scores and response times. The results are shown in Figure 2. For A' scores, a 2-way repeated-measures ANOVA with presentation time and set size as factors revealed a significant main effect of set size [$F(3, 45)$ = 19.427, p = .000] and a significant interaction between pronunciation length and set size [$F(3, 45) = 3.127$, p < .05], but no main effect of pronunciation length (p >.5). For response time, only a main effect of set size was significant [$F(3, 45) = 43.064$, p < .001].

Combining data from the long ISI condition in this experiment and data from the short presentation time condition in Experiment 1A, the 3-way mixed ANOVA revealed a main effect of set size on A' scores [$F(3, 90) = 55.747$, $p = .000$] and significant interaction effects between set size and ISI [$F(3, 90) = 3.146$, $p < .05$] and between pronunciation length and set size [$F(3, 90) = 3.098$, $p < .05$]. For response time, the 3-way mixed ANOVA revealed a significant main effect of set size [$F(3, 90) = 83.196$, $p = .000$] and a significant interaction between pronunciation length and set size [$F(3, 90) = 2.787$, $p < .05$]. Other effects for A' scores and response time did not reach significance.

Contrary to our prediction, the results from Experiment 1B revealed that longer ISI did not significantly increase behavioral differences between short and long name conditions as shown in Experiment 1A. Furthermore, closer inspection of the significant interaction between pronunciation length and set size indicated that this effect is different from the interaction pattern observed in Experiment 1A. In Experiment 1A, although it failed to reach significance ($p = .056$), an interaction between pronunciation length and set size seemed to be driven by the behavioral difference only at set size 6 with no difference in the other set sizes. In Experiment 1B, however, this significant interaction between pronunciation length and set size was driven by a reversed pronunciation length effect at set size 4. Paired t-tests applied to each set size further confirmed that performance for the long name condition was significantly higher than that for the short name condition at set size 4 [$t(15) = -2.829$, $p < .05$]. Therefore, the pronunciation length effect observed in Experiment 1A seemed to be driven by presentation time, rather than ISI between study items.

Experiment 1C

Results from Exp. 1A indicated that a phonological coding effect during visual object working memory is observable with longer encoding time, revealing a significant pronunciation length effect which disappears when presentation time is short. However, studies have indicated that phonological processing continues even though a visually presented word is no longer present. In a study of attentional blink, processing of words preceded by longer words was more difficult than those preceded by shorter words (Olson et

al., 2001). In Experiment 1C, we tested the hypothesis that visual objects, even though briefly presented, would be phonologically recoded and rehearsed if a longer delay between the study phase and test phase is allowed.

Method

Participants

Sixteen right-handed young adults (11 females, mean age: 19.6 years) were additionally recruited from the Stony Brook community to participate in the long delay condition. Other characteristics of participants were the same as in Exp. 1A.

Apparatus and Materials

Apparatus and materials for Exp. 1C were the same as in Exp. 1A.

Design and Procedure

A 2 (pronunciation length: long vs. short names) X 4 (set size: 1, 2, 4, and 6) X 2 (delay: long vs. short) mixed design was used, where pronunciation length and set size were within-subjects factors and delay was a between-subjects factor. The procedure of Exp. 1C was the same as in Exp. 1A, except that the delay between the study and probe phases was lengthened to 7800 ms while only short presentation time of study stimuli (200 ms) was used.

Results and discussion

Results are shown in Figure 2. For A' scores, a 2-way repeated-measures ANOVA for the long delay condition revealed a significant main effect of set size [$F(3, 45) = 19.945$, $p < .001$] but neither the main effect of pronunciation length nor the interaction was significant. With response times, however, we found a significant main effect for set size [$F(3, 45) = 17.829$, $p < .001$] and a significant interaction between pronunciation length and set size [$F(3, 45) = 4.832$, $p = .005$]. Paired t-tests indicated that the difference between short and long name conditions was significantly different only for a set size of 6 [$t(15) = 3.282$, $p = .005$].

According to a 3-way mixed design ANOVA, for A' scores, there was a significant main effect of set size [$F(3, 90) = 47.58$, $p < .001$] and a main effect of delay approached significance, $F(1, 30) = 3.017$, $p = .093$. With reaction time, we found a significant main effect of set size [$F(3, 90) = 54.095$, $p < .001$] and a significant interaction between pronunciation length and set size [$F(3, 90) = 5.261$, $p = .002$]. A 3-way interaction among pronunciation length, set size, and task approached significance [$F(3, 90) = 2.315$, $p = .081$].

The results from Exp. 1C indicate that verbal coding influences visual object working memory performance when the delay period between the study phase and the test phase is lengthened, even though the visual objects were presented for a brief period of time. This pronunciation length effect, however, was again observable only when the memory demands were large (i.e., set size of 6). Although the pronunciation length effect is present at this larger set size, we may have missed the effect of phonological coding at a smaller set size because the task may be too easy overall, resulting in a ceiling effect. In Exp. 2, we increased task difficulty by manipulating "no" probes and examined the pronunciation length effect in the context of greater visual demands, which we refer to as the difficult probe task.

Experiment 2

Experiment 2 was similar to Experiment 1A except that we increased task difficulty by manipulating negative probes in the task. Thus it increased the visual processing demand during encoding and maintenance of visual representations in working memory. We hypothesized that a weaker or no pronunciation length effect would be observed if participants completed the task by relying on visual information processing. However, we would observe a stronger pronunciation length effect if verbal coding occurs automatically and the ceiling effect is removed.

Method

Participants

Ten right-handed young adults (8 females, mean age: 20.5 years) were additionally recruited from the Stony Brook community to participate in the difficult probe task. Other characteristics of participants were the same as in Exp. 1A.

Apparatus and Materials

Visual stimuli consisted of 6 categories as used in the experiments described above. However, in this experiment, each object name had 4 unique pictures, resulting in 96 unique pictures for each pronunciation length condition. Other characteristics of visual stimuli and object names were the same as in the experiments described above.

Design and Procedure

A 2 (pronunciation length: long vs. short names) X 4 (set size: 1, 2, 4, and 6) within-subjects design was used. A major change in the task was made with the selection of negative probes. Unlike in the experiments reported above, negative probes were selected among pictures of the same object name. Therefore, remembering the study stimuli by their name should not help recognition performance; subjects may rely more on retaining visual information for correct recognition. Each picture was selected equally often as a study stimulus and a probe item.

Results and Discussion

Results are shown in Figure 2. In contrast to our prediction that the pronunciation length effect would be reduced with increased visual demand through difficult probes, the 2-way repeated-measures ANOVA revealed significant main effects of pronunciation length [$F(1, 9) = 25.437$, $p = .001$] and set size [$F(3, 27) = 10.953$, $p < .001$] on A' scores. Paired t-tests further showed that performance in the long name condition was significantly lower than that in the short name condition at set sizes of 1 and 4 ($p < .05$ for each). For response

time, only a main effect of set size was significant [$F_{(3, 27)} = 23.039$, $p < .001$], suggesting that there was no trade-off between accuracy and response time.

The present results clearly indicate that verbal processing influences visual object working memory even in a situation where explicit verbal coding is not supposed to help visual object working memory. The present results further show that the pronunciation length effect was more pronounced when the ceiling effect was removed.

Experiment 3

In Experiment 2, we observed a stronger pronunciation length effect with difficult probes, even though verbal coding was not explicitly required. Findings from the last two experiments may suggest that participants may have used both verbal and visual coding of visually presented objects. In Experiment 3, we tested this possibility by using different probe types, requiring participants to perform the task by focusing only on visual aspects of study stimuli (i.e., shapes or colors; "Picture" condition), only on verbal aspects of study stimuli (i.e., names of visual objects; "Word" condition), or on both visual and verbal aspects ("Either" condition). If the pronunciation length effect observed thus far occurred due to participants' strategies of actively holding both verbal and visual aspects of study stimuli, we expected to observe the pronunciation length effect only in the "Word" and "Either" conditions. If the pronunciation length effect observed thus far reflected automatic verbal coding without involvement of the central executive, we expected to observe a similar pronunciation length effect with the "Picture" probe condition. When pictures were presented as a test probe, we expected that magnitude of the pronunciation length effect would be greater in the "Either" condition compared to the "Picture" condition, because involvement of the central executive to maintain both verbal and visual representations may play a role in increasing the pronunciation length effect. Overall performance in the "Either" condition was expected to be lower than that in the "Picture" and "Word" probe conditions.

Method

24

Participants

Fifteen right-handed young adults (9 females, mean age: 19.2 years) were recruited from the Stony Brook community to participate in Exp. 3. Other characteristics of participants were the same as in Exp. 1A.

Apparatus and Materials

Apparatus and materials for Exp. 3 were the same as in Exp. 2. Therefore, characteristics of the visual stimuli were the same as in Exp. 2.

Design and Procedure

A 2 (pronunciation length: short vs. long) X 3 (probe type: word vs. picture vs. either) within-subjects design was used. A schematic diagram is shown in Figure 3A. Probe type was grouped by block, and each block started with 2-s presentation of the letter "P," "W," or "E." "P" indicated that the test probe would be a picture. "W" indicated that it would be a word, while "E" indicated that it could be either a picture or a word. The meaning of these instruction letters was learned by participants during practice prior to the experimental task. Following the instruction letter, 4 study stimuli were presented consecutively at the center of the screen for 900 ms with an ISI of 100 ms followed by a mask for 200 ms at the end of the fourth item. A delay period of 2800 ms of a fixation cross followed until a test probe appeared. The test probe was presented for 1 s and subjects were required to make a yes/no judgment to indicate whether the test probe was one of the study items. An ITI of 1 s followed before the start of the next trial. For half of the trials of the longer name object condition, the test probes were pictures, and for the other half, the test probes were words. In the picture probe condition, a "yes" probe was the exact picture of one of the studied items, whereas a "no" probe was a different picture of the same object that was studied in the study phase. In the word probe condition, a "yes" probe was the name of one of the studied items whereas a "no" probe was a name of a different object that was not studied in the study phase but falls in the same category (e.g., the word "pineapple" could be a "no" probe for a study item depicting a banana). The experimental task consisted of 2 blocks of each probe type with 24 trials in each block. The order of the blocks was counterbalanced across subjects.

There were 48 trials per probe type, with half of the trials having a matching probe and the other half having a nonmatching probe. Subjects completed a naming task and a practice session for 20 min prior to the experimental task. A post-experiment questionnaire was administered to determine whether subjects labeled the study items while performing the task and, if so, what words were used as labels. The experimental session lasted about an hour.

Results and Discussion

The results are shown in Figure 3B & C. A 2 (pronunciation length: short vs. long name) X 4 (probe type: picture, word, either-picture, either-word) repeated-measures ANOVA with pronunciation length and probe type revealed a significant main effect of pronunciation length [$F(1, 14) = 33.649$, $p < .001$] and probe type [$F(3, 42) = 33.706$, $p < .001$] on A' scores, showing that recognition performance was higher with short name objects and word probes compared to long name objects and picture probes. An interaction between pronunciation length and probe type did not reach significance [$F(3, 42) = 2.095$, $p = .115$]. Paired t-tests further showed significant differences between short and long name conditions across all probe conditions (picture probe: $t(14) = 2.744$, $p = .016$; word probe: $t(14) = 3.226$, $p = .006$; either-picture probe: $t(14) = 2.709$, $p = .017$; either-word probe: $t(14) = 4.889$, $p < .001$). According to post-hoc t-tests (with Bonferroni correction), difference in accuracy between the picture probe condition and the word probe condition ($p < .001$), between the picture probe condition and the either-word probe condition ($p < .001$), between the word probe condition and the either-picture probe condition ($p < .001$), and between the either-picture probe condition and the either-word probe condition ($p < .001$) was significant. For response time, only a main effect of probe type was significant [$F (3, 42) = 9.851$, $p < .001$]. Post-hoc t-tests (with Bonferroni correction) comparing the 4 conditions resulting from the combination of task instruction and probe type revealed significant differences in response time between the word probe and the either-picture probe conditions ($p = .002$) and between the word probe and the either-word probe conditions ($p < .001$). Paired t-tests applied to each probe type further showed a significant difference in response time between short and long name conditions for the word probe condition only [$t(14) = - 2.860$, $p = .013$].

In Experiment 3, we tested whether the pronunciation length effect we observed in previous experiments reflects automatic processing between verbal and visual storage buffers without engagement of the central executive. According to Baddeley's working memory model, holding both verbal and visual information in working memory requires coordination by the central executive. Results from Exp. 3, however, cannot be incorporated into the current multi-component working memory model proposed by Baddeley because the pronunciation length effect was observed even in the "Picture" condition in which verbal coding was not explicitly required and not useful. Consistent with our hypothesis, the pronunciation length effect was less variable or more consistent across subjects in the "Word" condition and "Either" condition, suggesting that participants had to actively retain verbal labels of the study items in these conditions. Verbal coding of visual objects demonstrated by the presence of the pronunciation length effect during visual object working memory ("Picture" condition), however, cannot be explained by the episodic buffer, because the nature of the observed pronunciation length effect in this condition is rather automatic than attention-demanding. Attention may be required in order to bind information from different domains and to keep the bound information in the episodic buffer (Allen et al., 2006; Stefurak & Boynton, 1986; Wheeler & Treisman, 2002). The episodic buffer has been proposed as a capacity-limited temporary storage mechanism of multi-modal information that is bound by the central executive (Baddeley, 2000, 2003). In the picture condition, however, there is no need to actively bind verbal (i.e., object names) and visual (object images) information to form integrated information. Rather, viewing visual object images may make object names activated, which in turn influences visual object working memory performance. Therefore, the present data are more consistent with a view of automatic verbal processing of visual objects.

General Discussion

In the present study, we aimed to examine what task factors modulate the level of phonological coding in visual object working memory. To this end, we manipulated presentation time of study stimuli, ISI, duration of delay, probe difficulty, and probe type and

examined the extent to which the pronunciation length effect changes as a function of these task parameters. We found that the pronunciation length effect emerged when the presentation time of visual objects was 900 ms while it disappeared with a 200-ms presentation time. However, the pronunciation length effect was significant only with larger set size (i.e., 6 study items), suggesting that the effect is observable only when set size is beyond the previously proposed visual capacity of 4 items. The pronunciation length effect was also unnoticeable with lengthened ISI, indicating presentation time rather than ISI is more critical in revealing the level of phonological coding of visual objects in working memory. Furthermore, we have shown that the pronunciation length effect is observable with visual objects presented for 200 ms when the duration of delay between the study and test phases is lengthened from 2800 ms to 7800 ms. This result suggests that the visual objects are verbally recoded and maintained if a longer rehearsal period is given even though encoding time was not long enough to verbally code visual objects.

The current findings cannot be explained solely by pure verbal or visual rehearsal of visual objects. The results from Experiment 2 indicated that the pronunciation length effect is still present even though phonological coding did not appear to have an obvious advantage in completing the task employed in the current study. Interestingly, the pronunciation length effect was much larger than that seen in Experiments 1A-1C when the demand of maintaining visual representations was increased. These results indicate that phonological coding of visual objects is engaged in visual working memory performance regardless of the task demand. Results of Experiment 3 further confirm that phonological coding of visual objects is automatic rather than depending on the central executive to coordinate,

Verbal coding of visual objects in working memory

Previously, verbal coding of visual objects has been examined using a recall paradigm (Schiano & Watkins, 1981), which, by nature, requires subjects to encode, rehearse, and retrieve visual information by their verbal labels. As expected, the results from recall tasks confirmed verbal coding of visual objects by demonstrating similar effects shown in verbal tasks (e.g., phonological similarity effect, articulatory suppression, and word length

effect). More recently, however, evidence supporting verbal coding of visual objects has come from studies where verbal coding was not expected to be shown, because unnamable visual objects were used (Postle et al., 2005; Postle & Hamidi, 2007). By manipulating the phonological aspects of visual items in 3 experiments, our results demonstrate that verbal coding is involved in holding visual objects in working memory but that the level of verbal coding changes as a function of task factors.

Task factors influencing the level of verbal coding

Previous studies demonstrated that the degree of verbal coding of visual information changes depending on the task engaged in. In Zelinsky and Murthy's study (2000), subjects were trained pair-associations between unfamiliar faces and names with different pronunciation length, and then were tested on studied face stimuli under either delayed recognition or visual search paradigms. When the task involved delayed recognition, subjects' eye gaze during encoding tended to stay longer with face stimuli associated with longer names than with those associated with shorter names. This object name length effect was observed during the delayed recognition task but not in the search task, suggesting that verbal coding of visual objects may not occur automatically during search. However, differences in oculomotor behavior as a function of object name length might have occurred due to other task-related dimensions, because oculomotor behaviors were measured during encoding in the delayed recognition task while measured during search in the visual search task. Therefore, it is possible that the observed oculomotor differences might have resulted due to the difference in task stages (i.e., encoding vs. search) rather than in task type. It is unclear whether verbal coding of visual objects during encoding changes as a task goal changes (i.e., recognition of studied items vs. search of a target). Nevertheless, the object name length effect observed during the delayed recognition task in Zelinsky and Murphy's study (2000) closely ties with the pronunciation length effect observed during the delayed visual recognition task used in the present study, especially when visual objects were presented for a longer time (i.e., 900 ms). Tversky (1975) demonstrated that subjects changed their encoding strategies for visual objects depending on the modality of test stimuli. That is,

29

subjects remembered the visual objects by their pre-associated names if they were later tested using names, while remembering the visual objects by their visual features if they were tested using visual objects. Similar findings were observed in the present study (Experiment 3) in that behavioral performance was higher in the word probe condition than the picture probe condition and in the word probe condition than the either-word probe condition Given the higher working memory capacity for verbal compared to visual items (Luck & Vogel, 2001; Miller, 1956), this result indicate that subjects tended to encode visual items into pictorial, verbal, or both pictorial and verbal format depending on the modality of the probe. However, it is worth noting that probe type did not abolish verbal coding of visual objects.

In this study, we further examined other task factors that may influence the level of verbal coding of visual objects in a more systematic way in the context of visual working memory. Our findings extended the previous studies by showing that, in addition to task goal, other parameters influenced the pronunciation length effect. In visual working memory studies aiming to study pure visual information processing in working memory, Vogel et al. (2001) examined the effect of presentation time of study stimuli and duration of delay on visual object working memory. When presentation time was lengthened from 100 ms to 500 ms, behavioral performance in a change detection task did not change across set sizes of 4, 8, and 12. When duration of delay was varied from 900 ms, 2.9 s, and 4.9 s, behavioral performance did not change, either. On the contrary, we found significant effects of presentation time (Experiment 1A) and interaction between pronunciation length and delay duration (Experiment 1C) on visual object working memory performance in the present study. Although direct comparison between results from Vogel et al. (2001) and those from the present study cannot be made, it is clear that task factors such as presentation time and duration of delay can differentially affect behavioral performance across visual object working memory studies.

In the present study, we manipulated not only pronunciation length of visual objects but also set size of visual items to be studied. The effect of set size on visual object working memory is quite well established (Luck & Vogel, 1997; Vogel et al., 2001). Our focus was on how pronunciation length interacts with set size rather than the main effect of set size on the current task. Clearly, results from Exps 1A-1C indicate that pronunciation length does not

differentiate behavioral performance in remembering visual objects in smaller set sizes. However, this pattern changed when labeling did not help in remembering visual objects as in Exp. 2, which showed a significant behavioral difference between short and long name object conditions at a set size of 1. However, it is inconclusive whether or not increased demand of visual processing creates a larger pronunciation length effect at a smaller set size or due to lack of effect of pronunciation length at a set size of 2 in the present data.

Multiple representations in visual object working memory

Another aim of the present study was to elucidate the cognitive mechanism underlying the relationship between verbal and visual information held in working memory. According to Baddeley's multi-component working memory model, information in working memory is represented in a domain-specific manner, either verbal or visuospatial. Then, how does the model explain simultaneous processing of information composed of multiple domains? One possible solution, according to the modified version of the model (Baddeley, 2000), is that the central executive plays a role in binding information from different domains (e.g., object and location) and the bound information is held in episodic buffer for further processes (e.g., manipulation).

In the present study, we sought to compare behavioral differences in visual object working memory as a function of probe type. Results from Experiment 3 suggest that phonological coding occurs and influences visual object working memory even though it is neither explicitly required nor useful, although no significant behavioral difference between the picture and either-picture conditions. Furthermore, behavioral performance in the either condition was worse compared to the word condition, suggesting that maintaining both visual and verbal aspects associated with visual object stimuli was more demanding than maintaining verbal information alone.

Additional supporting evidence for this distinction between attention-based binding process (i.e., involvement of the central executive) and automatic processing between verbal and visual information comes from several neuroimaging studies. Using fMRI, Prabhakaran et al. (2000) showed greater activity in the prefrontal cortex when remembering both letters

31

and locations of letters in a bound fashion than remembering each domain alone. Because the task required subjects to actively encode and maintain information from two domains (i.e., verbal and spatial) in an integrated manner, the results were interpreted as suggesting a neural basis of active binding of information from multiple domains in working memory. On the other hand, Postle and Hamidi (2007) showed enhanced activity in the left sylvian cortex during maintenance of visual objects along with the verbal interference task. The left sylvian cortex has been implicated in processing auditory and phonological information; therefore, the results can be interpreted as suggesting phonological processing during maintenance of unnamable objects. These latter results, however, cannot be explained by the prevalent view of binding information from multiple domains as interpreted in Prabhakaran et al. (2000). In Wheatley et al. (2005), two words, which were drawn from either the same object category (e.g., dog-lion) or different object categories (e.g., mango-chair), were consecutively presented while subjects underwent an fMRI session. This study showed that repetition suppression occurred not only in the brain regions implicated in verbal processing (i.e., left IFG) but also in the regions implicated in visual processing when the two consecutively presented words were drawn from the same object category compared to different object categories. Because the task simply required subjects to view the words presented on the screen at a relatively fast speed (i.e., 150 ms presentation per word with a 100 ms ISI), it is very unlikely that the task was demanding enough to require involvement of executive functions. These results thus suggest that the verbal semantic priming effect on visual areas occurs automatically, although it is unclear whether the priming effect is exclusively semantic and not phonological.

Similar to the experimental paradigms adopted in the aforementioned neuroimaging studies, in the present study, the either condition in Experiment 3 required subjects to actively maintain both verbal and visual information in working memory. Behavioral results showed that performance in the either condition was worse than that in the word condition. However, there was no performance difference between the either and picture conditions. Therefore, although speculative, the either condition required a binding mechanism similar to what Prabhakaran et al. (2000) suggested in their study, while the picture condition did not.

Not only concurrent representations of pictorial and verbal information, but also concurrent representations of visual word forms in working memory have been reported. Fiebach et al. (2006), using fMRI, demonstrated that not only phonological codes but also visual word forms of word stimuli were processed during verbal working memory tasks by showing increased activity in regions known as "visual word form areas" (i.e., inferotemporal cortex). These findings further suggest concurrent processing of information from different domains, even though the maintenance of information from the other domain was not explicitly required by the given task. The authors claimed that concurrent activation in the IT cortex during verbal working memory reflects the "frontally guided, sustained activity of pre-existing cortical language representations," which is consistent with the view that working memory representations are attention-based activations of long-term memory, as proposed by Cowan (1999). Therefore, this study further provides support for the possible two mechanisms (i.e., attention-based binding vs. automatic processing) underlying concurrent processing of information from different domains.

Taken together, the two storage buffers formulated in Baddeley's working memory model seem to be interconnected. One possible mechanism of this interconnectivity between storage buffers is attention-based binding achieved through the central executive and stored in the episodic buffer. Another potential mechanism is through automatic interaction between verbal and visual domains. That is, visual information in the visual storage buffer can be automatically transferred to verbal information in the phonological storage buffer, and vice versa. Research thus far has not provided enough evidence to differentiate these potential mechanisms underlying interaction between visual and verbal information. Results from the present study lend support for concurrent verbal processing of visual objects during visual object working memory, which occurs automatically, therefore, cannot be explained by Baddeley's working memory model. Based on the present findings, we argue that in working memory, information from different domains can be processed by spreading across domains by means of a previously established association, which may relate to the activation of long-term memory in working memory, as more recent theories have postulated (Cowan, 1995, 1999).

Limitations of the current studies

Although we examined the effect of presentation time, interstimulus interval, and duration of delay, we did not look at the effect of each task parameter in detail. Testing each task parameter at a finer scale would allow us to better understand changes in the level of phonological coding of visual objects.

By controlling for frequency of object names used in the present study, we sought to minimize any long-term memory effect that may have influenced the current data. However, it is still plausible that individual differences may have interacted with objects selected for the current study, although they may have occurred unsystematically between critical conditions (i.e., the length factor). Training subjects with newly associated sets of visual objects and names may elucidate the topics of interest while minimizing the influence of individual differences.

Conclusion

In the present study, we sought to investigate whether the level of phonological coding of visual objects is modulated by task parameters of presentation time, interstimulus interval, delay of duration, and probe type. We questioned the validity of three common methodologies adopted in visual working memory to minimize or eliminate any verbal coding. Therefore, we directly manipulated the verbal aspects that may play a role in visual working memory, instead of relying on methodologies that leave the possibility of an unsystematic effect of verbal confounds.

Our results indicate that longer presentation time enforces verbal coding of visual objects, but is not a sufficient factor. When we examined the effect of other task factors such as duration of delay and ISI on verbal coding of visual objects that were presented briefly (200 ms), pronunciation length effect was observable with lengthened delay duration, but not with lengthened ISI. This verbal coding does not seem to be engaging the central executive. While Baddeley's model has been guiding and stimulating a large body of research efforts on working memory, a suggested modification of the model in order to incorporate the

spreading of activation between different domains may encourage more communication regarding different working memory theories, and result in a better understanding of the nature of working memory.

III. Differential modulation of frontal, inferior temporal, and lateral occipital regions by pronunciation length during visual object working memory

3.1. Background

Visual object working memory is commonly referred to as a system for holding visually presented information such as colors, shapes, and objects for a short period of time (Baddeley, 1986). The prefrontal cortex (PFC) has long been viewed as the neural substrate supporting working memory (Goldman-Rakic, 1987; Miller & Cohen, 2001). A significant amount of research has supported the domain-specific maintenance mechanisms in working memory by demonstrating functional organization of the prefrontal cortex by material type (Courtney et al., 1998; Gruber & von Cramon, 2001; Leung et al., 2002; Levy & Goldman-Rakic, 1991; McCarthy et al., 1996; Mohr & Linden, 2005, 2006; Smith et al., 1996). Inconsistent findings, however, have been reported suggesting that the functional specialization associated with material type within the PFC is not clear (Gruber & von Cramon, 2001; Nystrom et al., 2000). Part of the reason, especially in human studies, has been attributed to the involvement of language in diverse cognitive domains (Gruber & von Cramon, 2001; Nystrom et al., 2000). Indeed, recent studies have demonstrated that brain regions implicated in phonological processing are engaged in non-verbal cognitive processes such as visual object working memory (Postle & Hamidi, 2007). Although a potential influence of language in humans on diverse aspects of cognitive processes, including working memory, has been noted (Gruber & von Cramon, 2001; Nystrom et al., 2000; Postle & Hamidi, 2007), few attempts have been made to directly examine the influence of verbal processes on visual object working memory. The present study aimed to investigate the neural correlates underlying the influence of verbal processes on visual object working memory using functional magnetic resonance imaging (fMRI).

3.1.1. Neural basis of visual object working memory

A hallmark of visual object working memory at the neuronal level is delay-spanning sustained activity in the brain without any simultaneous perceptual input. Single-cell and lesion studies in nonhuman primates have indicated that this delay-spanning sustained activity is observed in the posterior brain regions, specifically the inferior temporal cortex (IT), as well as the ventral part of the PFC (Gaffan & Murray, 1992; Iwai & Mishkin, 1969; Miller et al., 1993, 1996; Petrides, 2000). In neuroimaging studies with humans, similar results were obtained, indicating that both posterior brain regions and the ventral part of the PFC subserve the maintenance of visual object information during the short delay in working memory (Courtney et al., 1997; Druzgal & D'Esposito, 2003). However, the exact nature of the functions of each brain region in relation to working memory is under debate. Some studies suggest a role for the PFC in maintaining domain-specific information such as location and shapes (Courtney et al., 1997; Levy & Goldman-Rakic, 2001; Wilson et al., 1993), while others suggest the critical role of posterior brain regions such as the IT in maintaining domain-specific information (Postle et al., 2003). On the other hand, others suggest that the PFC is not involved in maintenance of domain-specific information per se, but resistance to distracters by exerting a top-down signal on task-relevant information that may be primarily maintained in the posterior brain regions (Miller et al., 1996).

3.1.2. Activation in brain regions traditionally considered to be involved in phonological processing during visual object working memory

Although nonhuman and human studies have been conducted in parallel in order to elucidate the role of the PFC and posterior brain regions in working memory, inconsistent findings exist between species. Unlike the results from the nonhuman literature, within the prefrontal cortex, regions implicated in both visual object processing and verbal information processing have been frequently observed in visual object working memory studies (Courtney et al., 1997; Linden et al., 2003; Mohr & Linden, 2006; Munk et al., 2002; Nystrom et al., 2000; Rama et al., 2001). In one such task using faces, Courtney et al. (1997) reported activation in the left posterior mid-frontal/inferior frontal gyrus (BA 9/44) and anterior insula/inferior-frontal gyrus (BA 45/47). These regions have been shown to be

involved in verbal processes including phonological and semantic aspects of verbal information (Awh et al., 1996; Schumacher et al., 1996; Poldrack et al., 1999; Wagner et al., 2001). Nystrom et al. (2000) reported greater activity in the left inferior frontal gyrus (BA 44/45) in response to shapes compared to location in working memory tasks. Comparing verbal and nonspatial visual information, Rama et al. (2001) showed no differential activation in prefrontal regions between famous names and faces, while hemispheric dissociation in this region was observed with non-famous names and faces. These findings suggest that there is a hemispheric dissociation in prefrontal regions depending on domains to be processed. For information that can be processed both verbally and visually, however, an activity pattern in the PFC becomes similar. More recently, Postle and Hamidi (2007) showed greater activity in the left sylvian region, the region traditionally considered to be involved in phonological processing, during visual object working memory with nonverbalizable abstract shapes, suggesting involvement of nonvisual codes in maintaining visual objects in working memory. Taken together, the observation of activations in brain regions implicated in both verbal and visual information processing during visual object working memory suggest potential involvement of concurrent verbal processing during visual object working memory. The underlying neural mechanisms of concurrent processing of verbal information during visual object working memory, however, are largely unknown.

3.1.3. Activation in brain regions traditionally considered to be involved in visual processing during verbal information processing

Accumulated evidence suggests that an influence of verbal processing on brain activity is not limited to brain regions that have been traditionally considered to be involved in verbal processing. Visual association areas such as the IT have been activated in verbal tasks where only words are presented as task stimuli (Binder et al., 2005; D'Esposito et al., 1997; Wheatley et al., 2005). Compared to abstract words, concrete words activated posterior brain regions, including the left angular gyrus, middle and inferior temporal gyri, and the dorsal PFC, to a greater extent (Binder et al., 2005). Using a mental imagery task, D'Esposito and colleagues (1997) showed greater activation in the left IT with concrete versus abstract

words. Furthermore, a repetition suppression effect in the IT regions was observed when sequentially presented words were drawn from the same category (Wheatley et al., 2005).

Regions involved in verbal processing have been further observed during object naming tasks. By asking subjects to name colored objects, Chao and Martin (1999) found activation in the right middle and superior temporal gyri and the left PFC, as well as posterior regions including the lingual gyrus bilaterally and the left fusiform gyrus. The encoding and maintenance of man-made objects, such as tools, were found to activate brain regions involved not only in visual features but also motion-related attributes associated with those objects (Mecklinger et al., 2002). Together, these studies suggest that activation in the IT is observed not only by active representation of visual information per se but also by naming visual objects. Furthermore, it is suggested that processing of visual objects involves cognitive processes other than visual processing per se, such as semantic processing. Therefore, existing evidence suggests that brain regions traditionally considered to be involved in visual processing are activated by processing other domains such as verbal information.

3.1.4. Neural basis of the word length effect

Previous neuroimaging studies on verbal working memory suggested that the left inferior frontal gyrus is involved in articulatory rehearsal, while the left supramarginal gyrus is involved in phonological storage (Awh et al., 1996; Paulesu et al., 1993; Schumacher et al., 1996), These regions have been further implicated in phonological processes in contrast with semantic processes of verbal information (Poldrack et al., 1999; Wagner et al., 2001). A recent meta-analysis on neuroimaging results from speech/word production studies identified a set of right hemispheric regions including the mid-superior temporal gyrus, medial and lateral cerebellum, and supplementary motor area (SMA) associated with the speech-production process (Indefrey & Levelt, 2004). More widespread activation was found in the left hemispheric regions including posterior inferior frontal gyrus, ventral precentral gyrus, SMA, and mid and posterior superior and middle temporal gyri, although precise language processes (e.g., syllabification, phonetic encoding, and articulation) were further associated

with each brain activation (Indefrey & Levelt, 2004), In relation to the word length effect, Chein and Fiez (2001) observed an increased, although non-significant, activity in the left inferior frontal gyrus during the maintenance of 3-syllabic words compared to the 1-syllabic ones. Using a picture naming task, Wilson et al. (2009) showed a positive correlation between the number of phonemes produced during picture naming and activity in primary auditory areas bilaterally, left superior temporal gyrus and left superior temporal sulcus. Okada et al. (2003) also observed increased activity in the left superior temporal gyrus during picture naming using the number of syllables as a word length measure.

3.1.5. Hypotheses of the present study

Previous studies have implicated the lateral PFC and posterior regions, particularly the IT, as neural substrates supporting visual object working memory. However, it is unclear what exact role each brain region plays in supporting visual object working memory. It is also unknown whether differential engagement of verbal regions modulates the level of activity in brain regions implicated in processing visual information. Using fMRI, we investigated whether activation in visual association areas is modulated by phonological processing, which was manipulated by pronunciation length of visual object names. Specifically, we hypothesized that activity in brain regions including the left inferior frontal gyrus (IFG) and left supramarginal gyrus (SMG), which are typically associated with maintenance of phonological information in verbal working memory, will be modulated by pronunciation length of visual objects. Furthermore, we hypothesize that activity in the ventral visual association areas will be modulated by pronunciation length of visual objects to be held in working memory.

Method

Participants

Sixteen healthy young adults (10 females, age 18-26 years with a mean of 20.8 years) were recruited from the Stony Brook campus. All participants had no history of neurological

disorder, psychiatric disorder, or drug abuse according to self-report and had normal or corrected vision. Prior to participation, they gave informed consent that was reviewed and approved by the Institutional Review Board of the State University of New York at Stony Brook.

Visual stimuli

Between-category namable colored objects from the Hemera database (Hemera Technologies Inc., Gatineau, Quebec, Canada) and internet were used as stimuli. We selected between-category items consisting of a study list because they were thought to reduce interference, which may influence performance beyond our manipulation of interest, compared to within-category stimuli (Wickens, 1970). Objects were common objects and comprised 6 categories with 4 exemplars within each category. Half of the categories were living objects such as four-footed animals, vegetables, and fruits. The other half were non-living objects such as furniture, tools, and vehicles. Although the living or non-living nature of visual object categories was not our primary interest in the present study, an equal number of living and non-living categories were chosen as stimuli, based on findings from recent neuroimaging studies which have reported category-sensitive organization of the ventral occipitotemporal regions (Martin, 2007; Moore & Price, 1999). Each object name had 4 unique pictures, resulting in 96 unique pictures for each length condition. All object pictures were equivalent in terms of brightness and size.

We conducted two pilot studies. The first pilot study was conducted in order to control visual complexity of pictures between the two length conditions using the name-picture agreement task (Murphy & Brownell, 1985). Briefly, participants were given an object name visually followed by an object picture and had to respond to the object picture as quickly as possible to indicate whether the picture was representative of the object name presented immediately before it. Reaction times were dependent measures used to reflect participants' rating of the representativeness and visual complexity of a given picture. This task was conducted with a larger set of pictures initially and the finalized set was not different between conditions in terms of reaction time (M = 461 ms, SD = 138.33 ms, for short names; M = 462 mc, SD = 123.17 ms, for long names, p > 0.5). Because one

41

independent variable in the current study was pronunciation length (i.e., objects were grouped into longer name (2-4 syllables) or shorter name (mostly 1 syllable) objects), the second pilot study was conducted to ensure pronunciation length differences between the two conditions according to a method adopted in previous studies (Boutla et al., 2004). The results showed a significant pronunciation length difference between the two length conditions (M = 352 ms, SD = 48 ms, for short names; M = 438 ms, SD = 74 ms, for long names, p = .006). Stimulus categories and characteristics of the finalized stimulus set are listed in Appendix.

In addition to these measures, written word frequency of object names were equated between conditions based on values determined from the English Lexicon Database to control for the influence of long-term memory (Balota et al., 2007; See Appendix). In order to balance visual features of objects selected for each condition, colors and shapes of visual objects were chosen to maximize similarity between pictures from each condition. Each picture was resized to be equivalent by area. Object pictures were subtended by a visual angle of 2° X 2° on average.

Working memory task

We used a 2 (pronunciation length: long vs. short names) X 2 (set size: 1 vs. 4) within-subjects design for the study. The experimental paradigm was a delayed recognition task shown in Figure 4A. After an initial fixation (a black cross) for 2.3 s followed by a 200 ms warning sign (a green cross) and a 500 ms fixation (a black cross), either 1 or 4 visual objects were presented consecutively at the center of the screen. Each study stimulus was presented for 900 ms with a 100 ms interstimulus interval (ISI). At the end of the last item, a color-noise pattern mask was presented for 200 ms. Then, a delay period of 10.8 s followed until a test probe appeared. The test probe was presented for 1 s with an additional 1.5 s response window during which subjects were required to make a yes/no judgment to indicate whether the test probe was one of study items. Responses were made by pressing one of two keys on the button box with either index or middle finger of the subject's dominant hand. Finger-response mapping was counterbalanced across subjects. For each pronunciation-length condition, 24 objects and 4 exemplars of each object were used repeatedly throughout

the experiment and drawn equally often as a study item and test probe. For study stimuli in the set size 4 condition, no category was repeated and half of the study stimuli were drawn from living categories while the other half were drawn from non-living categories in order to control for any category effect on neural activity of ventral occipitotemporal regions, as previously mentioned. A test probe for a "yes" response was selected from one of the studied items with the constraint that a study item from each serial position was selected equally often as a "yes" probe. A test probe for a "no" response was selected from the rest of the exemplars of the same object (i.e., a different picture of a strawberry was presented at the probe stage if a strawberry was shown during the study phase). The experimental task was conducted in 8 blocks with 12 trials in each block. There were 24 trials total for each condition. The number of "yes" and "no" trials was counterbalanced across blocks and conditions. The inter-trial interval (ITI) varied between 6.5 and 11 s with a mean of 8.5 s.

Localizer task

We conducted two localizer tasks to define the phonological and visual object processing regions for each subject (Friston et al., 2006; Saxe et al., 2006). To define phonological processing regions, we modified a task used by Poldrack et al. (1999). Subjects were presented a sequence of pseudo-words with which they performed either a syllable judgment (i.e., two syllable or not-two-syllable items) or a case judgment (i.e., uppercase or lowercase). Each task block began with a 2-s instruction cue (i.e., the word "syllable" or "case"), indicating what task subjects would perform with subsequently presented pseudo-words, followed by sequentially presented 8 pseudo-words. Each pseudo-word was presented for 1.2 s and separated by an ISI of 800 ms. Subjects were instructed to press an index or a middle finger key for 2 syllable or not-2-syllable items for the syllable judgment task and for uppercase or lowercase items for the case judgment task. Finger-response mapping was counterbalanced across subjects. Four blocks of each type of task were interspersed with an 18-s fixation period within a localizer run. The presentation order of the two tasks was counterbalanced across subjects such that they were presented in the order ABBABAAB for half of the subjects and presented in the order BAABABBA for the other half of subjects. In order to prevent any effect associated with verbal stimuli used for each type of task, pseudo-

43

words used for the syllable judgment task for half of the subjects were also used for the case judgment task for the other half of subjects. Each item appeared only once during the localizer task for each subject.

To define visual object processing regions, we used a standardized fMRI localizer task involving two conditions with intact objects and scrambled images (Kourtzi & Kanwisher, 2000). Intact objects consisted of two categories of living objects (i.e., fruits and four-footed animals) and two categories of non-living objects (i.e., furniture and tools). Each category constituted one block among the 4 intact object blocks. Scrambled images were created by dividing intact images in a 5 X 5 square grid and repositioning each of the resulting squares randomly. Four blocks of each condition were interleaved with an 18-s fixation period. Each block began with a 2-s instruction cue (i.e., the word "objects" or "patterns"), indicating what image subjects would encounter subsequently, and consisted of 8 images presented for 800 ms each with an ISI of 1.2 s. Subjects performed a 1-back task by pressing an index or a middle finger key for a match or nonmatch image, respectively. Finger-response mapping was counterbalanced across subjects.

Post-experiment tasks

After the fMRI session, 2 additional behavioral tasks were conducted. First, participants completed the naming test. Due to time constraints, one image out of 4 images was used for each object (24 short name and 24 long name objects, 48 images in total). In this test, each visual object was presented on a computer screen and participants were asked to name it out loud as soon as it appeared. This test was conducted in order to reexamine what label subjects associated with a given visual object during the main task. Second, we measured pronunciation rate with object names used in the main task by asking each subject to read a list of object names separately for each length condition as fast as they could (Boutla et al., 2004).

General experimental procedure

Visual stimuli were presented using E-Prime 2.0 (Psychology Software Tools, Pittsburgh, PA). Prior to the fMRI session, participants completed a name-picture agreement

test and practice sessions which took about 30 min together. The name-picture agreement test provided subjects with an opportunity to view each study stimulus once along with an object name prior to the main task. This test also allowed us to ensure that there was no difference in visual complexity between short and long name objects. In the fMRI session, subjects performed 8 blocks of the main task followed by 1-block verbal localizer and 1-block visual object localizer tasks. After the fMRI session, the two post-experimental tasks described above were conducted. A post-experiment questionnaire was administered to ask what strategy subjects used to perform the task and whether they noticed that the study stimuli were grouped by pronunciation length of object names.

FMRI data acquisition

Whole brain images were acquired using a Philips 3-Tesla Achieva system (Cleveland, OH) at the Stony Brook University Hospital. High-resolution anatomical images were acquired with a T1-weighted three-dimensional turbo field echo sequence (repetition time (TR) = 9.9 ms; echo time (TE) = 4.6 ms; flip angle (FA) = 25°; Field of View (FOV) = 256 X 256 mm). A T1-weighted inplane anatomical volume of 24 axial-oblique slices, parallel to the anterior commissure-posterior commissure (AC-PC), was acquired (TR = 300 ms, TE = 5 ms, FA = 60°, FOV = 220 X 220 mm, matrix size = 256 X 256, slice thickness = 5 mm). For the main task, volumes of functional images were acquired using a T2*-weighted EPI sequence (TR = 1.5 sec; TE = 30 ms; flip angle = 80°; FOV = 220 X 220 mm). Each functional volume consisted of 24 axial slices in the same orientation as the inplane images. Six dummy volumes were acquired at the beginning of each functional run to allow the MR signal to reach equilibrium; these images were discarded from the data set before image processing and analysis.

Image processing and analysis

All pre-processing and statistical analyses were conducted using SPM2 (Wellcome Department of Cognitive Neurology, London, U.K.). Functional images were corrected for differences in slice timing. Head motion was corrected using a 6-parameter rigid body correction to realign each image to the first volume of the middle run. Inplane and high-

resolution images were segmented into gray and white matter and coregistered with the mean functional image. Images were then normalized to the Montreal Neurological Institute (MNI) gray matter template brain using a 12-parameter affine registration followed by nonlinear transformations (Friston et al., 1995). Lastly, images were smoothed with a Gaussian kernel of 8 mm at a full width at half maximum (FWHM). For region of interest (ROI) analyses of individually defined brain regions, separate pre-processing without a normalization step was conducted for all functional images of each subject. For this non-normalized data, images were smoothed with a Gaussian kernel of 12 mm at a FWHM.

For the whole brain analysis, we used the General Linear Model (GLM) to construct a design matrix for each individual data set. We defined the onset times of stimulus presentation, early-delay, mid-delay, and probe as events for each condition in the working memory task. A regressor for the mid-delay period was entered after 4.5 s since the offset of the stimulus presentation in order to minimize any residual effect from the encoding-related activity (Postle et al., 2000). Motion regressors were included in the design matrix as variables of no interest.

We also conducted ROI analyses to examine cortical activity during the working memory task. To define ROIs for phonological and visual object processing, a separate GLM was constructed for each individual using data from the localizer tasks. For the verbal localizer task, epochs of each task (syllable judgment or case judgment) were modeled with a boxcar function and convolved with a canonical hemodynamic response function (HRF). ROIs for phonological processing were defined as clusters of activated voxels with a median of 100 voxels within regions of interest including left IFG and left SMG for most of subjects using the contrast of syllable vs. case judgement task blocks. For a few subjects, clusters for the left IFG region spread towards a premotor region and clusters for the left SMG region spread towards the angular gyrus. Regions identified as involved in phonological processing were obtained in all but one subject who did not take part in the localizer runs. As a replacement for beta values for the left IFG for this subject, a 13 X 13 X 13 cubic ROI was made centered around the group mean coordinates (x = -45, y = 12, z = 28) and a beta value was extracted from this ROI for that subject. For lSMG ROI for this subject, we made a 12 X 12 X 12 cubic ROI centered at the group mean coordinates (x = -40, y = -47, z = 35) and a

beta value was extracted from this ROI. ROIs for visual object processing were defined as activated voxels with a median of 200 voxels within posterior regions spanning IT, fusiform and lateral occipital gyri. The activated voxels in the posterior regions were identified from the visual localizer task by contrasting intact objects vs. scrambled images 1-back tasks. For 2 subjects who performed the visual localizer task run, no voxel within the specified posterior regions showing greater activity in the contrast of intact objects vs. scrambled images was identified. For these two subjects, we used the contrast of intact objects vs. baseline to define clusters with a size of 200 voxels and extracted a beta value from this cluster for the visual ROI analysis. Voxels were identified on each individual subject's non-normalized space. We used the MarsBar Matlab toolbox (Brett et al., 2002; http://marsbar.courceforge.net) to extract beta values from the activated voxels in each region of each individual and calculated group averaged beta values for each region by working memory task event (i.e., stimulus presentation, delay, and probe).

Results

Behavioral Results

Both response time and A' score data were analyzed by repeated measures analysis of variance (ANOVA) and *post hoc* paired t-tests. The results are summarized in Figure 4B and 4C. Data points 3 standard deviations away from the mean were not included in the response time analysis (less than 1% of total trials). Average response time for the short name condition (851.22 ms) tended to be faster than that for the long name condition (885.56 ms) [$F(1,15) = 4.306$, $p = .056$]. Paired t-tests indicated no response time difference between the two pronunciation length conditions at both set size 1 (757.72 ms for the short name condition; 794.27 ms for the long name condition) and set size 4 (944.72 ms for the short name condition; 976.86 ms for the long name condition). The main effect of set size was significant, $F(1, 15) = 51.155$, $p < .001$, but the interaction between length and set size was not significant ($p > .1$).

With regard to A' scores, performance in the short name condition was slightly better than that in the long name condition, but the performance difference did not reach

significance (F[1, 15] = 2.937, p = .107). The main effect of set size was significant (F [1,15] = 57.366, p < .001), while there was no significant interaction between pronunciation length and set size [F(1, 15) <1].

Pronunciation rates

Pronunciation rates were measured for each participant according to the methods by Boutla et al. (2004). For 2 participants, pronunciation rates were not measured due to unexpected experimental circumstances. With data from 14 participants, paired t-tests revealed a significant pronunciation rate difference between short name objects (M: 359.09 ms, SD: 66.60 ms) and long name objects (M: 421.93 ms, SD: 61.22 ms) [t(13) = -7.28, p < .001].

fMRI results: Group activation results – the pronunciation length effect

Results are shown in Figure 5A. In order to examine the pronunciation length effect on the delay-related activity, first we contrasted activity in the short name condition with the long name condition. With set size 1, greater activity in response to the short name condition was observed in the middle occipital gyrus (MOG), inferior and middle frontal gyri, left middle temporal gyrus (MTG), right superior temporal gyrus (STG), right lingual gyrus (LG), and left cuneus (p < 0.05, FDR corrected, cluster size ≥ 9). With set size 4, greater activity in response to the short name condition was found in the left middle frontal gyrus (MFG), left IFG, and right inferior and middle temporal gyri at a lower threshold (p < 0.001, uncorrected, cluster size ≥ 9). Across the 2 set size conditions, greater activity was found in multiple visual areas including the MOG and cuneus, temporal regions including the middle and superior temporal gyri, and frontal regions including the left precentral gyrus and left IFG (p < 0.001, uncorrected).

On the contrary, few regions showed greater activation during long name compared to short name conditions. Greater activity was observed in the thalamus at set size 4 at a lower threshold (p < 0.001, uncorrected, cluster size ≥ 9), while no suprathreshold activity was found at set size 1 with a even lower threshold (p < 0.005, uncorrected, cluster size ≥ 9). Across the 2 set size conditions, weak activity was found in the right IFG at a lower threshold

(p < 0.005, uncorrected, cluster size ≥ 9). Table 1 lists the peak coordinates and z values of the suprathreshold activations.

fMRI results: Group activation results – the set size effect

Results are shown in Figure 5B. In order to examine the effect of set size on delay-related activity, we contrasted activity to set size 4 with the set size 1 condition. For the short name condition, greater activity in response to set size 4 was observed in the left IFG, MFG bilaterally, left precuneus, left SMG, and left angular gyrus (p < .001, uncorrected, cluster size ≥ 9). For the long name condition, greater activity in response to set size 4 was observed in the left IFG, MFG, left inferior parietal lobe, right superior parietal lobe, and left precuneus (p < .001, uncorrected, cluster size ≥ 9). Across the 2 pronunciation length conditions, greater activity in response to the set size 4 condition was observed in the IFG, left SMG, left IPL, and left precuneus (p < .05, FDR corrected, cluster size ≥ 9). Table 2 lists the peak coordinates and z values of the suprathreshold activations.

fMRI results: Visual area ROIs

In order to examine whether activation in brain regions implicated in visual processing is modulated by pronunciation length of visual objects, we extracted a beta value from functionally defined ROI for each subject. This ROI for visual object processing was defined as voxels showing greater activity in response to objects than scrambled images from the visual localizer task. Because the regions that were identified as object processing regions from the localizer task varied across subjects, each ROI defined for each subject consists of different regions with a constraint that all regions belong to ventral posterior regions including inferior temporal, fusiform, and lateral occipital gyri. Based on previous studies, we were particularly interested in examining activity in these regions. Figure 6 shows the results. A 2-way ANOVA with pronunciation length and set size as factors revealed a significant main effect of pronunciation length showing greater activity during the delay period in response to the short name compared to the long name conditions, $[F(1, 14) = 8.538, p = .011]$. However, effects of set size and interaction were not significant (p's > .1).

fMRI results: Left IFG and SMG ROIs

In order to examine whether activation in brain regions implicated in verbal processing is also modulated by pronunciation length of visual objects, we extracted beta values from each functionally defined ROI for each subject. These ROIs were defined as showing greater activity in response to syllable judgment than case judgment during the verbal localizer task. Figure 6 shows the results. For the left IFG, a 2 (pronunciation length: short vs. long name objects) X 2 (set size: 1 vs. 4 items) ANOVA revealed a significant main effect of set size during the delay period [$F(1, 15) = 5.951$, $p = .028$], while neither main effect of pronunciation length [$F(1, 15) = 3.268$, $p = .091$] nor interaction between pronunciation length and set size [$F(1, 15) = .127$, $p = .727$] reached significance. For left SMG, similar patterns were obtained as for left IFG, although a main effect of set size approached significance, $F(1, 15) = 5.216$, $p = .06$, while neither main effect of pronunciation length [$F(1, 15) = 1.751$, $p = .207$] nor interaction between pronunciation length and set size [$F(1, 15) = .014$, $p = .907$] reached significance.

fMRI results: Correlation with behavioral data (pronunciation length effect)

In order to examine whether activation in any brain region correlated with individual differences in the pronunciation length effect, we conducted a correlation analysis using response time and accuracy rates as covariates. Two measures of the pronunciation length effect (pronunciation length effect index; PLE index) in each individual were used for this analysis: (1) response time data obtained by subtracting reaction time in the short name condition from reaction time in the long name condition and (2) accuracy data obtained by subtracting percent correct in the long name condition from percent correct in the short name condition. Therefore, a larger PLE index with response time indicates that subjects responded to short object names faster than long object names, while a larger PLE index with accuracy indicates that subjects' recognition performance in the short name condition was higher than that in the long name condition. These indices were correlated with activation contrast of short name > long name. Results from this correlation analysis revealed that PLE index of RT at set size 1 was correlated with greater activation contrast in the right middle frontal gyrus, precentral gyrus bilaterally, and left posterior cingulate gyrus at a lower threshold (p

< .001, uncorrected, cluster size ≥ 6). At set size 4, correlations were found in regions including middle frontal gyrus bilaterally, left superior frontal gyrus (SFG), and anterior cingulate gyrus at a lower threshold (p < .001, uncorrected, cluster size ≥ 6). Activation in the MFG and left SFG was significant at a cluster level (MFG, p = .002; SFG, p = .03). PLE index of accuracy data at set size 1 was correlated with greater activation contrast in the IFG bilaterally, right precentral gyrus, and right inferior parietal lobe at a lower threshold (p < .001, uncorrected, cluster size ≥ 9). Among these regions, activation in the left IFG was significant at a cluster level (p = .005, corrected; Figure 7). At set size 4, greater activity in the right precentral gyrus was observed at a very lenient threshold (p < .005, uncorrected, cluster size ≥ 9).

fMRI results: Correlation with pronunciation rates

In order to examine whether activation in any brain region correlated with individual differences in pronunciation rates, we conducted a correlation analysis using pronunciation rate difference between long name and short name objects (time to pronounce a short name object subtracted from time to pronounce a long name object) for each subject as covariates. With 14 subjects whose pronunciation rate was measured after the fMRI session, we found that faster pronunciation rate with short name objects was correlated with greater activation contrast between short name and long name conditions in the left inferior parietal lobe (IPL) and right precuneus at a lower threshold (p < .001, uncorrected). Activation in the left IPL was significant at a cluster level (p = .002, corrected). At set size 4, we found greater contrast activity in the left IPL at a lower threshold (p < .001, uncorrected).

Discussion

In the present study, we examined the neural fate of concurrent verbal processing during visual object working memory. Results revealed the modulatory effect of pronunciation length of visual object names on several brain regions including visual association regions, which has been implicated in processing visual information. Contrary to our prediction, no significant modulatory effect of pronunciation length was found in brain

51

regions traditionally considered to be involved in verbal processing. Although several working memory studies examined concurrent processing of information from different domains (Prabhakaran et al., 2000), no study examined the effect of verbal coding of visual objects during working memory tasks especially when the phonological aspects associated with visual information were systematically manipulated. Therefore, the current results are novel in that, when verbal information is systematically manipulated, its effect is not limited to brain regions implicated in phonological processing.

Maintenance of visual object representations in working memory

Neurophysiological studies with nonhuman primates have indicated that the inferior temporal cortex is involved in higher-level visual processing including object perception and object recognition frequently referred to as the ventral "what" stream (Mishkin et al., 1983). This ventral visual pathway includes the primary visual cortex such as V1/V2 and extends along the posterior regions such as the IT. While the lower-level early retinotopic visual areas such as V, V2, V3, V4/V8 were shown to be sensitive to visual properties such as orientation, luminance, color, and contrast, the higher-level object-sensitive regions such as the IT were shown to be sensitive to shapes (Miyashita & Chang, 1988). Neuroimaging studies with humans have identified brain regions homologous to the ventral object-sensitive regions in nonhuman primates and those regions were suggested to include the lateral occipital cortex and posterior fusiform/occipito-temporal sulcus which constitute so called "lateral occipital complex (LOC)" (for a review, Grill-Spector & Malach, 2004). The degree of activation in the LOC has been shown to be correlated with the degree of recognition performance (Grill-Spector et al., 2004). Furthermore, the lateral occipital gyri are shown to respond to pictures of common objects to a greater extent than to pictures of other categories such as faces or scenes (Grill-Spector et al., 1999, 2000). Ishai et al. (2000) have shown this category-sensitive activity in IOG and MOG during both perception and mental imagery of visual stimuli from different categories (e.g., faces and chairs).

Previous research on visual object working memory has indicated that these ventral visual pathways implicated in visual object perception are also involved in maintaining visual

information in working memory (Drutzgal & D'Esposito, 2003; Gazzeley et al., 2005; Postle et al., 2003; Xu & Chun, 2006). Furthermore, delay-related activity in these regions was shown to be sensitive to the task-related information. Using faces and scenes, Gazzeley et al. (2005) showed that activity in the fusiform gyrus (FG) and parahippocampal gyrus (PHG) was enhanced or reduced depending on the task-related category (either faces or scenes). Similar modulation of activity in the FG and PHG as a function of task-related category (i.e., either faces or scenes) was also observed in Lepsien and Nobre's study (2007), in which activity in the FG and PHG changed following the cue instruction (e.g., switch to remembering face) presented after the study stimuli in a delayed recognition task. Oh and Leung (in press) also showed category-sensitive activity in the PHG but more generalized activity in the FG for visual information. These studies together indicate that these posterior regions are involved in maintaining visual information in a relatively category-sensitive manner. However, it is worth mentioning that the cause of category-sensitive organization of these regions is under debate (Bedny et al., 2008; Caramazza & Mahon, 2006; Martin et al., 1996).

In the visual object working memory literature, the cognitive and neural mechanisms underlying verbal-related processes in maintaining visual object information in working memory have not been investigated, although the potential of concurrent verbal processing during visual working memory performance has been frequently reported (Linden et al., 2003; Nystrom et al., 2000; Munk et al., 2002 ; Postle & Hamidi, 2007). When we manipulated a phonological aspect associated with visual objects to be held in working memory, we found that widely spread brain regions including right IFG, right middle and superior temporal gyri, cuneus, and middle occipital gyrus showed greater activity in response to the short name than long name objects. Results from the individually defined ROIs contained in the posterior visual association area including IT, FG, IOG, and MOG further indicate that pronunciation length influences the level of activity in these visual association areas during the delay period in support of maintaining visual objects in working memory.

The involvement of the visual association areas has been reported not only when the visual objects were shown but also when only object names were presented (Binder et al.,

2005; D'Esposito et al., 1997; Wheatley et al., 2005). Binder et al. (2005) showed greater activity for concrete compared to abstract words in the left angular gyrus, middle and inferior temporal gyri, and the dorsal PFC. D'Esposito et al. (1997) found greater activity in the left IT with concrete than abstract words during a mental imagery task. Wheatley and colleagues (2005) showed that activity in the IT was reduced when consecutively presented words were semantically related (e.g., pig – fox) than when they were semantically unrelated (e.g., hat – gun). These studies clearly suggest a possible connectivity between brain regions implicated in verbal and visual information processing. Our findings are consistent with these previous findings by suggesting a potential connectivity between verbal and visual information processing during visual object working memory at the neural level. However, functional connectivity analysis will be necessary to further examine this possibility.

Verbal processing of visual information in visual object working memory

Previous behavioral studies have indicated an influence of verbal coding on visual information processing during visual perception and visual working memory (Dent & Snyder, 2005; Postle et al., 2005; Winawer et al., 2007). In the neuroimaging literature, Postle and Hamidi (2007) examined whether verbal processing is involved in visual object working memory using a dual task paradigm while subjects underwent an fMRI session. They reported greater activity increase in the left sylvian cortex when the secondary verbal task was performed along with the primary object working memory task than the primary location working memory task. Activity in the frontal eye fields, which has been implicated in oculomotor processing, however, showed the opposite pattern. Although this might be a reverse inference (Poldrack, 2006), these neuroimaging results suggest potential involvement of verbal coding in order to maintain shapes in working memory, considering the relatively established role of the left IFG and left sylvian regions in verbal processing.

Not only the left IFG and left sylvian regions but also the inferior parietal cortex including the left SMG has been implicated in phonological processing, although their respective role has been suggested to be different from each other. The left inferior PFC (BA 44) and the anterior insula have been suggested to be responsible for phonological rehearsal

of verbal information (Awh et al., 1996; Schumacher et al., 1996; Poldrack et al., 1999). The inferior parietal cortex, including the inferior SMG, and the left angular gyrus have been implicated as the locus of short-term phonological storage (Awh et al., 1996; Paulesu et al., 1993; Markowitsch et al., 1999; Rypma & D'Esposito, 1999; Vallar et al., 1997). Recent neuroimaging studies have further implicated the role of superior-temporal-parietal junctions (STP) and left sylvian regions in verbal working memory storage (Buchsbaum et al., 2005; Postle & Hamidi, 2007).

In the present study, however, when we examined the activity in these regions considered to be involved in phonological processing, we did not find any differential activity according to pronunciation length of visual objects, although activity in both left IFG and left SMG showed a sustained activity during the maintenance of visual objects. In a verbal working memory study, Chein and Fiez (2001) found similar results using verbal items showing no differential activity in the left IFG (BA 44) and left SMG between 1-syllable and 3-syllable study word lists. In the current study, however, these regions showed a set size effect, which may relate to the difficulty effect as shown in Chein and Fiez's study (2001). Therefore, it is possible that activity in the left IFG and left SMG reflects the maintenance of increased amount of verbal information, while it is not differentiated by pronunciation length.

Interaction between verbal and visual information in visual object working memory

By what mechanism does phonological coding influence activity in the posterior brain regions? One possible mechanism is through concurrent activation of visual word forms of object names initiated by verbal coding of visual object names. Studies have indicated that part of the IT and FG, responds to visually presented words, but not to non-words or a string of letters (Cohen & Dehaene, 2004; Fiebach et al., 2006; Starrfelt & Gerlach, 2007; but see Price & Devlin, 2004). If visual objects with shorter names were more often rehearsed than those with longer names, visual word forms associated with short name objects might have been mentally repeated more often, which results in greater activity in the visual association areas including the IT and FG.

Another possibility is through concurrent activation of visual images associated with object names. As the dual coding theory (Paivio, 1971, 1991) suggests, visual images associated with words can be concurrently activated when words are processed. Therefore, a negative correlation between brain activity in the ventral visual areas and pronunciation length of visual object names fits well with this proposed mechanism. In the verbal working memory literature, words with shorter pronunciation length were remembered better and more often than words with longer pronunciation length, because the former can be rehearsed more within a given amount of time than the latter (Baddeley, 1986). On the other hand, words with longer pronunciation length may fade away before they are rehearsed (Baddeley et al., 1975). If similar phonological rehearsal occurs concurrently while holding visual objects in working memory, visual objects with shorter names may have more chance to be rehearsed than those with longer names. Therefore, memory representation of the former will be stronger than that of the latter. Greater activity in the visual association areas observed in the present study can be interpreted as neural substrates underlying the benefit of short pronunciation length in maintaining visual objects in working memory.

In the present study, we did not specifically examine activity in the anterior PFC. The anterior PFC has been shown to be activated in studies examining active integration of information from different domains (Prabhakaran et al., 2000). Therefore, this region can be examined in future studies to confirm whether interaction between verbal and visual information processing is automatic or attention-based binding process.

Differential activity in the frontal and visual association regions as a function of set size

An effect of set size on brain activity is quite well established (Leung et al., 2004; Rypma & D'Esposito, 1999). Leung et al. (2004) found the modulatory effect of set size on brain activation in the spatial working memory circuit including MFG, frontal eye field, precuneus, superior parietal lobe, and inferior parietal lobe, and middle temporal gyrus throughout the delay period during the spatial working memory task. Using a verbal working memory task, Rypma and D'Esposito (1999) found an increased activity in the dorsal and ventral PFC for the 6 letter condition compared to a baseline, while no significant increase in

brain activity for the 2 letter condition compared to a baseline. Consistent with previous findings, we also found increased activity in the IFG, left SMG, and precuneus in response to set size 4 compared to set size 1 across the 2 length conditions. Additionally, greater activity in the MFG and angular gyrus and greater activity in the MFG, IPL, and SPL were revealed for the short name and long name conditions, respectively. However, we did not observe the modulatory effect of set size in the visual association regions in which the effect of pronunciation length was revealed. Therefore, the present findings may suggest that the visual association regions are involved in representing visual information but that activity in these regions are not sensitive to the number of visual information to be maintained, which is consistent with other findings (Jha & McCarthy, 2000).

Summary

Using fMRI, we examined whether activity in brain regions previously implicated in phonological and visual information processing is modulated by phonological aspects of visual objects that are held in working memory. Results indicate that activity in the visual association areas is modulated by pronunciation length of visual objects, while activity in the left IFG and left SMG, which have been implicated in verbal processing, was not influenced by this manipulation in the present study. Although we manipulated pronunciation length of visual object names, it is unclear whether the effect of pronunciation length on neural activity has reflected further influence of semantic processing. Future studies may examine this possibility.

IV. General Discussion

Using behavioral and fMRI methods, we examined behavioral and neural correlates underlying the pronunciation length effect on visual object working memory. In the 3 behavioral experiments, we found that the effect of pronunciation length on visual object working memory changed as a function of presentation time of study stimuli, ISI, and duration of delay (Experiment 1). The effect of pronunciation length was further strengthened when task demand of visual information processing increased (Experiment 2). This effect was still observed even though it was not necessary to maintain verbal information (Experiment 3). In the fMRI experiment, we found that the effect of pronunciation length was accompanied by increased brain activity in the visual association areas including the IT, FG, IOG, and MOG during the delay period. Activity in brain regions traditionally considered to be involved in phonological processing, however, was not affected by pronunciation length difference of visual objects.

Findings from the present study suggest that verbal coding of visual objects during visual object working memory occurs even though it is not necessary, but that the level of verbal coding is influenced by task factors such as presentation time of study stimuli, ISI, and duration of the delay. The present results further suggest that the modality of information to be maintained in working memory is not fixed but changes depending on the task factors. Neural findings suggest that concurrent activation of information from a different domain (i.e., object names in the present study) during working memory influences activity in other brain regions that were not traditionally attributed to this cognitive process. Based on the findings from the behavioral experiments, this concurrent processing of verbal information during visual object working memory occurs automatically and supposedly does not involve the central executive to actively maintain both verbal and visual information. Therefore, the modulatory effect of pronunciation length on activity in the visual association regions may suggest a potential feedback from regions implicated in phonological processing to regions implicated in visual processing.

The present study was designed to test an interaction between phonological and visual processing during visual object working memory. The present behavioral and neural findings

of the concurrent phonological processing during visual object working memory cannot be explained by Baddeley's current framework of working memory, because the model does not postulate an interaction between phonological and visual processing that does not require an involvement of the central executive. Rather, the results support Cowan's (1999) embedded process model of working memory, which theorizes widespread activation of long term memory with a focus of attention directed to an item in working memory. The present findings of a modulatory effect of pronunciation length on visual association regions further support Postle's (2006) recent account of working memory as being an emergent process by attention-based recruitment of multiple brain regions which are primarily engaged in sensory-, representation-, and action-related functions.

To summarize, behavioral and neural findings in the present study can shed light into the current understanding of working memory at both behavioral and neural levels. At the behavioral level, the present results suggest that concurrent activation of information from different domains can occur without engagement of the central executive, but that the level of observable phonological coding is constrained by task factors. At the neural level, the present findings suggest a functional connection between brain regions associated with phonological and visual information processing, possibly through a feedback from the former to the latter. Future studies may examine functional connectivity between regions engaged in visual object working memory to further determine dynamic interaction between brain regions and cognitive processes during visual object working memory.

V. Conclusion

       Visual object working memory has been commonly studied from the perspective of visual information processing. In order to determine the potential influence of concurrent phonological processing on visual information processing in working memory, we studied behavioral performance and neural activity while subjects hold visual objects of short and long names. Based on the present findings, we conclude that phonological processing occurs automatically during visual object working memory, although the degree of phonological coding is constrained by various task factors, and that concurrent phonological processing of visual objects in working memory influences neural activity in the visual association areas.

References

Allen, R. J., Baddeley, A. D., & Hitch, G. J. (2006). Is the binding of visual features in working memory resource-demanding? *Journal of Experimental Psychology: General, 135*(2), 298-313.

Alvarez, G. A., & Cavanagh, P. (2004). The Capacity of Visual Short Term Memory Is Set Both by Visual Information Load and by Number of Objects. *Psychological Science, 15*(2), 106-111.

Attneave, F., & Arnoult, M. D. (1956). The quantitative study of shape and pattern perception. *Psychological Bulletin, 53*(6), 452-471.

Avons, S. E., Wright, K. L., & Pammer, K. (1994). The word-length effect in probed and serial recall. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*(1), 207-231.

Awh, E., Jonides, J., Smith, E. E., Schumacher, E. H., Koeppe, R. A., et al. (1996). Dissociation of storage and rehearsal in verbal working memory: Evidence from positron emission tomography. *Psychological Science, 7*(1), 25-31.

Awh, E., Barton, B., & Vogel, E. K. (2007). Visual working memory represents a fixed number of items regardless of complexity. *Psychological Science, 18*(7), 622-628.

Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends Cogn Sci, 4*(11), 417-423.

Baddeley, A. (2003). Working memory: looking back and looking forward. *Nat Rev Neurosci, 4*(10), 829-839.

Baddeley, A. (1986). *Working memory*. Oxford: Oxford University Press.

Baddeley, A. D., Lewis, V., & Vallar, G. (1984). Exploring the articulatory loop. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 36*(2), 233-252.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. Bower (Ed.), *The psychology of learning and motivation (*Vol. 8, pp. 47-90). New York: Academic Press.

Baddeley, A. D., & Logie, R. H. (1999). Working memory: The multiple-component model. In A. Miyake & P. Shah (Eds.), *Models of working memory* (pp. 28–61). New York: Cambridge University Press.

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning & Verbal Behavior, 14*(6), 575-589.

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., et al. (2007). The English Lexicon project. *Behavior Research Methods, 39*(3), 445-459.

Bedny, M., Caramazza, A., Grossman, E., Pascual-Leone, A., & Saxe, R. (2008). Concepts are more than percepts: the case of action verbs. *J Neurosci, 28*(44), 11347-11353.

Binder, J. R., Westbury, C. F., McKiernan, K. A., Possing, E. T., & Medler, D. A. (2005). Distinct brain systems for processing concrete and abstract concepts. *J Cogn Neurosci, 17*(6), 905-917.

Boutla, M., Supalla, T., Newport, E. L., & Bavelier, D. (2004). Short-term memory span: Insights from sign language. *Nature Neuroscience, 7*(9), 997-1002.

Brener, R. (1940). An experimental investigation of memory span. *Journal of Experimental Psychology, 26*(5), 467-482.

Brett, M., Anton, J. L., Valabregue, R., & Poline, J. B. (2002). *Region of interest analysis using an SPM toolbox [abstract].* Paper presented at the Presented at the 8th International Conferance on Functional Mapping of the Human Brain.

Buchsbaum, B. R., Olsen, R. K., Koch, P., & Berman, K. F. (2005). Human dorsal and ventral auditory streams subserve rehearsal-based and echoic processes during verbal working memory. *Neuron, 48*(4), 687-697.

Caplan, D., Rochon, E., & Waters, G. S. (1992). Articulatory and phonological determinants of word length effects in span tasks. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*(2), 177-192.

Caramazza, A., & Mahon, B. Z. (2006). The organisation of conceptual knowledge in the brain: The future's past and some future directions. *Cognitive Neuropsychology, 23*(1), 13-38.

Chao, L. L., & Martin, A. (1999). Cortical regions associated with perceiving, naming, and knowing about colors. *J Cogn Neurosci, 11*(1), 25-35.

Chein, J. M., & Fiez, J. A. (2001). Dissociation of verbal working memory system components using a delayed serial recall task. *Cereb Cortex, 11*(11), 1003-1014.

Cohen, L., & Dehaene, S. (2004). Specialization within the ventral stream: the case for the visual word form area. *Neuroimage, 22*(1), 466-476.

Collette, F., & Van der Linden, M. (2002). Brain imaging of the central executive component of working memory. *Neuroscience & Biobehavioral Reviews, 26*(2), 105-125.

Coltheart, V., & Langdon, R. (1998). Recall of short word lists presented visually at fast rates: Effects of phonological similarity and word length. *Memory & Cognition, 26*(2), 330-342.

Coltheart, V., Mondy, S., Dux, P. E., & Stephenson, L. (2004). Effects of Orthographic and Phonological Word Length on Memory for Lists Shown at RSVP and STM Rates. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(4), 815-826.

Courtney, S. M., Petit, L., Maisog, J. M., Ungerleider, L. G., & Haxby, J. V. (1998). An area specialized for spatial working memory in human frontal cortex. *Science, 279*(5355), 1347-1351.

Courtney, S. M., Ungerleider, L. G., Keil, K., & Haxby, J. V. (1997). Transient and sustained activity in a distributed neural system for human working memory. *Nature, 386*(6625), 608-611.

Cowan, N. (1995). *Attention and memory: An integrated framework*. New York, NY: Oxford University Press.

Cowan, N. (1999). An Embedded-Processes Model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control.* (pp. 62-101). New York, NY: Cambridge University Press.

Cowan, N., Nugent, L. D., Elliott, E. M., & Geer, T. (2000). Is there a temporal basis of the word length effect? A response to Service (1998). *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 53*(3), 647-660.

Cowan, N., Day, L., Saults, J. S., & Keller, T. A. (1992). The role of verbal output time in the effects of word length on immediate memory. *Journal of Memory and Language, 31*(1), 1-17.

Dent, K., & Smyth, M. M. (2005). Verbal coding and the storage of form-position associations in visual-spatial short-term memory. *Acta Psychologica, 120*(2), 113-140.

D'Esposito, M., Detre, J. A., Aguirre, G. K., Stallcup, M., Alsop, D. C., Tippet, L. J., et al. (1997). A functional MRI study of mental image generation. *Neuropsychologia, 35*(5), 725-730.

D'Esposito, M., Detre, J. A., Alsop, D. C., Shin, R. K., Atlas, S., & Grossman, M. (1995). The neural basis of the central executive system of working memory. *Nature, 378*(6554), 279-281.

Druzgal, T. J., & D'Esposito, M. (2003). Dissecting contributions of prefrontal cortex and fusiform face area to face working memory. *Journal of Cognitive Neuroscience, 15*(6), 771-784.

Fiebach, C. J., Rissman, J., & D'Esposito, M. (2006). Modulation of inferotemporal cortex activation during verbal working memory maintenance. *Neuron, 51*(2), 251-261.

Friston, K. J., Ashburner, J., Frith, C. D., Poline, J. B., Heather, J. D., & Frackowiak, R. S. J. (1995). Spatial registration and normalization of images. *Human Brain Mapping, 3*(3), 165-189.

Friston, K. J., Rotshtein, P., Geng, J. J., Sterzer, P., & Henson, R. N. (2006). A critique of functional localisers.[see comment]. *Neuroimage, 30*(4), 1077-1087.

Funahashi, S., Bruce, C. J., & Goldman-Rakic, P. S. (1989). Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *Journal of Neurophysiology, 61*(2), 331-349.

Gaffan, D., & Murray, E. A. (1992). Monkeys (Macaca fascicularis) with rhinal cortex ablations succeed in object discrimination learning despite 24-hr intertrial intervals and fail at matching to sample despite double sample presentations. *Behav Neurosci, 106*(1), 30-38.

Gazzaley, A., Cooney, J. W., McEvoy, K., Knight, R. T., & D'Esposito, M. (2005). Top-down enhancement and suppression of the magnitude and speed of neural activity. *Journal of Cognitive Neuroscience, 17*(3), 507-517.

Goldman-Rakic, P. S. (1987). Circuitry of primate prefrontal cortex and regulation of behavior by representational memory. In V. B. Mountcastle & F. Plum (Eds.),

Handbook of Physiology: The Nervous System, Higher Functions of the Brain (Vol. 5, pp. 373-417). Bethesda: American Physiological Society.

Grill-Spector, K., Knouf, N., & Kanwisher, N. (2004). The fusiform face area subserves face perception, not generic within-category identification. *Nat Neurosci, 7*(5), 555-562.

Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzchak, Y., & Malach, R. (1999). Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron, 24*(1), 187-203.

Grill-Spector, K., Kushnir, T., Hendler, T., & Malach, R. (2000). The dynamics of object-selective activation correlate with recognition performance in humans. *Nat Neurosci, 3*(8), 837-843.

Grill-Spector, K., & Malach, R. (2004). The human visual cortex. *Annu Rev Neurosci, 27*, 649-677.

Gruber, O., & von Cramon, D. Y. (2001). Domain-specific distribution of working memory processes along human prefrontal and parietal cortices: a functional magnetic resonance imaging study. *Neuroscience Letters, 297*(1), 29-32.

Hockley, W. E., & Murdock, B. B. (1987). A decision model for accuracy and response latency in recognition memory. *Psychological Review, 94*(3), 341-358.

Howard, D., & Franklin, S. (1987). Three ways for understanding written words, and their use in two contrasting cases of surface dyslexia (together with an odd routine for making 'orthographic' errors in oral word production). In A. Allport, D. MacKay, W. Prinz & E. Scheerer (Eds.), *Language perception and production: Relationships between listening, speaking, reading, and writing* (pp. 340-366). London: Academic Press.

Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition, 92*(1-2), 101-144.

Intraub, H. (1981). Rapid conceptual identification of sequentially presented pictures. *Journal of Experimental Psychology: Human Perception and Performance, 7*(3), 604-610.

Intraub, H. (1984). Conceptual masking: The effects of subsequent visual events on memory for pictures. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*(1), 115-125.

Ishai, A., Ungerleider, L. G., & Haxby, J. V. (2000). Distributed neural systems for the generation of visual images. *Neuron, 28*(3), 979-990.

Iwai, E., & Mishkin, M. (1969). Further evidence on the locus of the visual area in the temporal lobe of the monkey. *Exp Neurol, 25*(4), 585-594.

Jha, A. P., & McCarthy, G. (2000). The influence of memory load upon delay-interval activity in a working-memory task: an event-related functional MRI study. *Journal of Cognitive Neuroscience, 12 Suppl 2*, 90-105.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99*(1), 122-149.

Kourtzi, Z., & Kanwisher, N. (2000). Cortical regions involved in perceiving object shape. *J Neurosci, 20*(9), 3310-3318.

Lepsien, J., & Nobre, A. C. (2007). Attentional modulation of object representations in working memory. *Cereb Cortex, 17*(9), 2072-2083.

Leung, H. C., Gore, J. C., & Goldman-Rakic, P. S. (2002). Sustained mnemonic response in the human middle frontal gyrus during on-line storage of spatial memoranda. *Journal of Cognitive Neuroscience, 14*(4), 659-671.

Leung, H. C., Seelig, D., & Gore, J. C. (2004). The effect of memory load on cortical activity in the spatial working memory circuit. *Cogn Affect Behav Neurosci, 4*(4), 553-563.

Levy, R., & Goldman-Rakic, P. S. (2000). Segregation of working memory functions within the dorsolateral prefrontal cortex. *Experimental Brain Research, 133*(1), 23-32.

Lian, A., Karlsen, P. J., & Winsvold, B. (2001). A re-evaluation of the phonological similarity effect in adults' short-term memory of words and nonwords. *Memory, 9*(4-6), 281-299.

Linden, D. E., Bittner, R. A., Muckli, L., Waltz, J. A., Kriegeskorte, N., Goebel, R., et al. (2003). Cortical capacity constraints for visual working memory: dissociation of fMRI load effects in a fronto-parietal network. *Neuroimage, 20*(3), 1518-1530.

Logie, R. H. (1995). *Visuo-spatial working memory*. Hillsdale, NJ, England: Lawrence Erlbaum Associates Inc.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390*(6657), 279-281.

MacAndrew, D. K., Klatzky, R. L., Fiez, J. A., McClelland, J. L., & Becker, J. T. (2002). The Phonological-Similarity Effect differentiates between two working memory tasks. *Psychological Science, 13*(5), 465-468.

Markowitsch, H. J., Kalbe, E., Kessler, J., von Stockhausen, H. M., Ghaemi, M., & Heiss, W. D. (1999). Short-term memory deficit after focal parietal damage. *J Clin Exp Neuropsychol, 21*(6), 784-797.

Martin, A. (2007). The representation of object concepts in the brain. *Annu Rev Psychol, 58*, 25-45.

Martin, A., Wiggs, C. L., Ungerleider, L. G., & Haxby, J. V. (1996). Neural correlates of category-specific knowledge. *Nature, 379*(6566), 649-652.

Martin, R. C., Lesch, M. F., & Bartha, M. C. (1999). Independence of input and output phonology in word processing and short-term memory. *Journal of Memory and Language, 41*(1), 3-29.

McCarthy, G., Puce, A., Constable, T., Krystal, J. H., Gore, J. C., & Goldman-Rakic, P. (1996). Activation of Human Prefrontal Cortex during Spatial and Nonspatial Working Memory Tasks Measured by Functional MRI. *Cereb. Cortex, 6*(4), 600-611.

Mecklinger, A., Gruenewald, C., Besson, M., Magnie, M. N., & Von Cramon, D. Y. (2002). Separable neuronal circuitries for manipulable and non-manipulable objects in working memory. *Cerebral Cortex, 12*(11), 1115-1123.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience, 24*, 167-202.

Miller, E. K., Erickson, C. A., & Desimone, R. (1996). Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *Journal of Neuroscience, 16*(16), 5154-5167.

Miller, E. K., Li, L., & Desimone, R. (1993). Activity of neurons in anterior inferior temporal cortex during a short-term memory task. *Journal of Neuroscience, 13*(4), 1460-1478.

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review, 63*(2), 81-97.

Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision: Two cortical pathways. *Trends in Neurosciences, 6*(10), 414-417.

Miyashita, Y., & Chang, H. S. (1988). Neuronal correlate of pictorial short-term memory in the primate temporal cortex. *Nature, 331*(6151), 68-70.

Mohr, H. M., Goebel, R., & Linden, D. E. (2006). Content- and task-specific dissociations of frontal activity during maintenance and manipulation in visual working memory. *Journal of Neuroscience, 26*(17), 4465-4471.

Mohr, H. M., & Linden, D. E. J. (2005). Separation of the Systems for Color and Spatial Manipulation in Working Memory Revealed by a Dual-task Procedure. *Journal of Cognitive Neuroscience, 17*(2), 355-366.

Monsell, S. (1987). On the relation between lexical input and output pathways for speech. In A. Allport, D. MacKay, W. Prinz & E. Scheerer (Eds.), *Language perception and production: Relationships between listening, speaking, reading, and writing* (pp. 273-311). London: Academic Press.

Moore, C. J., & Price, C. J. (1999). A functional neuroimaging study of the variables that generate category-specific object processing differences. *Brain, 122 (Pt 5)*, 943-962.

Mueller, S. T., Seymour, T. L., Kieras, D. E., & Meyer, D. E. (2003). Theoretical Implications of Articulatory Duration, Phonological Similarity, and Phonological Complexity in Verbal Working Memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*(6), 1353-1380.

Munk, M. H., Linden, D. E., Muckli, L., Lanfermann, H., Zanella, F. E., Singer, W., et al. (2002). Distributed cortical systems in visual short-term memory revealed by event-related functional magnetic resonance imaging. *Cerebral Cortex, 12*(8), 866-876.

Murphy, G. L., & Brownell, H. H. (1985). Category differentiation in object recognition: Typicality constraints on the basic category advantage. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*(1), 70-84.

Nimmo, L. M., & Roodenrys, S. (2005). The phonological similarity effect in serial recognition. *Memory, 13*(7), 773-784.

Nystrom, L. E., Braver, T. S., Sabb, F. W., Delgado, M. R., Noll, D. C., & Cohen, J. D. (2000). Working memory for letters, shapes, and locations: fMRI evidence against stimulus-based regional organization in human prefrontal cortex. *Neuroimage, 11*(5 Pt 1), 424-446.

Oh, H., & Leung, H.-C. (in press). Specific and non-specific neural activity during selective processing of visual representations in working memory. *Journal of Cognitive Neuroscience*.

Okada, K., Smith, K. R., Humphries, C., & Hickok, G. (2003). Word length modulates neural activity in auditory cortex during covert object naming. *Neuroreport, 14*(18), 2323-2326.

Olson, I. R., Chun, M. M., & Anderson, A. K. (2001). Effects of phonological length on the attentional blink for words. *Journal of Experimental Psychology: Human Perception and Performance, 27*(5), 1116-1123.

Paivio, A. (1971). *Imagery and verbal processes*. Oxford, England: Holt Rinehart & Winston.

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology/Revue canadienne de psychologie, 45*(3), 255-287.

Paivio, A., & Csapo, K. (1973). Picture superiority in free recall: Imagery or dual coding? *Cognitive Psychology, Vol. 5*(2), 176-206.

Paulesu, E., Frith, C. D., & Frackowiak, R. S. (1993). The neural correlates of the verbal component of working memory. *Nature, 362*(6418), 342-345.

Petrides, M. (2000). Dissociable roles of mid-dorsolateral prefrontal and anterior inferotemporal cortex in visual working memory. *J Neurosci, 20*(19), 7496-7503.

Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends Cogn Sci, 10*(2), 59-63.

Poldrack, R. A., Wagner, A. D., Prull, M. W., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. (1999). Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. *Neuroimage, 10*(1), 15-35.

Postle, B. R. (2006). Working memory as an emergent property of the mind and brain. *Neuroscience, 139*(1), 23-38.

Postle, B. R., Druzgal, T. J., & D'Esposito, M. (2003). Seeking the neural substrates of visual working memory storage. *Cortex, 39*(4-5), 927-946.

Postle, B. R., & Hamidi, M. (2007). Nonvisual codes and nonvisual brain areas support visual working memory. *Cerebral Cortex, 17*(9), 2151-2162.

Postle, B. R., D'Esposito, M., & Corkin, S. (2005). Effects of verbal and nonverbal interference on spatial and object visual working memory. *Memory & Cognition, 33*(2), 203-212.

Postle, B. R., Zarahn, E., & D'Esposito, M. (2000). Using event-related fMRI to assess delay-period activity during performance of spatial and nonspatial working memory tasks. *Brain Research. Brain Research Protocols, 5*(1), 57-66.

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory, 2*(5), 509-522.

Potter, M. C. (1993). Very short-term conceptual memory. *Memory & Cognition, 21*(2), 156-161.

Prabhakaran, V., Narayanan, K., Zhao, Z., & Gabrieli, J. D. (2000). Integration of diverse information in working memory within the frontal lobe. *Nat Neurosci, 3*(1), 85-90.

Price, C. J., & Devlin, J. T. (2004). The pro and cons of labelling a left occipitotemporal region: "the visual word form area". *Neuroimage, 22*(1), 477-479.

Rama, P., Sala, J. B., Gillen, J. S., Pekar, J. J., & Courtney, S. M. (2001). Dissociation of the neural systems for working memory maintenance of verbal and nonspatial visual information. *Cogn Affect Behav Neurosci, 1*(2), 161-171.

Rao, S. C., Rainer, G., & Miller, E. K. (1997). Integration of what and where in the primate prefrontal cortex. *Science, 276*(5313), 821-824.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85*(2), 59-108.

Rudkin, S. J., Pearson, D. G., & Logie, R. H. (2007). Executive processes in visual and spatial working memory tasks. *The Quarterly Journal of Experimental Psychology, 60*(1), 79-100.

Rypma, B., & D'Esposito, M. (1999). The roles of prefrontal brain regions in components of working memory: effects of memory load and individual differences. *Proc Natl Acad Sci U S A, 96*(11), 6558-6563.

Saxe, R., Brett, M., & Kanwisher, N. (2006). Divide and conquer: a defense of functional localizers. *Neuroimage, 30*(4), 1088-1096; discussion 1097-1089.

Schiano, D. J., & Watkins, M. J. (1981). Speech-like coding of pictures in short-term memory. *Memory & Cognition, 9*(1), 110-114.

Schumacher, E. H., Lauber, E., Awh, E., Jonides, J., Smith, E. E., & Koeppe, R. A. (1996). PET evidence for an amodal verbal working memory system. *Neuroimage, 3*(2), 79-88.

Smith, E. E., Jonides, J., & Koeppe, R. A. (1996). Dissociating Verbal and Spatial Working Memory Using PET. *Cereb. Cortex, 6*(1), 11-20.

Snodgrass, J. G., Wasser, B., Finkelstein, M., & Goldberg, L. B. (1974). On the fate of visual and verbal memory codes for pictures and words: Evidence for a dual coding mechanism in recognition memory. *Journal of Verbal Learning & Verbal Behavior, Vol. 13*(1), 27-37.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments & Computers, 31*(1), 137-149.

Starrfelt, R., & Gerlach, C. (2007). The visual what for area: words and pictures in the left fusiform gyrus. *Neuroimage, 35*(1), 334-342.

Stefurak, D. L., & Boynton, R. M. (1986). Independence of memory for categorically different colors and shapes. *Perception & Psychophysics, 39*(3), 164-174.

Sternberg, S. (1966). High-speed scanning in human memory. *Science, 153*(3736), 652-654.

Sternberg, S. (1967). Retrieval of Contextual Information from Memory. *Psychonomic Science, 8*(2), 55-56.

Swanson, J. M. (1974). The neglected negative set. *Journal of Experimental Psychology, 103*(5), 1019-1926.

Tehan, G., Hendry, L., & Kocinski, D. (2001). Word length and phonological similarity effects in simple, complex, and delayed serial recall tasks: Implications for working memory. *Memory, 9*(4-6), 333-348.

Ternes, W., & Yuille, J. C. (1972). Words and pictures in an STM task. *Journal of Experimental Psychology, 96*(1), 78-86.

Tversky, B. (1969). Pictorial and verbal encoding in a short-term memory task. *Perception & Psychophysics, 6*(4), 225-233.

Vallar, G., Di Betta, A. M., & Silveri, M. C. (1997). The phonological short-term store-rehearsal system: patterns of impairment and neural correlates. *Neuropsychologia, 35*(6), 795-812.

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance, 27*(1), 92-114.

Wagner, A. D., Pare-Blagoev, E. J., Clark, J., & Poldrack, R. A. (2001). Recovering meaning: left prefrontal cortex guides controlled semantic retrieval. *Neuron, 31*(2), 329-338.

Wheatley, T., Weisberg, J., Beauchamp, M. S., & Martin, A. (2005). Automatic priming of semantically related words reduces activity in the fusiform gyrus. *J Cogn Neurosci, 17*(12), 1871-1885.

Wheeler, M. E., & Treisman, A. M. (2002). Binding in short-term visual memory. *Journal of Experimental Psychology: General, 131*(1), 48-64.

Wickens, D. D. (1970). Encoding categories of words: An empirical approach to meaning. *Psychological Review, 77*(1), 1-15.

Wilson, F. A., Scalaidhe, S. P., & Goldman-Rakic, P. S. (1993). Dissociation of object and spatial processing domains in primate prefrontal cortex. *Science, 260*(5116), 1955-1958.

Wilson, S. M., Isenberg, A. L., & Hickok, G. (2009). Neural correlates of word production stages delineated by parametric modulation of psycholinguistic variables. *Human Brain Mapping*.

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proc Natl Acad Sci U S A, 104*(19), 7780-7785.

Xu, Y., & Chun, M. M. (2006). Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature, 440*(7080), 91-95.

Zelinsky, G. J., & Murphy, G. L. (2000). Synchronizing visual and language processing: An effect of object name length on eye movements. *Psychological Science, 11*(2), 125-131.

Tables

Table 1. Results of t-tests for active regions in response to pronunciation length by set size and across the 2 set size conditions. (A) Regions showing greater activity for the short name compared to the long name condition at set size 1 (p < 0.05, FDR corrected). (B) Regions showing greater activity for the short name compared to the long name condition at set size 4 (p < 0.001, uncorrected). (C) Regions showing greater activity for the long name compared to the short name condition at set size 1 (p < 0.005, uncorrected). (D) Regions showing greater activity for the long name compared to the short name condition at set size 4 (p < 0.001, uncorrected). (E) Regions showing greater activity for the short name compared to the long name conditions across the 2 set size conditions (p < 0.001, uncorrected). (F) Regions showing greater activity for the long name compared to the short name condition across the 2 set size conditions (p < 0.005, uncorrected). Clusters were 9 contiguous voxels or larger. The table shows the peak coordinates in mm, t values (T), z scores (Z) and anatomical names of the major clusters. BA-Brodmann's area, Cu-Cuneus, CS-Calcarine sulcus, HG-Hippocampal gyrus, LG-Lingual gyrus, MOG-Middle occipital gyrus, PHG-Parahippocampal gyrus, HG-Hippocampal gyrus, AG-Angular gyrus, PoCG-Postcentral gyrus, PrCG-Precentral gyrus, MTG-Middle temporal gyrus, STG-Superior temporal gyrus, SFG-Superior frontal gyrus, MFG-Middle frontal gyrus, IFG-Inferior frontal gyrus, CG-Cingulate gyrus, Th-Thalamus, SPL-Superior parietal Lobule, PCu-Precuneus,OG-Occipital gyrus, CS-Calcarine sulcus. S1-short name, set size 1, S4- short name, set size 4, L1-long name, set size 1, L4-long name, set size 4.

Lobe	Region	BA	Cluster size	MNI coordinates			T	Z
				x	y	z		
A. S1 vs. L1								
Frontal	IFG	44	9	48	18	15	4.5	3.53
	IFG	47	14	33	27	-21	4.38	3.46

	IFG	47			27	21	-24	3.77	3.11
	MFG	8	15		39	24	45	4.91	3.73
Parietal	CG	31	499		15	-33	39	5.35	3.94
	POCG	2			66	-21	24	4.21	3.37
	SPL	7	13		33	-60	54	4.38	3.46
Temporal	PoCG	16			-48	-9	18	6.51	4.42
	PoCG				-36	-15	24	4.45	3.5
	Insula[a]				-30	-21	-6	4.7	3.63
	MTG	19			-45	-75	21	4.21	3.37
		21			12	-33	-36	5.53	4.02
	MTG	21	97		-42	-21	-3	5.44	3.98
	STG	22	102		60	-33	12	5.2	3.87
	STG	42			60	-15	15	3.68	3.06
	MTG	39	185		48	-54	12	5.15	3.85
	STG	22	48		39	-15	6	4.78	3.67
	STG	22			48	-15	0	4.72	3.64
	PoCG	40			39	-12	15	4.44	3.5
	MTG	21	11		60	-45	-9	4.03	3.27
Occipital	MOG	19	189		-39	-63	9	5.82	4.15
	MOG/MTG	19/37			-57	-63	3	4.59	3.57
	Cu	18	168		-15	-81	9	5.37	3.95
	MOG	19			-24	-72	9	4.82	3.69
	CG/Cu	31			-18	-57	15	4.01	3.25
	PCu	7			-15	-48	57	5.24	3.89
	PCu	7			9	-60	45	4.92	3.74
	MOG	19			42	-72	15	5.13	3.84
	MOG	39			54	-66	9	4.6	3.58
	LG	18	65		18	-72	3	4.56	3.55
	MOG	19			21	-78	9	4.51	3.53
	CG	31			12	-63	12	3.93	3.21

				x	y	z		
Subcortical	HG			-21	-30	-3	4.68	3.62
	TH	23		9	-21	0	4.31	3.42
B. S4 vs. L4								
	IFG/MFG	11/47	16	-24	30	-18	5.67	4.08
	ITG/MTG	20/21	14	33	6	-30	5.03	3.79
C. L1 vs. S1								
	None							
D. L4 vs. S4								
	Th		10	3	-6	-9	4.89	3.72
E. (S1 + S4) − (L1 + L4)								
Frontal	PrCG	6	11	-36	-3	42	5.51	4.01
	IFG	47	9	-27	33	-15	4.53	3.54
Parietal	CG	31	13	15	-33	39	4.81	3.69
Temporal	MTG	21	34	-36	-9	-18	6.91	4.57
	MTG	21		-42	-21	-6	4.76	3.66
	MTG[b]	21		60	-45	-6	4.78	3.67
	MTG/HG	21/35	29	30	6	-27	5.76	4.12
	STG	22	16	60	-30	15	4.1	3.3
Occipital	PCu	7	241	-6	-54	60	6.41	4.38
	PCu	7		-6	-66	48	5.23	3.89
	PCu	7		12	-60	54	4.52	3.54
	MTG/MOG	39/19	190	54	-69	12	6.24	4.32
	MOG	19		48	-75	6	4.94	3.75
	MOG	19	58	21	-78	6	5.84	4.16
	MOG	19		33	-75	6	4.27	3.4

	OG	19	47		-39	-75	27	4.98	3.77
	CS	17	37		-21	-69	12	4.81	3.68
	Cu	19			-15	-78	27	4.45	3.5
	Cu	18			-18	-72	21	4.12	3.32
	MOG	19	16		-48	-72	9	4.22	3.37
	Cu	17	26		-9	-81	9	4.2	3.36
	Cu	31	10		15	-63	9	3.92	3.2
Subcortical	HG				-33	-15	-12	4.39	3.46
	Th		14		-15	-27	3	4.54	3.54

F. (L1 + L4) - (S1 + S4)

Frontal	IFG	44/45	12		27	9	21	3.58	3

[a] From the MTG cluster

[b] From the MOG cluster

Table 2. Results of t-tests for active regions in response to set size by pronunciation length and across the 2 pronunciation length conditions. Regions showing the set size effect in (A) the short name condition, (B) the long name condition, and (C) both short and long name conditions. A higher threshold was used for the set size effect collapsing pronunciation length conditions (p < .05, FDR corrected) and lowered for the set size effect separately for the short name (A) and long name (B) conditions (p < .001, uncorrected). Clusters were 9 contiguous voxels or larger. The table shows the peak coordinates in mm, t values (T), z scores (Z) and anatomical names of the major clusters. BA-Brodmann's area, DFG-Dorsal frontal gyrus, SMG-Supramarginal gyrus, IPL-Inferior parietal lobe, Caud N-Caudate nucleus, Cblm-Cerebellum. S1-short name, set size 1, S4- short name, set size 4, L1-long name, set size 1, L4-long name, set size 4. See Table 1 for other abbreviations and notations.

Lobe	Region	BA	Cluster size	MNI coordinates			T	Z
				x	y	z		
A. S4 - S1								
Frontal	DFG	6	13	-3	15	48	5.53	4.02
	IFG	44	125	-42	9	30	5.48	4
	PrCG	6		-39	-6	27	4.97	3.76
	MFG	6	16	-30	6	57	4.61	3.58
	IFG	45	14	-42	27	15	4.56	3.55
	MFG	6	13	36	9	57	4.42	3.48
	MFG	6		36	0	57	4.06	3.29
Parietal	SMG	40		-33	-57	33	5.51	4.02
	AG	39		-30	-66	33	4.95	3.75
Occipital	PCu	19	172	-30	-72	42	5.72	4.1
B. L4 - L1								
Frontal	IFG	45	36	-24	36	3	5.64	4.07

		BA	k	x	y	z		
	MFG	46	38	42	36	27	5.33	3.93
	MFG	10	10	-30	51	12	4.14	3.33
Parietal	IPL	40		-36	-57	39	5.11	3.83
	SPL	7		36	-69	48	4.59	3.57
Occipital	PCu	19	293	-24	-78	42	7.34	4.71
				-30	-51	27	6.42	4.39
	PCu	7	20	-6	-75	39	5.63	4.07
	PCu	19	35	12	-72	42	5.31	3.93
	Pcu	19	43	33	-66	39	4.62	3.59
Subcortical	CaudN	42		-21	-18	21	4.76	3.66
	Clbm	30		-39	-60	-27	7.63	4.81
	Clbm		14	39	-66	-27	4.58	3.57
			9	-21	12	27	4.45	3.5

C. (S4 + L4) − (S1 + L1)

		BA	k	x	y	z		
Frontal	IFG	44	26	-42	9	30	5.99	4.22
Parietal	SMG	40	70	-30	-51	27	6.21	4.3
	IPL	40		-36	-57	39	5.7	4.1
Occipital	PCu	19	40	-30	-72	42	6.09	4.26
	PCu	19		-21	-78	42	5.69	4.09
	PCu	19	11	33	-66	39	5.38	3.95

Figure captions

Figure 1. Schematic diagram of the delayed recognition task used in Exps. 1A-1C. Exps. 1A-1C differed in presentation time of study stimuli, interstimulus interval, and duration of delay, as specified at the bottom of the figure. All the probes were valid and pictures were not repeated except for the matched probes. For simplicity, the warning signal, visual mask (color noise pattern), and intertrial interval (2 s) are not shown in the figure.

Figure 2. Response time and A' scores in Exps. 1A-1C & 2. Line graphs show the mean response time (± standard error of the mean [SEM]) and the mean A' (± SEM) for each set size and pronunciation length. Legends: Blue lines, short name conditions; red lines, long name conditions.

Figure 3. Schematic diagram of the delayed recognition task and behavioral results from Exp. 3. (A) A schematic diagram of the delayed recognition task. The three task conditions differed only in the instruction letter displayed in the beginning of each task block. The instructions were "P," "W," and "E," indicating that study stimuli within that block will be tested by pictures, names, and either pictures or names, respectively. All probes were valid and all pictures were presented with equal probability as either study stimuli or test probes throughout the experiment. For simplicity, the warning signal, interstimulus interval (100 ms), visual mask (color noise pattern), and intertrial interval (2 s) are not shown in the figure. (B & C) Behavioral results from Exp. 3. Bar graphs show the mean response time (± standard error of the mean [SEM]) and the mean A' (± SEM) for each task condition as a function of pronunciation length of visual object names. Legends: Blue bars, short name condition; red bars, long name condition.

Figure 4. Schematic diagram of the delayed recognition task and behavioral results from the fMRI experiment. (A) A schematic diagram of the delayed recognition task. At the beginning of each trial, either 1 or 4 visual items were presented, followed by a 10.8 s delay interval until a test probe appeared. The ratio of match and non-match test probes was 1:1, and all probes were valid. All 192 pictures were presented equally often as either study stimuli or test probes throughout the experiment. For simplicity, the warning signal, interstimulus interval (100 ms), visual mask (color noise pattern), and intertrial interval (6.5–11 s) are not shown in the figure. (B & C) Behavioral results from the fMRI experiment. Bar graphs show the mean response time (± standard error of the mean [SEM]) and the mean accuracy (± SEM) for each set size as a function of pronunciation length of visual object names. Legends: Blue bars, short name condition; red bars, long name condition.

Figure 5. Effects of pronunciation length (A) and set size (B) on brain activity during the delay period. Group composite maps are shown on the right and left lateral, dorsal, and ventral surfaces of the rendered MNI single-subject brain for responses to pronunciation length (A) and set size (B). Using individual contrast images constructed with contrast weights (S1 + S4 − L1 − L4) for the delay period, a one-sample t-test was used at the 2nd

level analysis to reveal suprathreshold activations in response to pronunciation length (i.e., greater activity for the short name compared to the long name conditions) across 2 set size conditions (A). Using individual contrast images constructed with contrast weights (S4 + L4 − S1 − L1) for the delay period, a one-sample t-test was used at the 2nd level analysis to reveal suprathreshold activations to set size across 2 pronunciation length conditions (B). Threshold at p <0.001 (uncorrected) is used for visualization and clusters were 9 contiguous voxels or larger. See Tables 1 and 2 for specific coordinates of suprathreshold activations and abbreviations.

Figure 6. Brain activity for phonological and visual object processing and mean beta values in the left IFG, left SMG, and visual association regions during the study, delay, and probe periods as a function of pronunciation length and set size. For visualization of brain regions involved in phonological and visual processing, group composite maps are shown on the right and left lateral, dorsal, and ventral surfaces of the rendered MNI single-subject brain for responses to phonological processing (green) and visual object processing (red). To show brain activity for phonological processing, a one-sample t-test was used to reveal suprathreshold activations in response to the syllable judgment in contrast to case judgment task in the verbal localizer task (p < .025, FDR corrected). To show brain activity for visual processing, a one-sample t-test was used to reveal suprathreshold activations in response to the intact object 1-back task compared to the scrambled image 1-back task in the visual localizer task (p < .05, uncorrected). Mean beta values for the stimulus presentation, delay, and probe periods of the task were extracted from individually defined ROIs including the left IFG and left SMG, with a median cluster size of 100 voxels and visual association regions with a median cluster size of 200 voxels from each individual subject's non-normalized space. Bar graphs of mean beta values during the 3 task periods across conditions are shown for each ROI. Legends: Blue bar, short name object; Red bar, long name object.

Figure 7. Regions showing significant correlation with the pronunciation length effect in individual subjects. Group composite maps are shown on sagittal, coronal, and axial slices of the MNI single-subject brain with a cross-hair marking the location of the average peak coordinate. The coordinates marked by the cross-hair are correlated with higher accuracy for the short name compared to the long name conditions at a set size of 1. Simple regression with correlation analysis was conducted on individual contrast images for the short name vs. long name condition during the delay period with an accuracy difference measure (% correct for the short name condition - % correct for the long name condition) as covariates. Correlation analysis with accuracy as a covariate revealed suprathreshold activity correlated with higher accuracy for the short name compared to the long name conditions at set size 1 (p < .001, uncorrected, cluster size ≥ 6 contiguous voxels). Activity in this left IFG was significant at a cluster level (p = .005, corrected).

Figures

A. Task Diagram

| Fixation | Study (1, 2, 4, or 6 items) | | | | Delay | Probe |
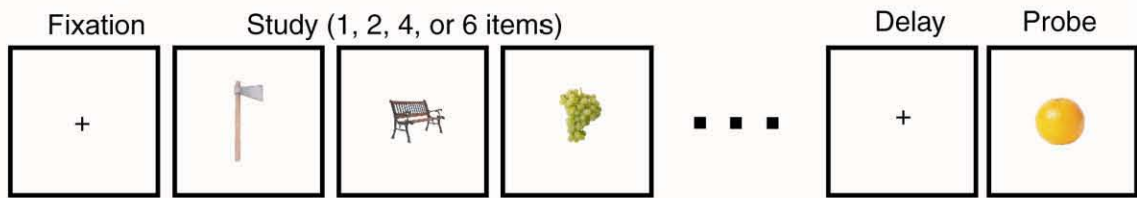
B. Task Parameters

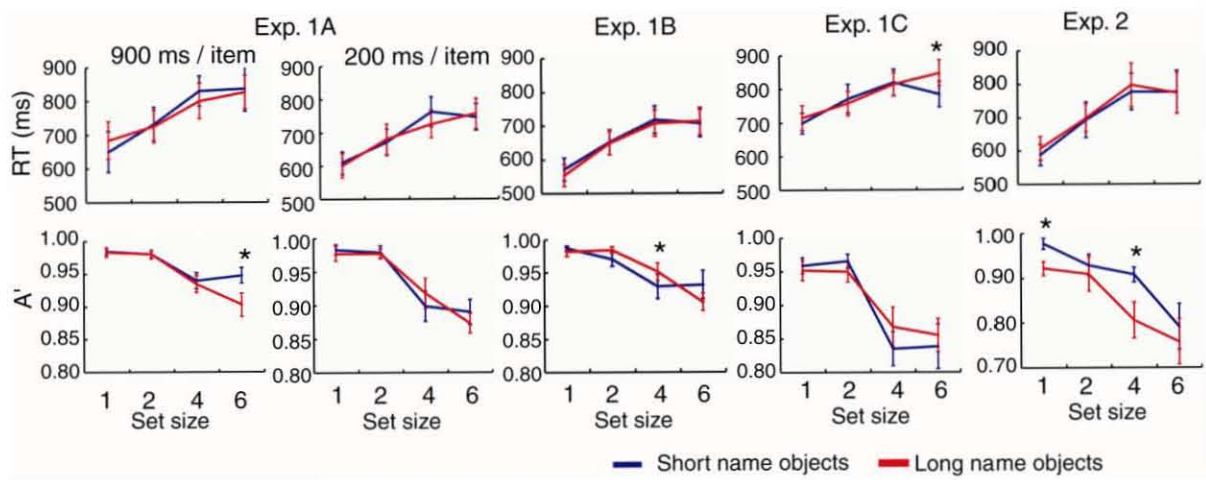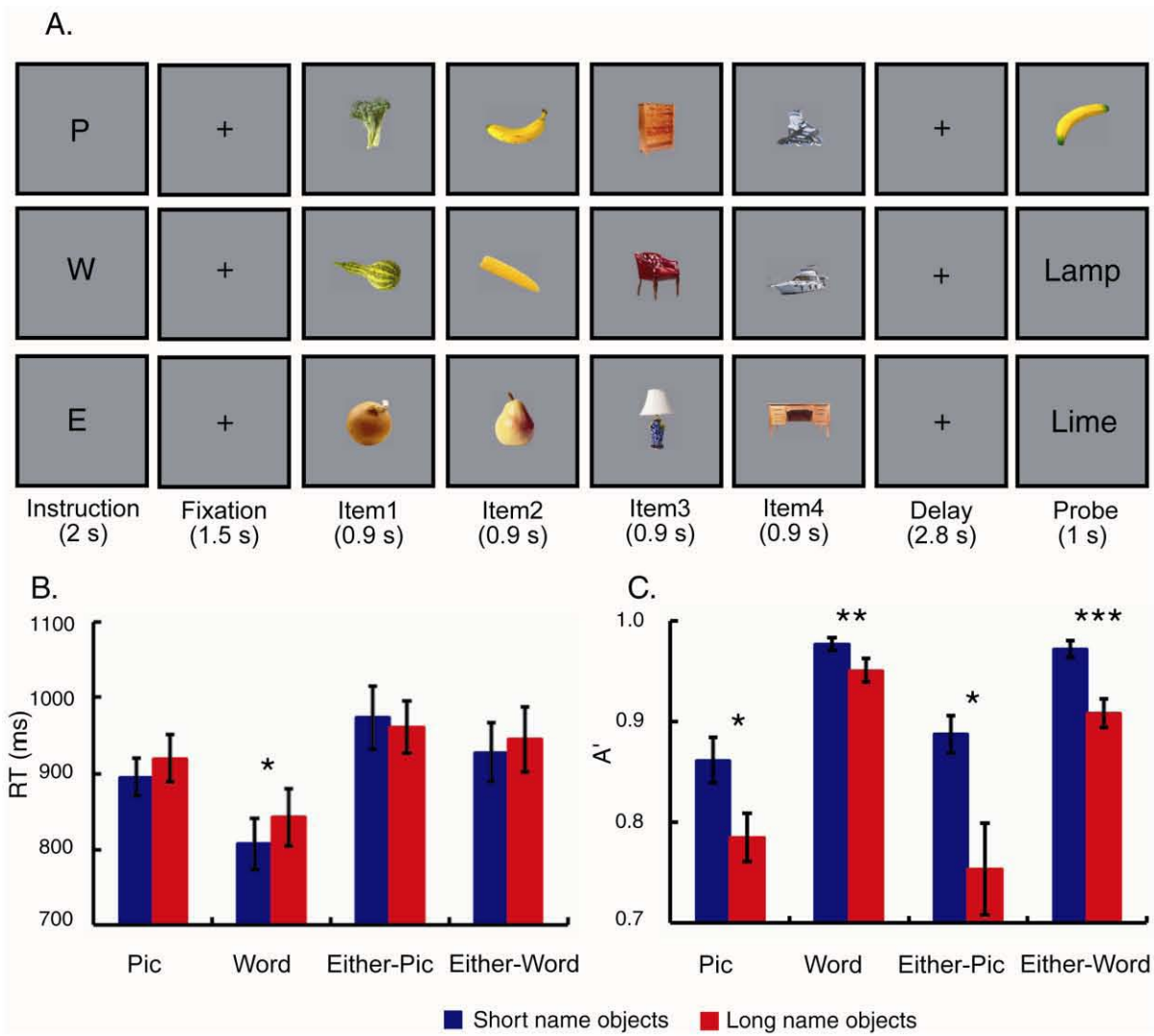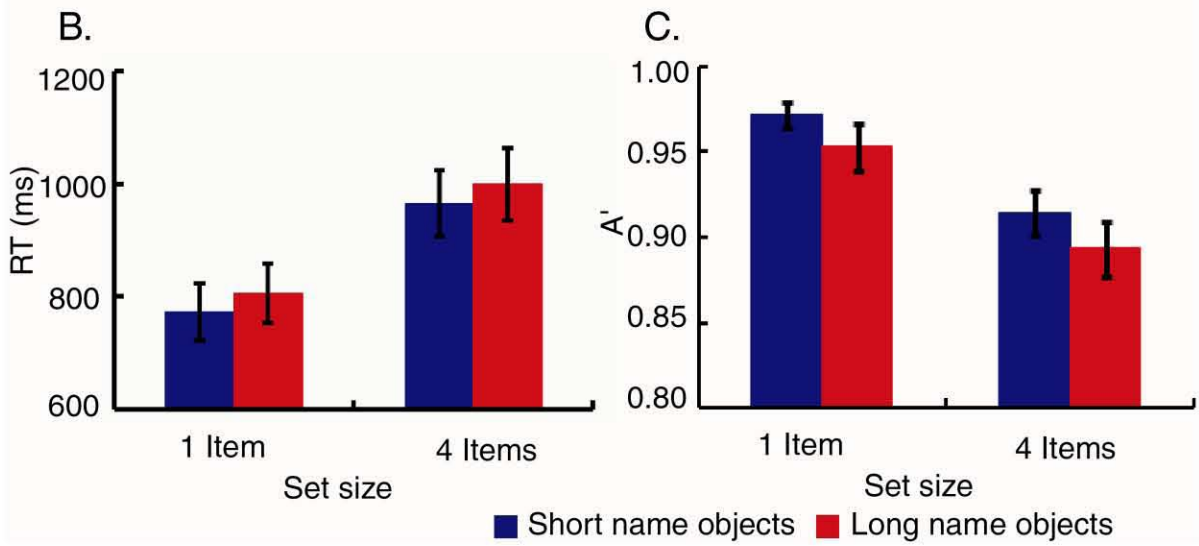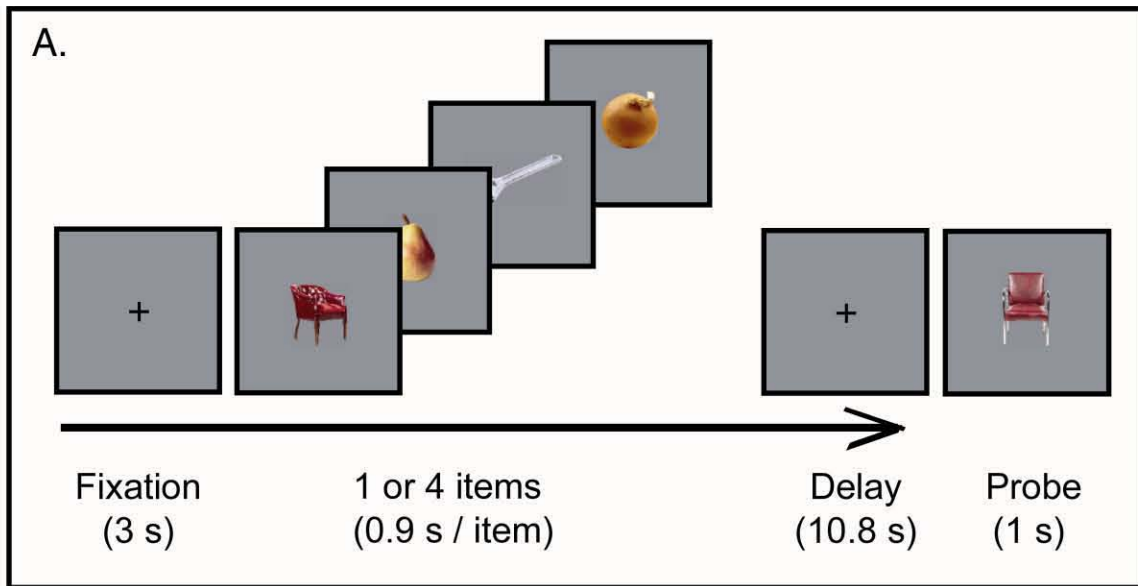|	Fixation	Presentation time per item	ISI	Delay	Probe
Exp. 1A	1.5 s	900 vs. 200 ms	100 ms	2.8 s	2 s
Exp. 1B	1.5 s	200 ms	800 ms	2.8 s	2 s
Exp. 1C	1.5 s	200 ms	100 ms	7.8 s	2 s
Exp. 2	1.5 s	900 ms	100 ms	2.8 s	2 s

Figure 1.
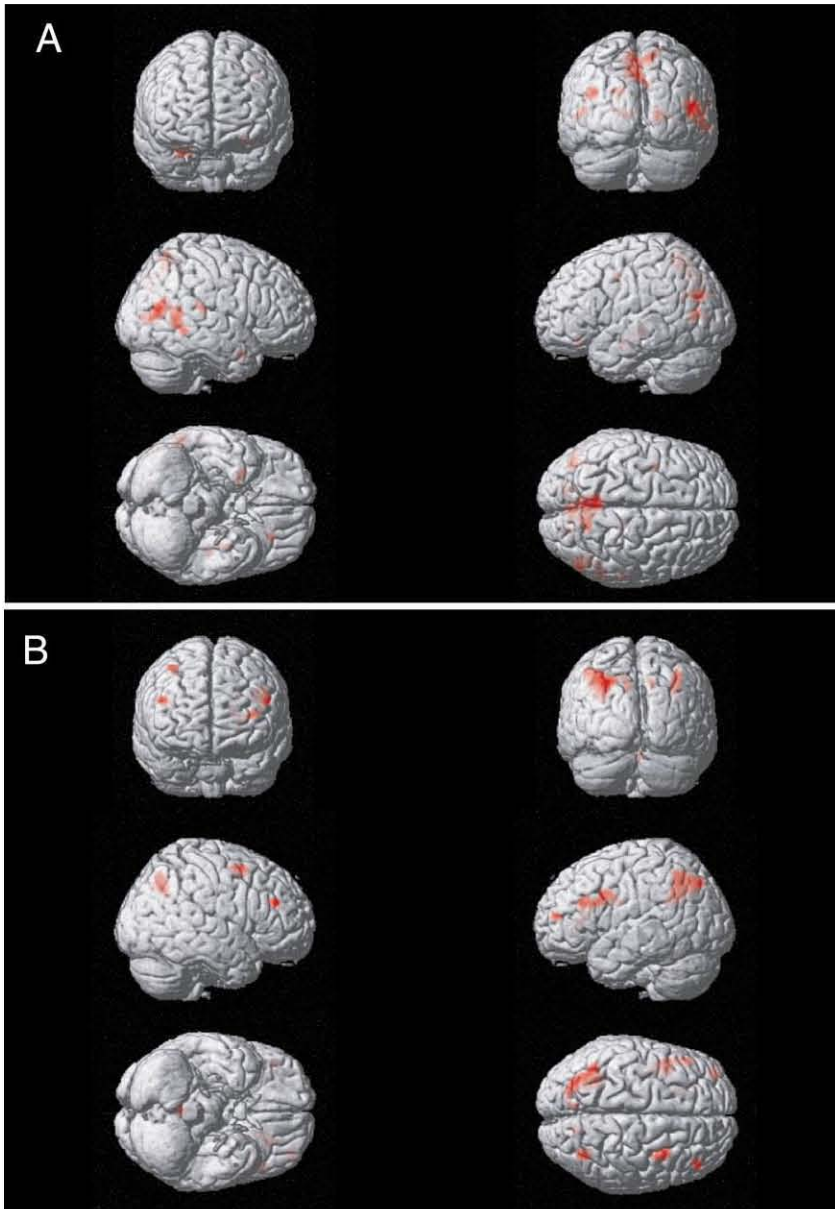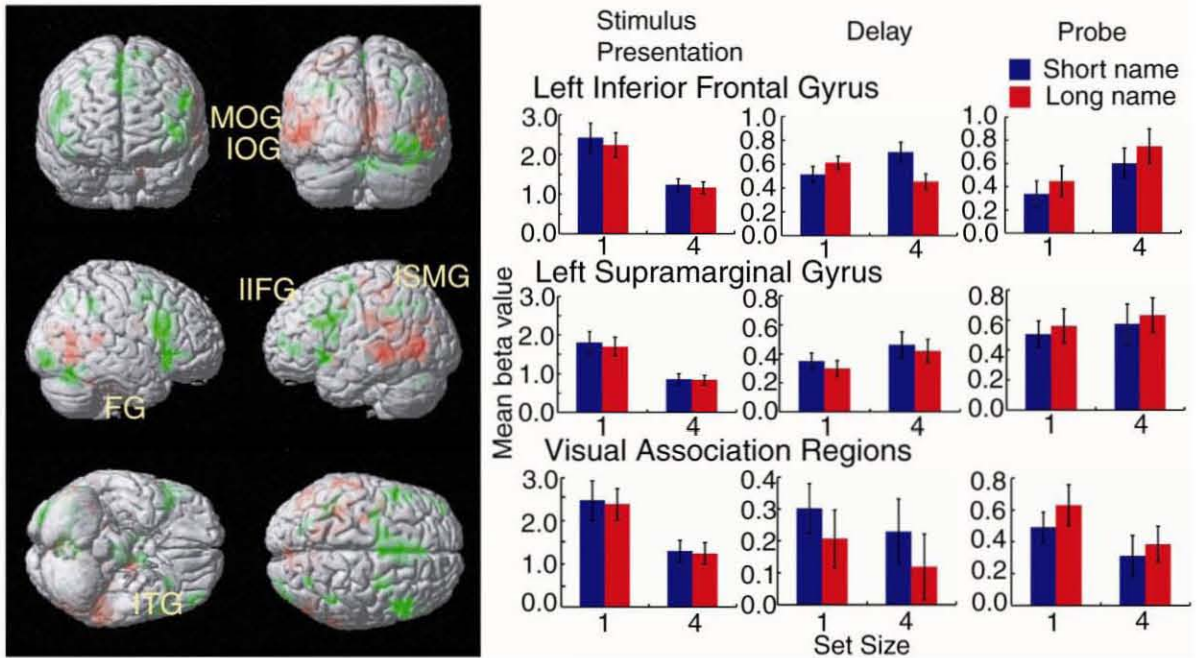
Figure 2.

Figure 3.

Figure 4.

Figure 5.

Figure 6.
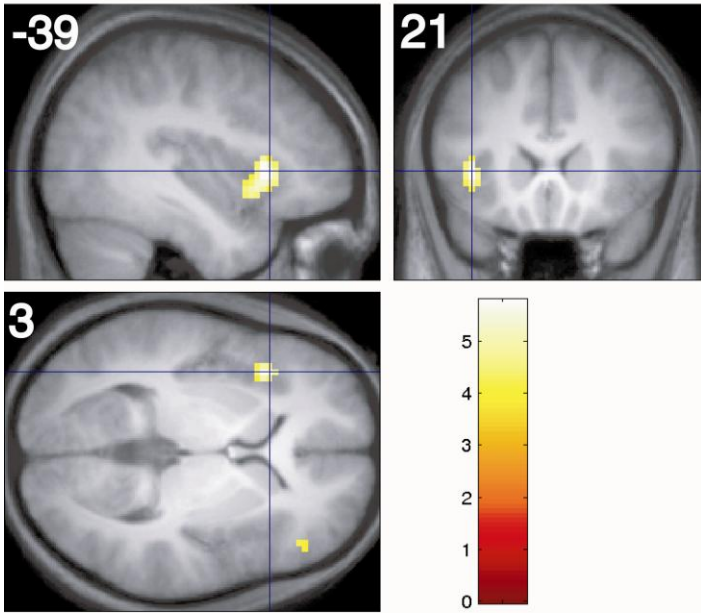
Figure 7.

Appendix

A table of stimulus information for the studies

Group	Object Name	Category	#Syllables	Word Freq_Raw	Word Freq_Log
Short	pear	Fruits	1	971	6.88
	peach	Fruits	1	1,812	7.50
	grapes	Fruits	1	1,281	7.16
	lime	Fruits	1	2,086	7.64
	knife	Tools	1	7,120	8.87
	nail	Tools	1	4,603	8.43
	axe	Tools	1	3,070	8.03
	wrench	Tools	1	1,389	7.24
	truck	Vehicles	1	10,203	9.23
	boat	Vehicles	1	15,857	9.67
	tank	Vehicles	1	17,481	9.77
	train	Vehicles	1	23,376	10.06
	chair	Furniture	1	18,589	9.83
	lamp	Furniture	1	6,657	8.80

	bench	Furniture	1	5,360	8.59
	desk	Furniture	1	13,624	9.52
	corn	Vegetables	1	4,988	8.52
	squash	Vegetables	1	1,217	7.10
	yam	Vegetables	1	577	6.36
	onion	Vegetables	2	2,587	7.86
	goat	Four-footed animals	1	2,562	7.85
	cow	Four-footed animals	1	7,262	8.89
	pig	Four-footed animals	1	6,375	8.76
	fox	Four-footed animals	1	13,959	9.54
		Average	1.04	7208.58	8.42
Long	banana	Fruits	3	2,879	7.97
	orange	Fruits	2	13,483	9.51
	pineapple	Fruits	3	637	6.46

strawberry	Fruits	3	1,622	7.39
pencil	Tools	2	3,021	8.01
ladder	Tools	2	2,970	8.00
hammer	Tools	2	6,714	8.81
scissors	Tools	2	1,369	7.22
scooter	Vehicles	2	744	6.61
rollerblade	Vehicles	4		
motorcycle	Vehicles	4	3,696	8.22
helicopter	Vehicles	4	3,201	8.07
table	Furniture	2	56,081	10.94
bookshelf	Furniture	2	1124	7.02
cabinet	Furniture	3	6,439	8.77
dresser	Furniture	2	874	6.77
celery	Vegetables	3	732	6.60
potato	Vegetables	3	3,178	8.06
lettuce	Vegetables	2	970	6.88
broccoli	Vegetables	3	915	6.82
elephant	Four-footed	3	7,010	8.86

		animals			
	buffalo	Four-footed animals	3	8,343	9.03
	squirrel	Four-footed animals	2	1,989	7.60
	rabbit	Four-footed animals	2	5,751	8.66
		Average	2.625	5814.87	7.92
		T-test (two-sided)	P < .001	P = .61	P = .12