Journal of Neurolinguistics 53 (2020) 100863

Contents lists available at ScienceDirect

Journal of Neurolinguistics

journal homepage: www.elsevier.com/locate/jneuroling

Research paper

Effects of encoding modes on memory of naturalistic events

Cong Liu[a,b], Ruiming Wang[a,*], Le Li[c], Guosheng Ding[c], Jing Yang[d,e], Ping Li[f,**]

[a] Guangdong Provincial Key Laboratory of Mental Health and Cognitive Science, Center for Studies of Psychological Application, School of Psychology, South China Normal University, Guangzhou, 510631, China
[b] Department of Psychology, Qingdao University, Qingdao, 266071, China
[c] State Key Laboratory of Cognitive Neuroscience and Learning & IDG/McGovern Institute for Brain Research, Beijing Normal University, Beijing, 100875, China
[d] Center for Linguistics and Applied Linguistics, Bilingual Cognition and Development Lab, Guangdong University of Foreign Studies, Guangzhou, 510420, China
[e] Bilingual Cognition and Development Lab, Guangdong University of Foreign Studies, Guangzhou, 510420, China
[f] Department of Psychology and Center for Brain, Behavior, and Cognition, Pennsylvania State University, University Park, PA, 16802, USA

ARTICLE INFO

Keywords:
Encoding modes
Memory retrieval
Naturalistic events
fMRI

ABSTRACT

The memory advantage of bimodal encoding in the retrieval of isolated stimulus have been extensively studied, but researchers have not investigated this advantage for naturalistic events. This study reports both behavioral and functional magnetic resonance imaging (fMRI) data on whether memory advantage of bimodal encoding exists for retrieval of naturalistic events. In Experiment 1, participants took memory tests after learning naturalistic events via three different encoding modes: (1) text reading, (2) story listening, and (3) video watching. The results showed that, at immediate recall, participants made few errors in the text reading and video watching conditions than the story listening condition; at delayed recall, these differences disappeared. In Experiment 2, participants similarly read texts, listened to stories, watched videos, and underwent fMRI scanning during a recall task. Our fMRI data showed stronger activation in the right angular gyrus for retrieving bimodal naturalistic events (i.e., video watching) than unimodal ones (i.e., text reading and story listening). These results suggest a memory advantage of the bimodal encoding for retrieving complex episodic memories, given the rich, multisensory events across encoding modes over time.

1. Introduction

People acquire new information by various means, including text reading, story listening, or video watching, which involve different encoding modes. Despite previous efforts in the literature to identify the relationship between the mode of encoding and memory performance (Thompson & Paivio, 1994; Winnick & Brody, 1984), it remains unclear which encoding modes can lead to better recall performance.

The dominant view suggests a memory advantage of bimodal encoding in terms of multimedia learning (Mayer, Moreno, Boire, & Vagge, 1999). In the context of multimedia learning, the cognitive theory of multimedia learning makes three main assumptions: there are two separate channels (auditory and visual) for processing information; there is limited channel capacity; and learning is an active process of filtering, selecting, organizing, and integrating information (Mayer, 2005, 2009). The principle known as the

* Corresponding author.
** Corresponding author.
E-mail addresses: wangrm@scnu.edu.cn (R. Wang), pingpsu@gmail.com (P. Li).

https://doi.org/10.1016/j.jneuroling.2019.100863
Received 15 January 2019; Received in revised form 2 July 2019; Accepted 7 August 2019

"multimedia principle" states that "people learn more deeply from words and pictures than from words alone", suggesting a bimodal advantage in multimedia learning.

The cognitive load theory follows the second assumption of the multimedia learning theory and states that situations that reduce cognitive load are more conducive to learning. It discusses two types of cognitive loads: intrinsic and extrinsic (Lan, Fang, Legault, & Li, 2015; Sweller, 2005, pp. 19–30). Intrinsic cognitive load is imposed by the nature of what is to be learned, including the number of information elements and their natural complexity or interactivity (Sweller & Chandler, 1994; 2010). Extraneous cognitive load is caused by the presentation manner of the information and is often the main source that is detrimental to learning (Sweller & Chandler, 1994). Multiple effects have been evaluated to identify what aids should be provided or what components should be removed from the learning materials to help reduce cognitive load during learning (for a review, see Sweller, 2010). Among the effects often discussed is modality effect, which refers to a cognitive load learning effect that occurs when a bimodal (both visual and auditory) presentation of information is more effective than a single-mode (either visual or auditory alone) presentation of the same information (Low & Sweller, 2005). Specifically, presentation of information in two modalities frees cognitive resources by increasing working memory capacity, so that extraneous cognitive load may be reduced and more attention can be paid to comprehension, which facilitates memory encoding and learning.

Consistent with the cognitive theory of multimedia learning and cognitive load theory, many studies have confirmed the additive effects of bimodal coding for memory (see Goolkasian & Foos, 2005; Santangelo, Mastroberardino, Botta, Marucci, & Belardinelli, 2006; Thompson & Paivio, 1994). For example, Thompson and Paivio (1994) made participants memorize three types of environmental items: picture-only, sound-only, or combination of pictures and sounds. Their results showed that participants' recall performance in the bimodal encoding condition (i.e., when pictures and sounds were presented together) was better as compared to the unimodal encoding condition (i.e., picture-only or sound-only), which suggests a memory advantage of bimodal encoding over unimodal encoding of items. However, the memory advantage of bimodal encoding was not found in all studies (Crooks, Cheon, Inan, Ari, & Flores, 2012; Winnick & Brody, 1984). For instance, Winnick and Brody (1984) found that words that are high in both auditory and visual imagery (e.g., clock) were not recalled better than single-modality words (e.g., click). Moreover, Tracy, Tracy, and Ramsdell (1985) found that bimodal imagery was mnemonically superior to single-modality imagery only when a between-subject design was used (see also Tracy, Roesner, & Kovac, 1988). Thus, the memory advantage of bimodal encoding cannot be taken as given and needs further investigation.

To date, previous studies in this domain have focused on isolated stimuli (e.g., pictures, sounds; see a review, Mastroberardino, Santangelo, Botta, Marucci, & Belardinelli, 2008), and few studies have examined whether the memory advantage of the bimodal encoding exists for naturalistic events. Compared to isolated stimuli used in previous studies, naturalistic events involve objects and scenes from different sensory modalities. For example, when we watch a movie, we see actors dancing and simultaneously hear the music. These naturalistic events in real life differ significantly from isolated words and pictures as stimuli used in cognitive experimental studies.

With regard to the neurocognitive mechanisms, the angular gyrus (AG) serves as a cross-modal integrative hub (Binder & Desai, 2011; Buckner, Andrews-Hanna, & Schacter, 2008), and plays a significant role in integrating different types of information. Previous work has also shown an involvement of AG in audio-visual speech integration in tasks such as passive speech perception (Bernstein, Auer, Wagner, & Ponton, 2008). There is also recent work that indicates specifically the right AG's engagement in multiple sources and knowledge integration during the processing of Chinese idioms (Yang et al., 2016). Moreover, recent studies have also indicated that the hippocampus and high-level cortical areas form a neural network (e.g., posterior medial cortex, medial prefrontal cortex, middle temporal gyrus, and angular gyrus) to jointly support real-world memory (Chen et al., 2016). Given the unique characteristics of naturalistic events (e.g., more vivid information across multiple sensory modalities) compared to isolated stimuli, we argue that it would be important for us to investigate the bimodal advantage using naturalistic events than isolated stimuli as in previous studies.

If bimodal encoding mode does have an advantage over unimodal ones in terms of memory performance, we would like to examine the robustness of the memory effect over time. There is ample evidence from neuroimaging studies suggesting memory change over time (e.g., Ritchey, Montchal, Yonelinas, & Ranganath, 2015; Takashima et al., 2009). For example, in the Furman, Mendelsohn, and Dudai (2012) study, participants were scanned during a memory test either hours, weeks, or months after viewing a documentary movie, and their results showed that recognition accuracy at a few hours decreased after weeks but then remained at the same level after months. Similarly, Takashima et al. (2006) asked participants to complete four recognition memory scans at four different times (1 day, 2 days, 30 days, and 90 days after learning). The results showed that for confident and correct recognition, activity in the hippocampal regions decreased over time, whereas activity in neocortical areas (particular in the ventral medial prefrontal cortex) increased. This pattern of decrease in the hippocampus versus increase in the ventral medial prefrontal cortex (vmPFC) was also observed in participants who underwent scanning for only two times, shortly after learning (15 min and 24 h) (Takashima et al., 2009). Given such characteristics of memory, in the present study, we aimed to explore the advantage of the bimodal encoding in both short-term and long-term memory, in contrast to the previous studies that have focused on such advantage only in short-term memory.

Overall, the present study is designed to examine the influence of encoding modes on memory performance using naturalistic episodes from real-life events instead of isolated stimuli (e.g., pictures, sounds). We use both behavioral and neuroimaging methods to study this issue. Experiment 1 examined the behavioral differences in recall performance of naturalistic events after learning stimuli via three different encoding modes: text reading, story listening, and video watching in which both visual and auditory information is available. Memory retrieval performance was also assessed at three time points: immediate test, delayed test one week later, and delayed test one month later. While text reading or story listening involves information processing in only one modality (i.e., visual or auditory), video watching provides a means for the integration of bimodal information at the same time, which

enhances memory performance according to predominant theories as discussed above. Specifically, according to the modality effect in cognitive load theory (Low & Sweller, 2005; Sweller, 2005, pp. 19–30), as auditory information and visual information are each processed in their respective system, the video watching condition (audio-visual encoding) has a lower load in visual or auditory working memory as compared to text reading and story listening. That is because in the latter cases (text reading or story listening), the total load induced by video watching is separately spread across the visual and the auditory components in the working memory system. Thus, we hypothesize that the memory performance in retrieval of bimodal encoding (i.e., video watching) will be better than unimodal ones (i.e., text reading and story listening).

In addition to the behavioral study, Experiment 2 is designed to identify the neural substrates of memory retrieval across different modes of encoding. The functional imaging data from Experiment 2 should further help us understand the neurocognitive basis of bimodal audio-visual information processing. On the basis of the cognitive load theory and the findings from previous studies that showed that angular gyrus is a semantic integration hub (Binder, Desai, Graves, & Conant, 2009) and play a critical role during real-life episodic memory retrieval (e.g., Chen et al., 2016), we hypothesize that the activation of the memory retrieval regions such as the angular gyrus will be stronger in bimodal encoding condition (i.e., video watching) than that in unimodal encoding conditions (i.e., text reading and story listening).

2. Experiment 1

2.1. Participants

Thirty participants (age range: 18–24 years old; 10 women) took part in this experiment. They had normal or corrected-to-normal vision, and were strongly right-handed as assessed by a modified Chinese version of the Snyder and Harris handedness inventory (Snyder & Harris, 1993). Informed written consent was obtained from the participants before the experiment. This research was approved by the Research Ethics Committee at South China Normal University.

2.2. Experimental materials

The experimental materials comprised study material and test questions. There were three types of study materials: text reading condition, story listening condition, video watching condition. The video watching materials included three video clips (duration: 255 s, 270 s and 335 s respectively) without subtitles, selected from the '*Rediscovering the Yangtze River*' documentaries from China Central Television (CCTV). The text reading materials were the subtitles of the video material (average 56 sentences), and the story listening materials were the narrations plus the background music of the videos. To sum up, there were three different materials for each condition. All three types of study materials were presented in Mandarin Chinese, and had equal length.

The test questions were comprehension questions with two choices, which were designed to measure participants' understanding of the learned materials. The multiple-choice questions were 13 Chinese characters, on average. Before the experiment, 20 participants were asked to evaluate the difficulty levels of the test questions, and 30 questions were finally selected. Sample test questions are included in Appendix A (see supplementary materials).

2.3. Procedures

There were two phases in this experiment: the learning phase and the testing phase.

Learning phase. Our participants were randomly assigned to three groups. Each group was presented with one of the three study material lists, and learning sequences were counterbalanced in a Latin-Square design. More importantly, every list contains the three different study materials with three different presentation modes (i.e., each group learns the stimuli in three different modes). Then they were given the same tests at three times (i.e., an immediate test after learning phrase, a delayed test one week later, and one delayed test one month later). Thus, it is a 3 (Tests: immediate test, delayed test one week later, and delayed test one month later) × 3 (Encoding modes: text reading, story listening, and video watching) within-subject design in this experiment. E-Prime 2.0 (Schneider & Zuccoloto, 2007) was used for stimulus presentation and data collection. During the learning stage of the text reading condition, participants were presented with subtitles of videos visually (i.e., visually-presented text with no visual scene), one sentence at a time; during the story listening condition, learning materials were presented via a headphone without visual stimuli. The video watching materials were presented with both visual and auditory inputs. All three learning materials were presented at the same speed as video broadcasts speed.

Testing phase. Each testing question started with a fixation point of 2000 ms and then a test question was presented. Participants were instructed to press the left or the right button to indicate which option might be correct. A red square frame would appear on the selected choice once the participant pressed the key, indicating this test question was completed. If participants failed to respond within 6000 ms after the test question began, the test question automatically disappeared and this particular trial would be excluded for later data processing. Participants' response accuracy was recorded via E-prime.

2.4. Results

Six participants, whose accuracy rates were lower than 50% (chance level) in the memory tests, were excluded from further analyses. The accuracy rates (ACC) and standard deviations (SD) during the retrieval of naturalistic information are presented in

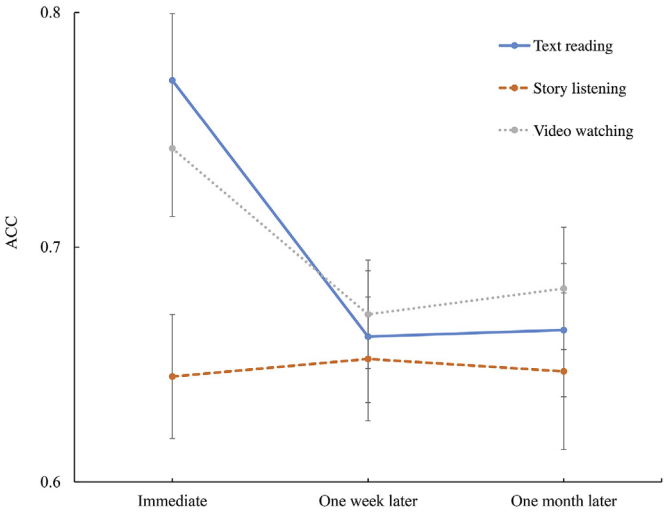

Fig. 1. Response patterns in ACC across three tests in Experiment 1.

Fig. 1.

A 3 (Tests: immediate test, delayed test one week later, and delayed test one month later) × 3 (Encoding modes: text reading, story listening, and video watching) repeated measures ANOVA was conducted on the accuracy. The results showed a significant main effect of tests, $F(2, 58) = 10.755$, $p < 0.001$, $\eta_p^2 = 0.271$, and a significant main effect of encoding modes, $F(2, 58) = 6.162$, $p < 0.01$, $\eta_p^2 = 0.175$. Moreover, we found a significant interaction between tests and encoding modes, $F(4, 116) = 3.917$, $p < 0.01$, $\eta_p^2 = 0.119$. Simple effects analysis for the 2-way interaction revealed that the difference in accuracy across the three encoding modes was significant in immediate test, $F(2,58) = 9.30$, $p < 0.001$, $\eta_p^2 = 0.241$, whereas these differences disappeared in the test one week later, $F(2,58) = 0.037$, $p = 0.690$, $\eta_p^2 = 0.021$, and one month later, $F(2,58) = 1.55$, $p = 0.220$, $\eta_p^2 = 0.047$. Post-hoc test for immediate test showed that, compared to the story listening, the participants made fewer errors in retrieving the information encoded by text reading ($p < 0.05$) and video watching ($p < 0.05$). However, there was no significant difference between retrieval of the information encoded by video watching and that of text reading ($p > 0.05$). Furthermore, simple effect analysis for the 2-way interaction indicated that for retrieving the information encoded by text reading and story listening, the participants made fewer errors in immediate test than the two delayed tasks ($ps < 0.05$), but such a difference was absent for retrieving the information encoded by story listening ($p > 0.05$).

2.5. Discussion

This experiment demonstrated that, in the immediate retrieval test, participants performed better in the video watching condition than in the story listening condition, which supports the hypothesis that retrieval of the information encoded by the bimodal mode is better than unimodal ones. In addition to the bimodal video encoding advantage, we also found a text encoding advantage, by which participants performed better in text reading condition than story listening condition, but not worse than video watching condition. This result was not expected. We think this finding might have resulted from the presentation formats of test materials: the same presentation format in both the learning phase and the test phase could have facilitated memory performance, compared to the other two conditions. As for why the differences across three encoding modes disappeared in delayed tests, it may be because the three conditions all suffered a decline in performance and reached the baseline level in the delayed test over time. To identify such differences more clearly, Experiment 2 was designed to explore the different neural mechanisms underlying these differences.

3. Experiment 2

3.1. Participants

Forty-eight participants (age range: 18–24 years; eight men) participated in this fMRI experiment. None of them participated in Experiment 1. They had normal or corrected-to-normal vision, and were strongly right-handed as assessed by a modified Chinese version of the Snyder and Harris handedness inventory (Snyder & Harris, 1993). Informed written consent was obtained from the participants before the experiment, and as in Experiment 1, this experiment and its scanning protocols were approved by the local Research Ethics Committee at the South China Normal University.

All participants were randomly divided into three groups (see description of procedures below). To avoid confounds due to individual difference variables, all participants completed a battery of behavioral tests, which include: *Raven Progressive Matrices*, a measure of nonverbal intelligence (Raven, Raven, & Court, 1998); *Letter-Number Sequencing,* a task to measure working memory (Crowe, 2000); and *the Attention Network Test* (ANT), a test of orienting, alerting, and conflict detection abilities (Fan, McCandliss,

Table 1
Participants' cognitive ability information.

Measure		Text reading	Story listening	Video watching	*p*
IQ (Raven)		57.31 ± 2.05	56.31 ± 2.54	56.56 ± 2.31	0.451
Working memory		0.44 ± 0.12	0.42 ± 0.12	0.40 ± 0.11	0.704
ANT	Alerting	47.93 ± 16.99	54.57 ± 22.48	46.38 ± 14.98	0.416
	Orienting	45.53 ± 19.23	47.59 ± 12.88	48.58 ± 21.86	0.893
	Conflict	86.45 ± 23.13	94.64 ± 27.76	96.77 ± 24.21	0.476

Sommer, Raz, & Posner, 2002). In Letter-Number Sequencing task, participants are instructed to type the numbers in ascending order, followed by the letters in alphabetical order, after hearing a letter-number sequence. In the ANT, there are four cue conditions (no cue, center cue, double cue, and spatial cue) and three target conditions (congruent, incongruent, and neutral). The ANT requires participants to determine whether a central arrow points left or right. The arrow appears above or below fixation and may or may not be accompanied by flankers. A set of cognitive subtractions is used to assess the efficiency of three attentional networks. The alerting effect is calculated by subtracting the mean RT (in milliseconds) of the double-cue conditions from the mean RT of the no-cue conditions. The orienting effect is calculated by subtracting the mean RT of the spatial cue conditions from the mean RT of the center cue. The conflict (executive control) effect is calculated by subtracting the mean RT of all congruent flanking conditions, summed across cue types, from the mean RT of incongruent flanking conditions.

Participant's performance in those measures was shown in Table 1. These results indicate no significant differences in the participants' cognitive ability information.

3.2. Experimental materials and procedure

The experimental materials, including study material and test questions, are similar to those used in Experiment 1. However, because participants were recruited to undergo fMRI scanning, we used only one long fragment of the study material (length: 849 s) with forty items of test questions in this experiment. A 2 (Tests: immediate test, delayed test) × 3 (Encoding modes: text reading, story listening, and video watching) mixed design was adopted in the study. In the learning session, each of the three groups of participants studied one type of material (text reading, story listening, and video watching). The experiment comprised two scanning sessions, with an interval of one week in between. During the first session, participants underwent the following scan sequence: resting-state fMRI scan (6 min), studying the materials inside scanner (14 min), resting-state fMRI scan (6 min), fMRI scan with test questions (6 min), and a structural MRI scan (5 min). In the second session, which was one week after the first session, the participants underwent the following sessions: resting-state fMRI scan (6 min), fMRI scan with test questions (6 min), and a structural MRI scan (5 min). For this study, we reported only the fMRI data and didn't include resting-state fMRI data. For the fMRI scan, each test included two tasks (i.e., the retrieval task and the baseline task). In the retrieval task, participants pressed buttons of the fMRI compatible response box to indicate the correct answer for the test question (i.e., pressing one button by left hand for choice A or another button by right hand for choice B). In the baseline task, the participants were instructed to judge which option contained the word in the question-stem by pressing the buttons. Sample test questions are included in Appendix A. We rephrased the test questions to make the questions in two tests different. It is noted that there was a potential confound variable that we cannot make sure whether or not the participants were watching the videos because we did not use eye-tracker in the scanner. However, we did believe the participants were really watching the video because they are all answering the test questions to a high level of accuracy (see 3.6.1).

3.3. Design

All stimuli were presented via E-Prime 2.0 (Schneider & Zuccoloto, 2007) to the participants through a mirror mounted on the head coil in the MRI scanner. The test (for both immediate test and delayed test) only included one run, containing 40 retrieval trials and 40 baseline task trials. All the trials were mixed in a rapid event-related design. In each trial the stimulus was presented for 4000 ms, followed by a fixation cross "+" which is a jittered inter-stimulus-interval (ISI) between 3000 ms and 5000 ms. The participants were instructed to do the retrieval task or baseline task by the colour of the fixation cross: red fixation cross indicates retrieval task and blue fixation cross links to baseline task. The jittered sequence was determined using a genetic algorithm approach that randomly created a series of numbers between 3000 ms and 5000 ms in steps of 250 ms for each participant (Wager & Nichols, 2003).

3.4. MRI acquisition

MRI images were acquired on a 3T Siemens Trio scanner with 12-channel phase array head coil at South China Normal University. Functional images were using T2-weighted gradient-echo planner imaging (EPI) sequence (TR = 2000 ms, TE = 30 ms, flip angle = 90°, FOV = 192 × 192 mm^2, matrix = 64 × 64, slice thickness = 3.5 mm, gap = 0.5 mm, voxel size = 3 × 3 × 3.5 mm^3). In addition, a T1-weighted structural image was acquired for each participant using the MPRAGE sequence (TR = 1900 ms, TE = 2.52 ms, flip angle = 9°, FOV = 256 × 256 mm^2, matrix = 256 × 256, slice thickness = 1 mm, voxel size 1 × 1 × 1 mm^3).

Table 2
Regions of interest.

Regions	Hemisphere	MNI coordinates		
		x	y	z
ParaHippocampal	L	−24	−42	−9
Angular	R	42	−69	36
Angular	L	−39	−75	42
Precuneus	R	9	−45	42
Inferior Frontal gyrus	R	30	27	−6

3.5. fMRI data analysis

All MRI data analyses were performed using Statistical Parametric Mapping (SPM8, http://www.fil.ion.ucl.ac.uk/spm) and DPABI (Yan, Wang, Zuo, & Zang, 2016). In the pre-processing, the first four volumes were discarded for each subject to minimize the transient effects of hemodynamic responses. Functional images were temporally corrected to the middle slice, spatially realigned, co-registered to the high-resolution T1-weighted structural image, normalized to the standard T1 template volume (MNI) and resampled with voxel size of $3 \times 3 \times 3\,mm^3$. The data were then smoothed with an isotropic 8 mm FWHM Gaussian kernel. Participants were excluded if their head movement during either run of the fMRI task exceeded 3 mm in translation or 3° in rotation for any axis, and using this criterion, two participants were excluded in the data analysis of the delayed test.

Statistical analyses were performed by modeling different conditions as explanatory variables within the context of the general linear model. In the first level analysis, we computed the *t*-images of parameter estimates for the comparison between the memory retrieval and baseline conditions (memory retrieval trials > baseline trials) at each voxel for every subject. Movement parameter estimates produced by the realignment procedure were entered as covariates of no interest in the first-level single-subject design matrices, in order to correct for potential movement artifacts. Then, at the second level analysis, we performed one-sample *t* tests for each encoding mode and two-sample *t* tests across different encoding modes according to the above assessment. Only the clusters larger than 30 voxels ($3 \times 3 \times 3$ mm for each voxel) activated above the height threshold of $p < 0.05$ (FDR corrected for multiple comparisons) were considered as significant.

Further, to confirm the differences across presentation formats, we select five regions of interest (i.e., ROI) based on individual subjects' peak of BOLD signal for ROI analysis. Specifically, for each individual's data, regions of interest were defined as a 6-mm-radius sphere around the coordinates of significant peaks of activity found for retrieval condition relative to the baseline condition (see Table 2). Then, the beta values were extracted from the defined ROIs, and these values were then used in a repeated-measures ANOVA.

3.6. Results

3.6.1. Behavioral results

All participants had high accuracy (> 0.96) in the baseline task, which indicated that they were actively engaged in the tasks. Two participants, whose accuracy rates were lower than 50% (chance level) in the memory tests, were excluded from further analyses. In the retrieval task, a 2 (Tests: immediate test, delayed test) × 3 (Encoding modes: text reading, story listening, and video watching) ANOVA was conducted on the accuracy. The results showed a significant main effect of tests, $F(1, 45) = 22.008$, $p < 0.001$, $\eta_p^2 = 0.328$, indicating that the accuracy in the immediate test was significantly larger than in delayed test. However, we found no significant main effect of encoding modes, $F(2, 45) = 0.981$, $p = 0.383$, suggesting that the differences in accuracy rates across the three presentation modes were not significant in immediate test and delayed test, but with similar overall pattern as Experiment 1 (see Fig. 2). In addition, there was no interaction effect between tests and encoding modes, $F(2, 30) = 1.478$, $p = 0.239$.

3.6.2. fMRI results

3.6.2.1. Whole brain analysis results. Brain activation for each encoding mode in both immediate and delayed tests. In immediate test, comparable activation patterns are seen in retrieving information encoded by text reading and video watching, including bilateral hippocampus, parahippocampal, medial prefrontal cortex (mPFC), precuneus/posterior cingulate cortex (PCC) and angular gyrus (AG), a commonly recognized memory retrieval network according to previous studies (see a review, Rugg & Vilberg, 2013). The subcortical striatal areas including the caudate nucleus and putamen were also implicated in both presentation formats. By contrast, only bilateral mPFC, left parahippocampal, left precuneus were active in retrieving information encoded by story listening (see Fig. 3).

However, in the delayed test, more brain areas were activated in retrieving information encoded by video watching (i.e., bimodal encoding) as compared to text reading or story listening (i.e., unimodal encoding). Specifically, the extensive memory retrieval network observed in the immediate test for video watching was still activated in the delayed test. By contrast, only limited brain areas such as mPFC and dorsolateral prefrontal cortex (DLPFC) were activated in text reading and story listening conditions.

Bimodal encoding vs. unimodal encoding in both immediate and delayed tests. Two-sample t tests (FDR corrected, $p < 0.05$) were

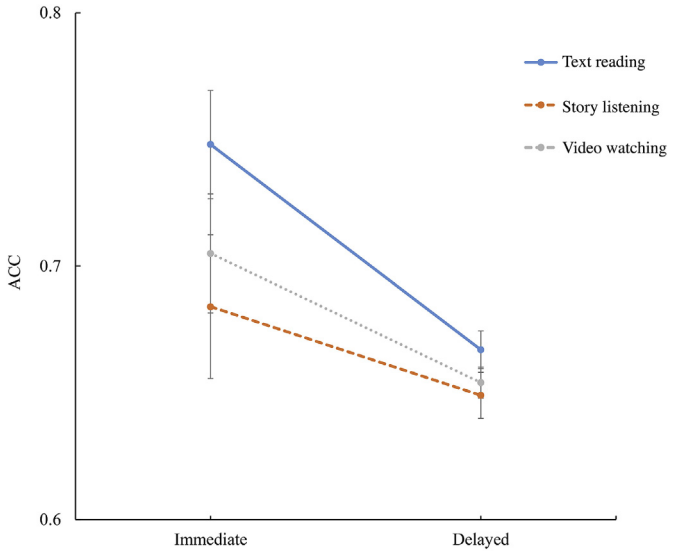

Fig. 2. Response patterns in ACC across immediate and delayed tests in Experiment 2.

conducted between bimodal encoding and unimodal encodings separately. The results showed no difference between the memory retrieval of information encoded by bimodal mode (i.e., video watching) and unimodal ones (i.e., text reading and story listening), though more brain areas such as bilateral parahippocampal and bilateral angular were activated in bimodal format as compared to unimodal format, in both immediate and delayed tests.

Immediate retrieval vs. delayed retrieval. As shown in Table 3, the paired t-tests (FDR corrected, $p < 0.05$) showed that, for retrieving information encoded by text reading, the activation in the left hippocampus, right parahippocampal, and bilateral precuneus/posterior cingulate declined in activation dramatically in the delayed test compared with the immediate test. By contrast, for video watching, only bilateral precuneus and right posterior cingulate declined in activation in the delayed test as compared to the immediate test. Moreover, there was no difference between immediate test and delayed test for story listening.

3.6.2.2. ROI results. Group differences within ROIs. We subtracted the baseline task from the retrieval task first. Then, the BOLD signals in the five ROIs were extracted from this contrast to test whether the neural basis underlying memory retrieval of information was affected by different encoding modes. These data were analyzed in the context of 2 (Tests: immediate test, delayed test one week later) × 3 (Encoding modes: text reading, story listening, and video watching) repeated measures ANOVAs. The results showed that, there were significant main effects of encoding modes in four of the five ROIs (i.e., left parahippocampal, right precuneus, left and right angular) (all $ps < 0.05$, FDR corrected), but not the frontal ROIs (i.e., right IFG). There was also a significant main effect of tests for each ROIs (all $ps < 0.005$, FDR corrected). In addition, significant interaction effects were found only in the left parahippocampal (all $p < 0.05$, FDR corrected), but not the other ROIs. However, to clearly identify the specific differences across three encoding modes in both tests, we further conducted the simple effects analysis and post hoc comparisons for each ROIs (Xie et al., 2015). As can be seen in Fig. 4, the results indicated that, in the immediate tests, retrieval of information encoded by video watching had more activation in the three ROI regions (right angular: $p < 0.001$; left angular: $p < 0.045$; right precuneus: $p < 0.01$) than that encoded by story listening, and retrieval of information encoded by video watching also had more activation than that of text reading in the right angular ($p < 0.05$). By contrast, in delayed test, retrieval of information encoded by video watching had more activation in left parahippocampal, right angular and right precuneus than both text reading and story listening (all $ps < 0.05$, FDR corrected), suggesting a bimodal encoding advantage in memory retrieval than unimodal ones (see Fig. 5).

Immediate retrieval vs. delayed retrieval. Simple effects analysis also showed that, for text reading, the BOLD signals declined significantly in four of the five regions (left parahippocampal: $p < 0.001$; right angular: $p < 0.05$; left angular: $p < 0.01$; right precuneus: $p < 0.01$; FDR corrected). For video watching, the BOLD signals declined significantly in only two regions (right angular: $p < 0.01$; left angular: $p < 0.01$; FDR corrected). However, there were only significant changes in left angular between immediate and delayed retrieval for story listening ($p < 0.05$) (see Figs. 4 and 5).

3.7. Discussion

In Experiment 2, we found no behavioral differences across the encoding modes in the immediate test, which is inconsistent with the results of Experiment 1. This lack of significant effects might be due to the small number of participants in Experiment 2 as compared to Experiment 1: only 16 participants in each group took part in the fMRI study. Another reason may be the fact that the background noise from the fMRI scanner in Experiment 2 had a detrimental effect on retrieval processing, as pointed out by a study showing the detrimental effect of background music on a memory task (Kämpfe, Sedlmeier, & Renkewitz, 2011).

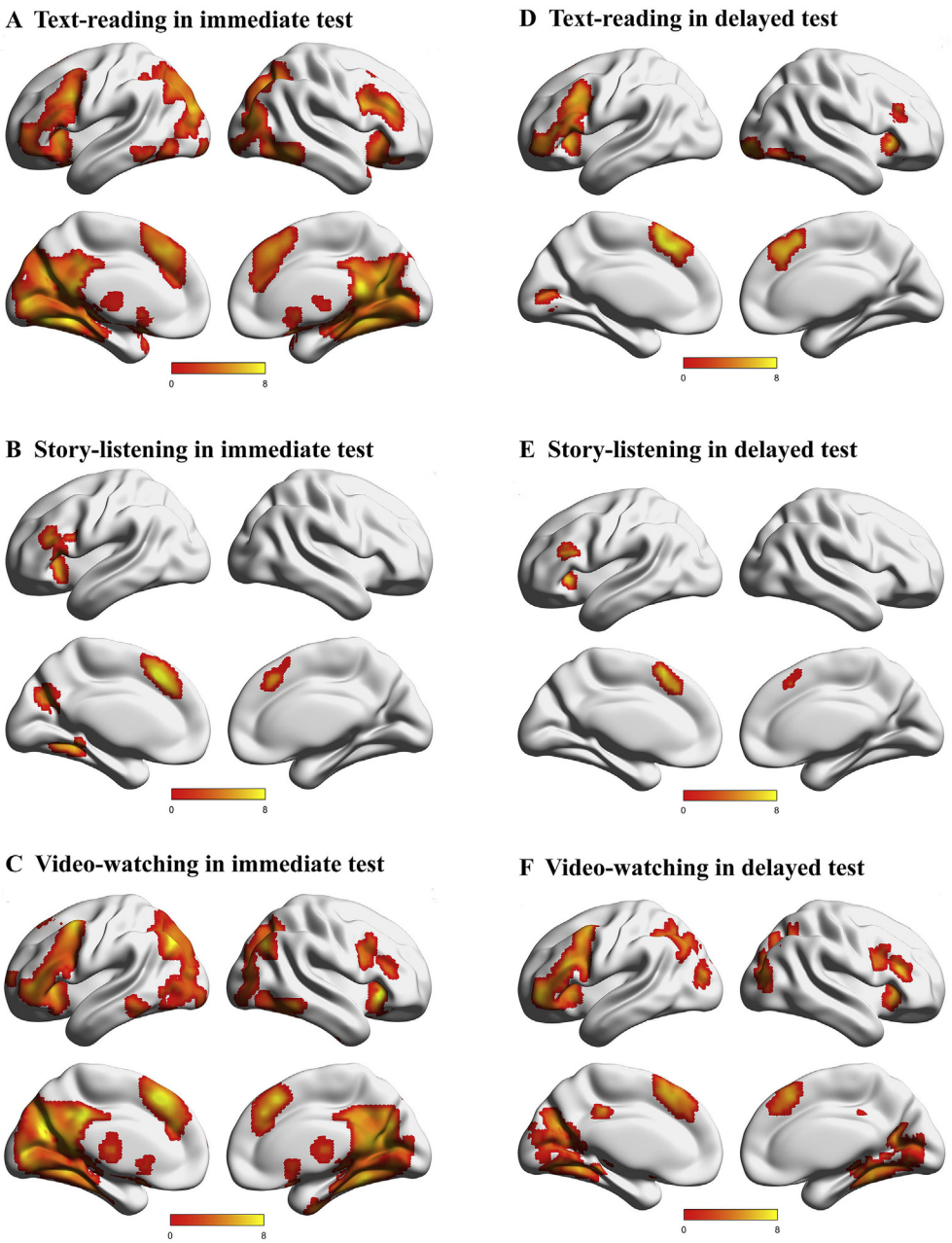

Fig. 3. Brain activation across three encoding modes in both immediate and delayed tests.

However, the ROI analysis results showed that the activation in some specific brain areas such as right angular gyrus was stronger in video watching than the other two encoding modes (i.e., text reading and story listening), in both immediate and delayed test, which suggested a bimodal encoding advantage in retrieving naturalistic information, though there were no behavioral differences. Since fMRI is a more sensitive way to measure individual differences than traditional behavioral methods, we have reasons to believe that the memory advantage of the bimodal encoding mode for retrieval of naturalistic episodes from real-life events could be better observed in fMRI data.

4. General discussion

The present study investigated the memory advantage of bimodal encoding as compared with unimodal encoding (auditory or visual) for retrieval of naturalistic information from real-life events and their changes over time. The behavioral results in Experiment 1 demonstrated that, in the immediate retrieval test, the participants behaved better in retrieving information encoded by video

Table 3
Contrasts between immediate test and delayed test for three encoding modes.

Regions	Hemisphere	MNI coordinates			T values	Voxels
		x	y	z		
Immediate - delayed						
Text reading						
Hippocampus	L	27	36	6	7.26	124
ParaHippocampal	R	33	33	12	4.77	35
Precuneus	L	6	66	45	4.41	297
Precuneus	R	9	66	36	7.99	322
Post Cingulate	L	6 -	42	21	6.08	50
Post Cingulate	R	6	39	18	6.43	53
Story listening						
None	–	–	–	–	–	–
Video watching						
Precuneus	L	9	63	33	9.82	203
Precuneus	R	12	51	15	6.46	118
Post Cingulate	R	12 -	45	30	5.55	53

* FDR corrected, $p < 0.05$.

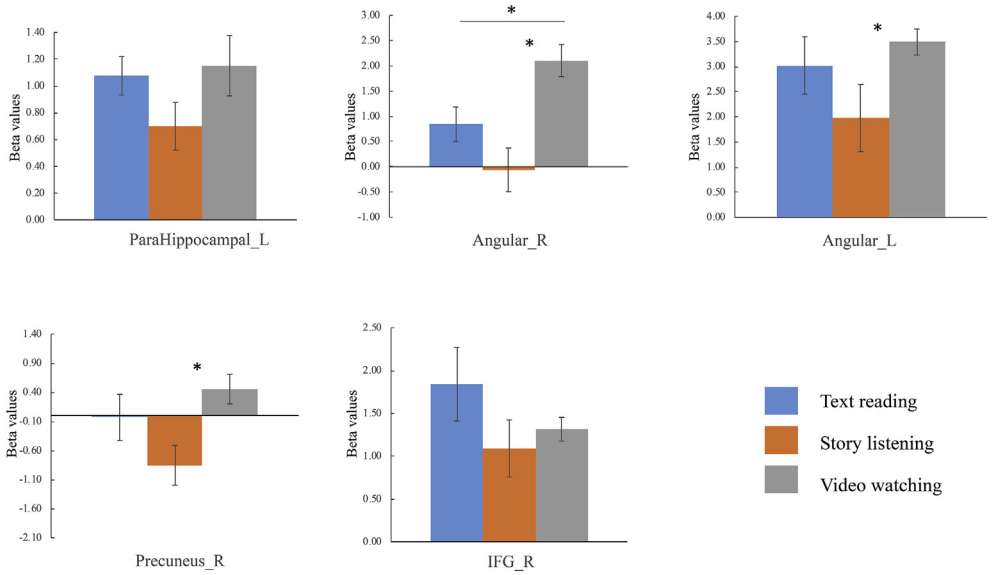

Fig. 4. Mean BOLD for each encoding mode in each of the five ROIs during immediate retrieval.

watching than story listening, which suggested a memory advantage of the bimodal encoding. But these advantages were absent in delayed retrieval test (Experiment 1 and behavior results in Experiment 2). Further, the fMRI results provided evidence for this advantage, which confirmed the memory advantage of the bimodal encoding for retrieval of naturalistic information.

4.1. Memory advantage in bimodal encoding

Whether the memory advantage of the bimodal encoding exists was debated in the literature, and one previous study found that this advantage for the retrieval of isolated stimulus was observed with a between-subjects design, but not with a within-subjects design (Tracy et al., 1988). In the current study, we conducted both between-subjects design in Experiment 1 and within-subjects design in Experiment 2 to investigate this issue. Our behavioral results in Experiment 1 indicated that retrieving episodic information presented by video watching was significantly better than story listening in the immediate test. In addition, although these behavioral differences were absent in Experiment 2, the fMRI results revealed the activation in some brain areas such as right angular gyrus were stronger in retrieving information presented by video watching (i.e., bimodal encoding) as compared to story listening and text reading (i.e., unimodal encoding), in both immediate and delayed tests.

Our results are consistent with the cognitive load theory. According to the modality effect in cognitive load theory (Low & Sweller, 2005; Sweller, 2005, pp. 19–30), the extraneous cognitive load is reduced by the use of dual modality, which increases working memory capacity. Compared with text reading and story listening in which only one modality is involved, the video

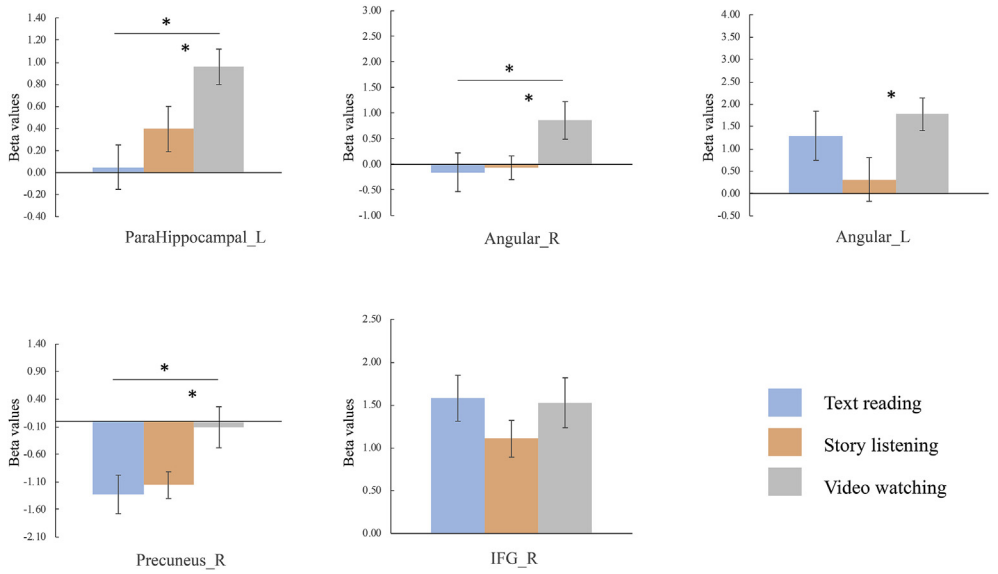

Fig. 5. Mean BOLD for each encoding mode in each of the five ROIs during delayed retrieval.

watching condition (audio-visual encoding) induces a lower load in visual or auditory working memory because auditory information and visual information are each processed in their respective system. The total load in the latter condition is spread across the visual and the auditory components in the working memory system. In other words, the video watching condition (audio-visual encoding) reduced cognitive load, freeing resources for comprehension and memory encoding (Mayer et al., 1999). Hence, the representation of audio-visual episodic information, when the information is acquired through bimodal encoding, would appear to be more robust and also be easier to recall than auditory or visual episodic information alone.

This memory advantage of the bimodal encoding has also been previously described with reference to the differences in memory vividness across different modalities (Gilboa, Winocur, Grady, Hevenor, & Moscovitch, 2004). Specifically, one assumption of the cognitive theory of multimedia learning indicates that multimedia learning is an active process of filtering, selecting, organizing, and integrating information (Mayer, 2005, 2009). Because video watching may involve rich features of audio-visual information during the process of information organization and integration, especially when reconstructing or re-experiencing is necessary, its facilitation to memory retrieval would be more significant than that in story listening or text reading. Thus, the current results extend previous findings from the isolated stimulus, suggesting a bimodal advantage in retrieving the naturalistic episodes from real-life events.

One other possible explanation is that the observed bimodal encoding advantage in the current study arises from the deprivation of multi-sensory modality. Specifically, when a naturally made bimodal documentary was deprived of either modal presentation, the unimodal formats were by no means natural and comfortable for the participants, eventually leading to our memory performance and the involved neural network being impacted. [1]

Our behavioral results demonstrated that retrieving information encoded by text reading was significantly better than retrieving that encoded by story listening in Experiment 1. This finding was consistent with previous behavioral results of better performance in the recall of visual stimulus (pictures) compared with the recall of auditory words (Buckner et al., 1996). As the visual episodic information is more accessible than auditory episodic information (Cohen, 1973), participants can easily catch the key points of the visual information, leading them to behave better in this presentation format. The neuroimaging results in our ROI analysis in Experiment 2 also support this finding, in that the activation in bilateral angular gyrus was stronger in retrieving information encoded by text reading than story listening, although these differences did not reach significance. The presentation formats of test materials may have played a role in this non-statistically significant effect: in the learning stage, the information was presented by text reading, story listening or video watching. However, all the test materials were only presented in a textual format during testing. It is likely that the same presentation format (textual) across the learning stage and test stage facilitated the performance in retrieval in the text reading.

In addition, compared to the immediate retrieval, participants' performance became worse significantly in delayed retrieval (see also Furman et al., 2012). This is similar to the findings from Furman et al. (2012) study, in which participants were scanned during a memory test either hours, weeks, or months after viewing a documentary movie (see discussion in *Introduction*). However, it is noted that the degree of decline in performance across three encoding modes was different. Specifically, retrieval of the information encoded by text reading become worse, since the activation in four of five ROIs declined dramatically, while only one ROI declined dramatically for video watching. Meanwhile, the poor behavioral performance and low activation for ROIs suggested that the

[1] We owe this explanation to one of the anonymous reviewers.

performance for story listening was at a low level across time. Thus, the bimodal advantage becomes more obvious in delayed retrieval, as the activations in left parahippocampal, right angular gyrus and right precuneus for video watching were stronger than those for text reading and story listening. These results might suggest that the bimodal encoding advantage in retrieving naturalistic information not only existed in short-term memory, but also become more pronounced for long-term memory.

4.2. Neural mechanisms for retrieving of real-life information

The memory-sensitive areas found in the current study have been noted in many previous fMRI studies, including the hippocampus, parahippocampal, posterior cingulate and lateral parietal cortices, and medial PFC (for review, see Vilberg & Rugg, 2008). However, most of these brain areas were mainly based on studies using isolated stimuli retrieval. Few studies have focused on the retrieval of real-life episodic information under naturalist conditions (except Chen et al., 2016). Recent functional imaging studies have suggested that natural memory function is subserved by a set of distributed networks, which include cortical regions in the medial PFC, the precuneus, and the angular gyrus (Buckner et al., 2008; Hasson, Chen, & Honey, 2015). Our results from the whole brain analysis extended these previous findings by showing that extensive brain areas including both cortical (e.g., bilateral hippocampus, parahippocampal, mPFC, precuneus, PCC and angular gyrus) and subcortical regions (e.g., the caudate nucleus and putamen) were involved in retrieval of the naturalistic information from real-life events.

Finally, our results suggested that the activation of some particular memory-related brain regions could be modulated by the modes of encoding. As the naturalistic stimuli encoded by video watching contain more vivid content, they require more mental resources to integrate, and therefore may engage more activation in brain regions such as angular gyrus than the other two encoding modes (i.e., text reading and story listening). The angular gyrus and the adjacent interior parietal lobe (IPL) have long been regarded as a semantic integration hub, according to a number of theories (see Binder & Desai, 2011 for review). Previous studies have found that the angular gyrus, especially in the right hemisphere, is consistently involved in semantic processing. For example, Binder et al. (2009) performed a meta-analysis of 120 functional neuroimaging studies, and found that the left angular gyrus is consistently activated in semantic integration tasks, and the right angular gyrus is also seen with stable but a less stronger level of activation. They also suggested that the angular gyrus may play a particular role in complex knowledge retrieval and information integration, and more importantly, the level of activation in the angular gyrus reflects the amount of information that can be successfully retrieved from a given input (Binder & Desai, 2011). A recent study by Yang et al. (2016) showed that the right angular gyrus plays a significant role in interpreting Chinese idioms: comprehension of these idioms involves the process of integration across multiple sources of information, including cultural and literal background information, and the right angular gyrus serves an important role underlying this integration. In sum, the involvement of the angular gyrus in the current study suggests the critical role of angular gyrus in bimodal integration during the retrieval of real-life events, and the different levels of activation in the right angular gyrus provide evidence for bimodal advantage as compared to unimodal encoding modes.

Conflicts of interest

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

Acknowledgements

This work was supported by the Project of Key Institute of Humanities and Social Sciences, MOE (17JJD190001), Guangdong Province Universities and colleges Pearl River Younger Scholar Funded Scheme (2016). Partial support for this work is also provided by the Guangdong Pearl River Talents Plan Innovative and Entrepreneurial Team grant #2016ZT06S220, National Natural Science Foundation of China (31500924) and Innovative School Project in Higher Education of Guangdong, China (GWTP-GC-2017-01).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jneuroling.2019.100863.

References

Bernstein, L. E., Auer, E. T., Jr., Wagner, M., & Ponton, C. W. (2008). Spatiotemporal dynamics of audio-visual speech processing. *NeuroImage, 39*(1), 423–435.
Binder, J. R., & Desai, R. H. (2011). The neurobiology of semantic memory. *Trends in Cognitive Sciences, 15*(11), 527–536.
Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex, 19*(12), 2767–2796.
Buckner, R. L., Andrews-Hanna, J. R., & Schacter, D. L. (2008). The brain's default network: Anatomy, function, and relevance to disease. *Annals of the New York Academy of Sciences, 1124*(1), 1–38.
Buckner, R. L., Raichle, M. E., Miezin, F. M., & Petersen, S. E. (1996). Functional anatomic studies of memory retrieval for auditory words and visual pictures. *J. Neurosci. 16*(19), 6219–6235.
Chen, J., Honey, C. J., Simony, E., Arcaro, M. J., Norman, K. A., & Hasson, U. (2016). Accessing real-life episodic information from minutes versus hours earlier modulates hippocampal and high-order cortical dynamics. *Cerebral Cortex, 26*(8), 3428–3441.
Cohen, G. (1973). How are pictures registered in memory? *Quarterly Journal of Experimental Psychology, 25*, 557–564.
Crooks, S. M., Cheon, J., Inan, F., Ari, F., & Flores, R. (2012). Modality and cueing in multimedia learning: Examining cognitive and perceptual explanations for the

modality effect. *Computers in Human Behavior, 28*, 1063–1071.
Crowe, S. F. (2000). Does the letter number sequencing task measure anything more than digit span? *Assessment, 7*(2), 113–117.
Fan, J., McCandliss, B. D., Sommer, T., Raz, A., & Posner, M. I. (2002). Testing the efficiency and independence of attentional networks. *Journal of Cognitive Neuroscience, 14*(3), 340–347.
Furman, O., Mendelsohn, A., & Dudai, Y. (2012). The episodic engram transformed: Time reduces retrieval-related brain activity but correlates it with memory accuracy. *Learning & Memory, 19*(12), 575–587.
Gilboa, A., Winocur, G., Grady, C. L., Hevenor, S. J., & Moscovitch, M. (2004). Remembering our past: Functional neuroanatomy of recollection of recent and very remote personal events. *Cerebral Cortex, 14*(11), 1214–1225.
Goolkasian, P., & Foos, P. W. (2005). Bimodal format effects in working memory. *American Journal of Psychology, 118*, 61–77.
Hasson, U., Chen, J., & Honey, C. J. (2015). Hierarchical process memory: Memory as an integral component of information processing. *Trends in Cognitive Sciences, 19*(6), 304–313.
Kämpfe, J., Sedlmeier, P., & Renkewitz, F. (2011). The impact of background music on adult listeners: A meta-analysis. *Psychology of Music, 39*(4), 424–448.
Lan, Y. J., Fang, S. Y., Legault, J., & Li, P. (2015). Second language acquisition of Mandarin Chinese vocabulary: Context of learning effects. *Educational Technology Research & Development, 63*(5), 671–690.
Low, R., & Sweller, J. (2005). The modality principle in multimedia learning. In R. Mayer (Ed.). *Cambridge handbook of multimedia learning* (pp. 147–158). New York: Cambridge University Press.
Mastroberardino, S., Santangelo, V., Botta, F., Marucci, F. S., & Belardinelli, M. O. (2008). How the bimodal format of presentation affects working memory: An overview. *Cognitive Processing, 9*(1), 69–76.
Mayer, R. E. (2005). Cognitive theory of multimedia learning. In R. Mayer (Ed.). *Cambridge handbook of multimedia learning* (pp. 31–48). New York: Cambridge University Press.
Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). New York: Cambridge University Press.
Mayer, R. E., Moreno, R., Boire, M., & Vagge, S. (1999). Maximizing constructivist learning from multimedia communications by minimizing cognitive load. *Journal of Educational Psychology, 91*, 638–643.
Raven, J., Raven, J. C., & Court, J. H. (1998). *Manual for Raven's Progressive Matrices and Vocabulary Scales.* San Antonio, TX: Pearson, Inc.
Ritchey, M., Montchal, M. E., Yonelinas, A. P., & Ranganath, C. (2015). Delay-dependent contributions of medial temporal lobe regions to episodic memory retrieval. *Elife, 4*, e05025.
Rugg, M. D., & Vilberg, K. L. (2013). Brain networks underlying episodic memory retrieval. *Current Opinion in Neurobiology, 23*(2), 255–260.
Santangelo, V., Mastroberardino, S., Botta, F., Marucci, F. S., & Belardinelli, M. O. (2006). On the influence of audio-visual interactions on working memory performance: A study with non-semantic stimuli. *Cognitive Processing, 7*(1) 187-187.
Schneider, E., & Zuccoloto, A. (2007). *E-prime 2.0 [computer software]: Pittsburg.* PA: Psychological Software Tools.
Snyder, P. J., & Harris, L. J. (1993). Handedness, sex, familial sinistrality effects on spatial tasks. *Cortex, 29*(1), 115–134.
Sweller, J. (2005). *Implications of cognitive load theory for multimedia learning.* The Cambridge handbook of multimedia learning.
Sweller, J. (2010). Element interactivity and intrinsic, extraneous and germane cognitive load. *Educational Psychology Review, 22*, 123–138.
Sweller, J., & Chandler, P. (1994). Why some materials is difficult to learn. *Cognition and Instruction, 12*(3), 185–233.
Takashima, A., Nieuwenhuis, I. L. C., Jensen, O., Talamini, L. M., Rijpkema, M., & Fernández, G. (2009). Shift from hippocampal to neocortical centered retrieval network with consolidation. *Journal of Neuroscience, 29*(32), 10087–10093.
Takashima, A., Petersson, K. M., Rutters, F., Tendolkar, I., Jensen, O., Zwarts, M. J., et al. (2006). Declarative memory consolidation in humans: A prospective functional magnetic resonance imaging study. *Proceedings of the National Academy of Sciences, 103*(3), 756–761.
Thompson, V. A., & Paivio, A. (1994). Memory for pictures and sounds: Independence of auditory and visual codes. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 48*(3), 380.
Tracy, R. J., Roesner, L. S., & Kovac, R. N. (1988). The effect of visual versus auditory imagery on vividness and memory. *Journal of Mental Imagery, 12*, 145–161.
Tracy, R. J., Tracy, J. K., & Ramsdell, C. L. (1985). The relationship between imagination and memory. *Journal of Mental Imagery, 9*(3), 91–108.
Vilberg, K. L., & Rugg, M. D. (2008). Memory retrieval and the parietal cortex: A review of evidence from a dual-process perspective. *Neuropsychologia, 46*(7), 1787–1799.
Wager, T. D., & Nichols, T. E. (2003). Optimization of experimental design in fMRI: A general framework using a genetic algorithm. *NeuroImage, 18*(2), 293–309.
Winnick, W. A., & Brody, N. (1984). Auditory and visual imagery in free recall. *Journal of Psychology, 118*(1), 17–29.
Yang, J., Li, P., Fang, X., Shu, H., Liu, Y., & Chen, L. (2016). Hemispheric involvement in the processing of Chinese idioms: An fMRI study. *Neuropsychologia, 87*, 12–24.
Yan, C.-G., Wang, X.-D., Zuo, X.-N., & Zang, Y.-F. (2016). DPABI: Data processing & analysis for (resting-state) brain imaging. *Neuroinformatics, 14*(3), 339–351.