HEAR THIS, READ THAT; AUDIOVISUAL INTEGRATION EFFECTS ON RECOGNITION MEMORY

By

NADIA POK AH WONG, B.SC

A Thesis Submitted to the School of Graduate Studies in Partial Fulfilment of the Requirements for the Degree of Masters of Science

Descriptive Note

McMaster University MASTERS OF SCIENCE (2014) HAMILTON, ONTARIO (Department of Psychology, Neuroscience & Behavior)

TITLE: Hear this, Read that; Audiovisual Effects on Recognition Memory AUTHOR: Nadia Pok Ah Wong, B.Sc (McMaster University) SUPERVISOR: Dr. D.I.Shore

NUMBER OF PAGES: viii, 45

Abstract

Our experience with the world depends on how we integrate sensory information. Multisensory integration generates contextually rich experiences, which are more distinct and more easily retrievable than their unisensory counterparts. Here, we report a series of experiments examining the impact semantic audiovisual (AV) congruency has on recognition memory. Participants were presented with AV word pairs which could either be the same or different (i.e., hear “ring”, see “phone”) followed by a recognition test. Recognition memory was found to be improved for words following incongruent presentations. Results suggest higher cognitive processes may be recruited to resolve sensory conflicts, leading to superior recognition for incongruent words. Integration may help in easing the processing of multisensory events, but does not promote the processing needed to make them distinctive.

Keywords: audiovisual presentation, congruency, multisensory, recognition memory

Preface

Original concept and development of Experiment 1 in this dissertation was developed as a third year undergraduate group project for the Psychology lab course: Psych 3EE3. The paradigm developed in Experiment 1 (and used in Experiment 2) was established by myself, J. Irvine, A. Tankel, M. Warren, M. L. Cadieux, M. Pachai and D. I. Shore. Experiment 2 was developed through discussion between J. Irvine, D.I. Shore and myself. Development of Experiment 3 was through discussions between myself and D.I. Shore.

The experimental script in the aforementioned studies (Experiment 1 and 2) was developed by myself and D. I. Shore. The experimental script for Experiment 3 was primarily my work with the contribution of D.I. Shore and M. Pachai in setting up noise presentation. Verbal stimuli presented in all experiments were contributed by M. Warren and C. Aruffo. Selection of verbal stimuli was done by A. Tankel, J. Irvine, M. Warren and myself.

Data collected by myself, J. Irvine, A. Tankel and M. Warren for the third year group project were included in Experiment 1a. Data collection for Experiment 1a was conducted by J. Irvine and myself. Data analysis in Experiment 1a was contributed by M.L. Cadieux, M. Pachai, J. Irvine and myself. Data collection and analysis was solely done by myself. M. Sadik was responsible for collecting the data presented in Experiment 3 with my contribution to the data analysis.

Acknowledgments

To begin, I would like to express my sincerest gratitude toward Dr. David I. Shore for his invaluable support, dedication and guidance throughout this research project.

I would also like to express my appreciation to my committee members: Dr. Bruce Milliken and Dr. Scott Watter, for their input, suggestions and thought provoking questions.

I would like to thank Jade Irvine, Andrea Tankel and Melanie Warren for their help in discussing and developing the paradigm for these experiments, providing and creating the stimuli used and collecting pilot data for this research project. I am indebted to them for allowing me to advance our third year group project to the research program it is now. I am grateful for Dr. Michelle Cadieux and Matt Pachai for their help in data processing and organization throughout the course of developing this research program. I would also like to thank Chris Aruffo for taking the time to help record the stimuli for these experiments. I am immensely appreciative for the time and effort Marina Sadik spent collecting data and keeping me on track.

I thank my lab mates: Anna Finkelshtein, Brendan Stanley, Kaian Unwalla, Amy Pachai, and Alex Gough for the stimulating discussions, deadline cram sessions, motivation, and all the fun story times and stress-relieving Rudshuk games this past year. I would also like to extend my thanks to the Milliken Lab for their questions, comments and discussions.

Finally, I am grateful to my parents and brother: Guinea Wong, Perly Fok and Berkley Wong for their continued support and faith in me through my academic career and life choices. I want to also extend my appreciation to my friends, most notably: Linda Liu, for her continuous belief in me, generous patience and motivation though my complaints, Isaac Ng, for his motivating critiques and Matthew Lai, for pushing me to work harder and for having unyielding faith in me.

Table of Contents

List of Figures

List of Abbreviations

AV = Audiovisual

ms = milliseconds

Introduction

Sensory coincidence and past experiences combine to produce perception. Multisensory processes influence this perception. Cues from multiple modalities are optimally integrated to generate a single conscious percept (Ernst & Bülthoff, 2004). Perception, in turn, determines the encoded memory trace. While there has been a great deal of recent research on the factors influencing multisensory integration, there has been less work on understanding the memorial consequences of those integrated percepts. There is an apparent divide in the literature regarding the movement between perception and memory. Multisensory research on perception would suggest integration of congruent modality signals would permit a richer memory trace (Ernst & Bülthoff, 2004; Alais, Newell & Mamassian, 2010). However, metamemory researchers would argue perceptual fluency at the time of encoding will be detrimental to memory formation (Bjork & Bjork, 2011; Besken & Mulligan, 2013). To reconcile these differences, the present study examines how multisensory integration and conflict affect memory for audiovisual (AV) stimuli.

Multisensory integration optimizes perception by streamlining information. Corresponding sensory information helps to reduce ambiguity of an incoming stimulus (Ernst & Bülthoff, 2004). Resolution of a signal stimulus sharpens as variance and noise is filtered by integrating sensory information. Noise occurring in separate modalities is independent of one another. When a sensory signal (e.g. Visual) carries too much noise, a separate sensory signal (e.g. Auditory) may be interpreted as more reliable to discriminate the stimulus (Ernst & Banks, 2002). By allowing for information to be provided by multiple modalities, one can continuously reduce the variance and ambiguity of a stimulus. Reduction of variance improves the resolution of the stimulus at the time of encoding, promoting a contextually richer and more distinctive memory trace. However, sensory modalities must correspond with one another in multiple dimensions for optimal integration to occur.

Prior assumptions about unity influences multisensory integration (Vatakis & Spence, 2007). Stimulus driven factors such as spatial concurrence and temporal synchrony encourage the observer to assume corresponding multisensory events are originating from the same source (Jackson, 1953; Welch, Dutton, Hurt & Warren, 1986). When AV cues appear close together, participants interpreted the spatial correspondence as strong evidence of the two cues originating from the same source (Jackson, 1953). Sensory events occurring synchronously in time can also be interpreted as originating from a single source (Massaro, Cohen & Smeele, 1996; Roach, Heron & McGraw, 2006; van Wassenhove, Grant & Poeppel, 2007). In a visual temporal order judgement (TOJ) task, presentation of a task-irrelevant auditory signal trailing the presentation of the second visual cue affected TOJ performance (Morein-Zamir, Soto-Faraco, & Kingstone, 2003). When sounds intercepted visual cues, performance declined. As the auditory signals appear further apart from each other, the perception of visual cues is also influenced. As sensory information becomes further removed from each other, they become separate events. Semantically learnt associations between complex sensory information also affects judgments of unity (Koppen, Alsius, & Spence, 2008; Laurienti, Kraft, Maldjian, Burdette, Wallace, 2004; Radeau & Colin, 2001; Vatakis & Spence, 2007). For example, in a TOJ task, participants were asked to determine whether the visual or auditory cue came first. Verbal syllables – recorded in both genders – were presented with along with conflicting or matching visual lip movements (i.e. male voice with male lips versus male voice with female lips). Participants had a harder time making a temporal judgment when auditory and visual cues match (e.g. lips – voice/voice – gender) (Vatakis & Spence, 2007). Performance improvement following semantic mismatching suggests the importance of semantic congruency on assumptions of unity. When unity of sensory information can be assumed, integration can act perception.

Integrated percepts may impact memory processes differently compared to unimodal percepts. In modality specific models of knowledge, memory representations are independently hosted within the processing modality (Barsalou, Simmons, Barbey & Wilson, 2003). When presented either bimodal or unimodal conflicting information, identification accuracy fell only when conflicting information was presented within the same modality (Larsen, McIlhagga, Baert & Bundesen, 2003). The cost observed in unimodal conflict suggests at the perceptual level, encoding may be separated. In the case of multisensory perception, information processed by modality specific input may be integrated in hierarchical associative areas (Damasio, 1989). Formation of multisensory memories may be attributed to activation of integrated areas rather than immediate perception. Congruency between modality inputs may serve to highlight shared associations and ideas, leading these properties to be more distinctive, thus easily retrievable. Cross modal priming studies demonstrate effect of cross modality interference on later categorization tests (Vallet, Brunel & Versace, 2010; Vallet, Riou, Versace & Simard, 2011). Participants were presented with auditory stimuli accompanied either with or without a visual mask and asked to make a categorization. Upon the test phase, participants were asked to perform the same categorization task using visual stimuli associated to the auditory stimuli presented during the learning phase. Slower response times were seen when participants judged visual stimuli related to auditory primes presented with the visual mask. Longer response times suggest visual masking interrupted the mechanisms needed to activate multisensory experiences thereby affecting the memory formation and retrieval (Brunel, Lebeye, Lesourd & Versace, 2009). Different mechanisms for memory encoding between integrated and unimodal percepts suggests retrieval performance for these traces will also be different.

Given differences in encoding, we asked if explicit memory is better for integrated percepts. Explicit memory is the storage of specific events and information. Accessibility to these memory traces is dependent on how information was perceived and encoded (Tulving & Thomson, 1973). Recognition memory involves the retrieval of these experiences through two means, familiarity and recollection (Yonelinas, 2001). For example, perceptual fluency is the speed and ease it takes to perceive information (Westerman, Lloyd & Miller, 1992; Whittlesea, Jacoby & Giard, 1990). Processing fluency invokes a sense of familiarity, enabling the observer to believe the current experience is a reoccurrence of one from the past. Recollection, the ability to retrieve a specific memory is depended on the strength of the trace during encoding (Craik & Lockhart, 1972). The strength of the trace is dependent on the initial processing a stimulus undergoes during encoding. For example, at time of encoding, semantically judged words (i.e. Is this word affiliated with the word tiger?) improves it's probability of being recognized opposed to a word that was simply judged for its physical characteristics (e.g. Does it have a letter 'A'?) in a subsequent recognition test (Jacoby & Dallas, 1981). Perception and processing of the stimuli at learning heavily impacts the availability of a specific memory trace. As more information is perceived and encoded, more avenues to retrieve the learnt episode can be manipulated.

Memories for multisensory experiences may be easier to retrieve due to the enriched information upon encoding, leading to a stronger memory trace (Lehmann & Murray, 2005; Murray et al. 2004). For instance, in a continuous recognition memory task (Murray. Michel, Peralta, Ortigue, Brunet, Andino & Schnider, 2004), participants were asked to discriminate whether a visual line drawing had appeared earlier. Upon the first presentation of the drawings, images were either presented only visually or accompanied by a semantically associated task-irrelevant sound (e.g. dog – bark). Recognition accuracy was higher for trials involving the appearance of the task irrelevant sound. In a follow-up study (Lehmann & Murray, 2005), participants were presented with visual alone (e.g. dog – none), congruent auditory (e.g. dog – bark) or incongruent auditory (e.g. dog – meow) trials. Again, accuracy was greater for images accompanied by a semantically congruent sound. Recognition rates did not differ between the semantically incongruent trials and the visual only trials. The lack of difference suggests that conflict detected upon encoding will separate, in memory, the individual unisensory events.

We examined recognition memory performance to address the movement of multisensory integration from perception into memory traces. Specifically, we asked whether memory performance between congruent and incongruent trials was due to the opportunity to integrate sensory cues during perceptual processing. Here, participants were passively presented congruent and incongruent AV sentences. Following the sentence presentations, participants were given a simple recognition task. The target words appeared at the end of either the visual or audio version of the sentences. We expected participants would recognize more words that appeared in congruent trials than incongruent trials. This expectation follows the results obtained previous studies. Identical AV information should promote integration between the AV cues, leading a stronger, richer and more distinct trace of the target word. When sensory information conflict, processing of either modality should act as a mask to the other (Vallet et al., 2011) or the conflict would be resolved under the pretense of two separate unisensory events occurring simultaneously. The lack of integration or the threat of interference should result in a less detailed version of the word, thereby increasing the difficulty in retrieving the word upon memory testing.

Experiment 1

Experiment 1A

Methods

Experiment 1a Participants. This experiment consisted of 44 recruited university students participating for class credit. Twenty-eight females and six males between the ages of 17 and 36 years old with (mean age =19.06 years old) participated in the study. All participants provided written informed consent and were naïve to the purpose of the study.

Experiment 1b Participants. Participants were recruited from the undergraduate McMaster University subject pool. Each participant received course credit for their participation. The participants consisted of 17 females and 5 males between the ages of 17 and 20 years old, with a mean age of18.5 years old. Participants provided written informed consent and were naïve to the purpose of the study.

Apparatus. A Røde NT-1000 condenser microphone and Adobe Audition 3 software was used to record auditory sentences and the test word heard in the experiment. MATLAB version rt2009a was used to program and run the entire experiment. Participants viewed the experiment using Mac Computers running Mac OS X, version 10.6. Visual stimuli were presented through a 24” BenQ (GL-2450) monitor at 1080 pi, with a refresh rate of 60Hz. A profile questionnaire (asking for their age, gender, time of experimentation and whether or not they wore corrective lenses) and a handedness questionnaire were completed by the participants before the experiment began.

Headphones for the experiment were provided by the participants themselves. Collected data was compiled and organized using Microsoft Excel 2007. Data analysis was done using IBM SPSS software.

Design and Procedure. This experimental study ran a 2 by 2 research design. We aimed to manipulate the conditions (congruent/incongruent) of audiovisual sentence presentations to investigate observers' ability to recognize previously presented information. Within incongruent AV trials, modality differences during the recognition phase were also explored. As the observers arrived, experimenters instructed them to be seated at any available computer, consent forms were signed, and the experiment was explained verbally to them. Observers were told that they were participating in an experiment that consists of simultaneous audiovisual sentence presentations broken down into 3 blocks, each block being immediately followed with a word recognition task before the next learning block began.

Sentence Creation. Sentences were created following the sequence 'subject-verb-noun'. Sentences were formatted with one constant stem per sentence, which included the subject and verb. Every stem sentence had two associated noun options to be used. In congruent trials only one noun was presented in both the visual and auditory streams. This word would be then be used in the subsequent recognition test. In incongruent trials, two different nouns were presented, one in the visual stream and the other in the auditory stream. One of the two nouns was selected as a target words for the recognition test. For example, in a congruent trial, participants will hear and see the sentence 'She picked up the keys'. The word key will be the target word that will appear in a following recognition task. In an incongruent trial, participants may hear 'She picked up the keys' while simultaneously read 'She picked up the phone'. Either the word 'phone' or 'keys' in this case will reemerge in a subsequent recognition test. Pronouns used were limited to only “he”, “she”, “they” and “we”. The difficulty level of the nouns chosen was around the Grade 5 level (http://www.momswhothink.com/reading/list-of-nouns.html). Semantically similar words were selected so both versions of a sentence made sense. Syllables were matched up between noun options. 72 sentences were created and presented during the learning phases with 144 words being the potential target words. Incongruent trials required the use of 2 different words per trial while congruent trials used 1. Words that were unused in the learning phased appeared in the testing phase as foils. An additional 36 foils were chosen to be presented in the recognition phases. Therefore, a total of 108 words appeared during the learning phases (36 appeared congruently, 72 appeared incongruently) and a total of 72 foils were used during testing.

Sentence and Word Presentation. During the experiment, 36 sentences were presented congruently, and another 36 sentences were presented incongruently. Sentences were presented aurally as well as visually. For incongruent trials, one version of the sentence was presented aurally while the other was presented visually (i.e. one sentence used one of the two possible nouns and the other sentence used the other one). Congruent trials presented the same version of a sentence aurally and visually. The presentation of target words and sentences were randomized in a MATLAB program prior to beginning the experiment.

In the learning phase participants watched the computer screen with headphones on. A white fixation cross appeared for 500ms, followed a white coloured sentence written on a black background was presentation. Each sentence was presented both visually and aurally for 3000ms before disappearing back to a fixation cross (Figure 1). After 24 presentations in the learning phase participants were entered into a recognition test. The learning phase and recognition test comprised 1 block, with a total of 3 blocks in the experiment. Each block was intermitted by a break duration determined by the participant.

In Experiment 1b, sentence stems were removed from presentation during the learning phase. Trial duration was reduced to 1000ms to accommodate this change. All other presentation and procedural steps were identical to Experiment 1a.

Recognition Task Procedure. At the end of each learning phase, a recognition test was conducted (Figure 2). Participants were instructed verbally at the beginning of the experiment that words would appear independently on the screen, and if they had any recognition of the word, regardless of its original presentation, they should press the 'A' key (old), and if they had no recognition of the word at all, they would press the 'L' key (new). Response keys were also presented during the trial in the form of "OLD (A)" and "NEW(L)" at the bottom of the screen. Letters within the brackets were presented to remind participants of which key to press for response. Visual instructions detailing the appropriate key to respond with were given between each learning and test phase. A white fixation cross appeared at the center of the screen between participant's response and the following trial. The word recognition task or 'test trials' consisted of the twenty-four target words that were presented in the prior learning block (twelve visual words, twelve auditory words, each originating from one of six congruent and six incongruent presentations), and twenty-four distractor words. After the three learning and test trials were completed the participants were debriefed.

Participant Instructions. Upon arrival, participants were settled in front of a computer. Prior to beginning the experiment, participants were informed there were three blocks in the test with a break intermitted between each block. Each block consisted of a learning and a testing phase. For the learning phase, participants were informed words would be presented both auditorily and visually to them that may or may not match one another. They were asked to do their best to attend to both words. Participants were informed a recognition test would follow after the learning phase was done.

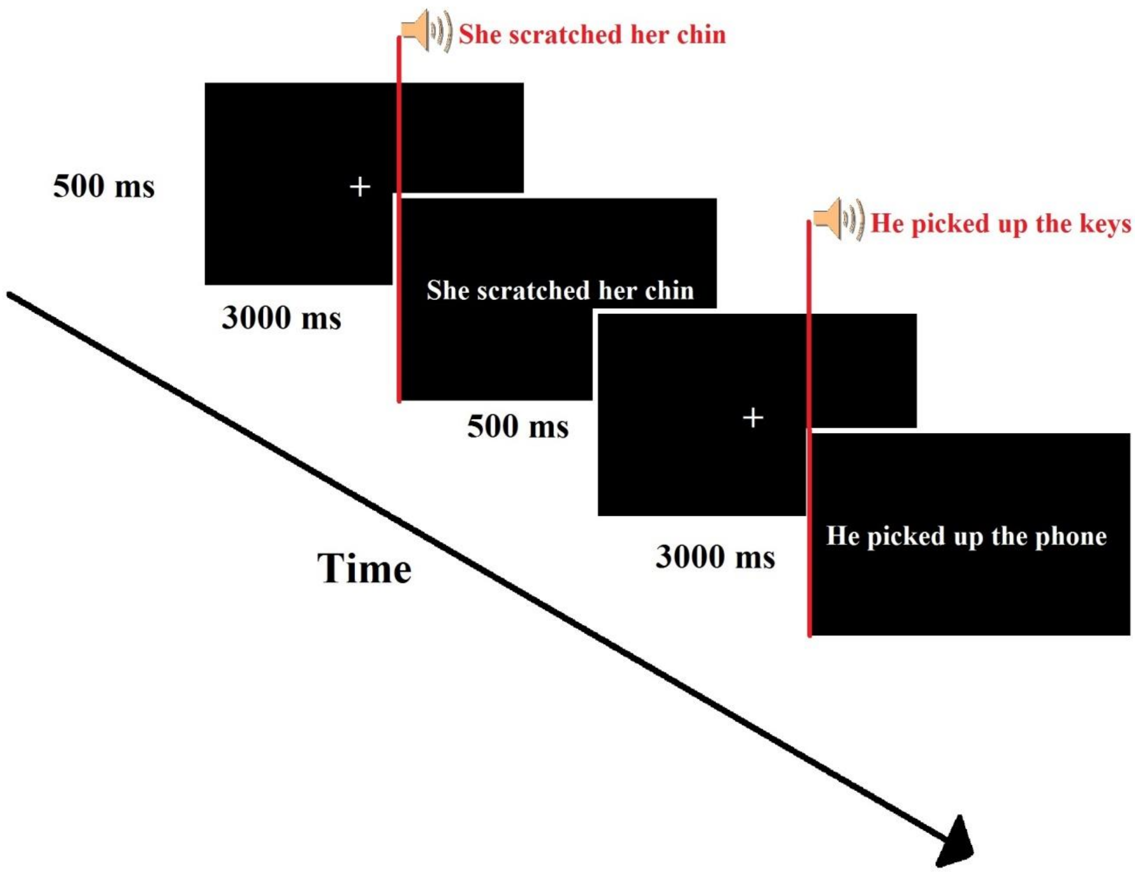

Figure 1. In a typical trial, participants were presented with a 500 ms fixation cross followed by an audiovisual sentence. Sentence stems were identical between modalities. In a congruent trial, the last word of either modality presentation was identical. For incongruent trials, the last word of the sentence differed. Words from either the congruent or incongruent modality reappeared during a subsequent recognition test. Each audiovisual sentence pairs were presented for 3000 ms.

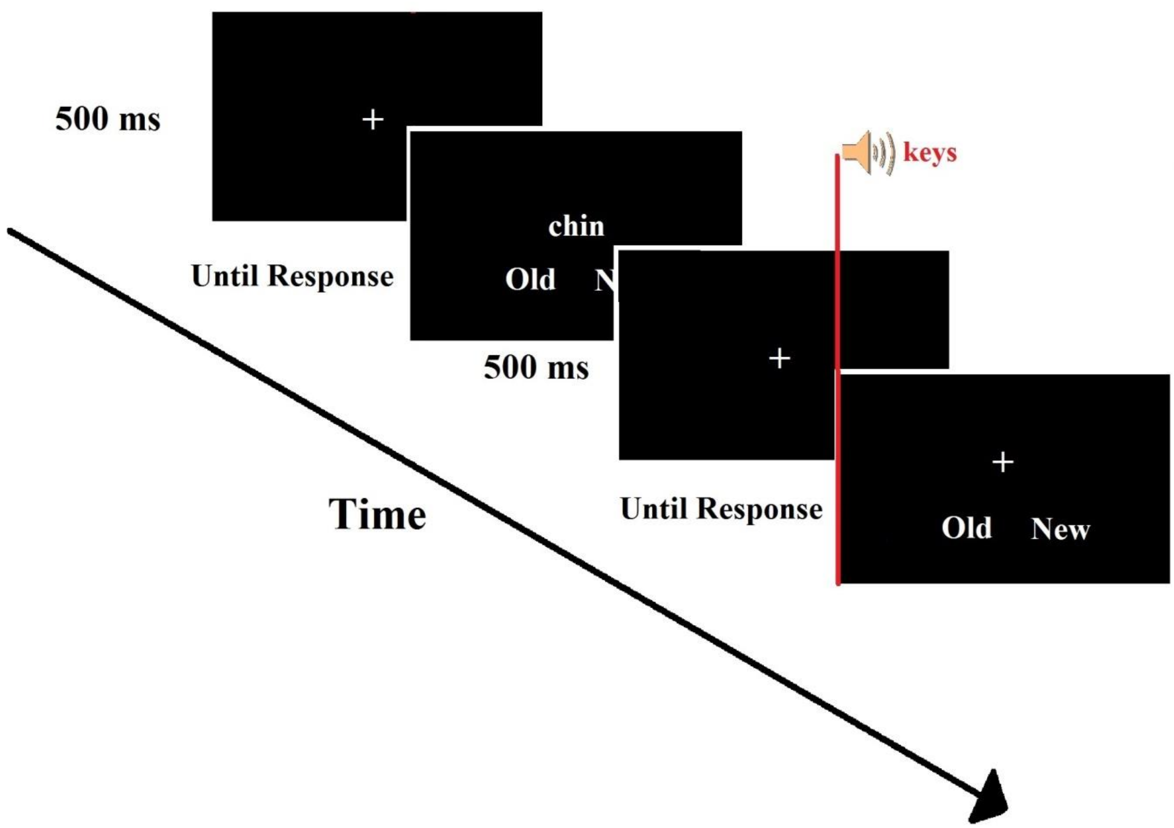

Figure 2. A 'Old-New' recognition task was presented to the participants after each learning phase. Participants were first presented with a white fixation cross for 500 ms. Following, a word would either appear on the screen or through the headphones and accompanied with the 'Old New' response cue. The response screen would remain until participants pressed either the designated "old" or "new" answer keys.

Results

Data Transformation. *d'* was calculated using Microsoft Office Excel 2007 for all participants over the 3 testing blocks. When calculating *d'*, we excluded trials for which the participants responded "new" and included only trials for which the participants responded "old". Sensitivity of word familiarity was determined using the hit rate and false alarm rate of the "old" responses. A hit was considered when participants responded "old" to an old word. False alarm rates were when participants responded "old" to a new word. We did not find any significant relationship between blocks and sensitivity to stimuli, so we collapsed the calculations for the four conditions across blocks for each participant. We then calculated means for these four conditions across participants.

Next, we organize the aforementioned means and calculate standard deviations and standard error for the error bars on the graph (Figure 1). Then we performed a 2x2 factorial ANOVA on the four conditions, by congruency and modality to explore any relationships.

Experiment 1a.

Analysis. There was no significant difference in false alarms made between modalities [$t(1, 43) = -0.133$, $p = 0.829$] suggesting differences observed in experiment are due to perceptual manipulation and not modality-specific memory retrieval mechanisms. Using a 2x2 factorial ANOVA, we found a main effect of congruency [$F(1,43) = 53.960$, $p<0.000$]. The participants were more sensitive to the words from the incongruent trials when they were presented in the test phase (Figure 2). There was also a main effect for modality [$F(1,43) = 6.261$, $p = 0.016$)]. Participants were better at recognizing auditory words regardless of congruency.

Experiment 1b.

Analysis. Analysis of the false alarm rates yielded no significant difference between recognition of auditory and visual words [$t(21) = -1.026$, $p = .316$]. The lack of significance between error rates suggests recognition was dependent on the perceptual manipulation of congruency rather than modality-specific retrieval differences. Removal of the sentence stems continued to yield an significant effect for congruency [$F(1,21) = 49.82$, $p<0.000$)](Figure 4). No modality effect was observed, however, a significant interaction for congruency*modality was found instead [($F(1,21) = 4.903$, $p =0.038$]. Further analyses of the interaction yielded a modality effect for the incongruent condition only. Auditory words were recognized more than visual ones [$t(21) = 2.428$, $p = 0.024$].

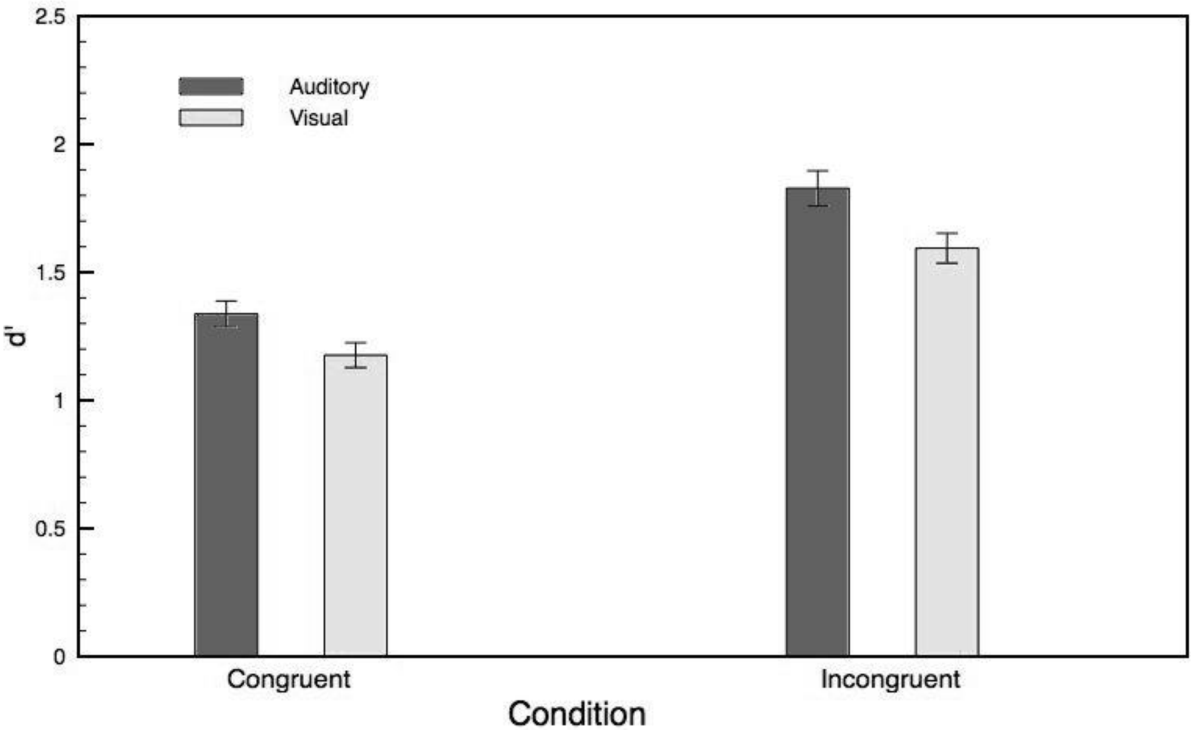

Figure 3. This figure represents the average sensitivity and standard error of 44 participants in recognizing words that had been previously presented to them in one of four conditions. These conditions were congruent visual, congruent auditory, incongruent visual and incongruent auditory. Words that had been presented incongruently were remembered more than those that were congruently presented. An insignificant trend of higher recognition rates of incongruent auditory versus congruent visual was also found. Upon further examination, there was no difference found between modalities in the incongruent conditions.

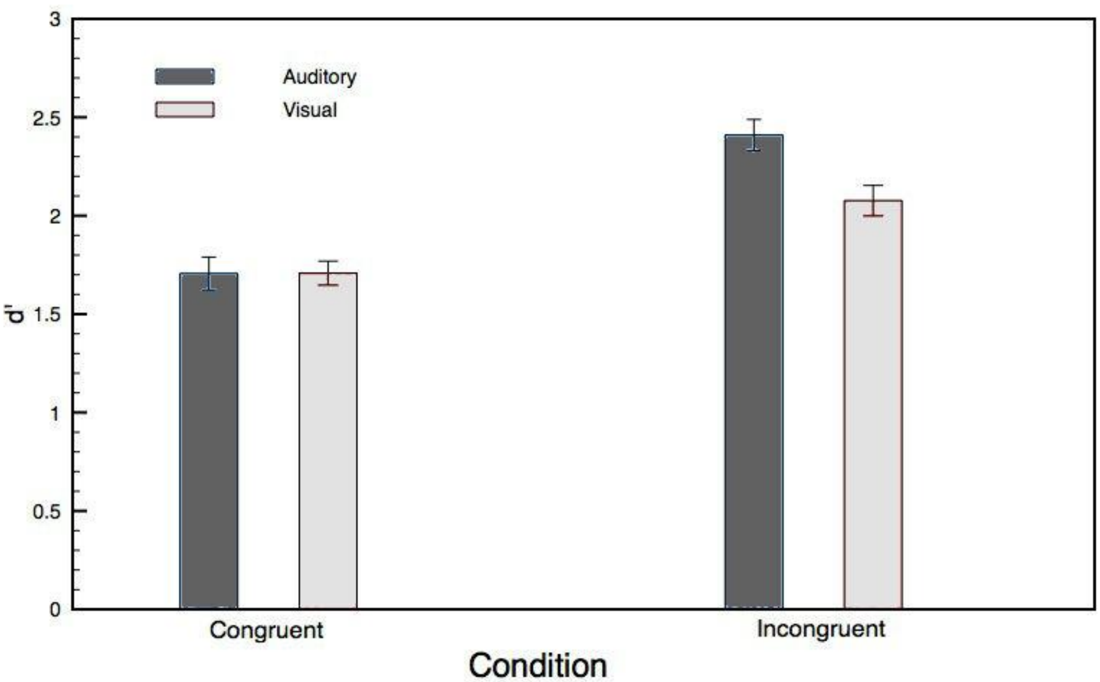

Figure 4. This figure represents the average sensitivity and standard error of 22 participants in recognizing words that had been previously presented to them in one of four conditions when the sentence stem was removed from the learning trials. These conditions were congruent visual (CV), congruent auditory (CA), incongruent visual (IV) and incongruent auditory (IA). Words that had been presented incongruently were remembered more than those that were congruently presented. A significant interaction was seen between congruency and modality. Upon further examination, there a modality effect was found in only the incongruent trials. When incongruent trials words were test, greater recognition of auditory target words was observed.

Discussion

In experiment 1A, audiovisual sentences were passively presented to participants. Sentences were either identical in both modalities or differed in the last word. The last word of either the audio and visual sentence was chosen to later reappear in a recognition test. The recognition test occurred following the end of a learning phase. We examined the effect congruency presentation of sentence stimuli had on recognition. We expected congruently presented words to be recognized more than incongruent ones due to stronger memory traces (Lehmann & Murray, 2005).

Results from experiment 1A indicated that words originating from incongruent presentations were recognized more than words that were presented congruently. A modality effect was also noticed. Auditory words were recognized more than visual words regardless of congruency presentation. No interaction between congruency and modality was observed. Lack of an interaction suggests recognition for words was not influenced by the combination of modality and congruency conditions.

We postulated auditory superiority resulted from increased attention paid toward auditory sentences. Participants may have been aware there was adequate time to focus on the more transient auditory signal before processing the visual one. Similarly, participants may have learnt to ignore the sentence stem altogether. By targeting the last word of the visual sentence first, they may have ample time to process the visual word then anticipate the auditory word. To confront these results, experiment 1B was conducted. Sentence stems were removed, leaving only the target word pairs. This aimed to reduce anticipatory effects. The same procedure was used.

Experiment 1B replicated the congruency effect observed in experiment 1A, suggesting incongruity improved item recognition. Interestingly, removal of the sentence stem also removed the modality effect. However, an interaction between congruency and modality was observed. Auditory words were better recognized than visual words in incongruent trials, whereas no modality differences were seen in congruent trials.

Our results were inconsistent with previous research by Lehmann and Murray (2005) and Murray et al. (2004) and rejected our hypothesis. Lehmann and Murray (2005) demonstrated memories for semantically congruent information were enhanced. Presentations of semantically incongruent information led to recognition performances identical to visual only presentations. Results from Lehmann and Murray (2005) suggest multisensory memory traces may be more distinctive and easily retrievable compared to their unisensory counterparts. Regarding our manipulation, incongruent audiovisual information should have disrupted the processing of information in both modalities. Interrupted processing and superior retrieval of multisensory events should reserve enhanced recognition for words appearing in congruent trials.

The appearance of an interaction between congruency and modality is interesting. Increased recognition for auditory words only in the incongruent condition adds evidence toward our increased processing argument. We argue the lack of a modality performance difference in our congruent condition suggests multisensory integration occurred during encoding. Upon congruent trials, integration occurred, creating a fluent perception of the target word. Both auditory and visual retrieval accessed the same memory trace. In the incongruent condition, detection of conflict between AV information may have led to integration failure. Focus on resolving the incoming information increased the processing of these two modalities. Since auditory information is more transient, more focus was placed on processing the auditory signal, leading to enhanced recognition in incongruent trials.

Experiment 2

The Congruency x Modality interaction effect from experiment 1 suggested congruity in our paradigm affected the integration of multisensory stimuli. We proposed to involve encoding-specificity and test-specificity manipulations to understand how multisensory events are stored after encoding. If integrated events are stored and retrieved as a singular memory trace. We predict there will be no differences in memory retrieval between modalities for words in the congruent condition. Likewise, failure of integration should promote modality specific memory traces for words in the incongruent condition. Retrieval for incongruent words would reflect the modality specific traces of these words. Therefore, we would expect a modality switching at test will decrease the ability to retrieve words. For example, participants may have difficulty recognizing words that appeared in the auditory condition, but tested in the visual condition compared to words both learnt and tested in the auditory condition. Differences between switch and no-switch conditions for incongruent words may help provide evidence for a modality-specific memory based system.

Methods

Experiment 2 was constructed and presented to participants identical to experiment 1b with the exception of the test phase.

Participants. Eighteen volunteers from the McMaster University undergraduate subject pool participated in the study. Each participant received course credit for their contribution. The participants consisted of 15 females and 3 males between the ages of 17 and 20 years old, the mean age being 19.22 years old. Handedness was not accounted for. Participants were provided written informed consent and ignorant to the purpose of the experiment.

Recognition Test Procedure. Participants performed a recognition test between every learning block. Within the test, participants were presented with a white fixation cross for 500ms. Following, a word would either be presented aurally through the headphones or visually on the screen. 'Old' words from the learning phase would either be presented aurally or visually during test regardless of its initial presentation modality during the learning phase. Of the 24 old words in each test, 12 were randomly selected to be presented in the same modality as its initial appearance, and the remaining 12 were presented in the opposite modality. Alongside the presentation of the word, response keys of 'OLD(A)' and 'NEW(L)' would appear. Letters within the brackets were presented to remind participants of which key to press for response. Participants could not proceed to the next trial until a response was made.

Results

Data Transformation. Participant *d'* scores were calculated through Microsoft Office Excel 2007. *d'* is the calculate difference between normalized hit rate and false alarm rates. Hit rates were calculated by totaling all 'old' responses made by the participant when trials were old.

False alarm rates were calculated by tallying all 'old' responses made when words were new. Separate *d'* were calculated for each of the 8 conditions.

Standard error was calculated using the Cousineau-Morrey method, accounting for between subject variance. *d'* scores were subjected to 2x2x2 within-subject ANOVA.

Analysis. We ran a 2x2x2 factorial ANOVA looking at the effects of congruency, modality and testing (Figure 5). A main effect of congruency was observed [$F(1,17) = 7.373$, $p= 0.015$]. Regardless of presentation modality at learning, or presentation at test, incongruent words were recognized more than congruent words. A main effect for modality was also detected [$F(1,17) = 5.806$. $p=0.028$]. Aurally learnt words were better recognized than visually learnt words regardless of congruency and presentation at test. No significant difference of recognition memory was observed when words were presented in the opposite modality from they were learnt.

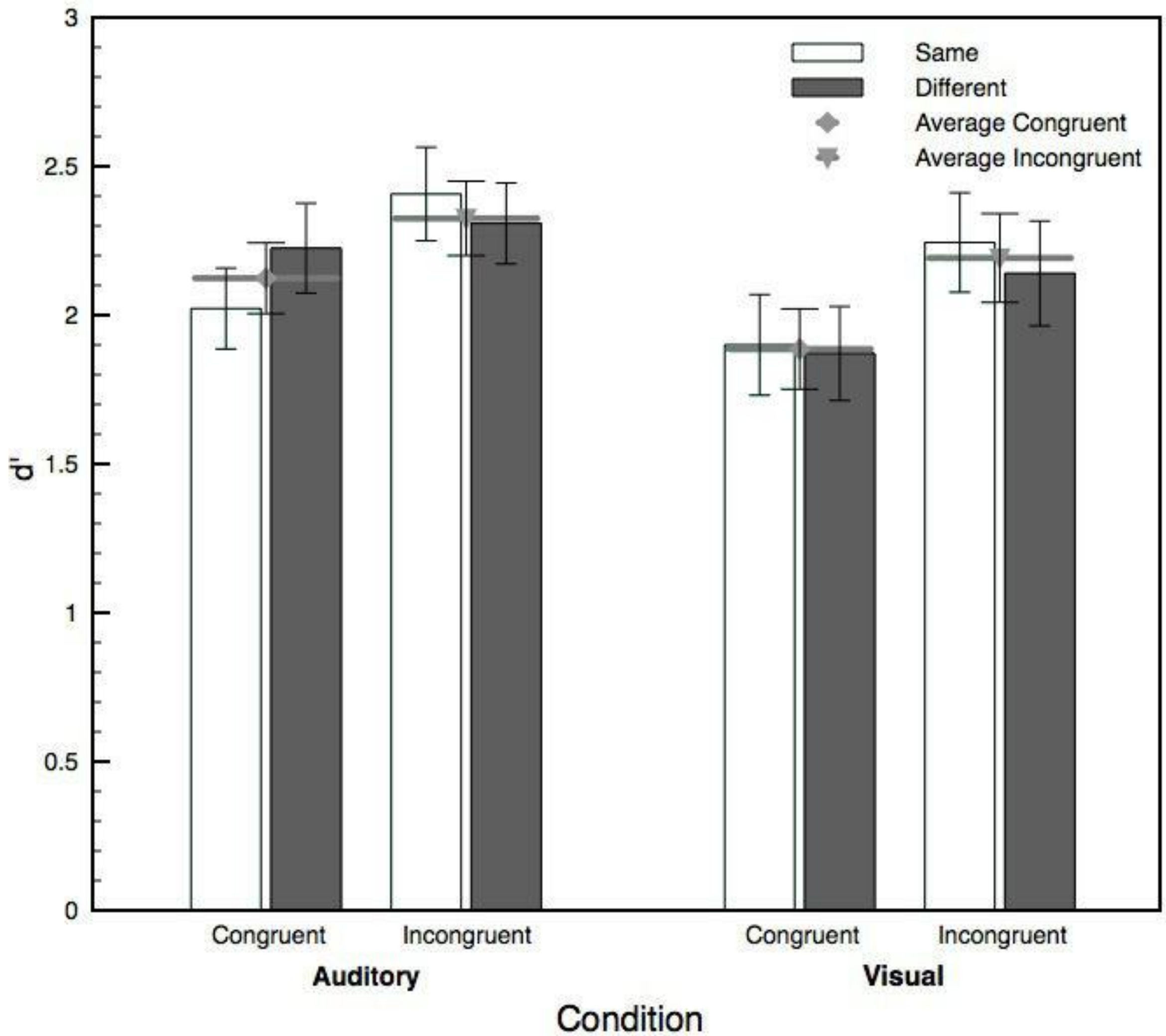

Figure 5. Eighteen participants were present audiovisual words at learning. An 'old-new' recognition test followed the learning phase. Target words at learning were manipulation of 3 independent conditions The above picture depicts the average sensitivity as a result of manipulations between modality (Auditory or Visual), congruity (Congruent or Incongruent) and test-switching (Same or Different) conditions across all 18 participants.. Bars on the graph represent condition of word presentation at the time of test. There were no significant differences between words presented in the same or opposite modality from which they were learnt. There was an overall modality effect of auditory words. Data points indicate the average d' sensitivity for congruency modality collapsed across test-switching. Overall, recognition for incongruent words was superior to congruent words. Error bars here represent standard errors around the mean.

Discussion

Experiment 2 was design to investigate the processing of information during multisensory presentations. Participants were presented the same learning phase as in experiment 1B. In experiment 2, we manipulated the testing phase by disregarding testing specificity effects. Words presented aurally during the learning phase had a 50 percent possibility of appearing in the visual modality during the recognition test and vice versa. It was hypothesized if conflict resolution allowed incongruent words to be processed more, incongruent words would only appear better recognized within their modality specific domain (e.g. aural target words would only be recognized more when tested aurally). Conversely, congruency enhances the singular word in either modality as it activates a common semantic system. Therefore recognition performance should be equivalent in either modality.

Here, we replicated the congruency effect observed in experiment 1; incongruent words were recognized more than congruent words. Additionally, a modality effect was observed suggesting more emphasis was placed in encoding auditory words at time of learning. Interestingly no test specificity effect was observed which rejected our hypothesis.

Appearance of a modality effect suggests participants made a voluntary shift in attentional resources towards words appearing the aural channel. Reliability of the visual channel may be used to disambiguate and reconfirm auditory signals passing through. For example, if participants were presented with the visual word ‘vase’, the auditory signal of the same word could be verified quicker. Likewise, if the visual word ‘’phone’ was presented, but the participant heard the word ‘keys’, this discrepancy could serve to highlight the mismatching the auditory word, leading to auditory dominance. A voluntary shift of attention toward the auditory channel does not necessitate a decline in visual word performance. The presence of an incongruent effect separate of the modality effect suggests that modality independent attentional systems (Ward, 1994) were employed to capture both words, but not necessarily how they were processed.

The lack of a testing effect in relation to the multisensory domain points toward an argument involving semantic congruency, a higher level process. We suggest detection of conflict recruits the semantic activation of both words. As a result, activation of the dual codes due to conflict resolution may have removed the testing effect.

Experiment 3

Increased recognition for incongruent words is counter-intuitive. Redundant information has been noted to decrease the cognitive load needed to encode information (Mayer & Moreno, 2002) allowing extra resources to process information. However, congruency may only provide a sense of fluency upon encoding, leading to a decrease in perceptual processing during encoding. Contrastingly, incongruent stimuli present conflict and promote perceptual processing for resolution (Kreb, Boehler, De Belder & Egner, 2013; Rosner, D'Angelo, MacLellan & Milliken, 2014). Perceptually difficult stimuli have been seen to encourage memory retrieval (Nairne, 1988; Hirshman & Mulligan, 1991). Under the assumption that conflicting words presented in incongruent trials can act as cross-modal masks, we should see the same beneficial effects on memory perceptual difficulty is driven by unimodal masks. For example, if the auditory words successfully mask the processing of visual words, perceptual processing mechanisms may be amodal. Therefore masks appearing in the same modality as the to-be-learnt item should generate the same beneficial effects as incongruent trials.

In experiment 3, participants were separated into two groups; the auditory condition group and the visual condition group. Auditory condition individuals learnt and were tested in the auditory modality only. Similarly, the visual condition participants only encountered visual-only trials in both learning and testing phases.

Methods

Experiment 3a

Participants. Observers from the McMaster undergraduate study pool were invited to participant in the study. An introductory psychology course credit was granted for their involvement. A total of 29 participants (23 females and 6 males) had partaken in the study. Mean age of participants was 18.5. Participants were given a one hour time frame with identical instructions and laboratory conditions. A maximum of three participants conducted the experiment concurrently. All participants completed a profile questionnaire (age, sex, visual acuity), a handedness questionnaire and were given a written informed consent. All observers were oblivious to the purpose of the study.

Apparatus. The experiment was assembled and ran using Mac computers on Mac OS X, version 10.6 through MATLAB version rt2009a. Visual stimuli were presented through a 24" BenQ (GL-2450) monitor at 1080 pi, with a refresh rate of 60Hz.

Collected data was assembled and arranged through Microsoft Excel 2007. IBM SPSS software was used for data analysis.

Design and Procedure. The experimental study ran a 9 by 1 research design. We investigated the differences in observer ability to recognize visual words presented within various levels of visual noise. Words selected to be used in the experiment were identical to those used in experiment 1. As the observers arrived, experimenters instructed them to be seated at any available computer, consent forms were signed, and then the experiment was explained verbally to them. Instructions of which response keys to press were re-presented to them visually between learning and test phases. Observers were informed the experiment was broken down into two blocks.

In the learning phase, participants were told to silently read the words presented. A white fixation cross appeared for 500ms, followed a white colored word with a visual noise mask presented on a black background. Each trial was displayed for 1000ms. Each learning phase consisted of 45 words, 5 words per noise level, randomized in appearance across each learning phase. Following 45 trials, participants were entered into a recognition test. The learning phase and recognition test comprised 1 block, with a total of 2 blocks in the experiment. A total of 180 words were presented, 90 being the learnt words from learning phases and 90 as foils. Recognition tasks and the onset of the next learning block were intermitted by a break. Duration of the break was determined by the participant.

Recognition Test Procedure. A recognition test appeared between each learning phase. Participants were presented with a white fixation cross for 500ms. Following, a word was presented visually on the screen. ‘Old’ words from the learning phase were presented visually without noise mask. In each block, all 45 words presented during the learning phase were presented in the recognition test alongside 45 new distractor words. Presentation order was randomized at the beginning of the experiment. Response keys of ‘OLD(A)’ and ‘NEW(L)’ appeared simultaneously with the tested word. A response was required to advance into the following trial.

Word Presentation. During the experiment, a total 180 words were presented. All words were presented visually in white at 32 point font. Words appearing in the learning phase were presented alongside a visual white noise mask. The mask was calculated from the area of a rectangular boundary around the word upon presentation multiplied by the noise level. Noise level was represented by a calculated percentage [15%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, and 85%] of the number of random pixel appearances the visual mask will have. For example, a visual noise level of 60 percent will involve the area of the boundary multiplied 0.6, to give the amount of pixels that will appear as the mask. Presentation of the target words were randomized in a MATLAB program prior to beginning the learning phase.

Experiment 3b

Procedure of experiment 3b was conducted in an identical manner as experiment 3a with exceptions in stimuli presentation.

Participants. The McMaster undergraduate subject pool was used to recruit participants. An introductory psychology course credit was granted for their involvement. A total of 15 participants (13 females and 2 males) engaged in the study. Mean age of participants was 18.87. Participants completed the study less than one hour with a maximum of three participants performing the experiment simultaneously. Laboratory conditions and experiment instructions were identical for all participants. They were given a profile questionnaire, handedness questionnaire prior to commencement of the study. Participants were unaware of the intention behind the experiment.

Apparatus. Auditory stimuli used were identical to those present in experiment 1b.Auditory stimuli were heard through headphones provided by the participants. Volume levels were adjusted by participants to a comfortable perceived level at the beginning of the experiment. Participants were instructed to not change volume levels once the experiment began.

Design and Procedure. The study involved a 9 by 1 research design. We investigated the effects of auditory noise presentation at encoding on observer ability to recognize auditory words.

In the learning phase participants were told to passively listen to the words presented. A white fixation cross appeared for 500ms, followed by a string of 5 white colored Xs. During presentation of the row of Xs, an auditory word was presented accompanied by an auditory static white noise mask. Each mask and word was presented for 1000ms.

Recognition Test Procedure. Participants were given a recognition test between learning phases. Within the test, participants were presented with a white fixation cross for 500ms. Following, a word was presented aurally through the headphones. 'Old' words from the learning phase were presented aurally without the noise mask. Presentation of the auditory word was escorted with a central white fixation cross and response keys of 'OLD(A)' and 'NEW(L)'.

Word Presentation. During the experiment, a total 180 words were presented. All visual information (e.g. XXXXX, OLD/NEW response keys) was presented in white at 32 point font. Words appearing in the learning phase were presented alongside an auditory static white noise mask. The auditory mask was calculated from the broadband spectrum of the target word multiplied by the noise level. Noise level was represented by a calculated percentage [15%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, and 85%] of the target word's broadband spectrum's amplitude. For example, a visual noise level of 60 percent will involve the amplitude of the mask to be multiplied by 0.6, to determine the loudness of the mask. Target word and noise presentations were randomized in a MATLAB program before commencing the study.

Results

Data Transformation. Observer d' scores were calculated using the identical methods as experiment 1 and experiment 2. The slope and y-intercept of the linear comparison derived from the mean d' for each condition was calculated using R ran with Rstudio. The Pearson correlation value, r, was calculated through R.

Experiment 3a

Analysis. The relationship between noise (pixilation) levels on recognition memory was examined through a 1x9 factorial ANOVA (Figure 6). A significant effect of noise level was found [$F(8,28) = 2.161$, $p = 0.031$]. Pearson correlation coefficient ($r = -0.84$. r^2=-0.7057) demonstrates a decreasing linear trend in regards to the effect of noise.

Experiment 3b

Analysis. The effect of static white noise on recognition memory was examined in a 1x9 factorial ANOVA (Figure 6). A main effect of noise was found [$F(8,14) = 3.273$, $p=0.002$]. As the noise mask increased during learning, recognition for auditory words decreased. Pearson's correlation coefficient ($r = -.9319$, $r^2= 0.8684$) demonstrates a negative correlation between noise presentation and recognition memory.

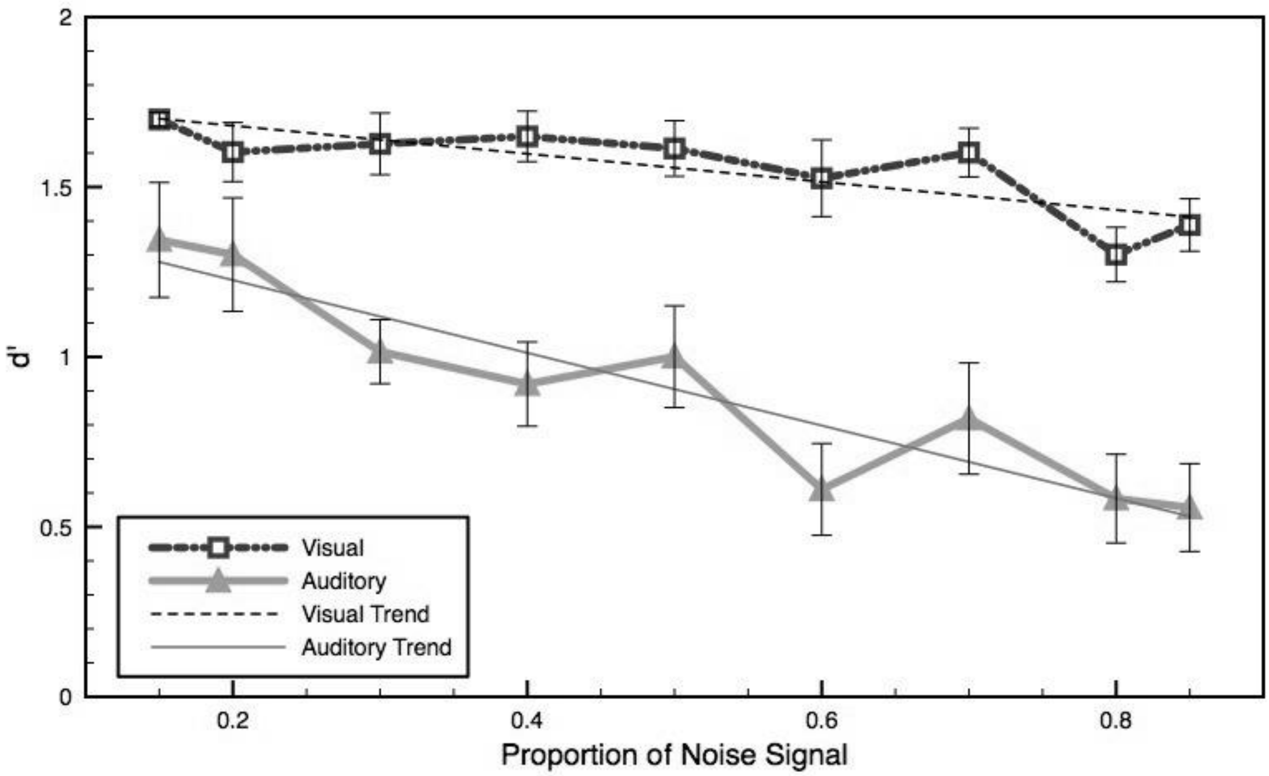

Figure 6. The above figure is a compiled average d' scores for all participants from two separate studies (15 for the auditory condition and 29 for visual). The dashed-line represents the linear trend of the visual mask condition. There was an effect of visual mask on recognition performance. Conversely, the gray solid line represents the linear auditory trend. As static white noise level increases, recognition for both visual and auditory words decreased. Error bars here represent 1 standard error around the mean.

Discussion

Experiment 3 investigated perceptual interference effects that may have occurred in experiment 1 and 2 during incongruent trials. Participants were presented either auditory or visual words. White noise (static white noise or black and white pepper dots) accompanied presentations of these words. Interference was conducted intra-modally. Participants performed an ‘old-new’ recognition test at the end of each learning phase. We predicted perceptual difficulty would improve word recognition in an inverted parabolic function. As difficulty increases, participants would increase processing leading to a strong memory trace for the word. As the separation reaches a peak level of difficulty, recognition will fall as a result of cognitive resource allotment.

Observation from experiment 3A in visual interference demonstrated a noise effect on recognition. Experiment 3B also yielded a significant decline in recognition ability for auditory words presented with increasing static white noise levels. Degradation of memory contradicts results reported by Nairne (1988) and Hirshman and Mulligan (1991). However, this may be due to the presentation interference type. Differences in Pearson’s correlation coefficient suggested the auditory modality may have a more sensitive nature to separating noise and signals.

In experiment 3, attentional resources must be allocated toward the modality-specific channel to resolve and separate noise presentations from the signal. Increasing distribution of resources toward noise-signal separation may result in having depreciated amounts left over for processing. Decreased processing of the signal may have led to the decline in recognition of target words.

General Discussion

The goal of this study was to explore the impact of perception during encoding on explicit memory in a multisensory context. Participants were presented with target words visually and auditorily (either at the end of stem sentences (Experiment 1A) or on their own (Experiment 1B and 2)) and tested with a recognition memory task. Critically, the target words could be the same in both modalities (congruent) or different in the two modalities (incongruent). We had predicted that congruent presentation, which is known to create clearer representations with less variability, would produce better explicit recognition memory. Contrary to our predications, we found that incongruent presentation produced better memory. Experiment 2 examined the role of reinstating the context and found no difference between presentations in the same modality at study and test and different modality presentation. This was also surprising given the robust literature on encoding specificity (Tulving, 1983). Finally, perceptual difficulty was manipulated by adding same-modality noise to single modality study and test. Contrary to the hypothesis that perceptual difficulty would enhance recognition memory, we observed a monotonic decrease in performance with increasing noise.

Although it is difficult to match salience across modalities, it is interesting to note that in Experiments 1 and 2, recognition memory was superior for auditory presentation at the time of study whereas in Experiment 3, where there was no intersensory conflict, recognition memory was superior for the visual modality. This implies that when attention can be directed singly to one modality, visual memory is superior, but when attention must be divided between modalities, auditory memory is superior.

In terms of our original hypothesis, it seems that stronger and cleaner representations do not necessarily produce more durable memories. This is in conflict with the findings of Murray et al. (2004) who found better implicit memory for integrated percepts. In that study, participants had to identify repeated images; performance was enhanced when a congruent sound accompanied the first presentation. This counterintuitive finding could be considered related to recent findings that stimuli requiring selective attention produce more durable memory traces (Rosner, D'Angelo, MacLellan & Milliken, 2014; Krebs, Bochler, DeBelder & Egner, 2013). In those studies, incongruent stimuli, at the time of study, produced better memory. The authors proposed that the need for selective attention to resolve the conflict resulted in deeper processing and a stronger memory trace. This line of thinking could account for the present findings if incongruent multisensory stimuli behave like conflicting unisensory stimuli.

Conflicting Information as a Method of Perceptual Difficulty

Intuitively, fluent perception of a stimulus decreases the effort needed during encoding leading to easier formations of episodic memories. However, conflict monitoring and resolution have been demonstrated to improve recognition of words in incidental memory tasks (Rosner et al., 2014; Krebs et al., 2013). As difficulty of the initial processing increases, more elaborative and extensive processing is needed to encode the stimulus (Nairne, 1988; Hirshman & Mulligan, 1991). Thus, the enhanced recognition memory following incongruent trials was due to their difficulty during initial processing. Recognition for tasks using the same level of semantic processing may have been influenced by cognitive effort (Jacoby, Craik & Begg, 1979). Increasing processing formulates stronger and more distinctive a memory traces. In congruent trials, redundant information increases the ease of processing. Sense of ease may have led to reduced effort on processing (Besken & Mulligan, 2013). Comparably, incongruent trials were harder to process. Due to conflict detection, incongruent words had a stronger memory trace compared to words that appeared in congruent trials. Extra processes were recruited to resolve the AV conflict and segregate incoming information (Simon & Berbaum, 1989). Increased focus at encoding highlighted the memory traces the words formed. Distinctiveness helped in recollection of the word (Dobbins et al., 1998; Jacoby, Craik & Begg, 1979; Hirshman & Mulligan, 1991), thereby enhancing recognition for it.

If provision of conflicting cross-modal information increased perceptual difficulty, this would suggest that resources involved in perceptual processing is amodal. If this is the case, we should observe identical benefits in unimodal settings. Comparisons between the effects of conflict in experiment 1 & 2 and unimodal masking in experiment 3 demonstrates that this may not be the full story. Experiment 3 demonstrated increasing noise levels during encoding impeded recognition. As the perceptual mask became more difficult, participants must work harder to separate the signal (i.e. the word) from the noise. Extensive effort used to separate the two may hinder encoding of the stimulus as result of exhausted resources available. Results of experiment 3 suggests that decline in recognition may be due to perceptual processing at time of encoding. Comparatively, cross-modal conflict received a benefit in recognition, suggesting the benefits (Nairne, 1988, Hirshman & Mulligan, 1991; Mulligan, 1996) and detriments of perceptual interference may be modality-specific. Comparisons of results between experiments 1, 2 and 3 suggests the benefits of multisensory conflict are accredited to higher-order, post-perceptual processes outside the presentation of perceptual difficulty.

The results of Experiment 2 highlight the amodal nature of the representations here. Reinstating the study modality at test did not improve memory performance. While there is evidence for modality-specific memory systems (Barsalou et al, 2003), the present findings highlight the possibility of an amodal memory store. This could also explain the contrasting results from Experiment 3 where difficulty within modality did not produce superior memory. Perhaps when the difficulty is confined to a single modality there is no need for conflict resolution, and no need for the application of selective attention, which appears to be the key element to produce superior memory in the current experiments.

The reversal of the auditory-superiority found in Experiments 1 and 2 in Experiment 3 may also shed light on this hypothesis. Recall that when there was multisensory conflict, the auditory presentation produced better memory overall, but that in the unisensory case, the visual modality was superior. Given the automatic attention grabbing nature of auditory stimuli and the need to attend voluntarily to visual stimuli (Posner, Nissen & Klein, 1976), it may be that when conflict arises, the auditory modality gains prior entry to an amodal representation whereas the visual modality lags, and produces a poorer memory trace. This is not the case with unimodal presentation, where chronic endogenous attention to vision produces superior memory performance. These speculations must be confirmed in future research to establish both the superiority of one modality presentation over another, and the amodal quality of memory when there is conflict.

Cross Modal Conflict as a Desirable Difficulty

Extension of our results may sit in the pedagogical literature. Current shifts in education have been aimed at integrating multimedia into the classroom. Informational slideshows have been used to support verbal instructor lessons. Multimedia formats have been investigated for their benefits in the delivery and retention (Sana, Fenesi & Kim, 2011) of information. It is expected that complementary audio and visual information will allow learners to obtain increased amount of information compared to a unimodal counterpart. Complementary or redundant information have been shown to improve recall (Lewandoski & Kobus, 1993; Treisman & Davies, 1973). Redundant signals increase discriminability and detection by reducing perceptual noise at time of encoding (Kinchla, 1974). Reduction of noise improves the resolution of the incoming stimuli. Processing of information increases as a result of reduction in efforts to separate signal from noise.

Alternatively, it has been suggested that fluency at encoding may be detrimental to learning. Learners attribute fluency to easily learnt items. As a result, effort to encode this information decreases due to subjective belief (Besken & Mulligan, 2013). To rectify this, desirable difficulty is a strategy use to encourage comprehension and learning through manipulation of encoding and retrieval processes (Bjork, 1994). Increasing difficulty by adding variability or contextual interference (Bjork, 1994; Battig, 1979) distort the sense of ease at encoding to promote effortful processing. Here, we argue that presentation of conflicting multisensory information may be another strategy to consider.

Multisensory information reduces the cognitive load needed to process incoming sensory information (Mayer & Moreno, 1998; Moreno & Mayer, 2002). By having redundant information processed in separate channels, information missed in the auditory cue may be substituted through processing of the redundant visual cue. When congruent information occurs, the necessity to capture complete information from either sensory cue is diminished as redundant information between sensory cues will compensate for any missed information. Upon incongruent presentations, auditory and visual information cannot support one another. Incompatibility between AV signals upsets processing fluency and participant’s sense of ease. Increase perceived difficulty in incongruent trials allows learners to reevaluate and analyze the conflict. Prevention of the illusion of comprehension (Bjork, 1994) permits learners to better retain incoming information.

Conclusion and Future Directions

Several limitations may have been present during the study. First, participants were asked to adjust the volume to a level they found comfortable. If volume levels were particularly high, auditory words may have been more salient. Loud auditory sentences would have taken attention away from the visual sentences, affecting recognition rates between the two modalities. However the lack of an auditory superiority effect in congruent trials refutes this claim. Therefore it is unlikely participants knew to increase the volume to hear auditory sentences better before the experiment began. Secondly, the duration of each trial was identical (3000 ms for Experiment 1A, 1000 ms for Experiment 1B). Since auditory words differed in signal length, there may have been excess time for participants to process each sensory signal separately. Another interpretation is that extra time between trials would have allowed for additional post-perceptual processing. A step further, it could be argued that the presentation onset between auditory and visual cues breached the temporal simultaneity needed for integration. However, the temporal binding window between audiovisual speech integration is more lenient at 300 ms where, asynchrony is emphasized when audio information precedes visual (Dixon & Spitz, 1980). The presence of the interaction effect seems to indicate that even with extra processing time this did not change the recognition ability between modalities in the congruent condition. Likewise, the lack of a modality difference between congruent trials seems to suggest that we did not violate the temporal binding of audiovisual information.

Results from these experiments suggest a situation of top-down processing of multisensory information. Multisensory conflict resolution may elicit increase processing as a quest to obtain multiple simultaneous distinct percepts. Congruent multisensory presentations promote fluency and the unification of a single percept via learnt associations. Taken apart, retrieval of multisensory memories may be stronger than unisensory ones (Lehmann & Murray, 2005; Murray et al., 2004), while perception during encoding of the multisensory event aims to sharpen the contextual information provided by redundant signals. Taken together, perception of multisensory events affects both retrieval factors of recognition memory. Specifically, congruent audiovisual presentation supports post-perceptual integration and promotes a sense of perceptual fluency. Conversely, incongruent presentations may recruit additional resources to resolve the conflict leading to a stronger memory trace aiding recollection (Boldini, Russo & Avons, 2004).

In our experiments, observers were informed presentations of the AV stimuli will be followed subsequently with a recognition test. Intentional recognition tests or instructions on an upcoming recognition test may allow participants to increase processing of incoming stimuli. For example, if participants were asked to make a response when AV conflict occurs but unaware of an upcoming recognition task, participants would have primarily focused on detecting and identifying conflicts without consciously resolving the conflict. Lack of resolution may result in decreased recognition accuracy. Likewise, when participants are informed of an upcoming recognition task, expectation may focus on detectable conflict and resolution to encode both auditory and visual stimuli. If so, presentation of an incidental recognition task may investigate the effect of familiarity of a stimulus over the actual recollection of it. Using the relationship between familiarity and perceptual fluency, we can examine the perceptual effects of congruent and incongruent multisensory perception and its relation to explicit memory.

Fluent perception reduces the need to extend further effort to capture all sensory information whereas conflicted perception recruits necessary effort for resolution. Increased processing may have enhanced the memory trace of the incongruent sensory events to compensate processing difficulty during encoding. Comparisons between incidental and intentional memory tasks in relation to multisensory encoding may help uncover avenues into understanding why multimodal presentations improve memory retrieval. Furthermore, current pedagogical tools rely on multidimensional presentations to deliver knowledge and information. By understanding how multisensory information is encoded and stored into memory, we may be able to maximize retention of information within students.

References

Alais, D., Newell, F. N., & Mamassian, P. (2010). Multisensory processing in review: From physiology to behavior. Seeing and Perceiving, 23, 3-38

Barsalou, L. W., Kyle Simmons, W., Barbey, A. K., & Wilson, C. D. (2003). Grounding conceptual knowledge in modality-specific systems. Trends in Cognitive Sciences, 7(2), 84–91.

Battig, W. F. (1979). The flexibility of human memory. Levels of Processing and Human Memory, Lawrence Erlbaum Associates, Hillsdale, NJ, 23–44.

Besken, M., & Mulligan, N. W. (2013). Easily perceived, easily remembered? Perceptual interference produces a double dissociation between metamemory and memory performance. Memory & Cognition, 41(6), 897–903.

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 56-64). New York: Worth Publishers

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. P. Shimamura (Eds.), Metacognition: Knowing about knowing (pp. 185-205). Cambridge, MA: MIT Press

Boldini, A., Russo, R., & Avons, S. E. (2004). One process is not enough! A speed-accuracy tradeoff study of recognition memory. Psychonomic Bulletin & Review, 11(2), 353–361.

Brunel, L., Labeye, E., Lesourd, M., & Versace, R. (2009). The sensory nature of episodic memory: sensory priming effects due to memory trace activation. Journal of Experimental Psychology: Learning, Memory, and Cognition, 35(4), 1081.

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. Journal of Verbal Learning and Verbal Behavior, 11(6), 671–684.

Damasio, A. R. (1989). Time-locked multiregional retroactivation: A systems-level proposal for the neural substrates of recall and recognition. Cognition, 33(1), 25–62.

Dixon, N. F., & Spitz, L. (1980). The detection of auditory visual desynchrony. Perception, 9(6), 719–721.

Dobbins, I. G., Kroll, N. E., Yonelinas, A. P., & Liu, Q. (1998). Distinctiveness in recognition and free recall: The role of recollection in the rejection of the familiar. Journal of Memory and Language, 38(4), 381–400.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. Nature, 415(6870), 429–433.

Ernst, M. O., & Bülthoff, H. H. (2004). Merging the senses into a robust percept. Trends in Cognitive Sciences, 8(4), 162–169.

Hirshman, E., & Mulligan, N. (1991). Perceptual interference improves explicit memory but does not enhance data-driven processing. Journal of Experimental Psychology: Learning, Memory, and Cognition, 17(3), 507.

Jackson, C. V. (1953). Visual factors in auditory localization. Quarterly Journal of Experimental Psychology, 5(2), 52–65.

Jacoby, L. L. (1983). Remembering the data: Analyzing interactive processes in reading. Journal of Verbal Learning and Verbal Behavior, 22(5), 485–508.

Jacoby, L. L., Craik, F. I., & Begg, I. (1979). Effects of decision difficulty on recognition and recall. Journal of Verbal Learning and Verbal Behavior, 18(5), 585–600.

Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. Journal of Experimental Psychology: General, 110(3), 306.

Kinchla, R. A. (1974). Detecting target elements in multielement arrays: A confusability model. Perception & Psychophysics, 15(1), 149–158.

Koppen, C., Alsius, A., & Spence, C. (2008). Semantic congruency and the Colavita visual dominance effect. Experimental Brain Research, 184(4), 533–546.

Krebs, R. M., Boehler, C. N., De Belder, M., & Egner, T. (2013). Neural Conflict–Control Mechanisms Improve Memory for Target Stimuli. *Cerebral Cortex.*

Larsen, A., McIlhagga, W., Baert, J., & Bundesen, C. (2003). Seeing or hearing? Perceptual independence, modality confusions, and crossmodal congruity effects with focused and divided attention. Perception & Psychophysics, 65(4), 568–574.

Laurienti, P. J., Kraft, R. A., Maldjian, J. A., Burdette, J. H., & Wallace, M. T. (2004). Semantic congruence is a critical factor in multisensory behavioral performance. Experimental Brain Research, 158(4), 405–414.

Lehmann, S., & Murray, M. M. (2005). The role of multisensory memories in unisensory object discrimination. Cognitive Brain Research, 24(2), 326–334.

Lewandowski, L. J., & Kobus, D. A. (1993). The effects of redundancy in bimodal word processing. Human Performance, 6(3), 229–239.

Massaro, D. W., Cohen, M. M., & Smeele, P. M. (1996). Perception of asynchronous and conflicting visual and auditory speech. The Journal of the Acoustical Society of America, 100(3), 1777–1786.

Morein-Zamir, S., Soto-Faraco, S., & Kingstone, A. (2003). Auditory capture of vision: examining temporal ventriloquism. Cognitive Brain Research, 17(1), 154–163.

Moreno, R., & Mayer, R. E. (2002). Verbal redundancy in multimedia learning: When reading helps listening. Journal of Educational Psychology, 94(1), 156.

Mulligan, N. W. (1999). The effects of perceptual interference at encoding on organization and order: Investigating the roles of item-specific and relational information. Journal of Experimental Psychology: Learning, Memory, and Cognition, 25(1), 54.

Murray, M. M., Michel, C. M., Grave de Peralta, R., Ortigue, S., Brunet, D., Gonzalez Andino, S., & Schnider, A. (2004). Rapid discrimination of visual and multisensory memories revealed by electrical neuroimaging. Neuroimage, 21(1), 125–135.

Nairne, J. S. (1988). The mnemonic value of perceptual identification. Journal of Experimental Psychology: Learning, Memory, and Cognition, 14(2), 248.

Posner, M. I., Nissen, M. J., & Klein, R. M. (1976). Visual dominance: an information-processing account of its origins and significance. Psychological Review, 83(2), 157.

Radeau, M., & Colin, C. (2001). Object identity is not a condition but a result of intersensory integration: the case of audiovisual interactions. Current Psychology of Cognition, 20(5), 349–358.

Roach, N. W., Heron, J., & McGraw, P. V. (2006). Resolving multisensory conflict: a strategy for balancing the costs and benefits of audio-visual integration. Proceedings of the Royal Society B: Biological Sciences, 273(1598), 2159–2168.

Rosner, T. M., D’Angelo, M. C., MacLellan, E., & Milliken, B. (2014). Selective attention and recognition: effects of congruency on episodic learning. Psychological Research, 1–14.

Sana, F., Fenesi, B., & Kim, J. A. (2011). A Case Study of the Introductory Psychology Blended Learning Model at McMaster University. The Canadian Journal for the Scholarship of Teaching and Learning, 2(1), 6.

Simon, J. R., & Berbaum, K. (1990). Effect of conflicting cues on information processing: the "Stroop effect"vs. the "Simon effect." Acta Psychologica, 73(2), 159–170.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. Psychological Review, 80(5), 352.

Vallet, G., Brunel, L., & Versace, R. (2010). The perceptual nature of the cross-modal priming effect: arguments in favor of a sensory-based conception of memory. Experimental Psychology, 57(5), 376.

Vallet, G., Riou, B., Versace, R., & Simard, M. (2011). The sensory-dependent nature of audio-visual interactions for semantic knowledge. In Proceedings of the 33rd annual conference of the cognitive science society (pp. 2077–2082).

Vallet, G. T., Simard, M., Versace, R., & Mazza, S. (2013). The perceptual nature of audiovisual interactions for semantic knowledge in young and elderly adults. Acta Psychologica, 143(3), 253–260.

Van Wassenhove, V., Grant, K. W., & Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. Neuropsychologia, 45(3), 598–607.

Vatakis, A., & Spence, C. (2007). Crossmodal binding: Evaluating the "unity assumption" using audiovisual speech stimuli. Perception & Psychophysics, 69(5), 744–756.

Ward, L. M. (1994). Supramodal and modality-specific mechanisms for stimulus-driven shifts of auditory and visual attention. Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale, 48(2), 242.

Welch, R. B., DutionHurt, L. D., & Warren, D. H. (1986). Contributions of audition and vision to temporal rate perception. Perception & Psychophysics, 39(4), 294–300.

Westerman, D. L., Lloyd, M. E., & Miller, J. K. (2002). The attribution of perceptual fluency in recognition memory: The role of expectation. Journal of Memory and Language, 47(4), 607–617.

Whittlesea, B. W., Jacoby, L. L., & Girard, K. (1990). Illusions of immediate memory: Evidence of an attributional basis for feelings of familiarity and perceptual quality. Journal of Memory and Language, 29(6), 716–732.

Yonelinas, A. P. (2001). Components of episodic memory: the contribution of recollection and familiarity. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 356(1413), 1363–1374.