THE DYNAMICS OF AUDIO-VISUAL INTEGRATION CAPACITY AS A FUNCTION OF ENVIRONMENTAL DIFFICULTY, STIMULUS FACTORS, AND EXPERIENCE

By

Jonathan Michael Paul Wilbiks

Master of Arts, Ryerson University, 2012

Master of Arts, University of Sheffield, 2009

Bachelor of Science, University of Toronto, 2008

A dissertation

presented to Ryerson University

in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

in the Program of

Psychology

Toronto, Ontario, Canada, 2016

AUTHOR'S DECLARATION FOR ELECTRONIC SUBMISSION OF A DISSERTATION

I hereby declare that I am the sole author of this dissertation. This is a true copy of the dissertation, including any required final revisions, as accepted by my examiners.

I authorize Ryerson University to lend this dissertation to other institutions or individuals for the purpose of scholarly research.

I further authorize Ryerson University to reproduce this dissertation by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

I understand that my dissertation may be made electronically available to the public.

THE DYNAMICS OF AUDIOVISUAL INTEGRATION CAPACITY AS A FUNCTION OF ENVIRONMENTAL DIFFICULTY, STIMULUS FACTORS, AND EXPERIENCE

Doctor of Philosophy, 2016

Jonathan Michael Paul Wilbiks

Psychology

Ryerson University

Abstract

The capacities of unimodal processes such as visual and auditory working memory, multiple object tracking, and attention have been heavily researched in the psychological science literature. In recent years there has been an increase in the amount of research into multimodal processes such as the integration of auditory and visual stimuli, but to my knowledge, there has only been a single published article to date investigating the capacity of audiovisual integration, which found that the capacity of audiovisual integration is limited to a single item.

The purpose of this dissertation is to elucidate some of the factors that contribute to the capacity of audiovisual integration, and to illustrate that the interaction of these respective factors makes the capacity a fluid, dynamic property. Chapter 1 reviews the literature coming from multimodal integration research, as well as from unimodal topics that are pertinent to the factors that are being manipulated in the dissertation: namely, working memory, multiple object tracking, and attention. Chapter 2 considers the paradigmatic structure employed by the single study on audiovisual integration capacity and breaks down the component factors of proactive interference and temporal predictability, which contribute to the environmental complexity of the scenario, in the first illustration of the flexibility of capacity of audiovisual integration. Chapter

3 explores the effects of stimulus factors, considering the effects of crossmodal congruency and perceptual chunking on audiovisual integration capacity. Chapter 4 explores the variability of audiovisual integration capacity within an individual over time by means of a training study. Chapter 5 summarizes the findings of the research within, discusses some overarching themes with regard to audiovisual integration capacity including how information is processed through integration and how these findings could be applied to real-life scenarios, suggests some avenues for future research such as further manipulations of modality and SOA, and draws conclusions and answers to the research questions.

This research extends what is known about audiovisual integration capacity, both in terms of its numerical value and the factors that play a role in its establishment. It also demonstrates that there is no overarching limitation on the capacity of audiovisual integration, as the initial paper on this topic suggests, but rather that it is a process subject to multiple factors, and can be changed depending on the situation in which integration is occurring.

Acknowledgements

First and foremost, I thank my supervisor Dr. Ben Dyson for his support and advice throughout my time at Ryerson, even after he found greener pastures across the world. Thank you for helping me integrate my thoughts and to keep looking at the forest amongst the trees. Along with Ben, thank you to Dr. Frank Russo for your valuable feedback and advice on preparing the dissertation for defense. Thank you to Dr. Raj Sandhu for putting up with many beeps and boops in helping me pilot test my paradigm, and for providing much wise counsel, both formally and casually. The experiments described within the dissertation would not have been possible if not for Carson Pun who went above and beyond in helping me with designing the stimuli and the experimental program that was employed.

I acknowledge the contributions of the examination committee, Dr. Maureen Reed, Dr. Margaret Moulson, Dr. Stéphanie Walsh Matthews, and Dr. Laurence Harris. Thank you for your availability and effort in serving on my committee, and for your critical contributions to the dissertation.

Finally, thank you to my family and friends for their love, support, and understanding during this lengthy process. To my parents and sister, thank you for not asking *too* often how much longer I will be a student. To my friends, thank you for providing a chance to not always worry about my dissertation. To my daughter, Amelia, thank you for always having a hug and a kiss for me, and for reminding me why I’m doing this. And to my wife, Hannah, thank you for all your love, support, and for pulling up all the slack I’ve left while sitting in the office endlessly.

Dedication

For the love of my life, my constant partner, and my best friend. The one who sacrifices so much of herself so that I can pursue this dream of mine. For putting up with all the lonely days, and all my late nights. For loving me through all of this, and for going through it all together.

Table of Contents

Author's Declaration ii

Abstract iii

Acknowledgements v

Dedication vi

Table of Contents vii

List of Tables xi

List of Figures xii

List of Appendices xiv

Introduction **1**

Chapter 1 – General Introduction **3**

- Audiovisual integration 3
- Visual working memory 9
  - Attention and visual working memory 12
- Multiple object tracking 13
  - Attention and multiple object tracking 17
- Capacity of audiovisual integration 19
- Research Questions and Hypotheses 21

Chapter 2 – Exploring the Van der Burg et al. (2013) Paradigm **25**

- General Introduction 25

Experiment 1 29

Introduction 29

Method 33

Results 38

Discussion 45

Experiment 2 47

Introduction 47

Method 48

Results and Discussion 49

Experiment 3 51

Introduction 51

Method 53

Results and Discussion 54

Experiment 4 55

Introduction 55

Method 57

Results and Discussion 58

Between-experiments comparisons 60

Cross-experimental analysis of proportion correct data 64

General Discussion 68

Chapter 3 – Stimulus-based Effects on Audiovisual Integration 73

General Introduction 73

Experiment 5 74

Introduction 74

Method 78

Results 79

Discussion 82

Experiment 6 84

Introduction 84

Method 86

Results 88

Discussion 91

General Discussion 93

Chapter 4 – Individual flexibility in audiovisual integration capacity 97

Experiment 7 97

Introduction 97

Method 104

Results 105

Discussion 108

Chapter 5 – General Discussion 113

Summary of findings 113

Data and resource limits 116

Capacity for simple vs. complex objects 120

Is there a maximum capacity for audiovisual integration? 122

Future directions 123

Multiple SOAs 124

Inverted modalities 126

Entrainment of neuronal oscillators 128

Application to alert system scenarios 130

General conclusions 131

Appendices **133**

References **143**

List of Tables

Table 1. Results of ANOVA on response accuracy data for Experiment 1. 39

Table 2. Estimates of audiovisual integration capacity (*K*) for each participant and experiment means for Experiments 1 – 4. 61

Table 3. Results of ANOVA on response accuracy data for Experiments 1 – 4. 66

Table 4. Results of ANOVA on response accuracy data for Experiment 5. 80

Table 5. Results of ANOVA on response accuracy data for Experiment 6. 88

Table 6. Results of ANOVA on response accuracy data for Experiment 7. 111

List of Figures

Figure 1. Factorial 2 x 2 design for Experiment 1 – 4. 29

Figure 2. Trial schematic for each of Experiments 1 – 4. 35

Figure 3. Capacity estimates (K) of audiovisual integration capacity for individual participants and experiment means for each of Experiments 1 – 4. 38

Figure 4. Response accuracy results for each SOA and number of locations to be tracked for Experiment 1. 40

Figure 5. Group average ERP response across the three stages of audiovisual integration as a function of SOA and to be tracked locations in Experiment 1. 42

Figure 6. Mean amplitudes for three stages of audiovisual integration as a function of SOA and number of locations to be tracked in Experiment 1. 43

Figure 7. Correlations between mean amplitude brain data and behavioural responding in Experiment 1. 44

Figure 8. Expected and observed values of interference and predictability in Experiments 1-4. 57

Figure 9. Interaction between capacity estimates (*K*) between Experiments 1 – 4, based on factors temporal predictability and proactive interference. 62

Figure 10. Proportion correct for each combination of interference, predictability, and SOA between Experiments 1 – 4. 65

Figure 11. Capacity estimates (*K*) for congruency effects in Experiment 5. 79

Figure 12. Proportion correct for each SOA, number of locations to be tracked, and congruency effects in Experiment 5. 81

Figure 13. Trial schematic for Experiment 6. 84

Figure 14. Capacity estimates (*K*) for vertex effects in Experiment 6. 89

Figure 15. Proportion correct for each SOA, number of locations to be tracked, and vertex effects in Experiment 6. 91

Figure 16. Capacity estimates (*K*) for Experiments 1 – 6. 98

Figure 17. Capacity estimates (*K*) for training effects in Experiment 7. 106

Figure 18. Proportion correct for each SOA, number of locations to be tracked, and training effects in Experiment 7. 107

Figure 19. Proportion correct within training blocks in Experiment 7. 109

List of Appendices

Appendix A – Consent form for Experiment 1 with EEG 133

Appendix B – Debrief form for Experiment 1 with EEG 136

Appendix C – Representative consent form for Experiments 2 – 7 137

Appendix D – Representative debrief form for Experiment 2 – 7 140

Appendix E – Review of manuscript from Dr. Erik Van der Burg 141

Introduction

Establishing the capacity of perceptual and cognitive systems has been a wide-ranging research interest for many years. Miller (1956) studied the capacity of short term memory, establishing it to be a 'magic number' of 7 plus or minus 2. Since then, many other capacities have been established, and a recent paper by Van der Burg et al. (2013) examined the capacity of audiovisual integration, establishing that there is a strict limitation of one visual item that can be integrated with one auditory stimulus. Based on wide-ranging evidence in unimodal perception, however, it seemed unlikely that there should be such a strict limit on the capacity of audiovisual integration. In order to further investigate the capacity of audiovisual integration, an experimental series was designed that would examine the environmental demands placed on participants by the Van der Burg et al. (2013) paradigm. Beyond a simple breakdown of their paradigm, the current research examines other potential factors known to influence unimodal perception – these include stimulus factors such as crossmodal congruency and perceptual chunking, as well as individual variability by means of training. Through these seven experiments, a case is built for the capacity of audiovisual integration as a dynamic function influenced by all the factors discussed above.

Chapter 1 provides a general introduction and a review of the literature across fields that pertain to elements of the experimental paradigm, and the contributing factors to the capacity of audiovisual integration at large. It proceeds to set up research questions and hypotheses that will guide the inquiry throughout the rest of the dissertation. Chapter 2 presents an initial set of four experiments that deconstruct the factors present in the Van der Burg et al. (2013) paradigm. By orthogonally varying the level of temporal predictability of the target and the level of proactive interference, the conditions under which the capacity of audiovisual integration is maximized is determined. Chapter 3 goes beyond the initial paradigm, considering additional stimulus factors that influence integration. Crossmodal congruency and perceptual chunking are explored in two experiments as potential facilitatory factors in the capacity of audiovisual integration. Chapter 4 presents a single study in which participants were trained to improve their audiovisual integration capacity, revealing that capacity is flexible not only through variation in stimulus factors but also in variability within individuals. Finally, Chapter 5 summarizes the findings from all the experiments, and discusses them in light of some wider issues in terms of audiovisual integration and perception in general.

Chapter 1 – Literature Review

Audiovisual Integration

While navigating our everyday lives, we are constantly stimulated by various sensory inputs in several different modalities. These sensory inputs are processed by separate sense organs, and they are subsequently either integrated (forming a single multisensory percept), or not (remaining multiple and separate sensory stimuli). During this process, we are also making frequent implicit decisions as to the potential unity of the sources of these inputs. For example, if a visual stimulus and an auditory stimulus reach your sensory organs at around the same time, we have to decide whether they came from the same source (one event creating both a sound and a visual stimulus) or from two separate sources.

The concept of multisensory integration is one that has been studied heavily since the mid-20th century. In one seminal empirical paper, Sumby and Pollack (1954) found that auditory perception of speech in a noisy environment is improved when one is able to see the lips of the person that is speaking, and this was followed by a long line of research into the effects of one sensory modality on others. Welch and Warren (1980) present multisensory integration as an implicit decision-making process, wherein an individual must decide whether two sensory inputs they receive are caused by the same event or multiple different events. Whether integration occurs or not is based on a number of factors, each of which will be discussed in this initial section, and it is important to firstly highlight what these factors are so that we can understand how to increase (or decrease) the likelihood of integration. For example, it was found that an individual 'driving' in a car simulator braked more quickly when exposed to a multisensory alert signal than when exposed to a unisensory signal (Ho, Reed, & Spence, 2007). For applications such as this one, it is important to also understand the elements that make up multisensory integration so that it can be used to increase our safety.

The majority of research into factors that influence integration has been into *audio-visual* integration rather than other modality combinations, although there has been interest in other sensory combinations as well. Spence (2011) reviews a number of these interactions, including those between vision and touch (Martino & Marks, 2000), audition and touch (Walker & Smith, 1985), taste and sound (Crisinel & Spence, 2009), vision and smell (Spence, 2010), vision and taste (Spence & Gallace, 2011). While these other combinations of modality interactions have been demonstrated experimentally, the focus of the literature review is on the processes surrounding audiovisual interactions and integration. The factors that influence integration can be generally classed into temporal, spatial, and crossmodal congruency factors, and these factors interact in making binding decisions. For the purposes of this dissertation, the focus will be on temporal and congruency factors, although for a discussion of spatial factors influencing audiovisual integration see Spence & Santangelo (2009)

In terms of temporal factors, the general finding is that there is a range of timing within which an auditory and a visual stimulus can be bound, referred to as the *temporal window of integration (TWI)*. At its most basic, the TWI extends from around 30 ms auditory lead to around 170 ms visual lead in sensation (Van Wassenhove, Grant, & Poeppel, 2007). This asymmetry is likely an artifact of the relative speeds of conduction of light and sound in the atmosphere, and in our bodies. That is to say, since light travels more quickly than sound in the air, a visual signal will reach us before an auditory signal, if both are emanating from the same event. For example, lightning and thunder are both *caused* by the same electric discharge in a stormcloud, but the lightning is usually seen before the thunder is heard. The preference for audio to lag behind vision allows binding to occur more accurately – actually binding two sensory inputs that originated together, despite their different times of arrival. Another example can be seen while playing the role of spectator at a baseball game. If one is in a seating location that is some distance from the batter, one will see the visual effects of the impact of bat on ball (the ball changing directions) tangibly earlier than one hears the auditory effect (a "crack" sound). In this case, the temporal delay between visual and auditory information can be overcome by the congruency between the two stimuli; if the sound was of a bubble popping instead of a crack, it would be much more difficult to presume the unity of those inputs. While specific estimates of this temporal window vary (Zampini, Shore & Spence, 2003; Spence & Squire, 2003; van Wassenhove, Grant, & Poeppel, 2007; Soto-Faraco & Alsius, 2009), most of them find that audiovisual binding between two stimuli is optimized when the visual stimulus occurs around 85-100 ms ahead of an auditory stimulus. Moreover, the window of integration has been shown to be flexible both between (Fujisaki, Shimojo, Kashino, & Nishida, 2004; Heron, Whitaker, McGraw, & Horoshenkov, 2007) and within individuals (Stone, Hunkin, Porrill, Wood, Keeler, Beanland, Port, & Porter, 2001).

Congruency factors have been shown to influence likelihood of binding as well, with auditory and visual stimuli that are related to each other in some way shown to be more likely to be bound than others (Spence, 2011). Spence puts forth three general types of crossmodal correspondences: structural, statistical, and semantically mediated correspondences. Structural correspondences are those which occur due to "intrinsic attributes of the perceptual system's organization" (Spence, 2011, p. 988). That is to say, if certain unimodal stimulus traits are processed in proximal areas in the brain, there is likely to be a correspondence between those traits (Ramachandran & Hubbard, 2001). Walsh's (2003) ATOM (A Theory of Magnitude) theory, proposes that there is a common coding for sensory phenomena that can be measured by magnitude. In this way, auditory loudness and visual brightness (for example) may share a common coding. Statistical correspondences are based on regularities in the environment, and our subsequent exposure to these regularities leading to an increased correspondence between two stimuli. The example Spence provides here is that, since resonance properties of objects require that a small object generate a high-pitched sound, there is a crossmodal correspondence between high pitch and small size (and low pitch with large size; and see also Marks, 1987; Evans & Treisman, 2010). Finally, semantically mediated correspondences relate to the use of common language to describe different sensory inputs. These are often idiosyncratic to certain cultures as they are based on language, but there are some universal findings to report. Spence presents the findings of Stumpf (1883), explaining that almost every language uses words similar to "high" and "low" to describe pitch of high and low frequency, respectively. This is presented as the cause of the crossmodal correspondence between pitch and height, which has been shown in more recent research (Rusconi, Kwan, Giordano, Umilta, & Butterworth, 2006; Leboe & Mondor, 2007). Walker (2012) disagrees with Spence's (2011) breakdown of the different types of crossmodal correspondence, providing evidence that all correspondences can be broken down as having semantic roots. For the purposes of this dissertation, however, what is of utmost importance is that crossmodal congruency can have an effect on audiovisual integration.

In our lab, it was found that when these temporal and congruency factors are put into a competitive binding setting with one another, temporal factors take on a primary role (Wilbiks & Dyson, 2013a). In two separate conditions, participants were presented with two visual stimuli (anchor stimuli) and asked which of them a single auditory stimulus (to-be-bound stimulus) was bound with (VAV; visual-weighted condition), and presented two auditory stimuli and asked which one a single visual stimulus was bound with (AVA; auditory-weighted condition). The stimuli varied on their temporal coincidence, with the to-be-bound stimulus being presented at a time overlapping with the first anchor, overlapping with the second anchor, or at an ambiguous time point that was not coincident with either of the anchors. There was an additional manipulation of congruency, with the visual stimuli being either large or small, and the auditory stimulus being loud or quiet. By orthogonally varying the temporal and stimulus congruency factors, it was possible to examine which factor had a stronger influence on binding. It was found that responding was modulated primarily by temporal factors, with a stimulus that was temporally coincident with another stimulus being highly likely to be bound there. There was also an asymmetry between binding when a visual stimulus is being bound to one of two auditory stimuli and binding in a situation when an auditory stimulus is being bound to one of two visual stimuli. Overall, there was a tendency to for participants to bind in such a way that auditory stimuli follow visual stimuli. Congruency factors (which, in this case, were within rather than between modality) played a role only when temporal information was ambiguous – when the to-be-bound stimulus took place at a time that was not simultaneous with either of the pair of stimuli.

Having summarized the factors at play in audio-visual integration in general, I will proceed to discuss some other elements that play a role in establishing the *capacity* of audiovisual integration. This is the critical question being investigated in this thesis, and is also an important open question for the literature on audiovisual integration at large. While unimodal processing limits have been tested and determined (such as the capacity of visual working memory that will be discussed below), there has only been a single study (Van der Burg et al, 2013) that has investigated the capacity of audiovisual integration. This study involved presentation of rapidly changing visual displays in which a certain subset of dots changes polarity, with a single auditory stimulus presented with which the dots might be integrated. Determining the capacity of integration with experiments such as this, as well as exploring which factors play a role in its determination, is important to the study of audiovisual display design, and creating alert systems that are more effective for their users. This is also an important question for our sensory processing in everyday life. The world in which we exist is a highly complex one, with myriad auditory and visual inputs stimulating our sensory organs simultaneously. The processes through which we attend to some, and block out others, has been researched in the field of selective attention (Moran & Desimone, 1985; Houghton & Tipper, 1994). Similarly, the factors that play a role in integration (or not) have been discussed in this chapter. The question that remains outstanding is how many of these visual and/or auditory objects can be bound to one another.

First, I will consider working memory capacity, as working memory can be seen as a prerequisite to audiovisual integration: in order to integrate successfully, you must first have been able to retain elements in working memory. Second, I will present research from multiple object tracking literature, including how attention plays a role in performing a multiple object tracking task. While in the current research we are not asking participants specifically to track objects, we are asking them to *keep track* of features of said objects, and as such multiple object tracking can contribute evidence to support the current research. Finally I will explore some of the existing literature on the capacity of audiovisual integration, considering paradigmatic issues, and ultimately leading to the research questions and hypotheses for the current research, which will be set out in full at the end of this chapter.

Visual Working Memory Capacity

In considering the capacity of audiovisual integration, it is important to understand the literature regarding the capacity of visual working memory. While these are not identical processes, it follows that the capacity of audiovisual integration should not be greater than that of visual working memory (Van der Burg et al., 2013), as working memory is required for tracking potential candidates for integration. In both behavioural and electrophysiological studies, it has been shown that the capacity of visual working memory is limited to an extent, but that the limit is dynamic, and is modulated by stimulus and individual factors. Alvarez and Cavanagh (2004) studied the capacity of visual working memory and found that rather than being a pure numerical limitation, it is a capacity based on both number of items and complexity of items to be kept in memory. They found that the capacity for simple stimuli (coloured line drawings) was around 4.4 items, while for complex stimuli ("3-D" cubes) capacity was 1.6 items.

This result was challenged by Awh, Barton, and Vogel (2007) who found that the capacity was around 4 items regardless of complexity. They proposed that there are two features of working memory for objects that work together to set capacity: the number of items to represent in memory, and the resolution of those representations. They reason that Alvarez and Cavanagh (2004) contained a confound wherein some of the stimulus types had greater target-distractor similarity than others, which would lead to an underestimation of capacity in these conditions (the complex stimulus conditions). These two opposing theoretical perspectives agree, however, on the fact that the capacity of working memory for *simple* stimuli is around 4 items. This corresponds with the argument put forth by Cowan (2001), who found that there is a capacity limit of between 3 and 5 items in visual working memory. Cowan used a variety of tasks, including searching a visual array, an unattended auditory channel, or overt repetition of words, and found that this limitation is consistent across modalities and tasks. Cowan (2010) theorizes that the reason for the existence of this limit is a combination of a temporal limit of neural firing (with each memory item needing to be maintained approximately every 100 ms) and an issue of interference when multiple memory items are active simultaneously.

The capacity and general function of working memory is also subject to effects of training programs, as discussed in a review by Klingberg (2010). He presents comparative evidence from animals which showed that visual working memory capacity can be improved after many trials (Recanzone et al., 1992). Further, he shows that children that have been diagnosed with hyperactivity (a symptom of which is low working memory capacity) are able to improve working memory span by being taught chunking and other metacognitive strategies (Abikoff & Gittelman, 1985). This research proposes that using working memory training along with medication of hyperactive children provides the best possible behavioural outcomes. Klingberg et al. (2002) also performed a training intervention study with a clinical population of children - in this case, ADHD - and found that 25 sessions over 5 weeks training in working memory tasks was enough to increase working memory capacity. They also showed that core working memory training leads to far reaching transfer effects. That is, when participants were trained on tasks that required domain-general working memory mechanisms to successfully complete them, this led not only to improvement in the trained task, but also to improvement across all working memory measures. So working memory capacity seems to be malleable within the individual, and can be trained to increase its capacity through techniques such as chunking.

In order to further explore the capacity of working memory, and specifically to examine the neural correlates of the capacity of visual working memory, Vogel and Machizawa (2004) used EEG recording while changing the number of visual items to be maintained in working memory. They presented participants with a visual display of between 1 and 10 visual items (coloured squares) in each visual hemifield, and participants were asked to memorise the contents of one of the arrays (either left or right hemifield). Participants were then presented with a second, similar array, and were to respond as to whether the test array was the same or different to the memory array. While presenting these arrays, EEG data were recorded at posterior parietal, lateral occipital, and posterior temporal electrode sites. They analyzed contralateral delay activity by averaging activity at electrodes in the opposite hemisphere to the hemifield in which the memory and test arrays were presented. They found that mean amplitude of contralateral delay served as a predictor of the number of items that were being held in memory, with increasing amplitude when faced with an increased number of items up to a limit of around 4 (although they also found high levels of individual differences with capacity ranging from around 1.5 to 6 depending on the person). Once the number of items exceeded a participant's capacity, their contralateral delay activity remained at the same level as the highest number that they were able to successfully maintain. This study provides valuable evidence that electrophysiological data can be used to index the capacity of visual working memory.

The findings related to visual working memory can be useful to the current research as they provide both a maximum value for the capacity of audiovisual integration, as well as a potential analogous methodology to be used. As discussed earlier, it should not be possible for audiovisual integration capacity to exceed the established capacity of visual working memory. It should also not be possible for the capacity of audiovisual integration to exceed the capacity of auditory working memory (which was found to be between 1 and 2; Saults & Cowan, 2007), but in the current research this is not an important factor. The paradigm being used in this experimental series always has a single auditory stimulus (which is sub-capacity), with multiple visual stimuli being employed, to test the capacity of integration.

Attention and Visual Working Memory

Entry into working memory has been known to be influenced by attention, both early and late in processing (Awh, Vogel, & Oh, 2006). Early in processing, responses to activation of attended stimuli showed amplification of P1 ERP response within the first 100 ms after presentation, as shown by Van Voorhis and Hillyard (1977). On the other hand, later in processing attention takes on the role of a gatekeeper, controlling which 4 items occupy the maximal capacity of working memory. Awh et al. (2006) outline the concept that if the four slots of working memory are occupied, it is not possible for new percepts to gain access to them. In this model, attention serves as a selection mechanism which either maintains the items that are already in working memory, or allows new information to take over a previously occupied slot. Sobel et al. (2007) agree with the dual process view of attention in working memory, wherein top-down and bottom-up mechanisms both play a role in processing, but state that we first use top down control to focus on the task at hand and to avoid other tasks - for example, choosing to focus on shape rather than colour - and that bottom-up mechanisms are only employed when the original top-down processing is not able to successfully allow for entry into working memory. As such, while attention has been shown to have early perceptual effects in terms of working memory, it plays an even more integral role in late selection of items, which can hold a significant influence over working memory capacity and function.

While conceptualising attention as a gatekeeping mechanism for visual working memory provides an apt description, Chun (2011) goes even further, suggesting that visual working memory *is* sustained visual attention over some time period. That is to say, visual working memory only works as one continues to focus attention on certain visual stimuli over some time course, while excluding the other stimuli that are present. He describes the similarities between working memory and attention, specifying that both systems have capacity limitations (Chun, Golomb, & Turk-Browne, 2011), and that both show similar patterns in their processing of simple features, and more complex objects (Alvarez & Cavanagh, 2004; Luck & Vogel, 1997). Kane, Bleckley, Conway, and Engle (2001) also subscribe to this view, in comparing working memory span to attentional control tasks. They found that individuals scoring in the bottom quartile for working memory span made slower and more erroneous saccades in an attentional control task, while those in the top quartile of working memory span were significantly more accurate. This evidence indicates that working memory capacity is a function of controlled attention, which fits suitably well with the ideas put forward by Chun (2011). McCabe et al. (2010) provide additional evidence to this end, as they compared scores on working memory capacity and executive functioning and found a very strong correlation between these scores in their participants ($r^2 = .97$). Given this evidence, his proposal of visual working memory being of the same stuff as attention seems a logical extension of the theory of it working as a gatekeeper to memory - rather than simply being one system affecting information entry in another, it is actually one and the same system.

Multiple Object Tracking

In order to track the state and movement of multiple objects, one must have a trace of them in working memory. Having established that visual working (or short term) memory has a capacity of around 4 items (Cowan, 2001), Cavanagh and Alvarez (2005) provide a comprehensive review of the literature in the field of multiple object tracking, and its capacity limit. One of the important studies that they discuss is the research of Oksama and Hyönä

(2004), who found that when participants were asked to track objects for 5 seconds, the average number of items that were able to be tracked successfully was around 4 items, which was in line with previous findings (Pylyshyn & Storm, 1988). When participants were asked to track items for 9 or 13 seconds, response accuracy decreased significantly at a set size exceeding 3 items. They also found that there was a large amount of individual difference between participants, with a uniform distribution between 2 and 6 items being successfully tracked.

As previously stated, in order to track multiple objects successfully, there first needs to be an available slot within working memory. Once they are held in a working memory slot, these individual objects need to be tracked (or have their features tracked) in order to be eventual candidates for integration. Bettencourt and Somers (2008) looked deeper into the limitations on the capacity of multiple object tracking, which they originally proposed to be around 4 objects. They discuss the likelihood that multiple object tracking can be limited by both a temporal resolution limit, and a spatial resolution limit. The temporal resolution limit pertains to the speed at which the target (and distractor) objects move and/or change as they move around the space. Holcombe and Chen (2013) discuss the maximum number of items that can be tracked based on the temporal frequency of change in the targets and distractors, finding that when only one object is being tracked, it could be tracked at any frequency below 7 Hz (~143 ms). However, with two objects to be tracked this frequency limit fell to 4 Hz (250 ms) and with three objects it fell further to 2.6 Hz (~385 Hz). This is a clear pattern of data showing that with increasing number of objects, the frequency limit at which it is not possible to reliably track the objects decreases. This finding was also shown to be independent of any spatial resolution limit, as was shown by Intriligator and Cavanagh (2001). They examined spatial resolution, finding that as objects were put closer to each other (for example, along an imaginary circle with a smaller diameter) it became significantly more difficult to track a greater number of them. Intriligator and Cavanagh (2001) showed that when the angle between targets subtended less than one degree of visual field, it was not possible to track the difference between targets and distractors. They also showed that this effect was independent of capacity of working memory, because there was no difference in results with one or three targets being displayed. In answer to the question of whether temporal or spatial resolution plays a greater role in multiple object tracking, Franconeri, Jonathan, and Scimeca (2010) considered the interplay of these factors. They presented dots moving in the central and/or peripheral fields of vision, and also had dots that may be close to one another or distant from one another. The length of time participants were asked to track the objects and the speed of movement of objects were also tested. The findings of this experiment indicated that the only factor that played a significant role was object spacing, with no evidence for modulation of object tracking by speed, time, or capacity of working memory. In sum, these finding show that there are multiple factors that can influence multiple object tracking. While Franconeri et al. (2010) found no effect of speed, time or capacity, it is also important to keep in mind that factors such as these do not necessarily work in isolation from one another. Rather, it is likely that they work in combination, with the potential for certain factors (e.g. visual stimulus factors) to only reveal themselves as a modulator under certain, ambiguous temporal conditions (or vice versa). In the current research, these factors take the form of visual perceptual load, as well as the speed of presentation as operationalized by SOA. For example, the number of locations changing may only play a role when SOA is of a sufficiently perceptible speed.

Drew, Horowitz, and Vogel (2013) looked at the types of errors participants made when performing a multiple object tracking task. Increasing distractor load by adding more distractor stimuli (that are present in the display, but are not to be tracked) leads to an increase in errors. Similarly, increasing speed of movement of targets also leads to an increase in errors. These errors could be caused by ‘swapping’ – inadvertently switching a target for a distractor – or ‘dropping’ – losing a target, and eventually tracking less than the number of targets you started with. Contralateral delay activity (CDA) is a slow wave evident in EEG recordings that is sensitive to the number of objects that are held in visual working memory (Luria, Balaban, Awh, & Vogel, 2016). By measuring CDA in a multiple object tracking task, it is therefore possible to index the number of objects being tracked (Vogel & Machizawa, 2004). As such, a swapping error should result in a CDA indicating the ‘correct’ number of objects being tracked, but with an increase in likelihood of making a response error, while a dropping error should show a decrease in CDA, as fewer items are now being tracked. Drew et al. (2013) performed a series of experiments in which they manipulated both distractor load and speed, while measuring CDA and testing participants’ behavioural responding as well. They confirm that both increasing speed (from 8.5 degrees/sec to 11.6 degrees/sec) and adding distractors (6 distractors rather than 3, with 1 or 3 targets) led to a decrease in response accuracy, but that only increasing of speed provided electrophysiological evidence that targets were being dropped. They conclude that speed increases lead to potential targets being dropped, while increased distractor load leads to potential targets being swapped (mistaken for distractors).

While these data pertain to multiple object tracking, and not audiovisual integration directly, there is a parallel to be drawn, and as such the findings can be used to make predictions in our study. In multiple object tracking participants are asked to track the location of objects, updating their locations repeatedly as they move around a display, in the same way that we must track and integrate may stimuli in our everyday existence.. In the task we will use to test capacity of audiovisual integration, a participant will be asked to keep track of a feature (colour) of a number of stimuli as they change, and then respond to whether a particular stimulus changed in synchrony with an auditory tone (after van der Burg et al., 2013). Bahrami (2003) investigated the respective tracking of elements of objects, such as colour and shape, during a simultaneous multiple object tracking task. While objects were being tracked as they moved in space, they also sometimes changed colour or shape, and did so while visible or occluded. He found that participants were able to identify colour changes for targets, but not distractors, and that this was as reliable as tracking the objects themselves when the colour changes occurred while the object was visible to the participant. Given this evidence, tracking the location of multiple objects can be used as a rough analogue for tracking the state (colour) of them, and in fact the capacity of participants to notice colour changes in multiple object tracking (performing both tasks simultaneously) has been shown to have a capacity closer to 2 (Bahrami, 2003). So it might be expected that the maximum capacity of audiovisual integration would be somewhere at or above the measure of 2 found by Bahrami.

Attention and Multiple Object Tracking

There also exist data indicating that early attentional selection plays a role not only in working memory entry, but also in a multiple object tracking task. Drew et al. (2009) had participants track two targets moving amongst four stationary and four moving distractors. They measured P1 and N1 components when flashes were presented on targets or distractors, finding enhancement of both P1 and N1 magnitude for probes that occurred on targets when compared to flashes on distractors. Since flashes on targets (that were meant to be attended to) led to a greater visual evoked response than did flashes on distractors (that were meant to be ignored), this provides evidence that attention was being employed to focus on the targets and ignore the distractors.

Sears and Pylyshyn (2000) presented participants with a multiple object tracking task in which targets or distractors could change form during the task. They found that, when tracking was completed successfully (evidenced by a correct identification at the end of the task), participants also showed a higher level of detection of the form change. This was true for targets, and not for distractors, and as such this experiment showed enhanced processing for targets and not for distractors in a multiple object tracking task. Sears and Pylyshyn (2000) argue that this means participants are able to contribute an *a priori* attentional focus on a certain number of targets, and are able to track both their movements and their nature throughout a task. Using attention to focus on targets is similar to what was discussed by Drew et al. (2009) above, while the ability to allocate attention to specific objects ahead of time (rather than allocating focus to certain areas) is a novel finding from Sears and Pylyshyn (2000). Doran and Hoffman (2010) also present ERP evidence from multiple object tracking research, showing N1 differences when tracking two targets among two distractors. This further reinforces the findings of Drew et al. (2009), providing additional evidence in support of the function of attention in multiple object tracking. Participants are able to successfully track targets and suppress the effects of distractors when the visual load is sufficiently low (2 targets and 2 distractors), and this is possible by attentional functions, as indexed by ERP.

Multiple object tracking has been shown to use similar processes and has similar response patterns to tracking properties of objects (Bahrami, 2003; Sears & Pylyshyn, 2000). If anything, the capacity of tracking properties seems to be smaller (~2) than the capacity for tracking objects themselves (~4). Given that the capacity for both these properties exceeds one item, and given that the presence of an auditory tone has been shown to increase perceptibility of a visual stimulus (Van der Burg, Olivers, Bronkhorst, & Theeuwes, 2008), we would certainly expect it is possible for the capacity of audiovisual integration to exceed one item.

Capacity of Audiovisual Integration

The specific phenomenon which this dissertation will investigate is the capacity of audiovisual integration. This capacity is a function of perceptual processes, working memory processes, and decision-making processes, as discussed in the review. It may be determined by the likelihood of information being accurately perceived by sensory systems, successfully encoded into working memory, and / or clearly retrieved from memory.

The capacity of audiovisual integration is related to the "pip and pop" effect, in which a visual stimulus, or change in an attribute of a visual stimulus, is more easily perceived when accompanied by an auditory tone (Van der Burg et al., 2008; Matusz & Eimer, 2011). This effect shows us that auditory signals can increase the likelihood of perceiving a visual event, which is a type of audiovisual integration (albeit low-level). More recently, Van der Burg, Awh, & Olivers (2013) asserted that regardless of any stimulus factors, there is a strict limit of one item on the capacity of audio-visual integration. However, in light of the literature discussed above, it seems highly unlikely that there should be such a strict, impenetrable limit on the capacity of audiovisual integration. Given that around 4 items can be held in visual working memory (Cowan, 2001), 2 items can be held in auditory working memory (Saults & Cowan, 2007), there is no account from working memory that would require capacity to remain below one. Additionally, Holcombe and Chen (2013) provide temporal guidelines that would allow for up to three objects to be tracked at an SOA of 385 ms, and Bahrami (2003) showed that colour can be tracked on two items simultaneously. This evidence, along with that coming from unimodal perceptual literature on effects of visual load (Lavie, 2005), SOA (Marois & Ivanoff, 2005), and training (Klingberg et al., 2002) indicates that the capacity of audiovisual integration should be flexible, affected by perceptual and attentional factors, and that it need not be limited to a single item.

In a series of experiments, Van der Burg et al. (2013) adapted their own pip-and-pop paradigm to determine capacity of audiovisual integration. They presented participants with a number of locations indexed by dots, arranged along an imaginary circle, each of which could be either black or white. Then a subset of locations between 1 and 8 changed from white to black, or vice versa. Participants were asked to keep track of which locations were changing, and to remember which locations changed in synchrony with an auditory tone. After the changes had been completed, a probe display was presented wherein one location was marked in red, and participants responded as to whether that location did (or did not) change in synchrony with the tone. The experimental series employed a number of manipulations, including set size (16 or 24 dots), speed of presentation (150 or 200 ms SOA), and binding type (audiovisual or visuo-visual (where a ring appeared around the dots to indicate the critical trial)). Participants' raw proportion correct scores were subject to model fitting to a variation of Cowan's (2001) *K*, which is an estimate of the capacity of integration. The model holds that if $n \leq K$, then $p = 1$, and if $n > K$, then $p = K/2n + .5$, where n is the number of visual events, and p is the probability of an observer giving a correct response. This model yields an estimate of capacity of audio-visual integration for each condition. The results of Van der Burg et al.'s (2013) study show that the capacity of audio-visual integration was limited to 1 item, regardless of the variety of factors tested. Both set sizes of 16 and 24 yielded capacities of less than 1 item (.84 and .71, respectively). Using a 200 ms SOA resulted in a capacity of .71 and 150 ms SOA resulted in a capacity of .61. Finally, capacity was greater when using an audio-visual signal (.58) than when using a visuo-visual signal (.13).

The first section of this dissertation will examine the specific paradigm used by Van der Burg et al. (2013) and will attempt to show that by reducing the difficulty of the paradigm and by varying specific parameters, it is possible for the capacity of audio-visual integration to exceed one item. Van der Burg et al.'s (2013) experiments already hint at variation in audiovisual integration capacity, as it shows increases in capacity at a 200 ms rather than a 150 ms SOA, and when visual set size is 16 rather than 24. I believe that this is indicative of audiovisual integration capacity being on a continuum rather than something that has a strict limit of one. While Van der Burg et al.'s (2013) experimental parameters were not sufficient for the capacity to exceed one, but that does not mean it is not possible for this to occur.

Research Questions and Hypotheses

Taking into account the literature reviewed above, the main research question for this series of experiments is: To what extent is the capacity of audio-visual integration fixed, and if it is malleable what are the factors that influence its capacity? The series of experiments outlined below will seek to answer this question, and we expect will illustrate the dynamic nature of the capacity of audiovisual integration and the factors that play a role in its flexibility.

In the current research, one of the aims is to reduce the perceptual load such that it falls within a range that a participant can perceive without requiring top-down control (Sobel et al., 2007). This means limiting the number of visual stimuli that are to be tracked such that it is at least possible for all of them to be tracked by a participant. This will be accomplished by reducing the number of visual stimuli in the paradigm to 8 items (whereas Van der Burg et al. (2013) had either 16 or 24 stimuli).

Logie et al. (2011) found that at SOAs of less than 500 ms, there was a decrement in working memory capacity. Additionally, Holcombe and Chen's (2013) findings indicate that the minimum SOA at which two objects can be tracked is 250 ms (and 385 ms for three objects). This is important as the previous studies by Van der Burg et al. (2013) used SOAs of less than 500 ms (150 and 200 ms, respectively). This would likely have a detrimental effect on encoding into working memory, and could then cause a subsequent decrement in the apparent capacity of audio-visual integration. In Experiment 1, we will also record EEG data in the encoding, maintenance, and recall phases of an audio-visual integration task, to see which stage is critical to increasing integration capacity. This will provide an index as was done by Luck and Vogel (1997), demonstrating the degree to which there is neural discrimination of the various conditions presented and which neural responses serve as indices of successful and unsuccessful integration.

The first set of four experiments will seek to break down the paradigm employed by Van der Burg et al. (2013), and to examine the experimental factors that affect the capacity of audiovisual integration. Experiment 1 will establish that rather than being fixed at 1 (as proposed by Van der Burg et al., 2013), slowing of the stimulus presentation allows the capacity of audiovisual integration to exceed 1 item. The main purpose of Experiment 1 is to show that the capacity of audiovisual integration is dynamic, and not limited to one item.

Having shown that the capacity of audiovisual integration is flexible, it is necessary to investigate further factors that may play a role in audiovisual integration of stimuli. In addition to visual load (number of objects; Lavie, 2005) and SOA (Marois & Ivanoff, 2005), it is possible that other factors influenced the findings of Van der Burg et al (2013) to limit it to a capacity of 1 item. In their experimental paradigm, there was a relatively high degree of proactive interference

(8 presentations before the critical stimulus) and a high degree of temporal uncertainty as to when the critical stimulus will be presented. Experiments 2, 3, and 4 (along with the findings of Experiment 1) will serve to examine the respective effects of these factors in isolation, and in combination. Based on previous research, I expect a greater amount of proactive interference to reduce the capacity of audiovisual integration. Similarly, I expect low temporal predictability of critical stimuli to further reduce capacity of audiovisual integration.

Experiments 5 and 6 will look at stimulus factors that may affect capacity of audiovisual integration. Experiment 5 will incorporate previous research we have done (Wilbiks & Dyson, 2013a), looking at the effect of crossmodal congruency on the capacity of audiovisual integration. That is to say, does the capacity for audiovisual integration increase if the auditory and visual inputs that have the potential to be integrated are congruent with one another or not. I expect that just as congruent stimuli are more likely to be bound to each other in a competitive binding scenario, and are more resistant to asynchronies in presentation (Wilbiks & Dyson, 2013a), using congruent stimuli will increase the capacity of audiovisual integration. Experiment 6 will test for the possibility of employing perceptual chunking as a way to increase the capacity of audiovisual integration. The research question here is whether creating ‘chunks’ in the form of lines or polygons will allow participants to integrate 2, 3, or 4 vertices with a single tone, while 3 or 4 locations have been shown to exceed the capacity of audiovisual integration. I expect that this will be the case, based on Miller’s (1956) research on working memory capacity and chunking, as well as later work into perceptual chunking (Gobet, Lane, Croker, Cheng, Jones, Oliver, & Pine, 2001).

Experiment 7 will examine the potential for individuals to increase their capacity of audiovisual integration via training. Having shown that capacity is dynamic, and having explored the effects of proactive interference and temporal predictability, I will finally look at whether audiovisual integration capacity is malleable *within* an individual in repeated testing sessions. The prediction here is that capacity will increase through repeated training sessions.

Chapter 2 – Exploring the Van der Burg et al. (2013) paradigm

General Introduction

Van der Burg et al. (2013) propose that there is a strict limit to the capacity of audiovisual integration, asserting that it is not possible for it to exceed one item. The suggestion of an upper bound for audio-visual integration is consistent with the idea of visual short term memory (VSTM) limits, but somewhat at odds with the reality of individual differences and the range of values often reported for VSTM capacity (1.5 – *6* reported by Vogel & Machizawa (2004)) and, indeed, AV capacity (0.70 – *1.56* reported by Van der Burg et al. (2013) in their Experiment 1c [200 ms SOA] condition, or, 0.30 – *1.36* in their Experiment 2 [150 ms SOA] condition). The data are also at odds with the multi-modal integration literature in general, which tends to emphasize its dynamic nature (e.g., Chan, Pianta, & McKendrick, 2014; Fujisaki, Shimojo, Kashina, & Nishida, 2004; Wilbiks & Dyson, 2013b). Finally, it seemed likely that the factors that influence capacity in the visual domain like perceptual load (Lavie, 2005) and rate of presentation (Marois & Ivanoff, 2005) should also modulate capacity in the audio-visual case.

Firstly, a central tenet of perceptual load theory states that attention can be deployed to both target and distractor information under conditions of low rather than high load (Lavie, 2005). Perceptual load can be manipulated in many ways, including set size (Lavie & Cox, 1997), the degree of difference between targets and distractors (Lavie, 1995), and difficulty of the search task itself (conjunction of features rather than simple feature; Treisman & Gelade, 1980). Given the paradigm that was used by Van der Burg et al. (2013), which is being explored and manipulated here, the most appropriate form of load manipulation looks to be set size. In the current research, between one and four of eight possible locations will be changing colour rapidly between black and white, in contrast to Van der Burg et al. (2013), who noted in their

Experiment 1b that decreasing set size from 24 to 16 items led to a non-significant increase in audio-visual capacity. However, with a reported *p* > .1 using 9 participants, it is likely that the importance of set size has been underestimated. Secondly, the number of locations or objects that can be reliably tracked is in part determined by the rate of stimulus presentation. For example, Holcombe and Chen (2013) presented participants with two or three rings, each of which contained squares that rotated about a fixation point. On different trials, one, two, or three of the squares were designated as targets, and the rest were designated as distractors. Participants were asked to track the targets as they rotated, and then indicate which square(s) were target(s) at the end of a trial. The researchers found that with a single target, the target could only be tracked at a period of rotation that was greater than 143 ms. For two objects this limit was 250 ms, and, for three objects it was 385 ms. These values are defined as a *temporal frequency limit*, meaning that with a period of rotation lesser than these values, the respective number of items cannot be reliably tracked. So it is likely that this temporal frequency limit makes it impossible for the capacity of integration to exceed one item at stimulus onset-asynchronies (SOAs) below 250 ms, since it is not possible to track more than 1 visual object. Thus, the limit is due to the constraints of visual processing rather than audio-visual integration. This aligns well with the finding of a capacity upper-bound of 1 using 150 or 200 ms. Van der Burg et al. (2013) also noted that decreasing the speed of presentation between successive frames (SOA) from 150 to 200 ms also led to a significant increase in performance, assumedly due a reduction in the number of incorrect audio-visual bindings (Van der Burg et al., 2013, p. 348). This lends further credence to the idea that the estimates of capacity established by Van der Burg et al. (2013) are likely a conservative estimate of actual performance. Given that both perceptual load and SOA are both continuous variables, there seems good reason to expect that with more distinct manipulations, the capacity of audio-visual integration would exceed 1.

The additional manipulations within this chapter will involve a comparison of the respective effects of proactive interference and the temporal predictability of the target stimuli. Lustig, May, and Hasher (2001) examined the degree to which proactive interference plays a role in affecting working memory span. They employed a reading task, where participants were asked to remember certain sentences while reading a different story. They found that, if the sentences being remembered were longer, it led to a greater interference effect on participants' recall of the story, while shorter sentences interfered less. This shows us that proactive interference can adversely affect working memory span, but it is not directly relatable to the kinds of stimuli and task being employed in the current research. Makovski and Jiang (2008) employed a task much more similar to the one that will be used in this dissertation, looking at participants' ability to perceive and remember a number of coloured discs, while manipulating the degree of proactive interference by means of a different colour having (sometimes) been presented at the same location on a previous trial (Experiment 1). The probe on each trial could be a match for that trial, could be a colour that is incorrect but was not presented at any location on trial *n* or trial *n-1,* could be a colour that *was* presented in a different location on trial *n-1,* or a colour that was presented in the identical location on trial *n-1*. They found that previously used locations / colours interfered with performance on current trials, indicating that proactive interference is affecting visual working memory performance. In an additional experiment (Makovski and Jiang, 2008; Experiment 3) they tested both the spatial and temporal resolution of these proactive interference effects by using different amounts of time between trials and be using four potential locations for their probe. This manipulation showed that location is not an important cue for interference, since mistakes were made equally at locations that matched or did not match the interfering stimulus. Especially interesting for the current research, they also show that temporal resolution of proactive interference did not play a significant role. That is, interference occurred equally regardless of whether 100, 400, 1000 ms passed between trials. In the current research, both the number of presentations and the SOA will be orthogonally manipulated to derive a sense of how these factors interact with one another in affecting our interference manipulation.

Temporal predictability is another factor that can contribute to successful perception of stimuli. Thomaschke and Dreisbach (2013) presented participants with target stimuli to be identified, and that could appear either at a time that was predictable within the experiment, or unpredictable. They found that participants responded more quickly when the presentation was predictable, and that this was also further facilitated when participants repeated blocks of temporally predictable (rather than unpredictable) trials. Shin and Ivry (2002) showed similar results, finding that when trials were presented with a predictable and structured temporal arrangement, participants were both better able to respond to stimuli more quickly and accurately, and showed an increasing sensitivity to spatial factors.

Over four experiments in this chapter, I will manipulate the parameters of Van der Burg et al.'s (2013) experimental paradigm for evaluating the capacity of audiovisual integration. In Experiment 1, a modified version of Van der Burg et al. (2013) was run, examining audiovisual integration capacity under both fast (200 ms) and slow (700 ms) SOA conditions, along with reduced perceptual load of 8 items. Simultaneous EEG recording during Experiment 1 also revealed the neural signatures associated with the encoding and retrieval phases of the paradigm. In particular, the repeated failure of AV capacity to exceed 1 under 200 ms SOA condition may be due to the inability of the visual cortex to successfully code the number of changing locations in frames prior to the critical one (as per Drew & Vogel, 2008; Culham, Brandt, Cavanagh, Kanwisher, Dale, & Tootell, 1998). In subsequent experiments, two further critical features were manipulated: the degree of proactive interference generated by non-critical frames and the temporal predictability of the critical frame. In Experiment 2, we reduced the amount of proactive interference, while maintaining the predictability of the critical stimulus. In Experiment 3, high proactive interference was paired with a temporally unpredictable (roving) critical stimulus, while in Experiment 4, a roving critical stimulus was presented with a low degree of proactive interference.

		Proactive Interference	
		High	Low
Temporal Predictability	High	Exp 1	Exp 2
	Low	Exp 3	Exp 4

Figure 1. Factorial 2 x 2 design for Experiments 1 – 4.

Experiment 1 – Longer SOA and EEG recording

Introduction

Given what is known about audio-visual integration and its dynamic qualities (Fujisaki et al., 2004; Vroomen et al., 2004), it seems highly unlikely that the capacity of audio-visual integration should be strictly limited to 1 item (Van der Burg et al., 2013). In an effort to test this hypothesis, I reasoned that if the capacity of audio-visual integration is absolutely fixed at 1, then it should be resistant to effects of factors that have been previously shown to modulate unimodal capacity limits. Experiment 1 served as a replication of the basic van der Burg et al. (2013) paradigm, with two important changes. By reducing the number of visual elements from 16 (or 24) to 8, we expected to see an increase in capacity, according to perceptual load theory (Lavie, 2005). Additionally, we included a slower rate of presentation (700 ms) in addition to one of Van der Burg's (2013) original SOAs (200 ms), which should lead to an increase in capacity, according to the principles of temporal frequency limits (Holcombe & Chen, 2013).

In an additional element of the study, scalp EEG was recorded to measure the level of activation as an index of resource use (Luck & Vogel, 1997), to titrate the paradigm into three stages: an encoding stage while the multiple displays of changing visual stimuli are presented to the participant, a maintenance stage, where participants are asked to remember which dot(s) changed simultaneously with the auditory tone, and a retrieval stage, when participants were asked to respond to a probe stimulus. In doing so, it was hoped that the stage(s) at which brain activity predicts behavioral performance would be revealed. This would also allow for identification of the stage(s) in which limitations on audiovisual integration capacity may be expressed.

During encoding of the non-critical frames, the neural component that could be most reliably compared between a trial running at 200 ms SOA and 700 ms SOA is the visual N1. The posterior/visual N1 has been shown to modulate according to spatial attention and discrimination (Vogel & Luck, 2000). Specifically, it is a component which is present and modulates based on discriminating between form- and colour-based changes, and is present in both presence and absence of motor responding. The visual N1 component can vary in location and latency as a function of specific stimulus and attentional factors. Generally speaking, however, it peaks at between 165 and 195 ms after stimulus presentation, between P1 and P2 components (Makeig, Westerfield, Townsend, Jung, Courchesne, & Sejnowski, 1999). Given that the visual N1 is sensitive to a number of characteristics including the magnitude of physical change, attentional allocation and the eventual requirement of a discriminatory response (Luck, Heinze, Mangun, & Hillyard, 1990; Vogel & Luck, 2000), we believed the N1 should index the ability of the visual cortex to discriminate between the number of to-be-tracked locations: a neural prerequisite of the task would be the coding of location change numerosity, confirming that the visual cortex is sensitive to the number of locations that change polarity at each trial. Successful performance on the current variant of the pip-and-pop paradigm (Van der Burg et al., 2013) relies on identifying which visual locations changed in polarity when an auditory cue is presented. Logically, one can only have a sense of which locations *changed* at any one frame by successfully registering the status of the various locations during preceding frames. Failure to discriminate between the number of changing locations (1, 2, 3, 4) in the frames leading up to the critical one would suggest that participants do not have the prerequisite perceptual information required to successfully identify which location(s) changed at the time the auditory cue was presented. Consistent in this regard is a previous report by Van der Burg et al. (2011), who show that with fast rates of presentation between 50 and 250 ms, basic exogenous responses such as P1 and N1 fail to generate in visual cortex. In Experiment 1, we further test the hypothesis that the AV capacity of 1 observed at fast SOA is due to poor quality sensory (visual) information entering working memory by examining neural responses at the encoding phases of the paradigm. In the current paradigm, task difficulty is operationalized by the SOA and the number of items to be tracked. So, within each SOA, as the number of items to be tracked increases, if they are being successfully tracked, this should be accompanied by an incremental increase in N1 amplitude for each additional item. If tracking is not successful, there should also be no increase in N1 magnitude associated with the number of items to be tracked.

During the retrieval phase in which participants responded to the probe, we also anticipated the observation of posterior N2 and P3b components as indices of visual selection (e.g., Mazza & Caramazza, 2011; Luck & Hillyard, 1994; Folstein & Van Petten, 2008) and the initiation of a response resulting from perceptual analysis (e.g., Verleger, Jaskowski, & Wascher, 2005). Posterior N2 arrives at approximately 280 ms, with its origin in posterior regions, while P3b originates from temporo-parietal regions, arriving at around 300 ms (O'Donnell , Swearer, Smith, Hokama, & McCarley, 1997). Consistent with the data in Van der Burg et al. (2011) we assumed that the components should be maximal at posterior sites, thereby locating one site of AV integration at posterior inferior cortex. In particular, the magnitude of N2 should increase as the number of successfully individuated visual objects increased (Mazza & Caramazza, 2011) and P3b should be reflective of the degree to which the location is behaviourally relevant (Van der Burg et al., 2011). Mazza & Caramazza (2011) examined the N2pc component, which is known to be an indicator of visual selection. They found that the magnitude of N2pc at posterior sites increases as a function of target numerosity when moving from 1 to 2 to 3 target items. In the current experiment, if tracking is successful this should be accompanied by a concomitant increase in N2pc magnitude during the encoding stage. Finally, by examining the correlation between neural response at encoding and retrieval phases with behavioural response, it should be possible to ascertain the degree to which successful eventual performance is associated with perceptual processes (encoding) and / or post-perceptual processes (retrieval).

This first experiment will begin to ascertain whether it is possible for the capacity of audio-visual integration to exceed one item. Behavioural measures of response accuracy (an estimate of audiovisual integration capacity (K), after Cowan's (2001) K) will show an index of the number of items being presented, and the shape of the curve these data produce will vary based on the SOA of the presentation of visual stimuli. These data will be further supported by ERP data, which will serve as a neurological measure of discrimination, indicating the number of items that are successfully able to be bound to an auditory stimulus.

Method

Participants. Informed consent was obtained from 25 participants prior to experimentation. All participants were recruited from an undergraduate research participant pool, and were compensated with partial class credit. The Research Ethics Board at Ryerson University approved the experimental procedure and recruitment practices. Before data analysis, a 95% confidence interval was calculated around 50% (chance responding) over 384 trials. Data from any participant who was performing within the 95% confidence interval on average across all 8 conditions was removed, due to them performing too close to chance on the task. A total of 9 participants were rejected due to chance responding. An additional 3 participants were rejected because of low quality EEG recording. As such, the final sample consisted of 13 participants – 3 males and 10 females – with a mean age of 18.5 years (SD = 1.2), and a total of 13 right handed individuals. All participants self-reported normal or corrected-to-normal vision and hearing.

Stimuli. Visual stimuli were presented on a Viewsonic VE175 monitor at a screen resolution of 1280 x 1024 and a refresh rate of 60 Hz, using an ACPI PC running Windows 7 Professional (Service Pack 1). Auditory stimuli were presented binaurally via Sennheiser HD 202 headphones, and were presented at approximately 74 dB(C) as confirmed by a Scosche SPL100 sound level meter. Stimulus presentation was controlled by Presentation (NBS, 2013) version 16.5, build 09.17.13, and behavioural responding was recorded with a Dell SK-8165 keyboard.

Dots were created using Presentation software. These dots were 1.3 cm in diameter, such that they subtended an angle of approximately 1.3° at a viewing distance of approximately 57 cm. Dots could be displayed in one of two colours: black (0, 0, 0) or white (255, 255, 255), and were displayed on a mid-grey background (128, 128, 128). Eight dots at a time were presented along an implied circle, which had a diameter of 13 cm, and the center of which was marked by a 1.5 mm fixation dot. A single, smaller probe dot was overlaid on a target dot at the end of each trial, and was red (255, 0, 0) with a diameter of 1 cm. The auditory stimulus used was a 400 Hz tone with 5 ms linear on-set and off-set ramps, which was created using SoundEdit 16 (MacroMedia).

Design and procedure. 16 individual conditions of stimulus were created, by orthogonally varying the SOA of visual stimuli (200 or 700 ms), the number of visual stimuli that changed on each alternation (1, 2, 3, or 4), and the validity of the probe stimulus (valid or invalid). These 16 conditions were each presented 3 times to create an experimental block with 48 trials. Each participant completed one practice block, and 8 experimental blocks, for a total of 384 experimental trials.

Figure 2 depicts the series of presentations for each of Experiments 1 through 4. Each trial began with the fixation point displayed in the center of the screen for 500ms. The sets of black and white dots were generated independently for each trial, and there was no restriction on which dot(s) could change colour at each alternation, nor was there a restriction on how many dots could be white or black at any one time. The first array of dots was presented for either 200 or 700 ms (dependent on condition), and subsequent arrays followed immediately at SOAs of

200 or 700 ms, for a total of 10 presentations. On the penultimate (9^{th}) presentation, the changing of the locations was accompanied by an auditory tone.

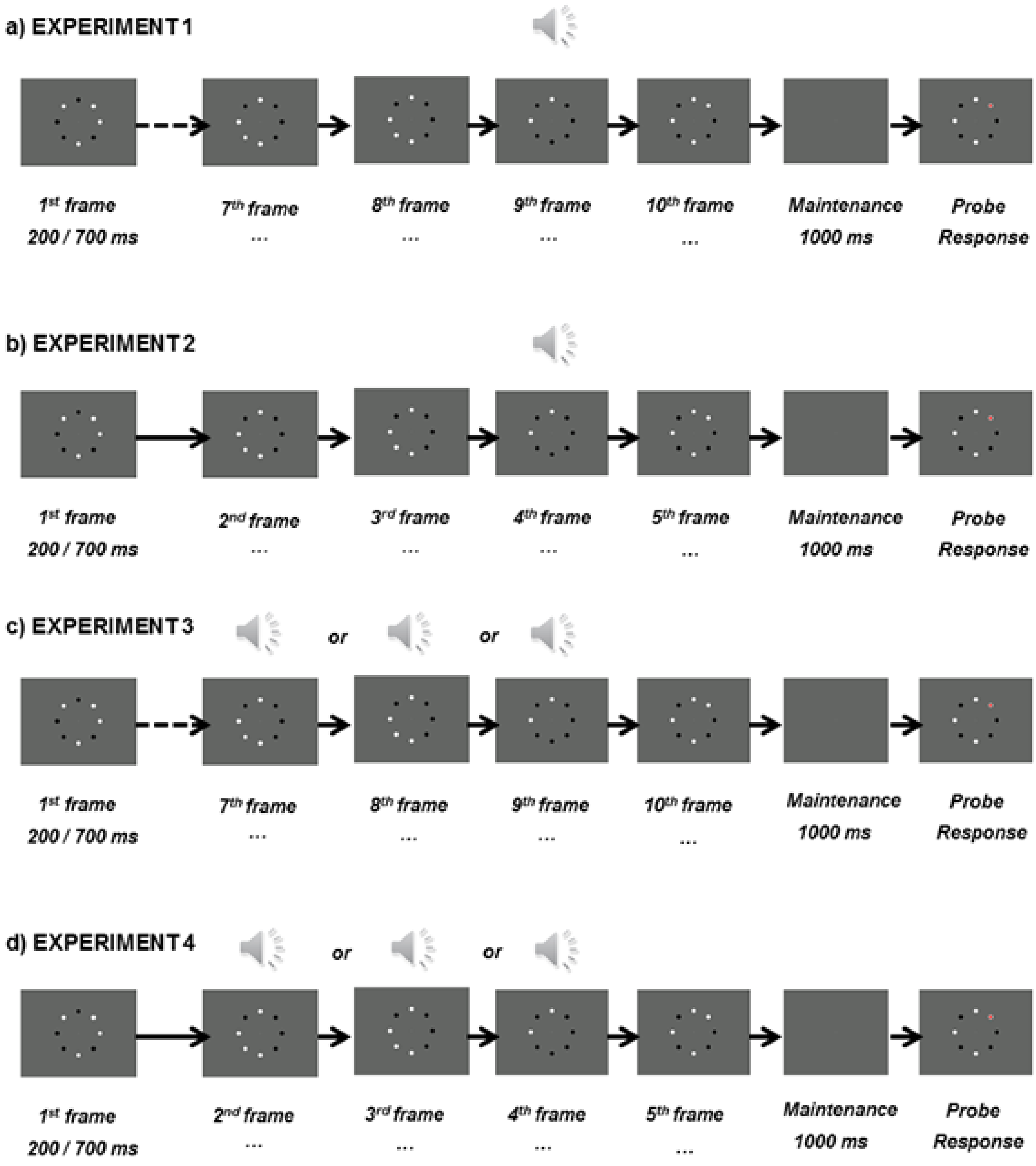


Figure 2. Trial schematic for each of Experiments 1 – 4, indicating locations of dots changing as well as timing of auditory stimulus.

Following the final presentation, a 1000 ms retention interval occurred during which only the fixation point was displayed on the screen. The tenth array of dots was then displayed again, along with an overlay of a red probe dot on one of the eight dots. This probe had a validity of 50%, meaning that for any given trial there was a 50% chance that the probed location did change, and a 50% chance that the probed location did not change. When the probe was invalid (probed location did not change), the location of the probe was randomly determined. Participants were asked to respond to whether the dot at the probe location had changed or not on the critical display (accompanied by the tone) by pressing the number 2 on the number pad if that dot did change, and by pressing the number 1 if that dot did not change. No feedback was provided, and the subsequent trial began immediately after a response was entered, with a variable time interval to disrupt any pattern of timing. Trial order was randomized in practice and in experimental trials.

Model Fitting. Data were modelled in the same manner as employed by Van der Burg et al. (2013). The proportion correct for each condition and for each participant was fitted to a model equivalent to Cowan's (2001) *K*, wherein if $n \leq K$, then $p = 1$, and when $n > K$, then $p = K/2n + .5$ (*n* represents the number of visual elements changing (1-4), *p* is the probability of correct responding, and *K* is an estimate of capacity of audiovisual binding). Data were fitted to this model by using Microsoft Excel Solver, and fitting was initiated from several starting values of *K*. The outcome with the smallest RMSE was selected, and this process was done independently for each participant and SOA condition. Successful model fit was confirmed by the low RMSEs observed in both the 200 ms SOA (range 0.003 – 0.028) and 700 ms SOA (range 0.002 – 0.036) conditions.

Electrophysiological recording. Electrical brain activity was continuously digitized using Acti-View (Bio-Semi; Wilmington, NC), with a band-pass filter of 208 Hz and a 1024 Hz sampling rate. Recordings made from FPz, F3, Fz, F4, C3, Cz, C4, PO7, P3, Pz, P4, PO8, T7, T8, POz, Oz, M1, M2, CMS and DRL were stored for off-line analysis. Horizontal and vertical eye movements were recorded using channels placed at the outer canthi and at inferior orbits, respectively. Data processing was conducted using BESA 5.3 Research (MEGIS; Gräfelfing, Germany). Following average referencing, the contributions of both vertical and horizontal eye movements were reduced from the EEG record using the VEOG and HEOG artefact options in BESA. Using a 0.1 (12 db/oct; zero phase) high-pass and 30 (24 db/oct; zero phase) Hz low-pass filter, epochs were rejected on the basis of amplitude difference exceeding 100 µV, gradient between consecutive time points exceeding 75 µV, or, signal lower than 0.01 µV, within any channel. Following average mastoid re-referencing, neural activity was averaged across PO7, POz, PO8, P3, Pz, P4, C3, Cz and C4 electrode sites. For the encoding phase, epochs in the 200 ms SOA condition were baseline corrected 100 ms prior stimulus onset and activity was examined 200 ms following stimulus onset. In the 700 ms SOA condition, baseline correction was established 200 ms prior to the stimulus and activity was examined 700 ms following the stimulus.[1] Irrespective of rate of presentation, maintenance and retrieval stages of audio-visual integration were defined according to a baseline of 200 ms prior to stage and 1000 ms following the onset of the stage.

[1] The reason for the different baseline epoch is a necessity of the different SOAs. Taking a 200 ms baseline in the 200 ms SOA would not be feasible, as it would encompass the entirety of the previous trial.

Results

In order to evaluate whether capacity of audiovisual integration can exceed one item, capacity estimates (*K*) were submitted to single sample t-tests against the test value of 1. Audiovisual integration capacity estimates of *K* were significantly smaller than 1 in the 200 ms condition (0.72; range = 0.17 – 1.52; t[12] = -2.34, *p* = .038; cf. van der Burg et al., 2013, Experiment 1c, 200 ms SOA, estimate of 0.71) but significantly larger than 1 in the 700 ms condition (1.91; range = 0.30 – 3.16; t[12] = 2.99, *p* = .011). The data in Figure 3 confirm that the capacity of audiovisual integration can be greater than one item, when slower rates of stimulus presentation are used.

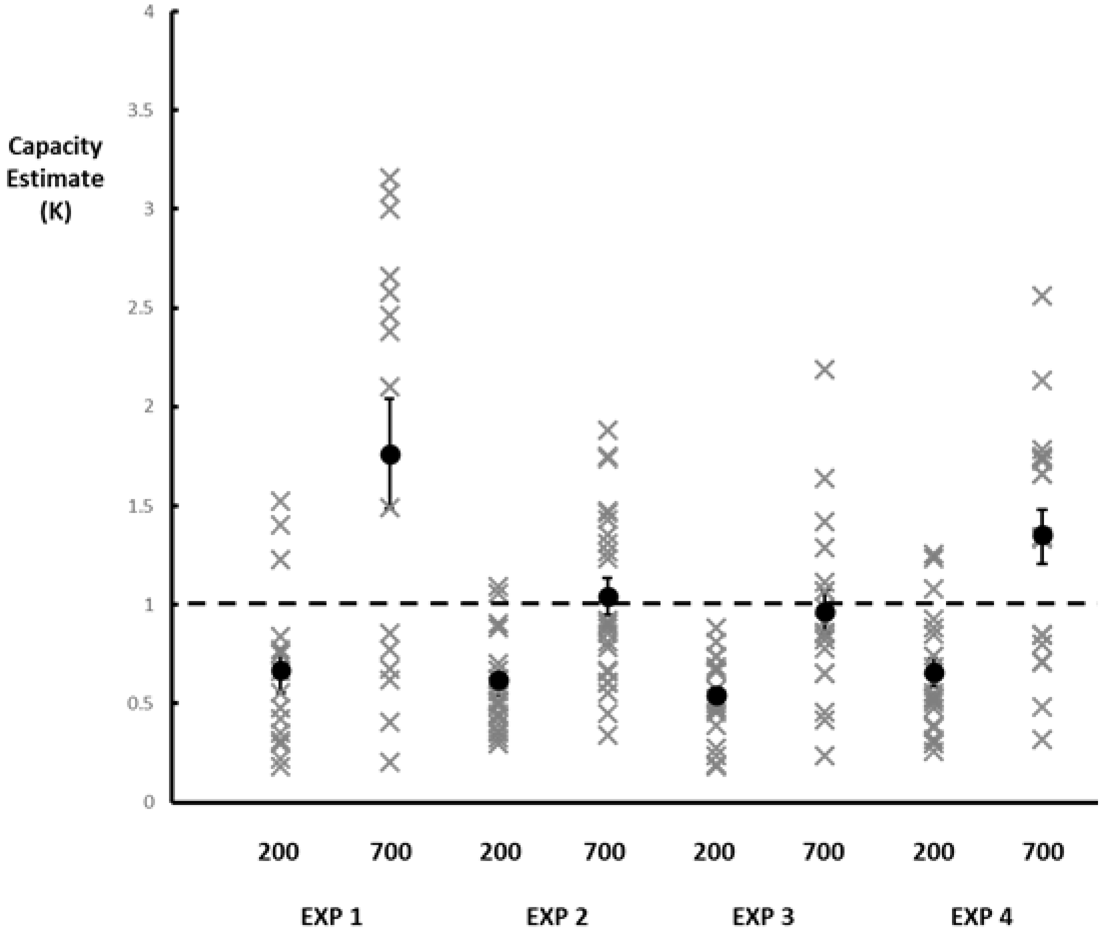


Figure 3. Capacity estimates (K) for each of Experiments 1 – 4. Solid dot represents experimental mean, with error bars representing standard error. Crosses represent individual participant scores.

Figure 4 shows proportion correct data as a function of both SOA (200 ms, 700 ms) and the number of to-be-tracked visual locations (1, 2, 3, 4). These data were submitted to a repeated measures ANOVA with within–participants factors of SOA (2) x number of objects to be tracked (4), the full results of which are displayed in Table 1. Performance was significantly better when presentation rate was slow (*p* < .001) and as the number of locations decreased (*p* < .001). An interaction (*p* = .006) revealed that performance was influenced by the number of to-be-tracked objects more in the 200 ms SOA than 700 ms SOA condition.

Table 1 Summary of repeated measures ANOVA on proportion correct for ERP sample of Experiment 1

Metric	*df*	F	MSE	*p*	η_p^2
SOA (S)	1,12	33.91	.017	**<.001**	.739***
Number (N)	1,12	28.24	.005	**<.001**	.702***
S x N	1,12	4.94	.003	**=.006**	.291***

Note: Statistical significance in bold

* = small effect size ($\eta_p^2 > .02$)
** = medium effect size ($\eta_p^2 > .13$)
*** = large effect size ($\eta_p^2 > .26$)

Figure 5 presents aggregated neural activity at each of the three stages of audio-visual integration (encoding, maintenance, retrieval) as a function of SOA and the number of to-be-tracked visual locations. An obvious constraint in the use of a 200 ms SOA at the time of encoding is the limited availability of neural components. Nevertheless, directly comparing the visual N1 (mean amplitude between 100 – 200 ms) response at the time of encoding between the two SOAs reveals significant main effects of presentation rate (F[1,12] = 7.41, MSE = 2.32, p = .019, ηp2 = .38), number of locations (F[1,12] = 5.99, MSE = 0.41, p = .002, ηp2 = .33) and an interaction (F[1,12] = 3.08, MSE = 0.15, p = .040, ηp2 = .20).

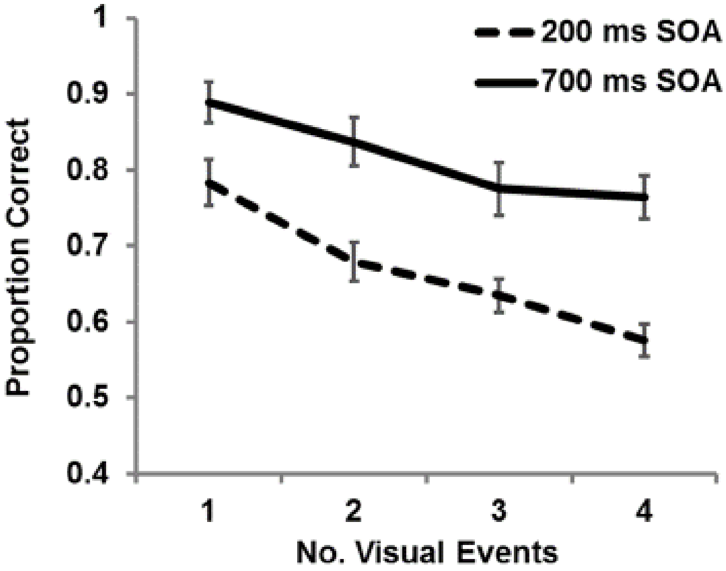


Figure 4. Response accuracy for each SOA and number of locations to be tracked in Experiment 1. Error bars represent standard error.

The interaction in the left panel of Figure 6 demonstrates a lack of statistical sensitivity to the number of visual locations in the 200 ms SOA condition. This is in contrast with the reduction of the visual N1 when participants were asked to track only one location using the 700 ms SOA relative to the other conditions (all Tukey's HSD test comparisons; *p* < .004). During the maintenance of visual locations, neural activity was characterized by a sustained negativity (500 -1000 ms) which is similar (though without the hemispheric implications) to the effect of CDA presented by Vogel & Machizawa (2004). The amplitude of this negativity was sensitive to the number of visual locations (F[3,36] = 6.42, MSE = 2.77, *p* = .001, η_p^2 = .35), with no main effect of rate of presentation, (F[1,12] = 0.26, MSE = 32.97, p = .618, η_p^2 = .02) and a trend towards an interaction (F[3,36] = 1.98, MSE = 4.13, p = .135, η_p^2 = .14) as shown in the middle panel of Figure 6.

Here, significantly larger sustained negativity was generated for trials in which participants had to retain a single visual location relative to 2 and 4 locations; this effect was only marginal when 1 location was compared to 3.

Finally, at the time of retrieval, neural activity including the N2pc and P3b (250 -600 ms) revealed increased positivity generated by the 700 ms SOA relative to the 200 ms rate of presentation (F[1,12] = 6.68, MSE = 11.53, p = .024, η_p^2 = .36), and increased positivity generated by a decrease in the number of visual locations that were tracked (F[3,36] = 16.17, MSE = 2.35, p < .001, η_p^2 = .57). Specifically, 4 locations produced significantly less positivity than all other conditions and 3 locations produced significantly less positivity than 1 location (Tukey's HSD; p < .05). There was no significant interaction (F[3,36] = 0.65, MSE = 3.97, p = .586, η_p^2 =.05).

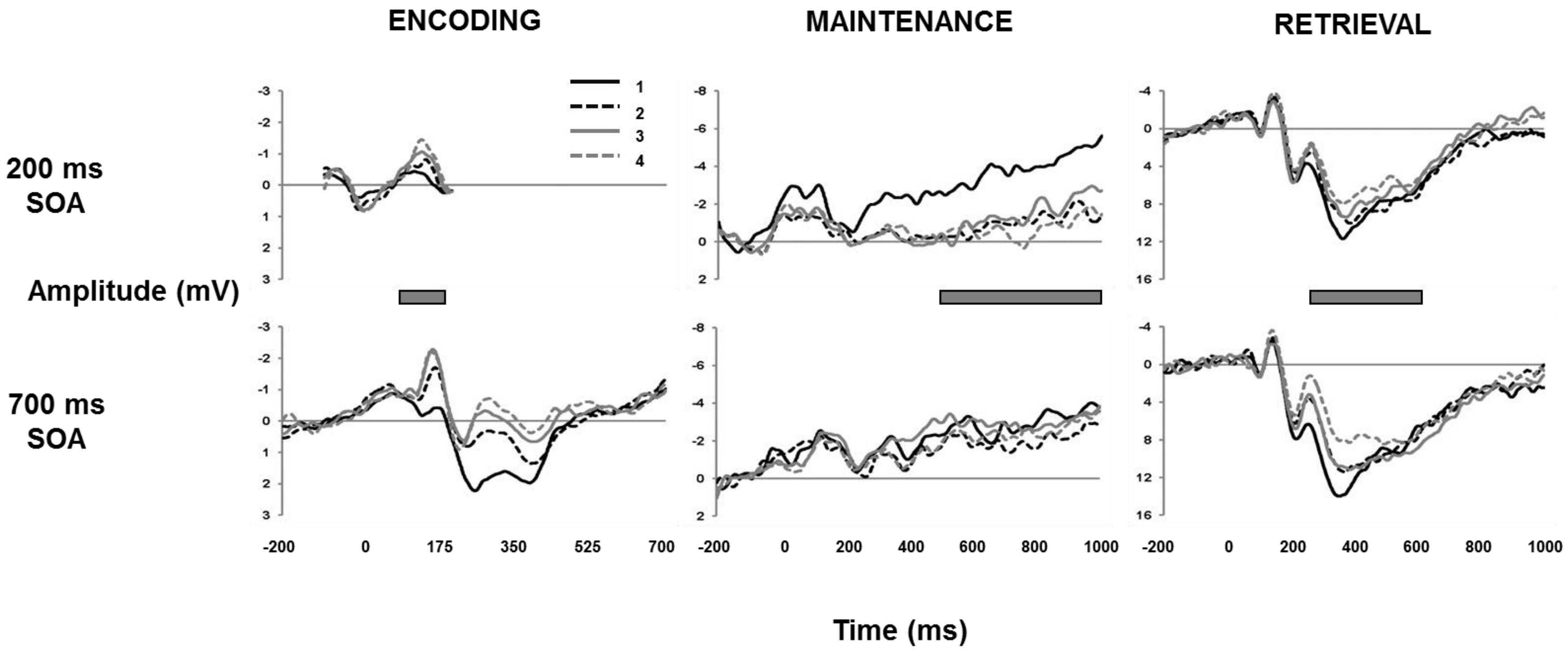


Figure 5. Group average ERP responses across the three stages of audio-visual integration as a function of SOA and the number of to-be-tracked locations.

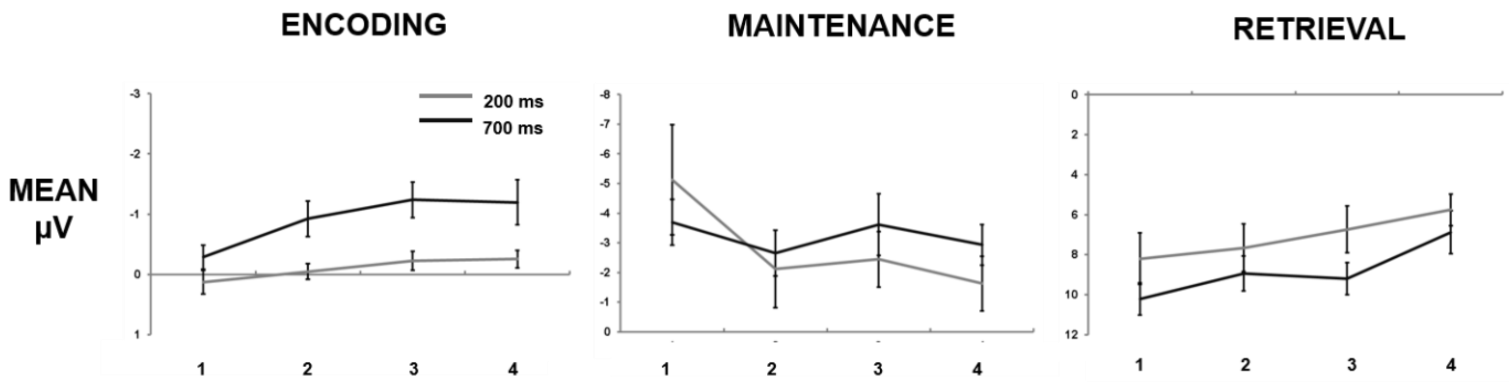


Figure 6. Mean amplitudes for the encoding (100 -200 ms), maintenance (500 -1000 ms) and retrieval (250 – 600 ms) phases.

Figure 7 summarizes the relationship between brain and behavior at each of the three stages of audio-visual integration according to SOA. Correlations were calculated between mean amplitude within the time ranges pertinent to each phase (encoding: 100-200 ms; maintenance: 500-1000 ms; retrieval: 250-600 ms) and proportion correct associated with the 4 levels of visual location, for each participant and for each stage of audio-visual integration. Average correlations across the thirteen participants were compared against zero in a series of one-sampled t-tests. The data reveal the significance of encoding (*r* = .445; *p* = .015) and retrieval (*r* = .482; *p* < .001) but not maintenance (r = -.062; *p* = .735) during 700 ms SOA presentation, and the significance of the retrieval stage (*r* = .435; *p* = .003) but not encoding (*r* = .108; *p* = .552) or maintenance (*r* = -.257; *p* = .173) during 200 ms SOA presentation. Therefore, the early correlation between brain and behavior observed in the 700 ms SOA condition appears to result from clearer perceptual information being extracted at the encoding phase. This appears to be one of the factors in driving the capacity of audio-visual integration beyond the previously suggested limit of 1.

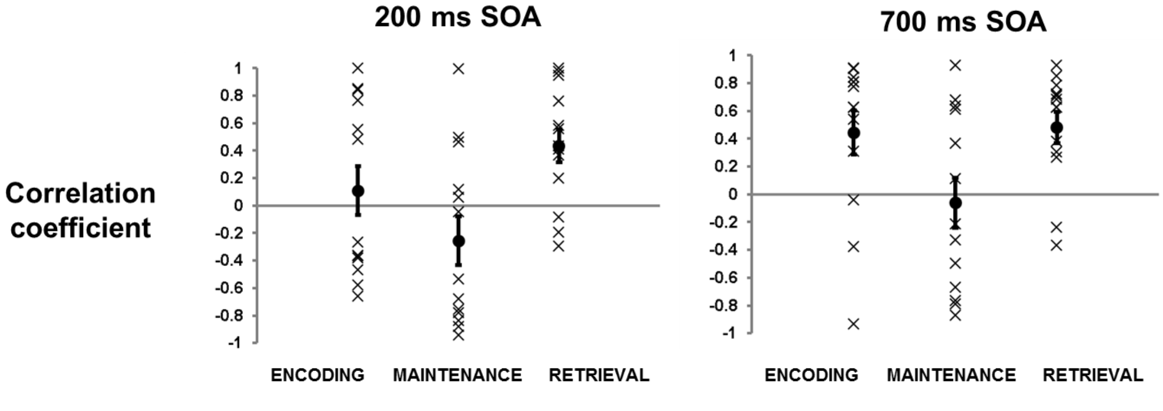


Figure 7. Individual (crosses) and group average (circle) correlation between mean amplitude brain data and proportion correct behavioural data for the four levels of visual location and the two levels of SOA. Error bars show standard error.

Discussion

The first important finding of Experiment 1 was that the capacity of audio-visual integration may exceed one item at slow (700 ms) but not fast (200 ms) SOA. By slowing down the rate of stimulus presentation (Marois & Ivanoff, 2005) we observed the average capacity of audiovisual integration to be closer to two (1.91). Based on the findings of Holcombe and Chen (2013), at 700 we might expect that as any as three items could be successfully tracked. However, there is likely more at play here than simple temporal frequency limits of object tracking. For example, Bahrami (2003) showed that the capacity for tracking features such as colours of multiple objects had a lower capacity (2, rather than 4) than tracking objects themselves. This is not to suggest that there is, necessarily, a limit of two on the capacity of audiovisual integration, but it is pertinent to the finding that participants are not reaching the limit they might be able to base on temporal factors alone. However, the findings here provide evidence contrary to the assertion that there is a "stricter, intersensory limitation such that attention is captured by only one audiovisual event at a time" (Van der Burg et al., 2013, p. 350). It is interesting to note again that their range of *K* estimates in Experiment 1c 200 ms SOA was 0.70 - 1.56, indicating that, even in their original study, some individuals may also have had an audio-visual capacity greater than 1.

In terms of the larger context of audiovisual integration, the data are particularly important as they show that the capacity of integration can exceed one when temporal and stimulus factors cannot adjudicate between visual candidates that may be bound to the auditory event: all changes occur at the same point in time and involved polarity shifts of the same nature. Both temporal and stimulus factors can affect integration. Parise and Spence (2009) showed that congruent auditory-visual pairs are more likely to be judged to be synchronous at larger time lags than were incongruent pairings, giving an example of stimulus factors affecting temporal binding. Alternately, we previously showed that temporal factors take precedence over stimulus factors when making an audiovisual binding decision (Wilbiks & Dyson, 2013a), with stimulus factors only playing a role when temporal information is ambiguous. However, in the paradigm being used for the current research, neither temporal nor stimulus factors provide any information that might help the perceiver to make a decision. All of the visual changes are simultaneous with each other and with the auditory tone, so none of them should be preferable based on temporal factors. In addition, the changing dots are randomly assigned to be black or white, and the tone is always of the same pitch, so stimulus congruency also provides no disambiguating information (although, see Experiment 6 in this series for a manipulation of congruency). In this case, with both temporal and stimulus factors being completely ambiguous with respect to audio-visual integration, it seems entirely plausible that the system should attempt to bind all of the visual stimuli with the auditory stimulus as source *candidates* – with the possibility that some post-hoc information may later emerge that might disambiguate them.

More importantly, our analysis of the time course of audio-visual integration using electrophysiology revealed that capacity limits can originate from indiscriminate brain responses at the time of encoding. While the visual N1 is sensitive to the magnitude of physical change, attentional allocation and the eventual requirement of a discriminatory response (Vogel & Luck, 2000) – three effects that may have independently contributed to N1 modulation during the encoding phase – the absence of visual N1 modulation to the number of visual locations in the 200 ms condition supports the notion that presentation rates can have severe consequences for the estimation of capacity limits (Marois & Ivanoff, 2005; Holcombe & Chen, 2013). Thus, the quality of incoming perceptual information also plays a role in defining the capacity of audiovisual integration. Additionally, an effect of numerosity can be seen to be affecting N2pc during the retrieval phase, with an increase in amplitude accompanying an increase in numerosity in the 700 ms condition, as per Mazza & Caramazza (2011). Brain-behaviour correlations for fast and slow presentation rates were equivalent during the retrieval phase, where participants were probed as to the validity of one specific location. Consistent with previous data (Van der Burg et al., 2011; Mazza & Caramazza, 2011; Verleger, Jaskowski, & Wascher, 2005), N2 decreased and P3b increased as the number of tracked locations decreased. In these respects, audio-visual integration constitutes a series of dynamic cognitive processes the limits of which may be determined by both internal (e.g. attention) and external (e.g. perceptual load) demands.

Experiment 2 – Reduced proactive interference

Introduction

In Experiment 1, it was found that by reducing visual load, slowing down SOA, and increasing temporal predictability, it was possible to increase the capacity of audio-visual binding to surpass the proposed limit of 1 item (Van der Burg et al., 2013). Having found that this capacity is dynamic, it stands to reason that audiovisual integration is likely subject to similar factors to those found in defining the quality of unimodal (visual) perception. The first of these factors to be considered is the role of proactive interference. When previously perceived information remains in working memory, it inhibits the perception of future stimuli (Crowder, 1976) (as opposed to retroactive interference, wherein subsequently presented stimuli can reduce the likelihood of remembering earlier ones). This effect has been shown to be stronger when using abstract visual stimuli such as dots, when compared to higher-level stimuli such as words or pictures (Luck & Vogel, 1997). Since proactive interference has been shown to affect visual working memory (Hartshorne, 2008) which, in turn, can affect integration, reduction of interference should lead to better integration. Given these findings, reducing the amount of proactive interference in the paradigm should lead to an increase in the capacity of audiovisual integration. As such, Experiment 2 involved reduction of proactive interference by eliminating the first 5 presentations of the stimuli. Rather than the 10 presentations we used in Experiment 1, there were only 5 presentations, such that there would be only 3 interfering presentations ahead of the critical stimulus (rather than 8). After having completed Experiment 1, revealing the neural underpinnings of audiovisual integration, the remainder of the experiments employ only behavioural research techniques. The reasons for reverting to behavioural testing only were manifold. It was a practical decision in that testing with behavioural methods is less work-intensive and more efficient than electrophysiological recording. This decision was also taken because the important contribution of the ERP components studied was in the difference between 200 and 700 ms SOAs, which was consistent across the experimental series.

Method

Informed consent was obtained from 24 participants prior to experimentation. All participants were recruited from an undergraduate research participant pool, and were compensated with partial class credit. According to the same procedure used in Experiment 1, 1 participant was rejected, so that the final sample consisted of 23 participants – 2 males and 21 females – with a mean age of 19.6 years (SD = 2.5), and a total of 21 right handed individuals. All participants self-reported normal or corrected-to-normal vision and hearing.

All stimuli were identical to those used in Experiment 1.

The design of the experiment was the same as in Experiment 1, orthogonally varying the SOA of visual stimuli (200 or 700 ms), the validity of the probe stimulus (valid or invalid), and the number of visual stimuli that changed on each alternation (1, 2, 3, or 4). The critical difference here was that the amount of proactive interference was reduced from 8 presentations to 3 (in addition to the critical trial, and the following trial; see Figure 2). These 16 conditions were each presented 3 times to create an experimental block with 48 trials. Each participant completed one practice block (16 trials), and 6 experimental blocks, for a total of 288 experimental trials. The reason for the reduction of trials was strictly a practical one. While Experiment 1 employed EEG recording, along with the setup and cleanup of participants, 384 trials fit into a 2 hour testing timeslot. Without EEG, it was decided to to reduce the number of trials in order to fit the experiment into a one hour time slot. This 288 experimental trial level was maintained through the rest of the experiments in Chapter 1.

Data were modelled according to the same procedure as in Experiment 1. Successful model fit was confirmed by the low RMSEs observed in both the 200 ms SOA (range 0.003 - 0.053) and 700 ms SOA (range 0.001 – 0.059) conditions.

Results and Discussion

K values for each SOA were compared with the previously proposed capacity limit of 1 by means of single sample t-tests. While the capacity for the 200 ms condition was, as expected, less than 1 (0.590 [range 0.296 – 1.093]; t(22) = -8.94, p < .001), the capacity for the 700 ms condition was not significantly greater than 1 (1.042 [range 0.341 – 1.880]; t(22) = 0.44 p = .661). This finding was unexpected, as it was predicted that reduction of proactive interference would increase, not decrease, capacity of audiovisual integration. Response accuracy across the 8 conditions (SOA x 2, objects x 4) will be examined below in order to look deeper into the phenomenology behind this decrease in capacity[2]. One possible reason for this reduction is not

[2] A full discussion of proportion correct data for Experiment 1 – 4 is presented later in this chapter. For experiment 2 – 4 there will be no discussion of proportion correct data within each individual experiment, but this discussion is presented and required for Experiment 1, because these data will be used in comparisons and correlational analyses with regard to the electrophysiological data that was collected in Experiment 1.

an effect of the removal of proactive interference, per se, but rather an effect of the removal of early stimulus presentations that gave a participant important information as to the number of objects to be tracked. Ma and Flombaum (2013) showed that performance is hindered in multiple object tracking when participants are not aware of how many items are to be tracked. I believe that a similar effect was occurring here, where the removal of the first 5 presentations meant that participants had less time to prepare for the task in terms of knowing how many items to track. Another possibility is that the overall trial length was shorter in this condition. Experiment 1 consisted of 10 total presentations of either 200 or 700 ms (total trial length = 2000 or 7000 ms), whereas Experiment 2 was half the length (total trial length = 1000 or 3500 ms). In the same way as information about the number of items to be tracked is reduced in this experiment, so might the shorter trial length result in less time to prepare for the response. Conversely, it is important to consider the findings of Oksama and Hyönä (2004), who found that participants are only able to reliably track 4 targets for a 5 second experiment, with the tracking capacity dropping to 3 once the experiment was extended to 9 seconds. According to these findings, it should be expected capacity should be greater in an overall shorter experiment (although it should also be noted that the paradigm used in the current research is still relatively short compared to the times tested by Oksama and Hyönä (2004)). Concerns of this nature could be tested in an future experiment where participants are given a numerical cue (1, 2, 3, or 4) on the screen such that they will know how many items to track regardless of amount of interference.

Experiment 3 – Decreased temporal predictability

Introduction

Experiment 2 considered the effects of proactive interference on the capacity of audiovisual integration and found that, surprisingly, capacity was not significantly greater than 1 with a low level of proactive interference, even in the 700 ms SOA condition. A second factor that was predicted to impact audiovisual integration capacity was the degree of temporal predictability of the crucial stimulus. Knowledge of when the critical trial is coming is important in being able to integrate auditory and visual information successfully. Wasserman, Chatlosh, & Neunaber (1983) used fixed or variable trial lengths in testing perception of a light appearing at the end of the trial, and found that performance was decreased in the variable trial length condition. Participants were found to be faster at perceiving the light when the trial was always of the same length. Additionally, presentation of an alerting stimulus ahead of the critical stimulus has been shown to increase response speed and accuracy (Fan et al., 2002). When an individual is aware that an important stimulus is forthcoming, they can prepare to attend to it, and in so doing increase their likelihood of perceiving that stimulus accurately. Taken together, these studies indicate that when visual stimuli are relatively high in temporal predictability (with two, unblocked SOAs, there is still some degree of variation in critical stimulus presentation timing), they are more accurately perceived by an individual. For the purposes of the current research, this would mean that if the critical presentation of the stimuli occurs at a time that can be predicted, there is a greater likelihood of perceiving (and subsequently, integrating) that presentation with the tone that it is presented in synchrony with.

In order to allow a visual stimulus to be bound with an auditory stimulus it must be, to some extent, attended to. Talsma, Senkowski, Soto-Faraco, and Woldorff (2010) provide a review of the relationship between multisensory integration and attention. They put forth a theory that attention works in both a top-down and bottom-up direction in influencing multisensory interactions. A temporally synchronized tone can have a bottom-up influence on multisensory integration which makes a visual stimulus appear to 'pop out' of a display of multiple stimuli (Van der Burg et al., 2008; Fujisaki, Koene, Arnold, Johnston, & Nishida, 2006). Alternatively, they say that when multiple stimuli are in competition for processing, top-down control may be required to process them effectively and allow them to be candidates for integration. This top-down attention is said to be more necessary when a secondary task is included, which takes attention away from the task at hand. It is also needed when overall perceptual and attentional load exceeds the capacity of an individual.

In Experiments 1 and 2 the critical stimulus always occurred at the penultimate visual presentation. To examine the effects of temporal predictability, in Experiment 3 the critical stimulus was less predictable, and could occur with equal probability on the 7th, 8th, or 9th presentation. Here, the prediction was that this would decrease performance, as it supplies the participant with less predictability in maintaining attention than in Experiments 1 and 2. While the overall length of a trial was the same, in order to respond correctly participants had to be attending through the 7th, 8th, and 9th presentation rather than only at the 9th presentation. Having to attend to the dots throughout a range of presentations that could be the critical stimulus increases the temporal range through which attention must be sustained, which should decrease the strength of the attention that a participant can apply. An additional reason for this manipulation is to test whether this unimodal factor will behave as expected in an audiovisual integration task. Experiment 2 did not have results as expected based on unimodal manipulations, so it is of interest to determine whether the same will be true for Experiment 3.

Method

Informed consent was obtained from 24 participants prior to experimentation. All participants were recruited from an undergraduate research participant pool, and were compensated with partial class credit. Participants were rejected based on the same criterion as was used in Experiments 1 and 2. A total of 6 participants were rejected in this way, so that the final sample consisted of 18 participants – 2 males and 16 females – with a mean age of 19.7 years (SD = 4.8), and a total of 16 right handed individuals. All participants self-reported normal or corrected-to-normal vision and hearing. None of the participants had taken part in Experiments 1 or 2.

All stimuli were identical to those used in Experiments 1 and 2.

The design of the experiment was the same as in Experiment 1, orthogonally varying the SOA of visual stimuli (200 or 700 ms), the validity of the probe stimulus (valid or invalid), and the number of visual stimuli that changed on each alternation (1, 2, 3, or 4). In this experiment, we included a variation in the timing of the critical stimulus. Rather than always occurring with the 9th presentation of visual stimuli, it could now occur at the 7th, 8th, or 9th presentation. The number of trials remained the same as in the previous experiments, but these trials were now split equally across the three critical stimulus timings. Each participant completed one practice block, and 6 experimental blocks, for a total of 288 experimental trials.

Data were modelled according to the same procedure as in Experiment 1. Successful model fit was confirmed by the low RMSEs observed in both the 200 ms SOA (range 0.001 - 0.056) and 700 ms SOA (range 0.001 – 0.059) conditions.

Results and Discussion

Comparing the K values to 1 as in the previous experiments, it was found that, again, capacity in the 200 ms condition was significantly less than 1 (0.529 [range 0.180 – 0.885]; t(17) = -9.36, p < .001), but that the 700 ms condition was again not significantly more than 1 (0.967 [range 0.241 – 2.187]; t(17) = -0.29, p = .772). This lack of capacity increase is easily explained as the introduction of a temporally roving critical stimulus (low temporal predictability) was a manipulation intended to make the task more difficult. As in the previous literature, taking away the predictability of the critical stimulus decreases performance, which in this case is indexed by a reduction in K (relative to Experiment 1), the capacity of audiovisual integration. Just as higher visual load, as found in Van der Burg et al. (2013) experiments, resulted in a capacity of audiovisual integration that was limited to less than one item, the inclusion of a temporally roving critical stimulus also limited the capacity to less than one item. Experiment 3 in the current series is the nearest analogue to Van der Burg et al's (2013) experiments, with a temporally roving critical stimulus and a high level of proactive interference. These observations in Experiment 3 raise questions about the interpretation of Experiment 1, in that participants may not need integration when the critical stimulus is predictable – that they are able to count the number of presentations and then attend to the critical stimulus only. However, participants weren't able to simply ignore (e.g. by closing their eyes) the early presentations because the SOA was mixed within each block. As such, they would not know whether to expect the critical stimulus at 1800 ms or 6300 ms from the beginning of the trial, even if the critical stimulus was fixed in terms of which presentation it came on.

Experiment 4 – Reduced proactive interference and decreased temporal predictability

Introduction

Experiments 2 and 3 revealed that it is possible for multiple factors to influence the capacity of audiovisual integration (although not always in the expected direction). Experiment 2 reduced the amount of proactive interference, from 8 pre-critical stimulus presentations to 3, and found that, surprisingly, capacity was decreased when interference was reduced. Experiment 3 used a high interference paradigm but with a temporally unpredictable (roving) critical stimulus. Here, it was found that capacity was numerically greater when the critical stimulus was predictable (Experiment 1) rather than unpredictable (Experiment 3), and this will be formally tested later in this chapter. In terms of multisensory integration in general, there is evidence to suggest that there are multiple factors that influence integration, such as temporal (e.g. van Wassenhove, Grant, & Poeppel, 2007), spatial (e.g Slutsky & Recanzone, 2000), and congruency (e.g. Gallace & Spence, 2006) variation. Studying interactions between these factors has not always yielded simple additive effects; that is, if two factors are tested that both increase the likelihood of integration, it is not always the case that the facilitative effect is equal to the sum of the effects of each factor on its own. Sandhu and Dyson (2013) considered the effects of task and modality switching on AV processing, both on their own and in combination. They found that the respective costs of task and modality switching when occurring alone did not simply add onto one another, but rather that there was a sub-additive pattern wherein the cost was less than the sum of the two independent costs. The potential reason given for this sub-additivity is that reorienting (with cost) through one type of switching may reduce the cost of the other type of switching. In the present research, we also have two factors (proactive interference and temporal predictability), each of which have effects on audiovisual integration capacity. It is possible that some task demands specific to one of the two factors may actually decrease the difficulty posed by the other factor. In that sense, we may also see a pattern of subadditivity here. Having looked at reduced proactive interference and temporal predictability each in isolation, we now wish to examine their combined effects on capacity. By comparing findings from Experiment 1, 2, and 3, we were able to generate predictions of what we should see in Experiment 4 *if* the effects of these two factors are additive. Any deviation from these predictions can be seen as an interaction between factors.

Using data from Experiment 1 and 3, and by subtracting the means at each SOA and number of items changing, we were able to determine the effect of having a temporally roving critical stimulus. In the same way, by subtracting data from Experiments 1 and 2, we found the effect of proactive interference. Combining these two effects, with reference to the reduced interference, roving condition, we were able to establish an expected value for each SOA and number of objects to be tracked for Experiment 4, which are displayed in Figure 8. If the individual effects of temporal predictability and proactive interference are additive, then we would expect to see an ever-increasing disparity between the 200 and 700 ms SOAs as the number of visual events to be tracked increases (as shown in the bottom left panel of Figure 8).

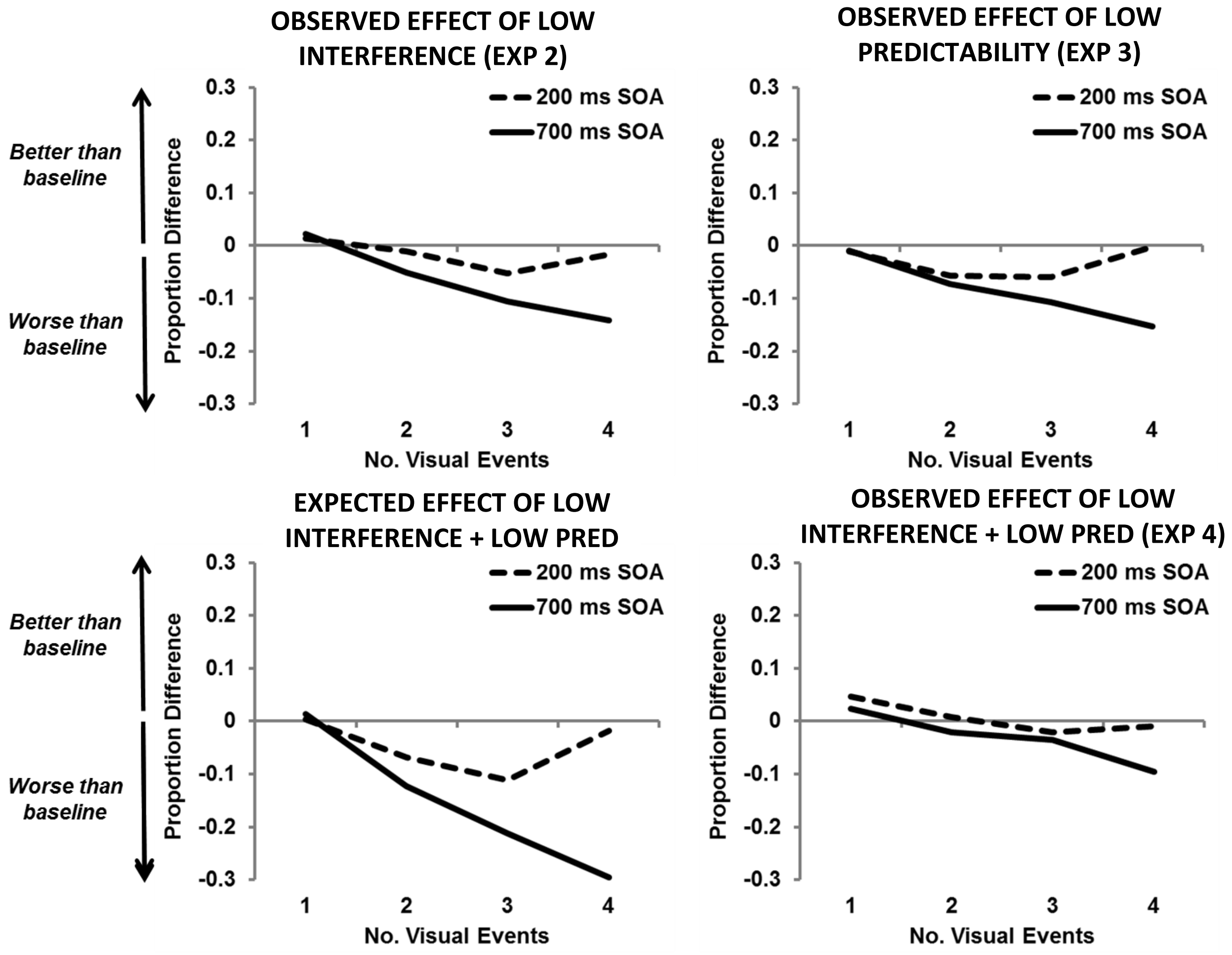


Figure 8. Observed effects of interference and predictability, as well as expected and observed effects of the combination of factors, yielded by subtracting proportion correct across Experiments 1-4.

Method

Informed consent was obtained from 23 participants prior to experimentation. All participants were recruited from an undergraduate research participant pool, and were compensated with partial class credit. A total of 4 participants were rejected based on their response criterion, so that the final sample consisted of 19 participants – 14 females – with a mean age of 23.4 years (SD = 6.7), and a total of 17 right handed individuals. None of the participants had taken part in any of the previous experiments.

All stimuli were identical to those used in Experiments 1, 2, and 3.

Again, the design here followed with the previous experiments, orthogonally varying the SOA of visual stimuli (200 or 700 ms), the validity of the probe stimulus (valid or invalid), and the number of visual stimuli that changed on each alternation (1, 2, 3, or 4). Experiment 4 also included both of the changes tested independently in Experiments 2 and 3. That means we had only 5 presentations, and that critical stimulus could occur at the 2nd, 3rd, or 4th of these presentations. Within each block, there were 2 (SOA) x 2 (validity) x 4 (number) x 3 (critical stimulus timing) trials, for a total of 48 (validity and critical stimulus timings were not considered in the analysis). Each participant completed one practice block, and 6 experimental blocks, for a total of 288 experimental trials.

Data were modelled according to the same procedure as in Experiment 1. Successful model fit was confirmed by the low RMSEs observed in both the 200 ms SOA (range 0.003 - 0.030) and 700 ms SOA (range 0.002 – 0.039) conditions.

Results and Discussion

K values for each SOA were compared by t-tests with the norm of 1. For the 200 ms condition, K was significantly less than 1 (0.658 [range 0.261 – 1.258]; $t(18) = -4.85$, $p < .001$), and for the 700 ms condition K was significantly greater than 1 (1.342 [range 0.317 – 2.562]; $t(18) = 2.49$, $p = .023$). This lends additional support to the finding from Experiment 1, that it is possible for the capacity of audiovisual integration to exceed 1 item. It is, however, unexpected that we find that this combination of two factors which, in isolation, reduce capacity leading to an overall increase in capacity. In the same vein as Sandhu and Dyson (2013), perhaps some element of one variable factor mitigates the negative effects of the other factor. While this is strictly speculative at this point, and would merit testing more formally, it seems sensible that a reduction in proactive interference (and along with it, a reduction in number of non-critical presentations) may increase perceptual vigilance at the potentially critical trials when there is a temporally roving critical stimulus. This potential explanation is supported by the work of Kane and Engle (2000), who investigated the effects of proactive interference and attention on working memory capacity. They found that high proactive interference limited participants' ability to successfully allocate attention to the task at hand. This could be tested by employing a condition with high proactive interference, but with a stimulus presented before the range in which a critical stimulus may occur. For example, a total of 10 presentations could be employed, but with a non-specific tone presented before the 7th stimulus (with presentations 7, 8, and 9 being equally likely to be the critical stimulus). In such an experiment, participants would have experienced a high level of proactive interference, but with a vigilance increase at the potentially critical stimuli. In the current Experiment 1, however, the high interference condition shows a high level of response accuracy and a high capacity. In light of our findings, it is possible that our high proactive interference condition reduced the degree to which participants could attend to the changing of dots, and specifically the degree to which they could remain vigilant for an extended period (with potential critical presentations on one of three switches rather than on one switch only). This maintenance of attention over a lengthy time span was not required in Experiment 1, when the critical stimulus always occurred on the same presentation of dots. When proactive interference was reduced, however, participants had a greater ability to allocate attention, and were not hindered to the same degree by temporal unpredictability.

Looking back at the predicted and observed values for Experiment 4 (displayed in Figure 8), it is clear that the predicted values have underestimated performance. This suggests that the effects of temporal predictability and proactive interference are not additive, but rather are two separate factors that influence the capacity of audiovisual integration to their own extent. Further to this argument, looking at the capacity values in the 700 ms condition, we see no variation based on roving in the low interference condition, but significant variation based on roving in the high interference condition. Roving is only having an effect in the presence of high interference, which means we do not have a simple additive relationship between these factors, but rather a dynamic system of factors (including ones we have not directly manipulated, such as knowledge of number of targets) that contribute to the capacity of audiovisual integration.

Between Experiments Comparisons

To compare capacity measures between the four experiments, K estimates were compared across Experiments 1-4. This analysis included 16 participants from the same sample of 25 as used in Experiment 1 above, with 9 participants rejected on the basis of their response criterion, but not due to the quality of their ERP signal. These 16 participants had a mean age of 18.5 years (SD = 1.1), 15 right handed individuals and 13 females. The proportion correct data for each SOA and number of items to be tracked are displayed, along with Experiment 2-4, in Table 2. Based on fitted data, the capacity of audiovisual integration was calculated for both SOAs.

The K values from each experiment, for each SOA, were submitted to a mixed ANOVA with between-subjects factors of Interference (2; high level, low level) and Roving (2; high predictability, low predictability), and a within-subjects factor of SOA (2; 200 ms, 700 ms). The full results of this ANOVA can be found in Table 3, and graphically in Figure 9, and important findings are discussed in detail here. There was a main effect of SOA ($p < .001$), with the capacities at 700 ms found to be significantly greater than those at 200 ms. There was a significant interaction between Interference x Roving ($p = .020$), which was probed further with a Tukey's HSD (*p* < .05). The only significant difference within the interaction was between the interference-predictable condition (Experiment 2) and the interference-non-predictable condition (Experiment 4), indicating that audiovisual integration capacity is significantly less when the critical stimulus occurs predictably rather than non-predictably, but only when there is a high level of proactive interference. Considering numerical values of K for each of these conditions, it seems that the presence of both interference and roving leads to the smallest capacity of integration, although this difference is only significant in the comparison described above.

Table 2 Summary of the three-way mixed model ANOVA comparing estimates of K across Experiments 1-4 (all dfs 1,71)

Metric	F	MSE	*p*	η_p^2
Roving (R)	1.82	.402	.182	.025
Interference (I)	0.49	.402	.486	.007
SOA (S)	**137.07**	**.119**	**<.001**	**.659**
R x I	**9.66**	**.402**	**.003**	**.120**
R x S	3.46	.119	.067	.046
I x S	3.07	.119	.084	.041
R x I x S	**15.23**	**.119**	**<.001**	**.177**

Note: Statistical significance in bold

* = small effect size ($\eta_p^2 > .02$)
** = medium effect size ($\eta_p^2 > .13$)
*** = large effect size ($\eta_p^2 > .26$)

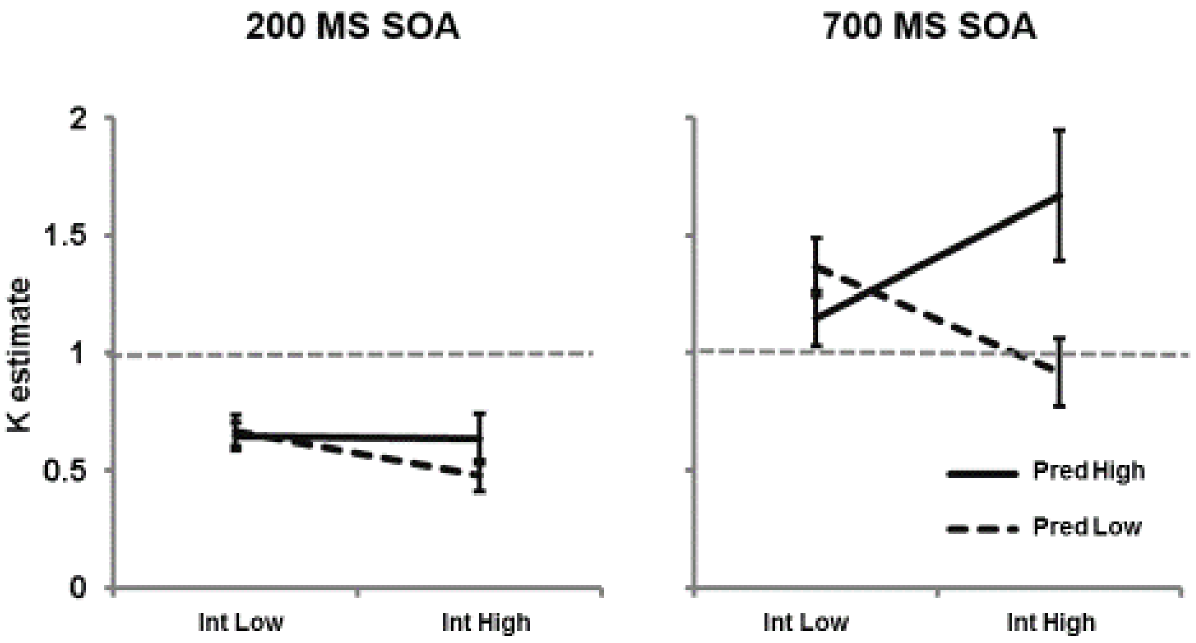


Figure 9. Interaction between capacity estimates by Interference (INT) x Predictability (PRED) x SOA. Error bars show standard error.

There was also a significant interaction between Interference x Roving x SOA (*p* = .040). Tukey's HSD revealed that K for the high interference, high predictability, 700 ms condition was the highest (K = 1.76, SE = .178), and that this was significantly greater than all 200 ms conditions, as well as the high interference, low predictability, 700 ms condition (K = .968, SE = .178). So when the perceptual load is decreased (via a slow SOA), proactive interference is present, and the critical stimulus is temporally fixed, the capacity of audiovisual integration is maximized. Of these factors, only the increase in capacity with proactive interference remains anomalous. However, as discussed before this is likely due to a separate phenomenon caused by the reduction of interference, namely information about the number of items to be tracked.

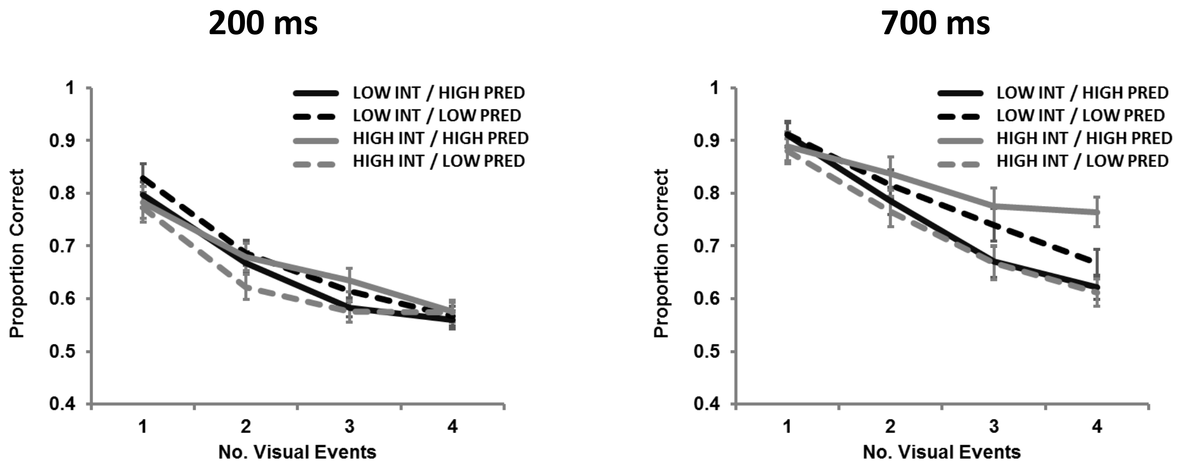
Another interesting finding within this three-way interaction is that there is no modulation of K within the 200 ms SOA condition (all *p*s > .95), which is consistent with the findings of Van der Burg et al. (2013). This lack of modulation based on the factors we have manipulated suggests that there is a barrier that does not allow the capacity of audiovisual integration to exceed 1 item when stimuli are presented 200 ms apart. In the discussion of Experiment 1, based on the correlation between ERP and behavioural data, I propose that there is a need for high quality information during encoding, and that at the 200 ms SOA there is a sensory limitation that does not allow this to occur.

The data are instructive in confirming that performance during 200 ms SOA was insensitive to experimental manipulation. While this would suggest the impenetrability of audio-visual integration capacity during fast presentation, the data from the 700 ms SOA condition clearly show capacity modulation and estimates that exceed 1. Therefore, the data underscore the importance for later experiments to consider task difficulty as a large scale influence of AV capacity, indexed by a relatively complex interaction between SOA, the degree of proactive interference, temporal roving of the critical frame, and perceptual load, to name but 4 factors. For the moment I suggest that these environmental designs represent intermediate levels of task difficulty allowing AV capacity to rise above 1. Specifically, audiovisual integration capacity is facilitated under 700 ms SOA when the combination of high temporal predictability and high proactive interference, or low temporal predictability and low proactive interference are present. Such cases stand in contrast to paradigms employing fast rates of presentation, temporal roving of the critical frame, and large degrees of proactive interference, which tend to generate conservative estimates of audio-visual integration capacity.

Cross-Experimental Analysis of Proportion Correct Data

While the main interest of this experimental series is to examine the factors that influence the capacity of audiovisual integration, it was also of interest to examine the proportion of correct responses supplied by participant in each of the conditions employed. This analysis will be instructive as to whether there were differential effects of the manipulated factors on the trials when there were 1 (or 2-4) items changing. It was expected that, in general, proportion correct would be lower in the 200 ms SOA as compared to the 700 ms SOA, and that it would decrease with an increase in number of events to be tracked. More specifically, as the task becomes of increasing difficulty (e.g. more locations to be tracked) there should be an increase in the facilitative effect of other factors, such as temporal predictability and proactive interference. So being of a moderate difficulty level (as discussed above) should exhibit the most facilitation under conditions where there are 4 locations to be tracked.

The proportion correct data from all 4 experiments were submitted to a mixed ANOVA with between-subjects factors of Interference (2: high, low) and Roving (2: high predictability, low predictability), and within-subjects factors of SOA (2: 200 ms, 700 ms) and Number of locations to be tracked (4: 1, 2, 3, 4). The full results of this ANOVA are displayed in Table 3, and graphically in Figure 10. There was a main effect of SOA ($p < .001$), with improved responding in the 700 ms condition than in the 200 ms condition. There was also a main effect of Number ($p < .001$), with each additional item to be tracked decreasing performance significantly. A trend towards an SOA x Number interaction ($p = .061$) revealed that while in the 700 ms condition there was incremental decreases in performance with added items to be tracked, in the 200 ms condition this was only true up to 3 items, with the 4th item not having added decrement beyond the 3rd. This follows well with the account of a sensory barrier, and with the finding that there was no modulation of K in the 200 ms condition while there was in the 700 ms condition.


200 ms

700 ms

Figure 10. Proportion correct for each combination of temporal predictability, proactive interference, and number of locations for 200 ms SOA (left panel) and 700 ms SOA (right panel) for Experiments 1-4. Error bars indicate standard error.

Table 3 Summary of the four-way mixed model ANOVA comparing raw proportion correct data across Experiments 1-4 (all dfs 1,71; apart from main effects and interactions with events [E] dfs 3,213)

Metric	F	MSE	*p*	η_p^2
Roving (R)	0.60	.053	.441	.008
Interference (I)	0.01	.053	.936	.001
SOA (S)	**195.19**	**.009**	**<.001**	**.733*****
Events (E)	**252.83**	**.006**	**<.001**	**.781*****
R x I	**5.39**	**.053**	**.023**	**.071***
R x S	1.41	.009	.239	.019
R x E	0.97	.006	.406	.014
I x S	1.44	.009	.234	.020*
I x E	**4.19**	**.006**	**.007**	**.056***
S x E	2.37	.005	.072	.032*
R x I x S	**4.39**	**.009**	**.040**	**.058***
R x I x E	**3.12**	**.006**	**.027**	**.042***
R x S x E	1.15	.005	.331	.016
I x S x E	0.25	.005	.860	.004
R x I x S x E	**4.19**	**.005**	**.007**	**.056***

Note: Statistical significance in bold.

* = small effect size (η_p^2 > .02)
** = medium effect size (η_p^2 > .13)
*** = large effect size (η_p^2 > .26)

A significant Interference x Number interaction (*p* = .001) showed that in high and low interference conditions there were different effects stemming from number of items to be tracked. While there was also no difference between interference conditions at any number of items to be tracked, performance was marginally better under conditions of no proactive interference when the number of tracked events was small (1 or 2) but that performance was marginally better under conditions of proactive interference when the number of tracked events was large (4; c.f., Yantis, 1992), but none of these pairwise comparisons were statistically significant.

Finally, there was a 4-way interaction between Interference x Roving x SOA x Number (*p* = .007). This interaction was probed by means of Tukey's (*p* < .05) post-hoc tests, and analyses were focused on the relationships between conditions that were present within a specific number of items to be tracked. When there is one item to be tracked, we see SOA effects at every combination of interference and roving, as well as a difference between the high interference, low temporal predictability, 200 ms condition and both low interference 700 ms condition points. We also see SOA effects at each interference/roving combination in the 2-change and 3-change conditions, and only in the low interference/low predictability and high interference/high predictability conditions in the 4-change condition. Beyond SOA effects, we see an effect of temporal predictability in the 2-change, high-interference condition, as well as in the 3-change, high-interference condition. To generalize these findings, it can be said that intermediate difficulty levels lead to the highest levels of integration capacity. Specifically, only in these intermediate difficulty conditions does SOA have an impact on response accuracy when four locations are changing – the hardest level of that factor – and therefore it can be concluded that it is under those conditions that capacity could be maximized.

General Discussion

Across four experiments, we have established some conditions under which the capacity of audio-visual integration may exceed 1. Specifically, visual set size should be low (comparing the current research to that of Van der Burg et al., 2013; Lavie, 2005) and stimulus change should operate at a slow rather than fast rate of presentation (Marois & Ivanoff, 2005). Capacity can go beyond 1 when there is temporal roving and low proactive interference (Experiment 4) or no temporal roving with high proactive interference (Experiment 1). Neither of these contributions predict increased AV capacity in isolation and so it is likely that intermediate task difficulty provides the appropriate levels of arousal for successful performance (Anderson, 1990; and for a similar example in the context of multitasking, see Adler & Benbunan-Fich, 2014). The single paradigm feature that does seem to be necessary for high AV capacity is the use of relatively slow (700 ms) compared to relatively fast (200 ms) SOA. While the impenetrability of audio-visual integration at 200 ms SOA might suggest some form of limit, the possibility remains that this is a data limit rather than capacity limit (Norman & Bobrow, 1975). This explanation is supported by the electrophysiological data and particularly the analysis of visual N1 during pre-critical frame presentations (Experiment 1). Here, visual cortex was insensitive to the number of polarity changes per frame in the 200 ms SOA condition but sensitive to similar changes during the 700 ms SOA condition. In this instance, the indiscriminate brain response during fast rates of presentation is taken to reflect poor quality sensory information entering working memory, and the failure of the brain to complete an initial tracking task that is a prerequisite for successful performance in the task.

It is probably also worth restating that despite their original claim, in the original Experiment 1c of Van der Burg et al. (2013), their range of *K* estimates was 0.70 - 1.56, indicating that some of their individuals also exceeded 1 (see also their Experiment 2, discussed below). The current data replicate the observation that certain individuals expressed capacity beyond 1, to 2 and even 3 (see Figure 3). To defend the position that ‘the capacity of audio-visual integration is limited to one item’ when there is data that some participants, at least on some of the trials, were able to bind two (or more) visual locations to a single auditory source appears contradictory. At the very least, the ranges cited above raise the clear need to further study individual differences in audio-visual capacity, in much the same way as it has received attention in the context of visual short term memory (e.g., Drew, Horowitz, & Vogel, 2013).

In light of criticisms put forward by Van der Burg (personal communication, 1 May 2015; See Appendix E), it is worthwhile to spend some time discussing three potential objections to the current data. First, the data of Van der Burg et al. (2013; Experiment 2) provide surface evidence against the idea that the reason why AV capacity cannot exceed 1 under 200 ms SOA conditions is due to the inability to successfully code the number of changing locations in frames prior to the critical one. Here, they show that under visual-only conditions running at an SOA comparable to the current research (150 ms), *K* was estimated to be around 3.34, whereas in an audio-visual condition presented at the same fast speed, *K* comes in around 0.78 (range = 0.30 – 1.36). A primary reading of the data would suggest that capacity was 3 when the task was visual-only, but capacity could not exceed 1 when the task was audio-visual. This apparently shows then under visual-only conditions, participants can track, on average, at least 3 locations. However, the comparison between the visual-only and audio-visual conditions is not an appropriate one. Specifically, the signal for the critical frame in the visual condition was marked by a unique color change (from white/black to green) at *specific dot locations*. The use of such a salient colour cue is likely to have given rise to perceptual pop-out (a location changed, and was made green; Treisman & Gelade, 1980) at target locations during the critical frame, additionally meaning that it would have been unnecessary for participants to track location changes in frames prior to the critical one. I believe these effects yield the high *K* in the visual condition and therefore do not support the idea that multiple locations can be tracked during particularly fast rates of presentation. Moreover, the lack of cueing to specific target locations in the auditory condition undoubtedly contributes to the observation of lower *K* in that condition. I also note that in the same paper, their previous Experiment 1d uses a non-location specific visual cue which was more comparable with the auditory case, by the authors own admission: "a cue that, like the sound cue, was not specific to any of the items" (Van der Burg et al., 2013, p. 349). Under these conditions in which the comparison between visual-only and audio-visual performance was more valid (both location non-specific), performance in the visual-only condition was poor (*K* = .56).

Second, there may remain opposition to idea that audio-visual integration capacity may exceed 1 since it only apparent during a slow (700 ms) rate of presentation. There are a number of responses to this, foremost the lack of evidence suggesting that qualitative changes in audio-visual binding should arise as the result of the manipulation of a continuous variable such as SOA: in contrast to the claim that 700 ms is 'too slow' to allow for integration across the senses there is evidence that neurons in the superior colliculus are sensitive to audio–visual integration at an asynchrony of 600 ms (Calvert & Thesen, 2004, cited in Koelewijn, Bronkhorst, & Theeuwes, 2010). Furthermore, it is not clear that faster rates of presentation (such as 150 and 200 ms) provide an error-free index of AV capacity. Van der Burg et al. (2013) note that slowing the rate of presentation from 150 ms to 200 ms SOA improved AV capacity probably as a result of "the reduced likelihood of misbindings" between auditory and visual events (Van der Burg et al., 2013 , p. 348). Given this significant increase in performance as a result of slowing SOA by

50 ms, there can be little surprise that audio-visual capacity exceeds 1 with further extension. Furthermore, these findings lead a move away from the possibility that the capacity of audio-visual integration is stuck at 1 due to spuriously poor performance caused by a high degree of illusory audio-visual binding. It is possible to empirically test the idea of illusory binding in future studies (see discussion of Experiment 6). Given the temporal preference for auditory-lag rather than auditory-lead in binding sound with vision (e.g., Vroomen & Keetels, 2010; Wilbiks & Dyson, 2013a) one prediction would be that under fast rates of presentation, participants incorrectly bind to the preceding rather than current visual frame. Therefore, there should be an increased number of responses that are ‘incorrect’ in accordance with the critical frame but ‘correct’ in accordance with the frame preceding it. Objections to slow delivery rates would also appear to confuse trial SOA with the importance of the degree of temporal separation between visual and auditory event at the critical frame. That is, during the critical frame, both visual and auditory on-sets occur simultaneously (0 ms difference), falling within the typical temporal window of integration required for audition and vision (e.g., van Wassenhove, Grant, & Poeppel, 2007). Kawachi, Grove, and Sakurai (2014) did not model capacity directly, but found that a single auditory tone can affect the perception of two visual stimuli. Their findings demonstrate that the delivery of auditory and visual events within a shared window of integration appears to be an essential characteristic if one wishes to associate a single auditory event to multiple visual events. Finally, if the mechanisms underlying performance in the present task qualitatively change at 700 ms SOA, then we would expect an exponential increase in capacity at 700 ms SOA, relative to the increases observed between 200 ms SOA and an intermediate 450 ms SOA (see Experiment 7).

A third and final objection to audio-visual integration capacity exceeding 1 may originate in an appeal to ecological validity, and the observation that in the real world unique sounds tend to have a single (visual) source. To wit: "From an ecological point of view, it would make sense to bind only one visual event to a specific sound. In natural scenes, individual, object-related sounds (unlike the sound of the wind or a babbling brook) come from a single source…" (Van der Burg et al., 2013, p. 345-346). Leaving aside the examples in the above quote, one might imagine a relatively large-scale auditory scene, such as an orchestra, where multiple visual events (e.g., violin section) give rise to a specific, streaming sound. Having shown the effects of the basic factors of temporal predictability and proactive interference, the next chapter will examine the effects on audiovisual integration capacity of stimulus factors such as perceptual chunking and crossmodal congruency will be examined. This will extend what is known about the interaction of features that lead to an increase or decrease in audiovisual integration capacity. For the moment, however, the data in Chapter 1 support the contention that the capacity of audio-visual integration is a dynamic process and reveal the environmental conditions and concomitant brain states under which it need not be limited to 1.

Chapter 3 – Stimulus-based effects on audiovisual integration

General Introduction

In the four experiments presented in Chapter 2, the paradigm of Van der Burg et al. (2013) was decomposed, and results showed that it was possible for the capacity of audiovisual integration to exceed one item, given the correct set of stimulus parameters. Now that it is clear that the capacity of audiovisual integration is malleable, it is of interest to consider if other stimulus-based factors that were not manipulated in Van der Burg's (2013) paradigm can also modulate capacity. Cross-modal congruency and perceptual chunking are both factors that have previously been shown to affect levels of unimodal perception and of audiovisual integration, and they will be employed to test their effects on the capacity of audiovisual integration in this chapter.

Within the 'biased competition' framework of attention (Desimone & Duncan, 1995), our attention is based on our internal goals. Under this model, if multiple objects are presented, both targets and nontargets will compete for processing capacity. This competition can be biased through both external (*pop-out*; Duncan & Humphreys, 1989) and internal (selective attention) factors, and these goals can lead us to alter the stimulus representations that we generate based on what we are exposed to (Talsma, Senkowski, Soto-Faraco, & Woldorff, 2010). When we are considering an audiovisual integration task, the way to maximize performance is to be able to successfully bind as many visual candidates as possible to the auditory stimulus, in hopes that one of the candidates you bound is the one that is probed. This is especially true when the potential candidates are equally likely to be the target – all visual stimuli that change do so simultaneously, and in locations that are equidistant from fixation. While this can be accomplished to some degree by means of sustained attention throughout the experimental task, it is also possible to use the paradigm itself to suggest strategies to participants.

The two experiments in this chapter manipulate stimulus factors that are known to facilitate unimodal processing. In Experiment 5 the effects of congruency will be manipulated, in hopes that crossmodally congruent stimuli demonstrate a higher capacity for audiovisual integration. This manipulation will provide participants with a piece of information that may draw attention towards a subset of dots that serve as potential targets. Experiment 6 will provide lines connecting dots as they change, leading participants to create perceptual chunks made up of multiple changing locations rather than individual locations themselves. This also should lead to an increase in capacity, as it should aid participants in perceiving the visual stimuli as a chunk.

Experiment 5 – Effects of congruency on capacity of audiovisual integration

Introduction

There is a great body of research showing that crossmodal congruency serves to increase the likelihood of multimodal binding (for informative reviews, see Spence, 2011, and Walker, 2012). The general concept of crossmodal congruency (or crossmodal correspondence) holds that certain stimulus factors in different modalities interact with one another in such a way that they can be more easily integrated with one another (or, conversely, in such a way that they are less likely to be integrated). These congruency relationships can have their root in certain perceptual commonalities such as size and pitch (Gallace and Spence, 2006), or in more abstract factors based such as height and pitch (Parise and Spence, 2009), which are determined either by second-level, statistical correspondences (Spence & Deroy, 2012) or by semantic labels (Walker, 2012).

Crossmodally congruent stimuli lead to increases in attentional capture across modalities (Shams & Kim, 2010), with an auditory stimulus increasing attention to a visual stimulus if they are crossmodally congruent to one another. Fiebelkorn, Foxe, and Molholm (2010) looked further into the mechanisms of leading to crossmodal integration and found that crossmodal integration was a combination of bottom-up feature binding that occurs regardless of congruency, as well as a function of top-down attentional spreading occurring in response to congruency. They did so by employing EEG recording while presenting pairings of congruent or incongruent stimuli, and analyzing the spread of attention during stimulus presentation as well as top-down control of attention. Further, they found that these two processes combined in an additive fashion when congruent stimuli are being used, with a top-down attentional contribution caused by stimulus congruency adding to the bottom-up effects present in any multisensory stimulus presentation. Sarmiento, Shore, Milliken, and Sanabria (2012) present similar findings, showing that attentional control set is influenced by crossmodal congruency. They asked participants to estimate the duration that a visual stimulus (a dot) was displayed on the screen, while simultaneously presenting an auditory stimulus. They found that the duration of the tone altered participants' perception of the duration of the visual stimulus being presented. They went further to examine the effects of congruency, with the finding that this effect was stronger when congruent stimulus pairings were rare within an experimental paradigm rather than common. Therefore, congruent stimuli can capture attention, increasing the likelihood that integration would occur when an auditory stimulus is presented in synchrony with the critical visual stimulus, and when characteristics of auditory and visual information match.

Beyond attentional capture, there is also evidence for crossmodal congruency influencing perceptual sensitivity. Marks, Ben-Artzi, and Lakatos (2003) found that including congruent crossmodal stimuli increases perceptual sensitivity on an auditory or visual stimulus discrimination task. Perceptual sensitivity to a stimulus in one modality was increased in the presence of a stimulus in the other modality, regardless of congruency, but the particular inclusion of a congruent stimulus increased perceptual sensitivity even further. This effect was not symmetrical across modalities, with a stronger effect found for an auditory stimulus accompanying visual perception, relative to when a visual stimulus was used to influence auditory perception. The inherent perceptual benefits afforded by crossmodally congruent stimuli also appear to assist with the re-development of cognitive architecture. Kim, Seitz, and Shams (2012) studied the benefits of using a congruent auditory stimulus in facilitating a visual motion detection task. They trained participants over five days using either visual only training, audiovisual congruent, or audiovisual incongruent training stimuli. They found that only audiovisual congruent training facilitated visual motion tracking, showing training effects over the five sessions. Audiovisual incongruent training was insufficient to show training effect, showing no statistical difference from visual-only training.

Given that crossmodal congruency between auditory and visual stimuli increases their respective perceptual sensitivity and degree of attentional capture, we would also expect crossmodally congruent stimuli to increase the capacity of audiovisual integration. To that end, this experiment employs a form of pitch-brightness congruency easily applicable to the current experimental paradigm. Marks (1987) found that light coloured (e.g. white) visual stimuli are congruent with high-pitched tones, and that dark coloured (e.g. black) visual stimuli are congruent with low-pitched tones, while Parise and Spence (2009) showed that crossmodally congruent stimuli using appropriate combinations of brightness and pitch can also increase the temporal window of integration between two stimuli. In previous work from our laboratory

(Wilbiks & Dyson, 2013a), we compared the relative effects of temporal alignment and crossmodal congruency in an audiovisual integration task. We found that when temporal information was informative (i.e. a stimulus in one modality was temporally coincident with a stimulus in the other modality), participants tended to make binding decisions based on these temporal factors. However, when the temporal alignment of the stimuli was uninformative, only then did stimulus factors influence binding judgments. Given this finding of stimulus factors playing a role only when temporal information is uninformative, it may seem like stimulus factors should not be expressed since the current experiment presents visual stimuli that are always simultaneous with the auditory stimulus to which they may be bound. However, since the auditory and visual stimuli are *always* simultaneous with one another, this makes the temporal alignment equally uninformative in attempting to decide between multiple candidates for integration.

Based on these previous findings, we expect that manipulating the crossmodal congruency of stimuli will increase the likelihood of successful binding, and will thus increase the functional capacity of audiovisual integration. Under the assumption that light visual stimuli and high pitched auditory stimuli are congruent, and that dark visual stimuli and low pitched auditory stimuli are congruent (as per Marks, 1987), Experiment 5 will include a factor of congruency in an attempt to measure the effects of this factor on capacity of integration. In Van der Burg et al. (2013), during valid trails, the ultimate probed location could switch between two states: white to black, or, black to white. We expected that the presentation of a low tone during the critical trial would promote binding to white to black changes whereas presentation of a high tone during the critical trial would promote binding to black to white changes. Therefore, correct responding should be higher during congruent relative to incongruent trials.

Method

Participants. 24 participants were recruited from the undergraduate research participant pool at Ryerson University, and compensated with partial class credit. Four participants were rejected from analysis who fell within a 95% confidence interval around 50% (chance responding) calculated across over all trials, as per all previous experiments. The final sample consisting of 20 participants with an average age of 20.8, with 16 females and 18 right handed individuals. None of the participants had taken part in any previous experiments in this series.

Design and Procedure. Stimuli were identical to those used in Experiments 1-4. 16 individual conditions of stimulus were created, by orthogonally varying the SOA of visual stimuli (200 or 700 ms), the number of visual stimuli that changed on each alternation (1, 2, 3, or 4), and the critical manipulation of crossmodal congruency of the dots and the tone (congruent or incongruent). A trial was deemed to be crossmodally congruent when the target dot changed to white in synchrony with a high-pitched tone (4500 Hz), or changed to black along with a low-pitched tone (300 Hz, both as per Parise & Spence, 2009). When a dot changing to black was paired with a high tone, or changing to white with a low tone, it was deemed an incongruent trial. These 16 conditions were each presented 3 times to create an experimental block with 48 trials. Each participant completed one practice block, and 8 experimental blocks, for a total of 384 experimental trials. The number of trials were once again increased in order to provide sufficient power for analysis of an additional factor (congruency in this case, and vertices in Experiment 6).

Model Fitting. Modelling was conducted in the same manner as in Experiments 1-4. Successful model fit was confirmed by the low RMSEs observed in both the 200 ms SOA (range

0.0028 – 0.1088) and 700 ms SOA (range 0.0001 – 0.1198) conditions (as compared to Van der Burg et al. (2013) Experiment 1c, where the RMSE which fell between 0.036 and 0.060).

Results

Capacity measures (*K*) were entered into a 2 x 2 ANOVA, with factors congruency (incongruent, congruent) and SOA (200, 700 ms). This analysis revealed a main effect of congruency, F(1,19) = 16.415, MSE = .092 *p* = .001, η_p^2 = .464, with capacity for crossmodally congruent pairings yielding a significantly higher capacity than incongruent pairings. There was also a main effect of SOA, F(1,19) = 52.948, MSE = .280 *p* < .001, η_p^2 = .736, supporting previous experiments in showing a higher capacity for 700 ms as compared to 200 ms. These two main effects were subsumed by a significant congruency x SOA interaction, F(1,19) = 13.615, MSE = .059, *p* = .002, η_p^2 = .417. Examining this interaction further by means of a Tukey's HSD (*p* < .05) revealed that congruency modulated audiovisual integration capacity at the 700 ms SOA, but not at the 200 ms SOA. Figure 11 shows that this interaction is present in 15 of the 20 participants in the experiment, indicating that the majority of participants showed facilitation through being exposed to crossmodally congruent pairings when compared to incongruent pairings.

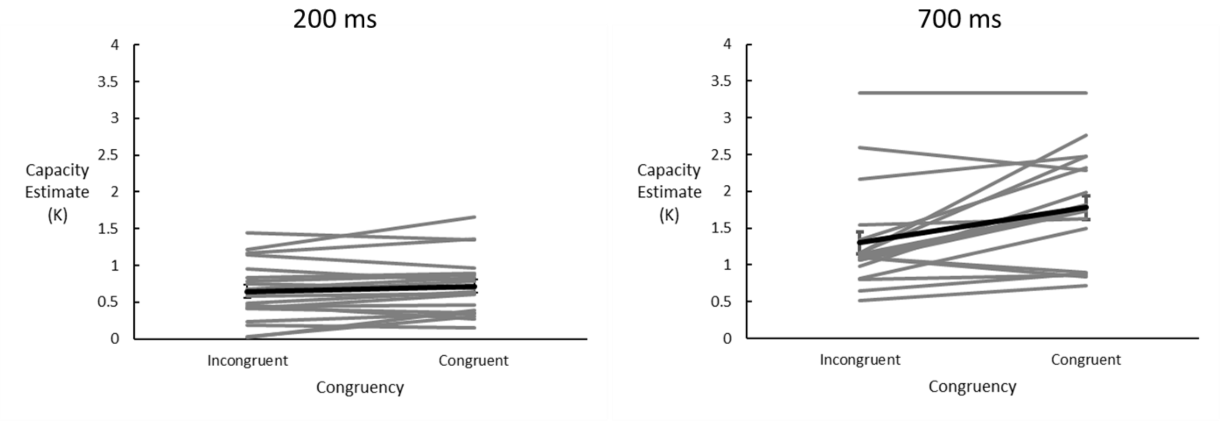


Figure 11. Capacity estimates for 200 ms (left panel) and 700 ms (right panel) SOAs for congruent and incongruent stimuli. Data shown for individual participants (grey) and experiment mean (black), with error bars representing standard error.

In order to answer the critical question regarding the conditions under which it is possible for the capacity of audiovisual integration to exceed 1 item, capacity measures were also subjected to single sample *t*-tests, against the fixed value of 1. These tests revealed that at 200 ms SOAs, capacity remains significantly less than 1 for both congruent (*t*(19) = -4.013, *p* = .001) and incongruent (*t*(19) = -37.029, *p* < .001). For 700 ms SOAs, capacity was significantly greater than 1 for the congruent pairings (*t*(19) = 4.928, *p* < .001), and was trending towards significance for the incongruent pairings (*t*(19) = 2.017, *p* = .058).

Table 4 Summary of the three-way repeated measures ANOVA comparing proportion correct data in Experiment 5

Metric	*df*	F	MSE	*p*	η_p^2
Congruency (C)	**1,19**	**17.74**	**.007**	**<.001**	**.483*** **
SOA (S)	**1,19**	**90.09**	**.019**	**<.001**	**.826*** **
Number (N)	**3,57**	**439.66**	**.002**	**<.001**	**.959*** **
C x S	**1,19**	**10.21**	**.004**	**.005**	**.349*** **
C x N	**3,57**	**7.20**	**.001**	**<.001**	**.275*** **
S x N	3,57	4.04	.005	.011	.175**
C x S x N	**3,57**	**12.31**	**.001**	**<.001**	**.393*** **

Note: Statistical significance in bold

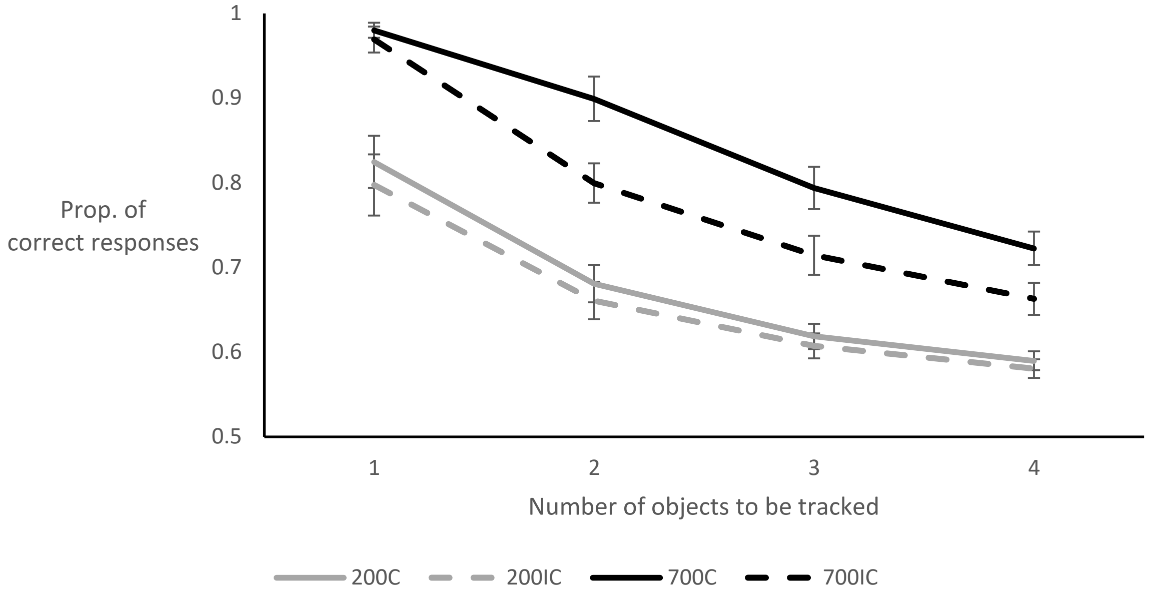


Figure 12. Proportion correct for each SOA (200, 700), congruency (congruent (C), incongruent (IC), and number of objects to be tracked in Experiment 5. Error bars represent standard error.

Finally, to take a more nuanced approach to performance in Experiment 5 and to examine *when* congruent relations between audio and visual information play a role, proportion correct data (means displayed in Figure 12) were entered into a repeated measures ANOVA with factors of congruency (2), SOA (2), and number of locations to be tracked (4). The full results of the ANOVA are shown in Table 4. There were main effects of SOA (p < .001) and of number (p < .001), with a greater likelihood of correct responding for 700 ms condition than in the 200 ms condition, and decreasing likelihood of correct responding as the number of to-be-tracked locations increased. In addition, a critical main effect of congruency (p < .001) indicates that when the colour change and the pitch of tone were congruent, a correct response was more likely than when they were incongruent. In addition to all two-way interactions being significant (all p's <.011), a significant congruency x SOA x number interaction (p < .001) was examined in detail by means of a Tukey's HSD (p < .05) comparison. In the 200 ms SOA condition, no significant differences were found between congruent and incongruent conditions at any number of items to be tracked. Within the 700 ms SOA condition, however, crossmodally congruent stimuli were more likely to be accurately integrated than incongruent ones when there were 2, 3, or 4 items to be tracked (see Figure 13).

Discussion

The data from Experiment 5 reiterate the finding from the earlier experiments – that audiovisual integration capacity can exceed one item when the rate of stimulus presentation is slowed from 200 ms to 700 ms SOA, consistent with the temporal frequency limits reported by Holcombe and Chen (2013). The reason for this may be that audiovisual integration cannot exceed one in the 200 ms condition as the result of a data limitation (Norman & Bobrow, 1975). That is, the speed of presentation is simply too fast, and the perceptual system is not able to process the incoming information adequately to be able to integrate more than one item. According to the estimates of Holcombe and Chen (2013), in order to adequately track 2 visual stimuli, the rate of change between them needs to be at minimum 250 ms (in their research: 4 Hz). Furthermore, in separating congruent and incongruent pairings, we see that congruent pitch-brightness pairings increase audio-visual integration capacity, consistent with the perceptual sensitivity effects shown by Marks et al. (2003), the facilitation of training in Kim et al. (2012), and, the widening of the temporal window of integration reported by Parise and Spence (2009). In directly comparing capacity measures, we find a critical 2 x 2 interaction, which indicates that at the 700 ms SOA, congruent pairings have a significantly higher capacity (1.78) than incongruent pairings (1.30), while there is no significant effect of congruency at the 200 ms condition (congruent: 0.72; incongruent: 0.64). Therefore, once we are no longer constrained by a data limit, we are able to see the facilitative effects of congruency.

In addition to the need for slow stimulus presentation, the more fine-grained examination of proportion correct data reveal one final caveat in the use of congruency in promoting audio-visual integration capacity: there must be more than one object being tracked. This fits well with the account in Experiments 1 – 4, wherein effects of stimulus predictability and proactive interferences were most pronounced when more than one object was being tracked. When only a single object is being tracked, integration is able to occur based only on the temporal coincidence (i.e. 0 ms) of visual and auditory stimuli. However, once more than one visual stimulus is being presented, which one should be integrated becomes ambiguous, and it is under these conditions that stimulus factors such as congruency begin to play a role.

It could be argued that tracking a single object, at a 700 ms SOA, is not a particularly difficult task, while tracking additional objects increases the difficulty. Therefore as the task difficulty increases, we look for additional factors that can help with integration, and in this case that supplementary information comes in the form of crossmodal congruency (see also Wilbiks & Dyson, 2013a). We also clearly find that tracking of more than a single object at a 200 ms SOA is seemingly impossible – the reason for this likely stems from a data limit wherein the visual information being received is not reliable with more than one object at such a fast SOA. This has been shown behaviourally throughout the experimental series thus far, and was also supported by ERP results from Experiment 1. So as long as we are looking at a resource limitation (as in 700 ms SOA), as the task increases we see additional information that help to adjudicate between visual locations as candidates for auditory binding (e.g. congruency) play an increasing role. Conversely, in a data-limited task (as in 200 ms SOA) we see no facilitation by congruency.

Experiment 6 – Increasing capacity through 'perceptual chunking'

Introduction

In terms of the uni-modal literature, working memory span has been shown to functionally increase by means of a technique called chunking. First described by Miller (1956), this technique involves combining multiple items to be held in working memory into more complex, but less numerous items, allowing for a greater amount of information to be maintained in working memory. For example, in the learning of language, chunking is implemented in both bottom-up (based on statistical regularities) and top-down (based on familiarity with words), allowing for more efficient reading (Jones, Gobet, & Pine, 2007). Through this chunking process, individuals are able to impose goal-directed perception on individual letters in order to maximize processing efficiency. While chunking has traditionally been discussed in terms of working memory, it has also more recently been shown to be an effective perceptual aid. Gobet and Simon (1998) considered expert chess players' perception of chess positions and found that, while non-experts perceive positions of each piece independently and then build a concept of the game situation, expert players perceive the chessboard as a chunk, a single situation including all piece positions.

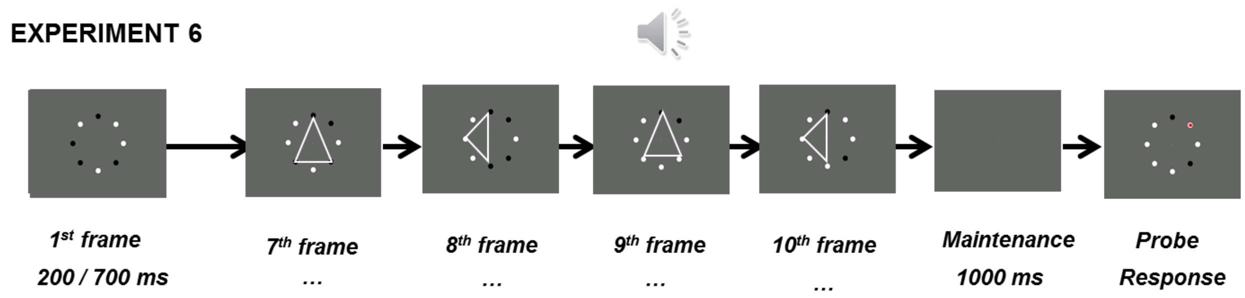


Figure 13. Trial schematic for Experiment 6.

Additional evidence from non-expert participants show that perceptual chunking is used in everyday contexts as well. Gilbert, Boucher, and Jemel (2014) discuss perceptual chunking as an online process, allowing for the combination and consolidation of domain-general information (different from the post-hoc working memory chunking described by Miller, 1956). They show behavioural evidence as well as electrophysiological modulation in the P300 event-related potential component as indicators of perceptual chunking being used in speech perception.

Gmeindl, Walsh, and Courtney (2011) presented participants with a display of scattered grey squares, with some designated targets (via a black outline) and others as distractors (no outline). After an indication of targets, participants were asked to indicate targets either by touching all targets or typing the locations on a keyboard. Their results indicated that people performed better when engaging in the spatial task of touching rather than typing them, and this effect was increased as a function of the nearness of the targets to one another in the display. The authors propose that this is evidence for the use of perceptual chunking, as participants are better able to perform the task when it is a spatial one, and when targets can be mentally joined to one another. Sargent, Dopkins, Philbeck, and Chichka (2010) provide similar evidence for perceptual chunking as a technique. Here, participants were exposed to targets arranged 360-degrees around them in a room. When attempting to identify them, results were improved if targets were closer to one another, within an arrangement that was seen multiple times within the experiment, and if they could be mapped onto a common object. This final explanation is most pertinent to the current research – using an object to chunk together disparate targets allows for processing cost to be decreased, and allow for more information to be successfully tracked.

The research discussed here provide a convincing account of perceptual chunking as a strategy for increasing effective perceptual span. Alvarez and Cavanagh (2004) qualify some of these findings in their study of the capacity of visual working memory, finding that rather than being a pure numerical limitation, it is a capacity based on both the number of items and the complexity of items to be kept in memory. While they do not explicitly discuss chunking as a strategy, they find that capacity for simple stimuli (such as coloured line drawings) falls at around 4.4 items, but for complex stimuli (three-dimensional cubes) capacity was reduced to 1.6 items. However, they argue, the amount of information being stored may be equal in both of these cases. Mathy and Feldman (2012) found that by using easily chunkable visual information, functional capacity was increased from 8, 12, 16, or 20 simple stimuli to hold 2, 3, 4, or 5 chunks of more complex information. By compressing information into fewer, more complex chunks, one increases the number of individual pieces of information that can be processed, and hopefully, remembered.

In Experiment 6 effects of perceptual chunking on capacity of audiovisual integration will be examined by- essentially- connecting the dots for the participant. By asking participants to track the type and orientation of lines or polygons created by connecting vertices overlaid on the dots as they changed, they should be able to perceive one, complex object rather than a greater number of simple objects. While the same amount of information needs to be tracked – the location of 4 vertices, for example, rather than 4 locations – we expect participants to be able to ‘chunk’ these vertices into a single representation of a polygon, and as such we should see an increase in the apparent capacity of integration. Like audio-visual stimulus congruency, this could again allow the functional capacity of audiovisual integration to exceed the previous estimate of one.

Method

29 participants took part in the study, and were compensated with partial class credit. Data were trimmed by the same method as in the previous experiments, with the final sample consisting of 24 participants with an average age of 22.0, with 20 females and 24 right handed individuals. None of the participants took part in previous experiments in this series.

Experiment 6 was similar to Experiment 5 in terms of the presentation of repeated visual stimuli, with a critical presentation accompanied by a synchronous tone. Note that the auditory stimulus was the same as the one used in Experiments 1 – 4, with no more congruency manipulation being in place here. However, in half of the trials, in addition to 1-4 locations changing at each alternation, a set of vertices was presented in a mid-grey colour (100, 100, 100) in the form of a dot with a diagonal slash on it (when 1 dot changed), a line (2-dot changed), a triangle (3-dot changed), or a quadrilateral (4-dot changed; see Figure 11). The number of locations / number of vertices to be tracked and SOA (again, 200 or 700 ms) were manipulated within blocks, while the critical comparison vertices (present, absent) was manipulated across blocks. Thus, a participant might first complete a practice block and 4 experimental blocks (48 trials in each) of the no-vertex condition, followed by a practice block and 4 experimental blocks of the vertex condition, for a 384 trial total. The order in which participants completed the two polygon conditions was counterbalanced throughout the experiment, and a comparison between participants doing polygons first or polygons second showed no significant practice effects (t(24) = 1.154, *p* = .260).

Table 5 **Summary of the three-way repeated measures ANOVA comparing proportion correct data in Experiment 6**

Metric	*df*	F	MSE	*p*	η_p^2
Vertices (V)	**1,23**	**46.17**	**.014**	**<.001**	**.667*******
SOA (S)	**1,23**	**125.79**	**.013**	**<.001**	**.845*******
Number (N)	**3,69**	**273.24**	**.003**	**<.001**	**.922*******
V x S	1,23	0.72	.015	.404	.030*
V x N	**3,69**	**23.27**	**.002**	**<.001**	**.503*******
S x N	**3,69**	**9.56**	**.003**	**<.001**	**.294*******
V x S x N	**3,69**	**4.53**	**.002**	**.006**	**.165****

Note: Statistical significance in bold

* = small effect size ($\eta_p^2 > .02$)
** = medium effect size ($\eta_p^2 > .13$)
*** = large effect size ($\eta_p^2 > .26$)

Results

Capacity measures (*K*) were calculated by means of the same fitting procedure as in the previous experiments, with goodness of fit confirmed by low RMSEs ranging from 0.0001 – 0.1581. Capacity measures were subjected to a 2 x 2 ANOVA with factors of vertices (present, absent) and SOA (200, 700 ms). This analysis yielded a main effect of vertices, F(1,23) = 59.34, MSE = .262, $p < .001$, $\eta_p^2 = .721$, and of SOA, F(1,23) = 106.07, MSE = .272, $p < .001$, $\eta_p^2 =$ .822, with no significant interaction (F(1,23) = 2.05, MSE = .149, $p = .165$, $\eta_p^2 = .082$). As shown in Figure 14, capacity was significantly greater with vertices than without vertices, and was greater for 700 ms SOA than for 200 ms SOA. Considering the data of all participants, we find that presence of vertices facilitates binding in all but 2 of the participants in Experiment 6 (see Figure 14).

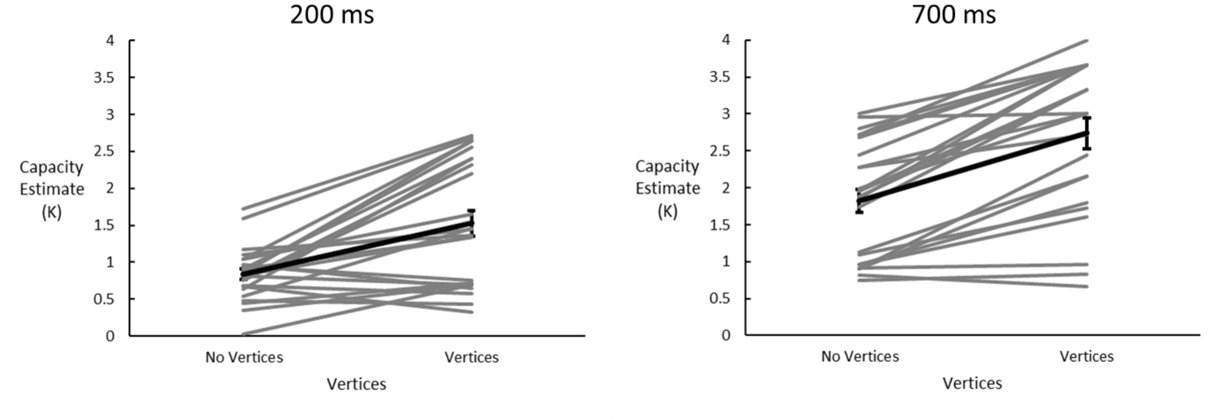


Figure 14. Capacity estimates for 200 ms (left panel) and 700 ms (right panel) stimuli for vertices or non-vertices for individual participants (grey) and experiment mean (black), with error bars representing standard error.

Capacity measures were compared to the norm of 1 item by means of single sample *t* tests. When SOA was 700 ms, it was significantly greater than one regardless of whether vertices were present (M = 2.739, SE = .208; *t*(23) = 8.348, $p < .001$) or absent (M = 1.821, SE = .159; *t*(23) = 5.148, $p < .001$). When SOA was 200 ms and without vertices, the capacity remained less than 1, (M = .837, SE = .074; *t*(23) = -2.195, $p = .038$, as in all previous experiments. Of particular interest, however, was the finding that the capacity of audiovisual integration was significantly greater than 1 in the presence of vertices, even when SOA was 200 ms (M = 1.529, SE = .174; *t*(23) = 3.043, *p* = .006). This is the first instance of capacity being significantly greater than 1 in a 200 ms SOA condition, which previously seemed to be a data limit that could not be breached.

Proportion correct data (Figure 15) were entered into a repeated measures ANOVA with factors of vertices (2), SOA (2), and number of locations to be tracked (4). The full results of the ANOVA are shown in Table 5. Main effects of SOA (*p* < .001) and number (*p* < .001) replicate the effects from Experiment 1, with an increasing likelihood of correct responding in slower SOA, and with fewer locations to be tracked. The critical manipulation in Experiment 6 of vertices showed a significant main effect (*p* < .001), indicating that presence of vertices assisted in tracking locations. A significant interaction (*p* = .006) was also observed between vertices x SOA x number, and decomposed using Tukey's HSD (*p* < .05). Here, the use of vertices in connecting together dot locations that changed in polarity increased the likelihood of correct responding at 200 ms SOA with any number (1-4) of locations, where as in the 700 ms SOA condition, vertices only enhanced correct responses when more than one location was to be tracked (2-4; as per Experiment 5).

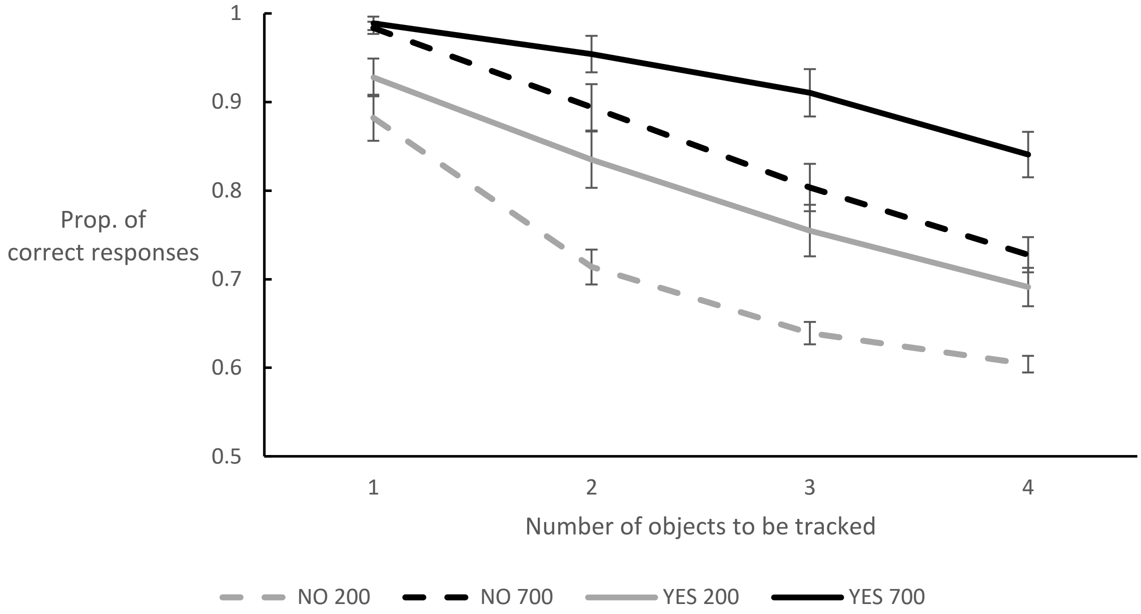


Figure 15. Proportion correct for each SOA, vertices, and number of objects to be tracked in Experiment 6. Error bars represent standard error.

Discussion

The addition of connected vertices to literally connect the locations that were changing within Experiment 6 served to increase the capacity of audiovisual integration. In terms of capacity measures, this was demonstrated by a main effect, without an interaction, and was the first time in the series of experiments in which capacity exceeded one item both at 700 and 200 ms. It was also found that in terms of the proportion correct data at a relatively slow rate of presentation (700 ms), vertices worked in a similar way to the congruency manipulated performance in Experiment 5, with their effect appearing also when the task was difficult enough that it could be done with simply the stimuli being presented themselves (i.e. not at the relatively easy 700 ms SOA / 1 location condition). Critically though, the effect of perceptual chunking extended to 200 ms SOA as well, once again taking capacity beyond 1 object. Bor and Seth (2012) found that traditional chunking lowers memory demands by compressing data and by increasing automisation. In this case, by reducing the level of difficulty placed on the perceptual system, we find that we can overcome the previously shown limit wherein no more than one visual location could be bound to a single auditory cue at an SOA of 200 ms. While an argument was previously made that this may be a data limit (cf. Norman & Bobrow, 1975), this appears to be ruled out by the finding that increasing the fidelity of the signal led to an increase in capacity of audiovisual integration. Rather than a data limit, perhaps what is being demonstrated here is a resource limit that requires facilitation to work at an SOA such as 200 ms. Alternatively, perhaps perceptual chunking is acting as a top-down factor, allowing for more information to be perceived by reorganizing the information while it is incoming. These arguments will be discussed further in the general discussion below.

Just as Gilbert et al. (2014) found perceptual chunking aided speech perception, Experiment 6 found that creating chunks that are more complex (connected vertices rather than dots) but less numerous (1 figure rather than up to 4 dots) increases effective integration capacity. In considering capacity of working memory, we can consider these findings in the context of the evidence provided by Alvarez and Cavanagh (2004). In the current research, we see that the effective capacity for audiovisual integration increases when integrating one, complicated chunk (shape) rather than multiple, simple stimuli (dots), whereas Alvarez and Cavanagh (2004) found that more individual items could be remembered if they were simple rather than complex. What this means, in this case, is that even though using a relatively complex item (e.g. a triangle) may reduce audiovisual integration capacity to one item, this serves a functional role that is greater than two simple (e.g. dots) items being integrated with an auditory stimulus. It also answers potential questions about how this experiment can show capacity for integrating two visual items, while Holcombe & Chen (2013) found that in order to track two items a minimum SOA of 250 ms was required. In the current experiment, participants are functionally binding a single complex stimulus rather than multiple simple stimuli. In a real life situation, what is of utmost importance in terms of audiovisual integration is the amount of information that can be bound across modalities. As such, it is more valuable to an individual to be able to bind as much information as possible, and binding a single complex stimulus provides a greater amount of data to an individual than does binding a single simple stimulus. So, while the true numerical capacity of integration may not be increasing – we may still be binding only a single stimulus – the functional capacity is increasing, in terms of the amount of information that can be integrated.

General Discussion

The findings from Experiments 5 and 6 support the idea that stimulus-based factors can modulate the functional capacity of audiovisual integration. While capacity has previously been thought to be limited to one item (Van der Burg et al., 2013), the data shown here enrich what we know about integration capacity even further. Experiment 5 manipulated the crossmodal congruency relationship between the auditory stimulus and the visual stimuli that were candidates for integration. This manipulation demonstrated that crossmodal congruency is able to influence the capacity of audiovisual integration by modulating the processing of information, increasing the ease of integration occurring. Experiment 6 provided participants with connections of vertices at each location that was changing polarity, and showed that perceptual chunking (as per Gobet et al., 2001) had a similar but more expansive effect, with a greater number of locations able to be tracked (and thus, an increased integration capacity) when those locations were joined by means of a polygon, which served as a cue to perceive the multiple locations as a single, chunked percept.

Over the course of these two experiments, a case has been constructed for the role of stimulus factors in increasing the capacity of audiovisual integration. In doing so, we have also identified specific factors that may contribute to modulating this capacity both in terms of a specific number of items, as well as the functional capacity in terms of amount of information that can be integrated, regardless of whether one simple or one complex object is integrated.

Looking to the future, it will be of interest to continue breaking down the factors that contribute to capacity of audiovisual integration even further. We have now shown that it is possible for capacity to exceed 1 item, even at the 200 ms SOA, but an interesting extension of this line of research would be to consider whether there is some speed of presentation under which it is not possible for capacity to exceed one. Under Holcombe and Chen's (2013) perspective, at a speed of presentation with an SOA less than 143 ms, it is strictly not possible for more than one visual item to be tracked. That having been said, it is also likely that just as audiovisual integration capacity is variable across participants the temporal frequency limit may also vary. Based on this, we might expect that using an SOA of 100 ms, for example, would be strictly unable to show a capacity of greater than one, even when stimulus factors like the ones employed here are manipulated to maximize capacity, although this may be different for specific individuals based on their own personal temporal frequency limits.

An additional idea stems from a potential criticism of Experiment 6; taking an alternate view based on Alvarez and Cavanagh's (2004) findings, one might argue that increasing functional capacity is not enough to demonstrate a true increase in capacity. The question here is what is truly being measured in establishing a capacity limit? In Van der Burg et al.'s (2013) work, and in Experiments 1 – 5, the dots all served as independent stimuli, with no need to differentiate between the number of objects that are being integrated and the amount of information being integrated. However, once locations are connected, we have such a difference. For example, if three locations are changing, in the non-vertices condition we have three pieces of information and three objects being integrated. On the other hand, in the vertices condition, three pieces of information (dots) are integrated and yet only one object (a triangle) needs to be integrated. This effect may be associated with findings regarding object-based attention in vision. Awh, Dhaliwal, Christensen, and Matsukura (2001) find that attention for multiple stimuli is modulated by their spatial relationship (e.g. distance between them) and whether they are located on the same 'object' or not. That is to say, two pairs of stimuli with the same distance between them would be perceived differently if one pair were located on the same 'object' while the others were not. Perhaps in this experiment the dots being perceived as being on the same 'object' led to them being processed more efficiently and thus led to a higher estimate of audiovisual integration. This distinction serves as an explanation as to why we see 200 ms show a capacity of greater than one only in this experiment, and only with vertices present – we continue to be able to integrate one visual object with the auditory stimulus in this condition, even though the visual object is more complex, and holds a greater amount of information. This hypothesis could be tested further by using perceived subjective contours induced by Kanizsa triangles (Kanizsa, 1976). These are triangles which are displayed with implied vertices, but no connecting lines. For the purposes of this experiment, it would induce perceptual chunking without providing participants with a single complex stimulus to attend to, and if the effect were to be replicated in such an experiment it would support the benefits of perceptual chunking.

These findings contribute to an ongoing account of audiovisual integration which illustrates that it does not seem to be qualitatively different from unimodal perception, in terms of capacity measures. We now have evidence showing that crossmodal congruency and perceptual chunking serve the same purpose multimodally as they do unimodally. Now that we have explored effects of stimulus factors on the capacity of integration, we will look at individual training as a final route to promoting audio-visual capabilities.

Chapter 4 – Individual flexibility in audiovisual integration

Experiment 7 – Improvement of audiovisual integration through training

Introduction

In Chapter 2, four experiments were conducted, and established that the capacity of audiovisual integration is flexible, can exceed one item, and is modulated by changes in proactive interference and temporal predictability. Chapter 3 demonstrated that the capacity of audiovisual integration is also subject to influences from stimulus factors such as crossmodal congruency and perceptual chunking. Within the first six experiments, in addition to exploring the effects that temporal and stimulus factors have on integration, it also became apparent that there was a large amount of variance between individuals in terms of their capacity. Figure 16 shows the capacity estimates for each participant in the first six experiments, and makes evident that there is large variation in terms of their respective capacities of audiovisual integration. We have also already discussed, in Chapter 1, the degree of variation present in visual working memory, both between and within individuals (Abikoff & Gittelman, 1985; Klingberg et al., 2002). Within the current data, for example, in each of Experiments 1, 3, and 4, there is at least one participant who shows a capacity greater than *two* items (in the 700 ms SOA), while there are also participants in all conditions who have capacities less than 0.5. A question arising from this pattern of data pertains to the mechanism of this variation between individuals. Given that there are large differences between individuals, it would also be of interest to consider differences in integration capacity within an individual. While we have not formally assessed participants' experience with audiovisual tasks, a potential explanation for the between-individuals differences would be that an increase in experience with a task will increase one's ability in it. For example, in a study of expertise in pianists, it was found that high levels of training in playing piano allowed participants to identify a musical excerpt being performed based solely on watching hand movements on the keyboard (Hasegawa, Matsuki, Ueno, Maeda, Matsue, Konishi, & Sadato, 2004). This training advantage is also indexed by activation of the left planum temporale via fMRI recording, an area which is related with integration of auditory and visual information. However, having no direct way to answer this question based on the current research, this chapter will explore the potential for individuals to increase their capacity of audiovisual integration by means of training.

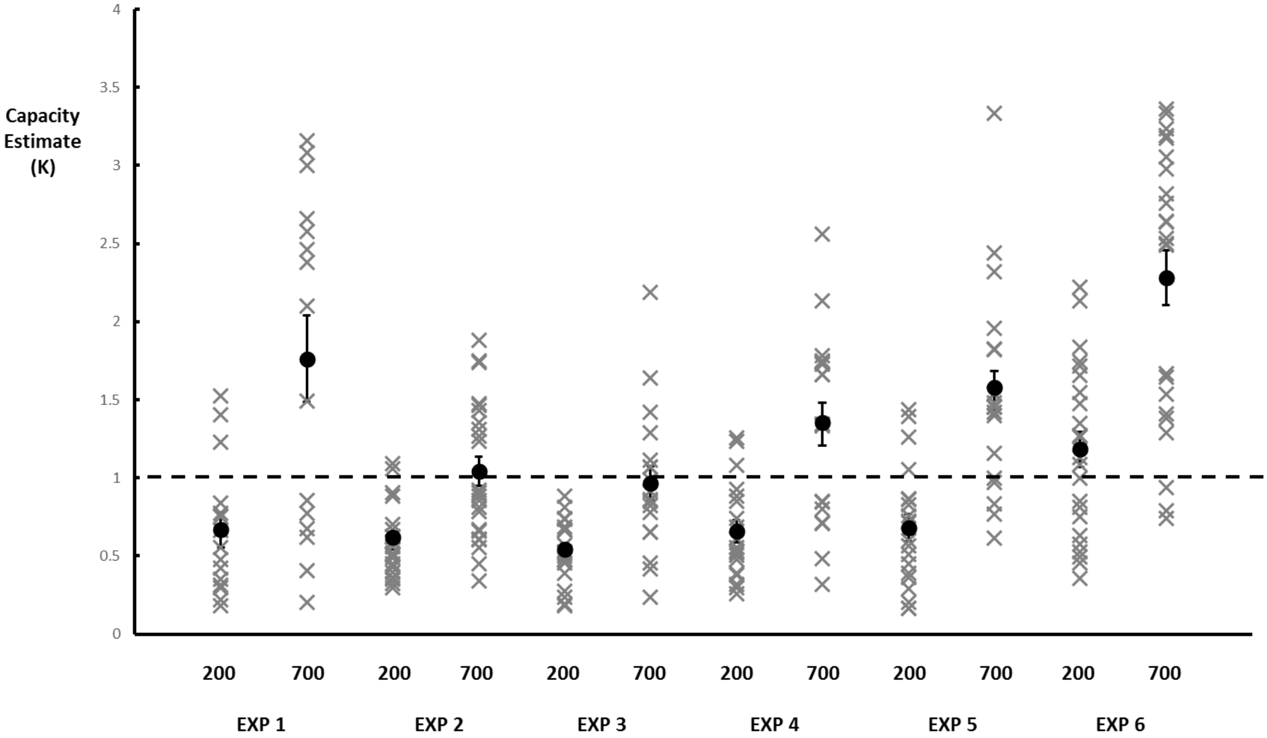


Figure 16. Capacity estimates for each individual participant (grey Xs), along with mean and standard error (black dot), for Experiments 1 – 6 as described in Chapter 2 and 3.

Training has been shown to have an influence on multimodal integration, often evidenced through recalibration of the temporal window of integration. Many studies have shown that over the course of an experimental session, participants find their perception of auditory and visual information to be changed via recalibration of their multisensory integration systems. Fujisaki, Shimojo, Kashino, and Nishida (2004) presented participants with an auditory and a visual stimulus and asked them to judge whether the two stimuli were presented simultaneously or not. They manipulated the lag between the visual and auditory stimuli systematically, and in doing so induced a recalibration of participants' point of subjective simultaneity such that it shifted towards the manipulated lag. That is to say, presenting a large number of trials where the visual stimulus preceded the auditory stimulus by, on average, 100 ms led participants to perceive that as being simultaneous. Heron, Roach, Hanson, McGraw, and Whitaker (2012) expand on this idea by showing that while recalibration within a set of stimulus presentations tends to be attractive (that is, move towards the preset lag prescribed by the experiment), there can also be repulsive aftereffects shown, wherein the newly calibrated system shifts away from the manipulated lag.

Work in our own laboratory (Wilbiks & Dyson, 2013b) found that the repulsive aftereffects described by Heron et al. (2012) show asymmetries with regard to the modalities in which stimuli are presented. Participants were asked to decide which of two potential 'anchors' of one modality (auditory or visual) were the likely cause of a single roving stimulus (of the opposite modality). In addition to main effects related to the relative order of presentation of stimuli, and stimulus congruency factors, it was found that in 'visual-rich' stimulus environments (i.e. 2 visual stimuli / 1 auditory stimulus), aftereffects occurred, while the same was not true for 'auditory-rich' environments (i.e. 2 auditory stimuli / 1 visual stimulus). We argued that this was due to the relatively higher reliability of auditory stimuli in the temporal realm, compared to vision. Essentially, when we have stimuli that are able to reliably provide the information required to complete a task such as the one described above we do not show recalibration of our binding window. Conversely, if the stimuli we are presented with do not provide enough information to allow a task to be completed, recalibration takes place in order to increase the likelihood of completing the task successfully. This finding is supported by research highlighting the respective dominance of auditory and visual stimuli in temporal and stimulus processing (e.g., Alais & Burr, 2004; Burr, Banks, & Morrone, 2009). While this research was looking at the window of integration, and the current research is looking at its capacity, it seems reasonable to expect similar results to occur. In a situation where stimuli are unreliable (e.g. 200 ms SOA), we might expect training to occur while with more reliable stimuli (e.g. 700 ms SOA) training should not occur. In discussing effectiveness of perceptual training, Ahissar and Hochstein (1997) propose a *reverse hierarchy model*, within which training is most effective on difficult trials, but that this training can only occur if training has previously been activated by using lower difficulty trials. In this framework, it is possible to conceptualize 700 ms trials as 'easy', and 200 ms trials as 'difficult', with 450 ms trials that fall somewhere in between. In order to maximize training efficiency, rather than presenting only 450 ms trials during the training blocks, it may be more useful to increase the difficulty during the training sessions. Pavlovskaya and Hochstein (2011) provide support for the findings of Ahissar and Hochstein (1997) and extend them by showing that transfer effects in perceptual learning are more likely to occur on easy than on difficult trials. In the current research, this may explain why no transfer effects were found into the 200 ms condition.

The findings discussed above show that the audiovisual integration system is malleable, and subject to alteration by presentation of repeated stimuli, as long as participants are exposed to large numbers of stimuli across multiple blocks. Recently, however, Van der Burg, Alais, and Cass (2013) presented participants with a visual and an auditory stimulus, separated temporally by between 0 and 512 ms (both visual lead and visual lag). Participants were asked to make a simultaneity judgment about the two modalities that were presented. If audition led on trial n-1, then on trial n participants were more likely to judge a slight auditory lead as being simultaneous. A similar effect occurred when vision led on trial n-1. In the context of the current research, the fact that even a single trial can show some recalibration effect gives a strong indication that only 2 training sessions will be ample time to lead to training effects being demonstrated.

Visual short term memory capacity has also been shown to be improved through training, which provides some impetus for attempting a training study on audiovisual integration capacity as well. Olesen, Westerberg, and Klingberg (2004) employed fMRI to show that through training, participants exhibited an increase in visuo-spatial working memory behaviourally, which was also supported by increased levels of activity in brain areas associated with working memory (specifically, superior and inferior parietal cortices, and middle frontal gyrus). They had participants practice three working memory tasks over the course of a 5 week period, and measured their working memory both before and after training. This task involved tracking dots displayed within a 4x4 grid, presented at 900 ms ISIs, and then responding by indicating each respective location in order. This task was administered both before and after training sessions, while the tasks to be practiced included a visuo-spatial working memory task, a backwards digit span task, and a letter span task, with 30 trials of each task being completed each day for 5 weeks. Participants showed improvement in terms of accuracy on the Span board task, and the backwards digit span task, during training. They also showed an improvement on the location marking working memory task, both behaviourally (response accuracy) and in level of BOLD response in working memory areas as described above.

The current Experiment 7 continues this line of thinking – that the capacity of audiovisual integration is dynamic, is subject to the same kinds of effects as in unisensory sensation and perception, and that it is possible for it to exceed one item with extended practice. Given that research has shown effects of training in experiments involving audiovisual integration and visual short term memory literature (Olesen et al., 2004; Heron et al., 2012; Fujisaki et al., 2004; Wilbiks & Dyson, 2013b), it is now of interest as to whether audiovisual integration capacity is fixed or flexible within an individual. This experiment will have participants repeatedly exposed to certain temporal parameters of our paradigm (SOA), and test whether their performance improves (and capacity increases) over time. There are three main types of training effects that are of interest in measuring the magnitude of training: criterion effects, near-transfer effects, and far-transfer effects (Brehmer, Westerberg, & Backman, 2012). Criterion effects involve an improvement on the specific task on which participants are being trained, in our case the SOA on which participants are being trained. Near-transfer effects are those in which training on a certain task yields improvement in a closely-related but not identical task, and which have some degree of biological or cognitive overlap, while far-transfer effects are those which occur in an unrelated field (Dyson, 2014). In this particular study, the interest is in whether near-transfer training effects can be demonstrated in proximal SOAs.

By considering a range of SOAs in this training experiment, it is now also possible to address an additional concern regarding rate of stimulus delivery on audiovisual capacity. Given that capacity at 200 ms is generally limited to one item (although, see Experiment 6 for a discussion of when it may exceed one), while at 700 ms it is possible to exceed one, there are two possible theoretical explanations that can be invoked. Van der Burg (personal communication, 1 May 2015) argues that there is a qualitative difference between the two rates of presentation – that an SOA of 200 ms or less represents “true” audiovisual integration, while a 700 ms SOA represents auditory cueing. For example, at 700 ms the task can be completed on the basis of simply tracking visual information, with the auditory stimulus serving simply as a cue to remember changing locations. First, if Van der Burg's perspective is to be accepted, one would expect capacity to approach that of visual short term memory, which has been shown to be between 3 and 4 items (Cowan, 2001), and this is not the case. Second, from this perspective, one would expect capacity to be limited to one item up to a certain threshold, and to exceed it beyond that threshold (at which point it no longer represents audiovisual integration).

On the other hand, the capacity difference between the two SOAs may simply be a function of a quantitative difference between the two conditions. Holcombe and Chen's (2013) temporal processing limits indicate that at 200 ms it should be possible to track 1, but not 2 or 3 items. In order to track 2 items, and in doing so demonstrate a capacity in excess of 1 item, we would expect to require an SOA of 250 ms or more. If the difference between 200 and 700 ms integration is qualitative (as per Van der Burg) then at a given SOA the capacity of integration should be static – based on the nature of the stimuli being presented. If the paradigm is subject to such a temporal processing limit, then we should expect no possibility of improvement in the 200 ms condition, regardless of the amount of training that is performed. If the difference is quantitative, however, then is should be malleable based on both the type and values of stimuli being presented, as well as effects of training. So if we see improvement then the limit that has led to the absence of modulation in the earlier experiments is not at the data processing stage, but rather specifically at the stage of audiovisual integration. Under this perspective, one would expect the possibility of integration capacity to be subject to improvement through training.

By employing an intermediate SOA of 450 ms, Experiment 7 will test whether the capacity of audiovisual integration can be increased through training. We expect that participants will show an increase in audiovisual integration capacity, specifically at an intermediate SOA that is used for training, but that this may also generalize to other SOAs that are included in the experiment. Specifically, we are expecting to find criterion effects wherein repeated training on a 450 ms SOA condition will show improvement in the 450 ms SOA, with the possibility of near-transfer effects to other proximal SOAs such as the original 200 and 700 ms conditions.

Method

Participants. 36 participants were recruited, but 10 of them failed to attend both testing sessions, or had a computer error during recording, meaning we were left with 26 complete and viable data sets. All participants were recruited from an undergraduate research participant pool, and were compensated with partial class credit. As per the procedure employed in the previous experiments, we calculated a 95% confidence interval around 50% (chance responding) over all trials. We then removed the data of any participant who was performing within the 95% CI on average across all 8 conditions. Five participants were removed in this way, so the final sample consisted of 21 participants, with a mean age of 20.2, including 17 right-handed and 16 females. None of the participants took part in any of the previous experiments. Each participant signed up for two 1-hour testing sessions, which were always scheduled for consecutive days. On Day 1, the participant initially completed a Test Block followed by a Training Block. On Day 2, the participants completed a Training Block, followed by a Test Block.

Stimuli. All stimulus and presentation parameters were identical to the previous experiments, with the exception of including an addition SOA of 450 ms.

Design and Procedure. Each test block orthogonally varied the SOAs (200, 450, 700 ms), the validity of the stimulus (valid, invalid), and the number of visual stimuli changing (1, 2, 3, or 4). Each block consisted of trials with the orthogonal combinations of factors, and participants completed 4 test blocks in each testing session, for a total of 96 trials.   The training block consisted of only a single SOA (450 ms), but still contained the combination of validity and number of stimuli changing as before. Each training block contained 3 repetitions of the 8 combinations of validity and number of stimuli, making for 24 trials in each block. Participants completed 10 training blocks in each testing session. As such, each testing session involved a total of 336 trials (96 in the test blocks, 240 in the training blocks). Participants were offered the chance to complete a practice block consisting of 12 randomly chosen trials before beginning their first test block and their first training block of each session. Trial order was randomized in practice and in experimental trials and validity was collapsed for analysis purposes.

Model Fitting. Data were modelled for each set of conditions in the same way as has been done for the previous experiments. The outcome with the smallest RMSE was selected, and this process was done independently for each participant and SOA condition. Successful model fit was confirmed by average RMSE of 0.068, 0.047, and 0.035 for 200, 450, and 700 ms conditions, respectively.

Results

Raw response rates were calculated for each combination of SOA and number of locations changing for each participant within each block (Test 1, Training 1, Training 2, Test 2). Means for each SOA and number of locations changing both before and after training were calculated, and estimates of audiovisual integration capacity were determined. As in all previous experiments, the first comparison involved single sample t-tests comparing each capacity to a normative value of 1 item, to show which of the conditions resulted in a capacity that exceeded one item (see Figure 17 for capacity estimates). Before training, the average capacity for 200 ms was found to be .751, t(20) = -2.87, *p* = .010, meaning the capacity was significantly *less than* 1.

The capacity for 450 ms was 1.219, t(20) = 1.62, *p* = .121, meaning capacity was statistically equivalent to 1. For 700 ms, capacity was 1.997, t(20) = 3.97, *p* = .001, significantly greater than 1 item. Looking at single sample t-tests on the post-training data reveal changes in the capacity of audiovisual integration relative to the reported limit of 1 object. The 200 ms capacity was now statistically equivalent to 1 (*K* = .996, t(20) = -.026, *p* = .979). Additionally, the 450 ms capacity increased to a point where it was now significantly greater than 1 (*K* = 1.800, t(20) = 3.284, *p* = .004), while 700 ms capacity remained significantly greater than 1 (*K* = 2.129, t(20) = 3.662, *p* = .002).

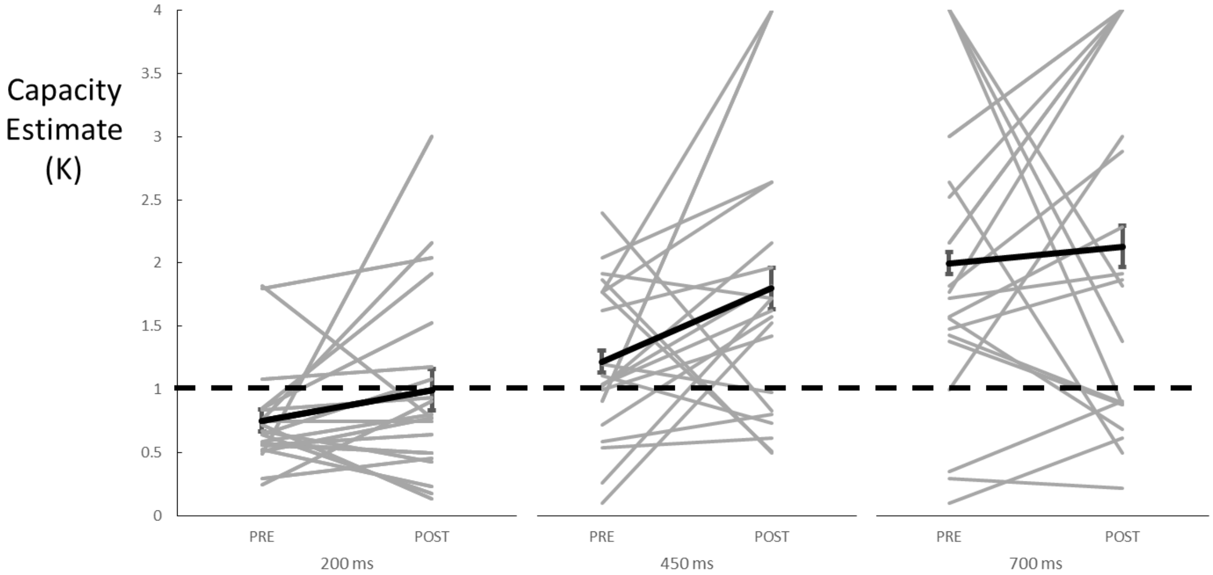


Figure 17. Capacity measures (*K*) for each participant in each SOA condition before and after training (in grey), as well as means (in black) and standard errors. Dotted line indicates the critical capacity limit of 1 item.

In order to ascertain whether training was effective in significantly increasing the capacity of audiovisual integration in any or all of the SOA conditions, paired-sample t tests were also conducted comparing the pre- and post- training scores with each other for each of the SOAs. As discussed in the introduction, I expect to find criterion effects leading to an increase in the 450 ms SOA condition, and examined the possibility of near-transfer effects to the other

SOAs. Paired sample t-tests on the 200 ms condition (t(20) = 1.498, *p* = .150) and 700 ms condition (t(20) = .386, *p* = .704) revealed no significant improvement, although Figure 18 shows that there was a numerical increase in capacity in both cases. On the critical 450 ms condition, there is a significant increase in capacity as a result of training (t(20) = 2.111, *p* = .048). So we find criterion effects, but no near-transfer effects in the capacity of audiovisual integration.

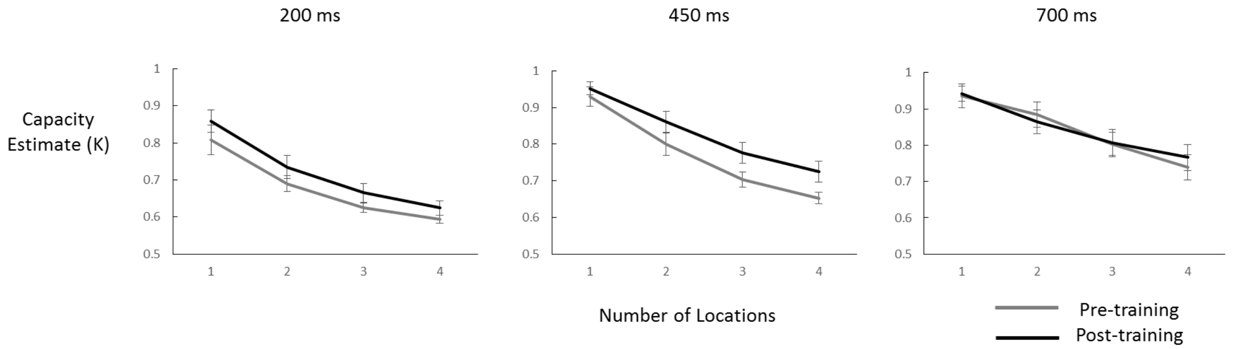


Figure 18. Proportion correct by training (pre, post) and number of objects to be tracked for 200, 450, and 700 ms SOAs in Experiment 7. Error bars indicate standard error.

In order to examine the conditions under which training was most beneficial, a 3-way ANOVA was performed on the proportion correct data, with factors of Training (2; Pre-training, Post-training), SOA (3; 200, 450, 700 ms), and Number of Locations (4; 1, 2, 3, 4). The full results of the ANOVA are displayed in Table 6. The data (shown in Figure 18) show main effects of SOA (*p* < .001) and Number of Locations (*p* < .001), both of which mirror findings from the previous experiments. That is to say, response accuracy increases as a function of increasing SOA and as a function of decreasing number of locations to be tracked. An SOA x Number interaction (*p* = .008) was decomposed by means of a Tukey's HSD (*p* < .05) test, which revealed in 450 and 700 ms SOAs, there was consistent decrease in accuracy with each additional location to be tracked, while at 200 ms proportion correct seemed to be approaching an asymptote as it approached 3 locations were to be tracked.

While the purpose of the training block was to provide repetition training to participants between the pre- and post-test sessions, looking at the data across training blocks allowed for the consideration of improvement during the sessions. Proportion correct data for each participant and for each number of locations was averaged across the first 5 blocks and the second 5 blocks of each session, yielding 4 bins of data according to day (one, two) and block (1-5, 6-10), which is displayed in Figure 19. While the capacity data we examined earlier show clear training effects between pre- and post- tests, there does not appear to be any improvement occurring during the training sessions themselves. In fact, submitting this data to a 2 (day; first, second) x 2 (half; first, second) repeated measures ANOVA shows that response accuracy decreases slightly (from .815 to .772; F(1,20) = 9.88, MSE = .016, p = .005, η_p^2 = .331) between the first and second halves of each session (which can be attributed to fatigue effects), and was slightly lower for the second day (.786) than the first day (.801; F(1,20) = .971, MSE = .021, p = .336, η_p^2 = .046), with no significant interaction (F(1,20) = .742, MSE = .017, p = .399, η_p^2 = .036).

Discussion

The findings clearly support criterion training effects in the capacity of audiovisual integration. When participants completed 20 short training blocks, over two days, with an SOA of 450 ms, their capacity of audiovisual integration was significantly increased. In light of previous research, this extends the findings from the previous experiments by showing that not only is the capacity of audiovisual integration sensitive to paradigmic factors, it also can vary within an individual by means of training.

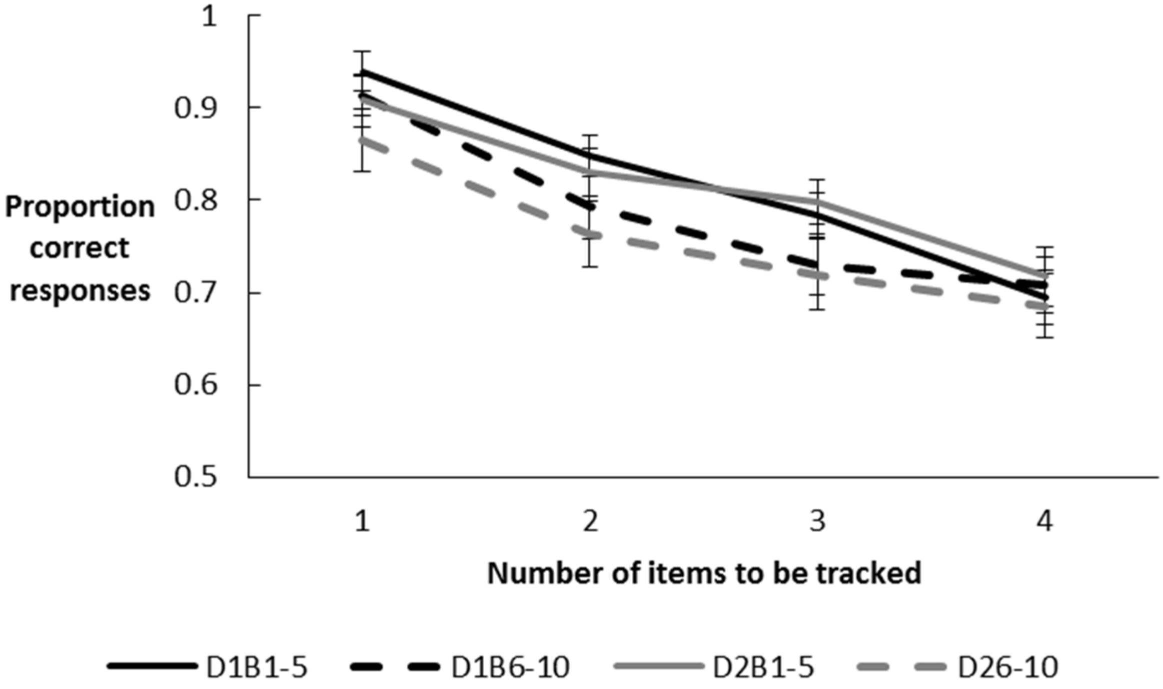


Figure 19. Proportion correct for training blocks, sorted into four bins (e.g. D1B1-5 = Day 1; Blocks 1-5). Error bars indicate standard error.

The failure to observe transfer effects into the 200 and 700 ms conditions can be explained by means of two separate mechanisms. In the 700 ms SOA condition, it seems that capacity was already at a high value, and therefore there may not have been enough room for improvement remaining. In the 200 ms condition, we were interested in whether there would be any movement from the baseline findings demonstrated in Experiment 1. While we do not observe significant improvement between pre- and post- training blocks for 200 ms SOA, we do note that the capacity itself moves from .751 (significantly *less* than 1) to .996 (statistically *equivalent* to 1), even though the difference between the two capacity values themselves is not significant. This indicates that there is some malleability present in the capacity of audiovisual integration, even at an SOA (200 ms) that has previously been shown to be beyond the ability of participants (with the possible exception of Experiment 6, although see the discussion there). If capacity is limited to a certain number of objects at a given SOA, then it should be restricted to that limit regardless of any training or practice effects. This would indicate a qualitative difference between the task which Van der Burg (personal communication, 1 May 2015) argues is an audiovisual integration task at 200 ms SOA, and is an auditory cueing task at 700 ms SOA. If the difference was qualitative, then at a given speed of data-limited presentation there should be no possibility of variation through training, or any other manipulation of stimuli, but the data show this is not the case. If it was a quantitative difference, however, then capacity at a given SOA should be malleable through training. The findings in the present research lend support to a perspective describing the difference between audiovisual integration capacity over various SOAs as quantitative rather than qualitative.

An additional piece of evidence to dispute Van der Burg's account of a qualitative difference between integration at 200 and 700 ms is provided by submitting the data from Experiment 7 to a linear trend analysis. Capacity estimates for each SOA before training were submitted to a one way ANOVA, along with a linear trend analysis. A main effect (F(2,60) = 13.374, MSE = .623, *p* < .001) indicated that there was a significant difference between SOAs, with Tukey's HSD post-hoc tests (*p* < .05) indicating a significant difference between 200 ms and 700s, as well as between 450 ms and 700 ms (with no significant difference between 200 ms and 450 ms). The linear trend model was a significant fit with the data (*p* < .001), indicating that with an increasing SOA we see an incrementally increasing capacity for audiovisual integration. For the post-training data, there was once again a significant main effect of SOA (F(2,60) = 5.643, MSE = .787, *p* = .006), but in this case the only significant difference was between 200 ms and 700 ms SOAs. The linear trend analysis was once again significant (*p* = .002). So, both before and after training, capacity can be modelled based on SOA, which is an argument in favour of the account for a quantitative difference in audiovisual integration capacity.

In order to further rule out Van der Burg's account of a qualitative difference it would be useful to test a greater number of individual SOAs. If the difference is qualitative, then it should be possible to determine a clear border at which the capacity of audiovisual integration changes from being necessarily less than one to being possibly greater than one. This difference should also be impervious to training. If, however, the difference is quantitative then we should observe a relatively linear trend between different SOAs which can be improved via training. The current findings provide some evidence for the quantitative account, but research of this kind could provide more direct responses to criticism. That being said, there is no evidence to show that it is possible for the capacity of integration at 200 ms to *exceed* one item, although future experimentation using 200 ms for the training blocks (or with increasing difficulty through the training as per Ahissar & Hochstein (1997)) could help in confirming or refuting this possibility.

Table 6 Summary of repeated measures ANOVA of fitted proportion correct data

Metric	*df*	F	MSE	*p*	η_p^2
Training (T)	1,20	1.97	.076	=.176	.090*
SOA (S)	2,40	34.42	.026	**<.001**	.632***
Number (N)	3,60	264.22	.004	**<.001**	.930***
S x T	2,40	0.93	.034	=.404	.044*
S x N	3,60	0.62	.003	=.604	.030*
T x N	6,120	3.04	.004	**=.008**	.132**
S x T x N	6,120	1.17	.004	=.330	.055*

Note: Statistical significance in bold

As shown in Figure 19, performance decreased when comparing the first five blocks of training on either day with the second five blocks. There was also a slight decrement comparing performance on Day 1 to Day 2. This seems counterintuitive since participants did show increase in performance via training during test blocks.

While training was observed when comparing between the test blocks before and after training, what was found within the training blocks themselves was somewhat more ambiguous. One possible reason for this is a fatigue effect, as continuous repetition of 10 training blocks may lead to drifting of attention of participants. The increase of performance once the second test block is underway comes from the re-inclusion of different SOAs, which increase interest and vigilance, allowing demonstration of what has been learned during the training blocks.

Chapter 5 – General Discussion

The purpose of this dissertation was to investigate the capacity and the nature of audiovisual integration. Over a series of seven experiments, I have found that the capacity of audiovisual integration is dynamic, subject to variation based on stimulus factors and between individuals, and can be greater than one object. This finding is in opposition to findings of Van der Burg et al. (2013) who proposed that there is a strict limit of one object on the capacity of audiovisual integration. However, as discussed throughout the dissertation, their results are based on an extremely high perceptual load (Lavie, 2005) and stimulus presentation rates that exceed what is possible for the human visual system to reliably process (Norman and Bobrow, 1976). By reducing load, and in doing so making the task of a more reasonable difficulty level, it has become clear that people can bind more than one visual stimulus to an auditory stimulus. Beyond that initial finding, I have provided a framework of individual difference and stimulus level factors that modulate the capacity of integration.

Summary of Findings

In Experiment 1, the basic paradigm used by Van der Burg et al. (2013) was altered by slowing down the presentation rate and reducing the number of visual stimuli present in the display. This manipulation was enough to show that when 8 (rather than 16 or 24) visual objects were presented, and when the rate of change of stimuli was 700 ms (slow SOA), participants were able to bind more than one object. At the same time, when presented with stimuli with reduced visual load but at 200 ms (fast SOA), the capacity of integration remained less than one item. Experiment 1 also employed EEG recording, which revealed that the reason for the limited capacity under fast SOA conditions was due to a lack of reliable information being received by the visual cortex. Given that the perceptual system was not receiving usable information, it would be strictly impossible to integrate more than one item due to a limitation of the availability of visual information, not a constrain of audio-visual capacity per se.

Having shown that the capacity of audiovisual integration *can* exceed one item, Experiments 2 through 4 were designed in order to fully break down the factors present in Van der Burg et al.'s (2013) experimental paradigm. This was done in order to compare the parameters affecting audiovisual integration to similar factors that are known to affect unimodal perception. Considering the degree to which perceptual systems are dynamic (e.g. Fujisaki et al., 2004; Marois & Ivanoff, 2005; Mazza & Caramazza, 2011), I wanted to test whether similar factors such as proactive interference (Lustig, May, & Hasher, 2001; Makovski & Jiang, 2008) and temporal predictability (Thomaschke & Dreisbach, 2013; Shin & Ivry, 2002) will play a role in multimodal integration. Experiment 1 had a critical stimulus that occurred at a predictable time and had a high level of proactive interference. In Experiment 2, the critical stimulus remained predictable, but the degree of proactive interference was reduced. Experiment 3 returned to high levels of interference, but with a temporally unpredictable critical stimulus. Finally, Experiment 4 had a low level of interference and an unpredictable critical stimulus. In sum, these four experiments revealed that capacity of integration is maximised under conditions of intermediate difficulty. When only one of the two factors was 'difficult' (i.e. high predictability / high interference OR low predictability / low interference), capacity was increased on 700 ms trials. Conversely, when both factors were difficult or both were easy, capacity was not increased.

Experiments 5 and 6 focused on stimulus factors that may play a role in audiovisual integration capacity. Experiment 5 used crossmodally correspondent stimuli (Marks, 1987; Parise & Spence, 2009) to test whether congruent stimuli would show higher levels of integration capacity than incongruent ones. Results show that this was indeed the case, although the facilitation was not sufficient to allow the capacity of integration to exceed one item at a 200 ms SOA. Experiment 6 focused on perceptual chunking (Gobet et al., 2001; Gilbert, Boucher, and Jemel, 2014), providing visual connections between changing visual stimuli to encourage participants to perceive visual stimuli as one complex object rather than multiple simple ones. In this case, integration capacity was increased at both 700 ms and 200 ms SOAs. This was the first case in which a 200 ms SOA was found to have a capacity exceeding one item, and called into question whether binding a single complex object was phenomenologically the same as binding multiple simple objects.

Finally, Experiment 7 looked at the potential for audiovisual integration capacity to increase within an individual via training (Garner et al., 2014; Brehmer et al., 2012). By training on an intermediate SOA of 450 ms, it was shown that participants could increase from having an integration capacity statistically equivalent to 1 to being significantly greater than 1. This indicates, once again, that the capacity of integration is flexible and can be influenced by many of the same factors as unimodal perception.

The remainder of this chapter will consider the relevance of this finding to other fields of study, as well as its implications in both theoretical and applied settings. First, I will discuss the findings in light of Norman and Bobrow’s (1975) work on data- and resource-limited processes. This will be followed by a consideration of how the interplay between temporal and stimulus factors are borne out in this experimental series. Capacity estimates will be discussed in terms of differences between simple and complex objects, in light of Alvarez and Cavanagh’s (2004) work on the same topic. I will then discuss the data with regard to the original research question – is there an overarching limit of one on the capacity of audiovisual integration, or is it fluid and dynamic, owing to various factors in the environment. Finally, some ideas for future research opportunities will be proposed in both basic and applied settings.

Data and Resource Limits

Norman and Bobrow (1975) discuss the difference between data-limited and resource-limited processes in perception. Their theory is based on the principle that for any system, processing resources are limited to some extent. If a process is resource-limited, then when available resources are not sufficient to allow for successful processing of stimuli, we observe what they call a "smooth degradation on task performance, rather than a calamitous failure" (Norman & Bobrow, 1975; p. 45). However, Norman and Bobrow go further with their argument, discussing the existence of data-limited processes. If a signal is difficult enough to perceive that, regardless of the amount of resources directed towards, it is impossible to sufficiently process it, then we have a data-limited process. Put another way, if using all of one's resources is still not enough to solve the problem, then it is not resource limited, but data limited. With data-limited signals, the issue is with the quality of the input signal – wherein the stimulus cannot be perceived regardless of how hard one tries to perceive it. An example of this would be a faint, near-threshold signal being presented in a noisy environment. Even if all perceptual and cognitive resources are devoted to attempting to hear the signal, it may be impossible to do so based on the relative loudness of the signal and the noise.

In Experiment 1, EEG recording was included within the experimental paradigm to examine the potential that the 200 ms SOA used by Van der Burg et al. (2013) represented a data-limited process, and that it was this temporal limit that set the upper bound of AV integration because of the paucity of visual information. The amplitude of visual N1 serves as an index of discrimination for attended visual stimuli (Mangun & Hillyard, 1991; Vogel & Luck,

2000) and a lack of modulation was observed in response to the number of polarity changes per frame in the 200 ms condition. In the 700 ms condition, however, modulation was apparent, with an increase in N1 amplitude when there were more locations changing on a given trial. Assuming that an individual completing this experiment has a certain amount of resources that they were able to contribute to the task, this disparity between the two SOAs can be used as evidence for the data-limited nature of the 200 ms presentation rate. In the 700 ms condition, a participant puts in as many resources as necessary to perceive, and keep track of the changing locations. The amount of resources they need to devote to the task increases with increasing difficulty (as indexed by number of locations changing), and this allocation of resources is exhibited by both behavioural (proportion correct and capacity) and electrophysiological data. In the 200 ms condition, however, no such modulation is observed, because participants are using all of their resources and are still unable to successfully perceive the stimuli. Given that the visual N1 indexes attention to stimuli, it seems strange that it should not show any index of attention allocation in the 200 ms conditions. One might have expected maximal attention allocation in an attempt to disambiguate the stimuli being presented. However, since the visual stimuli are being presented at a rate of speed that is not reliably perceived by participants (as per the behavioural data), it provides support for the argument that 200 ms trials are governed by a data-limitation, and that no amount of increased effort would be able to improve performance in this task, and as such no index of discrimination should be expected in the N1.

Experiments 2 through 4 reinforced these findings by manipulating proactive interference and temporal predictability, while maintaining the critical comparisons between 200 ms and 700 ms SOAs. It was found that, within 700 ms SOA conditions, there were effects of both factors. Specifically, when difficulty was at an intermediate level, performance was increased to the point where capacity of audiovisual integration was able to exceed one item. When difficulty was high or low, integration capacity did not exceed one. While there was modulation of capacity under changing conditions within the 700 ms SOA, under a 200 ms SOA there was no modulation, with the capacity remaining firmly limited to below one item (see Figure 9). This lack of modulation fits neatly with the perspective of 200 ms being governed by a data limit. Under a 700 ms SOA, changes in difficulty led to modulations in the amount of resources that participants had to devote to the task, leading to changes in the capacity of audiovisual integration. Under a data-limited 200 ms SOA, however, there was no modulation by any stimulus factors because regardless of how many resources are devoted to the task, the stimulus information was not clear enough to be perceived accurately.

Experiment 5 also provided findings compatible with the data-limit argument. Employing a manipulation of crossmodal congruency by matching high pitches with light coloured dots (and low pitches with dark coloured dots; as in Marks, 1987) provided additional information that allows for an increase in integration capacity, but only under 700 ms SOA. 200 ms remained data-limited to a capacity less than one. However, Experiment 6 employed perceptual chunking, providing participants with lines that connected the vertices of dots as they changed (as per Gobet et al., 2001) and seemingly breached the data-limited nature of the fast SOA, as the group average capacity of integration under the 200 ms condition exceeded one item. While this seems to be at odds with the argument that has been presented to this point, it is also possible that perceptual chunking occurred at an early level, and that participants were in fact binding one piece of complex information (e.g. a line, triangle, or quadrilateral) rather than a greater number of simple stimuli (Alvarez & Cavanagh, 2004). From this perspective, we can conceptualize the findings under a data-limit wherein it is still only possible to successfully perceive one item, but that the item is not limited to being simply a dot (see the discussion of simple vs. complex stimuli below for more on this).

Finally, Experiment 7 again revealed that 200 ms SOA is data limited, as it was unable to improve significantly via training, while the newly instated 450 ms SOA did show some degree of improvement, and 700 ms SOA was already significantly greater than one item. If a particular SOA is truly data-limited, then by definition it is not possible for it to be improved – not by changing stimulus factors and not by providing facilitatory stimulus factors. It could, however, be improved over a time by engaging in a long-term training regimen. Francis and Nusbaum (2009) showed that training with acoustic cues allowed participants to increase their ability to make sense of speech sounds, and in doing so eliminated what was previously shown to be a data-limited process. Additionally, we see data limits change over the life span, with increasing maturity of the perceptual and memory systems leading to a change in the threshold of data limitation (Karatekin, 2004). While Experiment 7 shows a capacity estimate of greater than one, it is also possible that this is an artefact of participants integrating one more complex object (e.g. a line) rather than two (or more) simple dots. Since 200 ms seems to be limited in this way, then we know there is no way for it to improve. As we saw a 450 ms SOA modulate through training, it follows that it is not data-limited, but resource-limited, although limited to a greater extent than is the 700 ms SOA.

Overall, the findings from this series of experiments support the argument of Norman and Bobrow (1975), that dependent on the stimuli that are being presented a task may be data-limited or resource-limited. This indicated that Van der Burg et al. (2013) set an artificial capacity limit of one item based on an experimental series that was affected by data-limited processing. This errant assignment of a limit did not consider the fidelity of incoming sensory information, and thus the data coming from the current set of experiments provide a more comprehensive view of the working of the capacity of audiovisual integration. Through seven experiments, six of them provide clear evidence in favour of this perspective, while Experiment 6 provides evidence that may be contradictory, but which can be easily interpreted in favour of Norman and Bobrow's (1975) argument. An interesting idea for future study would be to consider the point at which a data limit ceases to exist in this experimental paradigm. Preliminary results indicate that 200 ms is definitely data-limited, 700 ms is definitely not data-limited (limited only by available resources), and 450 ms also seems to be beyond the range of a data limit. To answer this question, an experiment should examine SOAs between 200 and 450 ms, considering the point at which it becomes possible for the capacity of integration to exceed one under normal conditions. Additional ideas for future study will be discussed in a later section.

The future studies discussed above could determine the point at which a data-limit is no longer affecting the capacity of audiovisual integration, and could also shed additional light on the extent to which these findings related to audiovisual integration capacity can be mapped onto Holcombe and Chen's (2013) findings on multiple object tracking. A finer scale of temporal resolution could allow for closer connections to be made between these two fields, and will permit closer analysis of how these temporal factors interact with the stimulus factors that have been the focus of this dissertation.

Capacity for simple vs. complex objects

There has been a debate in the literature about the nature of visual working memory span – namely, is it measured strictly by a number of objects, or rather by a combination of number of objects and complexity of those objects. As discussed in Chapter 1, Awh, Barton, and Vogel (2007) propose that the capacity of visual working memory is around 4 items, and that this limit is not affected by the level of complexity of items. Alvarez and Cavanagh (2004), on the other hand, provide evidence that the capacity of visual working memory is limited by both the number of objects, and the relative complexity of those objects. The data coming from Experiment 6 seem to support the findings of Awh, Barton and Vogel (2007). Experiments 1 through 5 all indicate that presentation of visual stimuli with an SOA of 200 ms results in a data-limited process that cannot be aided by resource allocation. Therefore, at 200 ms, it is simply not possible to track, and subsequently integrate, more than one visual item. In light of the findings from Awh, Barton, and Vogel (2007), however, it is possible that this single item that is being tracked could be a complex one, such as a line or a triangle rather than a single dot. In this way, the apparent capacity within the current paradigm could be increased, even while the actual enumeration of objects being tracked is still limited to one item.

In Experiment 6, however, the capacity of audiovisual integration at 200 ms does improve to the point where it is greater than one. Given the argument that has been made about data-limited processing, it would be surprising for the process to be released from these limits in Experiment 6 only. What makes more sense is that participants were indeed using perceptual chunking, as intended, and were tracking and integrating to a single more complex object: they were not tracking two dots, but rather the orientation of a line connecting those dots. Looking at the data from this perspective indicates that the true numerical capacity of integration is still one item at 200 ms, but that the functional capacity can be increased by means of perceptual chunking. This fits with Awh, Barton, & Vogel's (2007) conceptualization of working memory, wherein the same number of objects can be held in visual working memory (approximately 4), regardless of complexity. Here, it seems that at a 200 ms SOA, only 1 visual object can be integrated with an auditory stimulus, but that this object can be either simple (a dot) or complex (a line or polygon). As this is the case, we see that more information can be processed and integrated when the visual stimulus is complex rather than simple, even if the numerical value for the purposes of integration is still one. Additionally, the lack of a significant interaction between SOA and vertices in Experiment 6 indicates that the degree of perceptual chunking that is possible does not vary between 200 and 700 ms SOAs. This is an additional piece of evidence in the case being built against a qualitative difference between integration at 200 and 700 ms SOAs, as behaving in a similar manner would indicate that similar processes are occurring in each.

Is there a maximum capacity limit for audiovisual integration?

This dissertation has argued that there is not a general, overall limit of one item for the capacity of audiovisual integration. While this was proposed on the basis of some earlier research (Van der Burg et al., 2013), the current evidence support a more dynamic view of audiovisual integration capacity – one that is flexible, and based on differences in stimulus factors and individual experience. While this is more in line with previous unimodal research as well (Klingberg et al., 2002; Cowan, 2001), it is important to consider whether there is any maximum capacity limit, or whether the capacity of audiovisual integration is truly limitless.

The findings from this experimental series indicate that there may not be a clear limit to the capacity of audiovisual integration. For example, in Experiment 7 there are several participants who perform at ceiling – a capacity of 4, which is the maximum possible in this paradigm – in both the intermediate 450 ms SOA as well as the slower 700 ms SOA (see Figure 18). This experiment employed the ideal combination of high proactive interference and high predictability of the target stimulus, although it did not include any congruency factors. However, while these findings indicate that there is no apparent limit on the capacity of integration, it is still likely that there is some cap in existence. Just because a participant is able to perform at ceiling on a given task does not mean that, on a more difficult task, a limit would not be reached. Returning to the research done by Van der Burg et al. (2013), we could present a maximum of 8 changing locations to participants, and see if any of them had a maximum capacity of 8. This manipulation would eliminate any ceiling effect that may be in existence based on some participants performing at the maximum possible level, and in doing so would permit for more accurate estimates of audiovisual integration.

There is also evidence from unimodal research that it is not possible for capacity to be truly unlimited. In the case of visual working memory – which is a prerequisite for audiovisual integration – Cowan (2001) discusses a limit of around 4 items. Even in ideal stimulus conditions, a maximum of four visual objects can be held in visual working memory, and without being able to track more than four objects, it would not be possible to successfully integrate that many items with an auditory stimulus. While this evidence can be discussed in this light, it would be of interest to test this idea of a maximum limit directly. This could be done by presenting a greater number of changing locations, and observing whether there is a point at which capacity is limited, as well as examining the effects of training on visual working memory capacity.

Future Directions

While specific modifications to the current experimental series have been discussed within each respective chapter, there are some more general future research directions that will be discussed here. First, I will consider employing a greater number of SOAs in the paradigm, in order to answer questions pertaining to data-limited and resource-limited processes, as well as to test for whether there is a qualitative difference between integration at 200 or 700 ms (as proposed by Van der Burg). I will also discuss the potential for testing the number of auditory stimuli that could be integrated with one visual stimulus, since the current research has always tested for multiple visual candidates and a single auditory stimulus. Finally, applications to real-life scenarios in alert systems will be considered, with experimental suggestions arising in that field being proposed.

Multiple SOAs: Data vs. resource limits and qualitative vs. quantitative differences

One idea would be to look further into the effect of SOA on audiovisual integration capacity, in order to answer questions about two aspects of capacity. Firstly, it would provide an answer to the proposal by Van der Burg et al. (2013), that a 700 ms SOA is qualitatively different from a 200 ms SOA. The significant linear trend analysis discussed in Experiment 7 provides some preliminary evidence to this end, but an increase in the number of data points would lend additional support to the trend analysis. Secondly, it would allow for a closer examination of the stage at which this experimental paradigm begins to be affected by a data limit, rather than simply a resource limit. This experiment would involve essentially the same paradigm as the current research, but with a greater number of SOAs. Rather than using only 200 and 700 ms, participants would be presented with a number of SOAs between the two endpoints about which we already have information. Since the data derived from this experimental series have shown that 200 ms is likely a data-limited SOA, there is no reason to have any SOAs below that. Thus, for this proposed experiment, SOAs of 300, 400, 500, and 600 ms would be employed, so as to extract data that would be used to fill the space between our previous SOAs. By analysing the capacity figures at each SOA and submitting them to a linear trend analysis, an answer could be provided to whether there is some qualitative difference between 200 and 700 ms SOAs. That is to say, if there is an incremental increase in capacity with increasing SOA that is roughly linear in nature, that would be evidence for no qualitative difference – simply an increase in capacity as the task becomes 'easier' in terms of SOA. Additionally, by looking at the SOAs at the bottom end of this set, one could ascertain the point at which a data limit begins to affect integration. If, as shown in Experiment 7, there is a linear relationship between SOAs, with an increase in SOA associated with an increase in capacity, then as a decreasing SOA approaches a data-limited process, the slope of this line should flatten out. With the current data points of 200, 450, and 700 ms, this is difficult to look into, but with a greater number of data points it would be possible to examine the potentially changing slope at the bottom end of the SOA spectrum.

Further to this end, employing electrophysiological methods such as ERP analysis would be another way to access data regarding the data-limited nature of faster SOAs. In Experiment 1, ERP data contributed to the argument of a data-limited process because, while it modulated with the number of locations changing in the 700 ms condition, no such modulation was observed in the 200 ms condition. In addition to the behavioural results, this indicates that in the 200 ms condition, the incoming visual information was not sufficient to be processed and subsequently integrated with an auditory stimulus. By including ERP analysis with this multiple SOA experiment, similar data could be examined. At SOAs that show modulation by number of locations changing (or any other stimulus features), it can be concluded that the task difficulty is maintained by a resource limit which can be overcome by increasing the amount of mental resources allocated to the task. Conversely, if there are SOAs that show no modulation, as was seen in the 200 ms SOA of Experiment 1, it would seem that they are affected by a neutrally indexed data limit, which cannot be overcome by any amount of resource allocation.

Inverted modalities

The term ‘audiovisual integration’ refers to integration of any auditory and visual stimuli, and is not indicative of the direction of the binding. That is to say, there is no assumed difference between an auditory stimulus ‘being bound’ to a visual stimulus and a visual stimulus being bound to an auditory stimulus. In the current experimental series, while the capacity of audiovisual integration was being examined, all of the experiments presented participants with multiple (between 1 and 4) visual stimuli, with only a single auditory stimulus with which they could be integrated. That being said, it is also known that audition and vision do behave differently in binding situations, and so it would certainly be useful to repeat this experiment with the modalities inverted. Research into modality asymmetries in perceptual processing has shown that, generally, visual perception performs better than does auditory perception when the task is visual, but auditory perception has an advantage over visual perception when the task is temporal (Recanzone, 2003; Sandhu & Dyson, 2012; Wilbiks & Dyson, 2013a). As such, if a direct analogue to this study would be performed (with modalities inverted), I would expect the observed audiovisual capacity to be decreased, owing to the lesser capacity of auditory working memory (2; Saults & Cowan, 2007) when compared to visual working memory (4; Cowan, 2001), as well as the inferiority of spatial processing of auditory vs. visual stimuli.

The difficulty of this idea would be presenting a participant with multiple auditory tones simultaneously that they are able to easily parse from each other. Dyson and Quinlan (2003) used free-field speakers to present participants with auditory stimuli coming from 25 or 50 degrees to the left and right of their orientation, and also presented them at different pitches. This paradigm showed that location of an auditory stimulus serves as a feature of the stimulus itself, and as such could be differentiated from other locations. A similar setup would be employed in this proposed experiment. Within this setup, sinusoidal tones could be presented at each location that would alternate between two pitches, as an analogue for the black/white polarity shifts in the current research. A different type of tone (simulating a horn) would be presented from one of the speakers as a probe, with participants asked to respond as to whether the sound coming from that location changed, or not, at the same time as the visual cue was presented. An additional consideration related to this experiment would be an issue with free-field presentation of stimuli, wherein the perceived loudness of stimuli presented from different regions in space may be unequal, as well as being different for each participant based on their head size and shape (Sivonen and Ellermeier, 2006). This could be overcome by employing a head-related transfer function (HRTF), which models the presentation of stimuli from different locations around an idealized head, and creates a headphone presentation that simulates three-dimensional location presentations. This type of presentation eliminates many of the issues faced by free-field presentation, while maintaining the illusion of three-dimensionality (Drullman & Bronkhorst, 1999).

While this experiment seems largely analogous to what was done, auditory and visual stimuli do not always behave identically in experimental contexts or in practice. For one, stimulus effects tend to be preferentially used by people in visual stimuli than in auditory stimuli, while temporal effects are stronger in auditory stimuli (Burr, Banks, & Morrone, 2009; Wilbiks & Dyson, 2013a). For this experiment, that might mean that with an increase in the weighting of auditory stimuli (as compared to visual stimuli), temporal factors may be processed more efficiently. For example, research into the auditory mismatch negativity (MMN) has shown that people are sensitive to change in rapidly presented auditory stimuli at SOAs as fast as 100 ms (Sussman, Ritter, & Vaughan, 1998), while the current research has shown that rapidly presented visual stimuli cannot be accurately tracked at 200 ms SOA. Given this finding, one might expect the capacity of audiovisual integration in this case to be greater than in the current research, given the same SOA, while this may also be confounded by the differences in capacity of auditory working memory and visual working memory, as discussed previously. This would also mean that the point at which a data limit begins to affect the experiment would be at a lower SOA than 200 ms. However, given that auditory stimuli are inferior in terms of their utility with stimulus factors, it may be the case that as the number of locations to be tracked increases, the detrimental effect on response accuracy would be greater than was observed in the current paradigm with a greater number of visual stimuli and a single auditory stimulus. Whatever the specific findings, this study would be valuable in that it would show that audiovisual integration capacity can be measured both in terms of multiple visual candidates binding to a single auditory stimulus, as well as multiple auditory candidates to be bound to a single visual stimulus.

Entrainment of neuronal oscillators

An additional experimental idea stems from research into entrainment of neuronal oscillators. Repp and Su (2013) provide a review of this research, which share the underlying theoretical perspective that neurons in perceptual areas tend to experience times of higher and lower activation. Specifically, those neurons that track the frequency of stimulus occurrences are not always equally prepared to be stimulated, and go through oscillatory changes over time. Large and Snyder (2009) found that when a steady rhythm (or pulse) is present, the oscillatory activity of these neurons can be entrained to the frequency of presentation of those rhythms. This entrainment occurs optimally when the stimulus presentations are occurring at around 500 – 600 ms intervals. This speed of presentation is relatively close to the slower SOA used in the current research (700 ms), and as such an important question to consider would be whether rhythmic entrainment is playing a role in the increase in audiovisual integration capacity at 700 ms (as compared to 200 ms). Additional findings show that when stimuli are presented near oscillatory peaks, responses to them tend to be both faster (Lakatos, Karmos, Mehta, Ulbert, & Schroeder, 2008) and, more importantly to the current research, more accurate (Arnal, Doelling, & Poeppel, in press).

Given that presenting stimuli near these oscillatory peaks increases accuracy and speed of responding, it follows that doing so in the current paradigm would increase the capacity of audiovisual integration. If sensory cortical neurons are entrained to an appropriate rhythm, they should be better able to integrate stimuli that are presented with a tone than if the rhythm of presentation is not near an oscillatory peak. This might mean that capacity of integration at the 700 ms SOA in the current research is being aided not only by higher quality incoming perceptual information (as discussed in Experiment 1), but also by taking advantage of oscillatory peaks. To test this possibility, we could test a number of SOAs at varying distances from an oscillatory peak, with the expectation that loser proximity to a peak would lead to increased capacity. For example, assuming that oscillatory entrainment is optimal at a 600 ms SOA, we could present stimuli at SOAs of 500, 550, 600, 650, and 700 ms. If oscillatory entrainment is playing a significant role in the capacity of audiovisual integration, we would expect capacity to be greatest at 600 ms, and to drop off steadily as SOA moves away from 600 ms in both directions. Alternatively, if it does not play a role, we would expect an incremental increase in capacity with increasing SOA, as we have observed in the current research. In either case, it will be important to consider entrainment in the design of future experiments, as it may be an additional factor that is affecting the capacity of audiovisual integration.

Application to alert system scenarios

An interesting application of multisensory integration findings is in the field of designing alert systems that have a higher degree of salience to an operator. Providing signals that are present in both modalities – and that are integrated by a perceiver – allows for faster response to a potentially dangerous situation. This has been well researched within the field of redundancy gain, beginning with the findings of Biederman and Checkosky (1970), who found that reaction times were faster when participants were presented with two cues, either of which could provide enough information for a correct response, when compared to only having a single cue that could lead to a correct response. More recently, Girard, Pelland, Lepore, and Collignon (2012) found that redundant multisensory stimulation (visual and tactile) led to faster response times than unimodal stimuli, as well as being faster than when two stimuli were provided in a single modality. Ngo, Pierce, and Spence (2012) found that including an auditory or audio-tactile cue along with a visual signal led air traffic controllers to respond more quickly to situations requiring their attention (without any decrement in performance). Similarly, Ho, Reed, and Spence (2007) found that audio-tactile cues increased braking reaction time in a simulated car crash scenario, when compared with auditory or tactile cues alone. Given that the literature indicates that multisensory integration promotes quick responding in an alerting system, in the light of the current research it would be of interest to look at whether binding of multiple stimuli in one modality would also be useful in an alert signal context.

For example, in an air traffic control system scenario, there may be two (or more) pieces of visual information that require immediate attention at essentially the same time. In order to successfully alert an air traffic controller as to the multiple items needing attention, and audiovisual cue should be provided that will allow quick responding. Based on the findings from the current research, it is possible to make some predictions as to the type of stimuli that should be used. First, the two visual stimuli should flash at a rate of speed slower than 200 ms – perhaps at around 250 ms, as per Holcombe and Chen's (2013) findings. The auditory portion of the cue should be congruent with the visual stimuli, so for example if the visual stimuli are white, the tone should be high pitched. While these factors can be predicted based on previous research, a novel experiment should be designed that optimizes these factors for use in a real-life, air traffic control scenario.

General conclusions

The findings from the seven experiments in this dissertation indicate that it is possible for the capacity of audiovisual integration to exceed one item. Further to that end, the capacity of audiovisual integration is dynamic, subject to variation due to stimulus factors and training. When stimulus presentation was relatively fast, ERP data indicate that the incoming visual information is subject to a data limit (Norman & Bobrow, 1975), and that it is therefore impossible to improve one's integration capacity, regardless of the amount of resources devoted to it. Thus, the finding of Van der Burg et al. (2013) of a strict limit of capacity at no more than one item seems to be a simple artifact of this data limit rather than a true limit on audiovisual integration capacity. Only once the data limit has been alleviated can we observe the capacity of audiovisual integration based on stimulus factors and resource allocation.

Capacity is maximised when experimental stimulus factors are of a moderate degree of difficulty. When a relatively high number of stimuli are presented before the target of interest, and when the target of interest is presented at a predictable time, or when the opposite is true, it is possible for capacity to exceed one (at a relatively slow rate of presentation). Conversely, when both factors are difficult or easy, capacity does not exceed one. This is in line with findings of improved task performance under moderate levels of task difficulty (Anderson, 1990).

When auditory and visual stimuli are matched in terms of their pitch and colour, this provides an additional cue that increases audiovisual integration capacity, but again, only in the relatively slow condition. However, enabling participants to perceive multiple simple stimuli as one more complex stimulus increases the functional capacity of integration at both presentation rates. Finally, capacity of audiovisual integration can be trained within an individual. Participants moved from a capacity equivalent to one to a capacity greater than one at the speed at which they were trained.

While there may be some overarching maximum limit on the capacity of audiovisual integration, such a limit was not observed in this experimental series. Future research will provide definitive answers as to whether a maximum limit does exist, as well as to many of the questions asked above. What can be stated with certainty at this point is that there is not a maximum limit of one item, and that capacity varies based on stimulus factors and individual training effects. The capacity of audiovisual integration is dynamic, malleable, and flexible, dependent on the individual doing the integrating and on the specific situation into which they are placed.

Appendices

Appendix A – Consent form for Experiment 1 with EEG recording

Ryerson University Consent Agreement

AUDIO-VISUAL BINDING IN ADULTS:
CAPACITY OF AUDIOVISUAL BINDING

You are being asked to participate in a research study. Before you give your consent to be a volunteer, it is important that you read the following information and ask as many questions as necessary to be sure you understand what you will be asked to do.

Investigators: Jonathan Wilbiks, Ben Dyson; Department of Psychology

Jonathan Wilbiks
jwilbiks@psych.ryerson.ca
416-979-5000 x2186

Dr. Ben Dyson
ben.dyson@psych.ryerson.ca
416-979-5000 x2063

Purpose of the Study: This study is part of an ongoing research program where we hope to more fully understand the way in which the brain processes auditory and visual information, how information from different senses interact, and how those processes and interactions change as a function of age and expertise. We are hoping to test 24 individuals in this study, and wish to use only those individuals who self-report as having normal (or corrected-to-normal) hearing and vision.

Description of the Study: The study will take place in the HEAR Lab, located in the Psychology Research and Training Centre at 105 Bond Street, unless otherwise stated. Experiments will take approximately 2 hour so *please ask your experimenter now if you are unclear as to the time commitments of the current study*. Prior to the study, you will have been provided with an information sheet regarding how EEG (electroencephalography) is recorded, the study will have been explained to you and you will have been given the opportunity to take part in a practice block so you are familiar with the procedure. You will be given the chance to ask any questions you may have regarding the study, prior to reviewing the consent agreement. During the study, age, gender and handedness will be requested. After the study, you will be fully debriefed as to the purpose of the study, and given a further opportunity to ask questions.

You will complete 1 practice block, and 6 experimental blocks in all. In all of the blocks, you will see a series of dots arranged in a circle. At regular time intervals, a sample of these dots will change colour, alternating between black and white. On one of these alternations, you will hear a tone, and you should try to remember which dot(s) changed colour at the same time as you heard the tone. At the conclusion of the alternations, one of the dots will be highlighted in red, and you will then be to respond with the number **1** if that dot *did not* change colour at the same time as the tone, or respond with the number **2** if that dot *did* change colour at the same time as the tone. Please respond as accurately as possible. The number of dots that change could be 1, 2, 3, or 4, and will vary between trials. The timing of the alternations will vary between blocks.

What is Experimental in this Study: Previous research has examined how auditory and visual information integrate with one another. The study is experimental in the respect that we are investigating auditory and visual integration using a unique design. The study is also experimental in the respect that we manipulate the number of dots that change colour, as well as the timing of each alternation.

Risks or Discomforts: There are no known long-term risks associated with the recording of EEG (electroencephalography), although you might feel short-term discomfort as a result of wearing the electrode cap for a long period of time and a little messy as a result of the electrode gel application. ***If you have temporal-mandibular joint (TMJ) disease or any recurrent problems with your head or neck, then you should not take part.*** Effects of fatigue will be offset by providing breaks. *If you feel uncomfortable at any time during the preparation period or during the experiment itself, you may discontinue participation, either temporarily or permanently.*

Benefits of the Study: The potential benefits of the study for science and society are a greater understanding of how the cognitive processing of stimuli occurs. The studies may also offer avenues into how to tailor more aesthetically pleasing experiences as a result of understanding how the senses interact. However, there are no immediate benefits that you can reasonably expect from the study.

Confidentiality: Confidentiality will be maintained in all aspects of data dissemination. Only individuals involved in the research team will have access to a central password-protected electronic file with your data, but this data will not be linked to your personal information. All data will be stored for a minimum of 5 years after collection. Participants have the option of receiving a summary of their performance after participation, and should make this known to the researchers at the time of testing. Participants also have the option of removing their data from the study after participation, and should inform the experimenter before leaving the testing session if this is the case. Please note that after publication of a data set (usually no sooner than 3 months after participation) it is not possible to remove data.

Incentives to Participate: You will be completing the experiment for course credit, awarded either on the basis of participation or a walk-through in which the participant can take part in the study but not submit their data. Please indicate which incentive you require:

□ COURSE CREDIT (PARTICIPATION) □ COURSE CREDIT (WALKTHROUGH)

Voluntary Nature of Participation: Participation in this study is voluntary. Your choice of whether or not to participate will not influence your future relations with Ryerson University. If you decide to participate, you are free to withdraw your consent and to stop your participation at any time without penalty or loss of benefits to which you are allowed. At any particular point in the study, you may refuse to answer any particular question or stop participation altogether.
Questions about the Study: If you have any questions about the research now, please ask. If you have questions later about the research, you may contact.

Jonathan Wilbiks
jwilbiks@psych.ryerson.ca
001 416-979-5000 x2186

Dr. Ben Dyson
ben.dyson@psych.ryerson.ca
001 416-979-5000 x2063

If you have questions regarding your rights as a human subject and participant in this study, you may contact the Ryerson University Research Ethics Board for information.

Research Ethics Board
c/o Office of the Vice President, Research and Innovation
Ryerson University, 350 Victoria Street
Toronto, ON, M5B 2K3, Canada
001 416-979-5042

Agreement: Your signature below indicates that you have read the information in this agreement and have had a chance to ask any questions you have about the study. Your signature also indicates that you agree to be in the study and have been told that you can change your mind and withdraw your consent to participate at any time. You have been given a copy of this agreement. You have been told that by signing this consent agreement you are not giving up any of your legal rights.

Informed consent for study participation

____
Name of Participant (please print)

____________________________________ __________________
Signature of Participant Date

____________________________________ __________________
Signature of Investigator Date

Appendix B – Debrief form from Experiment 1 with EEG

Ryerson University **Debriefing Form**

Audio-Visual Binding in Adults: The Capacity of Audiovisual Binding

Dear Participant:

Thank you very much for you participation in our study. Your time and commitment to psychological research at Ryerson University is very much appreciated.

The study you took part in will contribute to ongoing auditory and visual research conducted in the H.E.A.R Lab. Our lab is dedicated to designing and implementing research studies that will help us better understand how the brain represents what we hear and see, and how this information is integrated.

The particular study you took part in was designed to assess how many visual stimuli can be bound with a single auditory stimulus. The number of visual stimuli that changed in any given trial could have been 1, 2, 3, or 4, but you were always asked to respond to one specific stimulus. Based on previous research, we know that when a tone is sounded you will bind a visual stimulus to that sound. The question we are looking at here is whether you can bind 2 (or more) visual stimuli to the same sound.

Our hypothesis is that you will be able to bind more than one, depending on the conditions that we set out. Previous research (Van der Burg et al., 2013) suggested that one is the maximum, but they used at least 16 circles (twice as many as we did) and 200 ms alternations (the fastest condition that we use). We expect that by removing distractors and slowing down the task, we will be able to increase the number of visual stimuli that you can bind to the auditory stimulus. Additionally, we believe that the mean amplitude of your brain activity during this task will increase as there are more visual switches to keep track of. Previous research (Vogel & Machizawa, 2004) into vision only showed that this amplitude increased from 1 up to 4, which they say is the limit of visual working memory. We believe that this increase in amplitude will indicate to us the capacity of audiovisual integration.

If you have any questions regarding your participation in this study, or would like to receive information about the results once they are available, feel free to contact Dr. Ben Dyson. We would be happy to provide you with your own data as well as the overall findings of our study.

Finally, if you are interested in taking part and learning more about visual and auditory perception research in the H.E.A.R Lab, feel free to contact Dr. Dyson.

References

Van der Burg, E., Awh, E., & Olivers, C. N. (2013). The Capacity of Audiovisual Integration Is Limited to One Item. *Psychological science*, *24*(3), 345-351.

Vogel, E. K., & Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature*, *428*(6984), 748-751.

Jonathan Wilbiks
PhD Student
Ryerson University
jwilbiks@psych.ryerson.ca

Dr. Ben Dyson
Professor of Psychology
Ryerson University
ben.dyson@psych.ryerson.ca

Appendix C – Representative consent form for Experiments 2 – 7

RYERSON UNIVERSITY

Ryerson University Consent Agreement

AUDIO-VISUAL BINDING IN ADULTS:
CAPACITY OF AUDIOVISUAL BINDING

You are being asked to participate in a research study. Before you give your consent to be a volunteer, it is important that you read the following information and ask as many questions as necessary to be sure you understand what you will be asked to do.

Investigators: Jonathan Wilbiks, Ben Dyson; Department of Psychology

Jonathan Wilbiks	Dr. Ben Dyson
jwilbiks@psych.ryerson.ca	ben.dyson@psych.ryerson.ca
416-979-5000 x2186	416-979-5000 x2063

Purpose of the Study: This study is part of an ongoing research program where we hope to more fully understand the way in which the brain processes auditory and visual information, how information from different senses interact, and how those processes and interactions change as a function of age and expertise. We are hoping to test 24 individuals in this study, and wish to use only those individuals who self-report as having normal (or corrected-to-normal) hearing and vision.

Description of the Study: The study will take place in the HEAR Lab, located in the Psychology Research and Training Centre at 105 Bond Street, unless otherwise stated. Experiments will take approximately 1 hour so *please ask your experimenter now if you are unclear as to the time commitments of the current study*. Prior to your participation, the study will have been explained to you and you will have been given the opportunity to take part in a practice block so you are familiar with the procedure. You will be given the chance to ask any questions you may have regarding the study, prior to reviewing the consent agreement. During the study, age, gender and handedness will be requested. After the study, you will be fully debriefed as to the purpose of the study, and given a further opportunity to ask questions.

You will complete 1 practice block, and 8 experimental blocks in all. In all of the blocks, you will see a series of dots arranged in a circle. At regular time intervals, a sample of these dots will change colour, alternating between black and white. On one of these alternations, you will hear a tone, and you should try to remember which dot(s) changed colour at the same time as you heard the tone. At the conclusion of the alternations, one of the dots will be highlighted in red, and you will then be to respond with the number **1** if that dot *did not* change colour at the same time as the tone, or respond with the number **2** if that dot *did* change colour at the same time as the tone. Please respond as accurately as possible. The number of dots that change could be 1, 2, 3, or 4, and will vary between trials. The number of alternations, and the timing of the tone will also vary between trials.

What is Experimental in this Study: Previous research has examined how auditory and visual information integrate with one another. The study is experimental in the respect that we are investigating auditory and visual integration using a unique design. The study is also experimental in the respect that we manipulate the number of dots that change colour, the number of total alternations, as well as the timing of each alternation.

Risks or Discomforts: There are no known long-term risks associated with this type of behavioural testing. The main discomfort you may experience will be tiredness, and this will be alleviated by being provided with several breaks between experimental blocks.

Benefits of the Study: The potential benefits of the study for science and society are a greater understanding of how the cognitive processing of stimuli occurs. The studies may also offer avenues into how to tailor more aesthetically pleasing experiences as a result of understanding how the senses interact. However, there are no immediate benefits that you can reasonably expect from the study.

Confidentiality: Confidentiality will be maintained in all aspects of data dissemination. Only individuals involved in the research team will have access to a central password-protected electronic file with your data, but this data will not be linked to your personal information. All data will be stored for a minimum of 5 years after collection. Participants have the option of receiving a summary of their performance after participation, and should make this known to the researchers at the time of testing. Participants also have the option of removing their data from the study after participation, and should inform the experimenter before leaving the testing session if this is the case. Please note that after publication of a data set (usually no sooner than 3 months after participation) it is not possible to remove data.

Incentives to Participate: You will be completing the experiment for course credit, awarded either on the basis of participation or a walk-through in which the participant can take part in the study but not submit their data. Please indicate which incentive you require:

□ COURSE CREDIT (PARTICIPATION)

□ COURSE CREDIT (WALKTHROUGH)

Voluntary Nature of Participation: Participation in this study is voluntary. Your choice of whether or not to participate will not influence your future relations with Ryerson University. If you decide to participate, you are free to withdraw your consent and to stop your participation at any time without penalty or loss of benefits to which you are allowed. At any particular point in the study, you may refuse to answer any particular question or stop participation altogether.

Questions about the Study: If you have any questions about the research now, please ask. If you have questions later about the research, you may contact.

Jonathan Wilbiks
jwilbiks@psych.ryerson.ca
001 416-979-5000 x2186

Dr. Ben Dyson
ben.dyson@psych.ryerson.ca
001 416-979-5000 x2063

If you have questions regarding your rights as a human subject and participant in this study, you may contact the Ryerson University Research Ethics Board for information.

Research Ethics Board
c/o Office of the Vice President, Research and Innovation
Ryerson University, 350 Victoria Street
Toronto, ON, M5B 2K3, Canada
001 416-979-5042

Agreement: Your signature below indicates that you have read the information in this agreement and have had a chance to ask any questions you have about the study. Your signature also indicates that you agree to be in the study and have been told that you can change your mind and withdraw your consent to participate at any time. You have been given a copy of this agreement. You have been told that by signing this consent agreement you are not giving up any of your legal rights.

Informed consent for study participation

____
Name of Participant (please print)

____________________________________ __________________
Signature of Participant Date

____________________________________ __________________
Signature of Investigator Date

Appendix D – Representative debrief form for Experiments 2 – 7

Ryerson University **Debriefing Form**

Audio-Visual Binding in Adults: The Capacity of Audiovisual Binding

Dear Participant:

Thank you very much for you participation in our study. Your time and commitment to psychological research at Ryerson University is very much appreciated.

The study you took part in will contribute to ongoing auditory and visual research conducted in the H.E.A.R Lab. Our lab is dedicated to designing and implementing research studies that will help us better understand how the brain represents what we hear and see, and how this information is integrated.

The particular study you took part in was designed to assess how many visual stimuli can be bound with a single auditory stimulus. The number of visual stimuli that changed in any given trial could have been 1, 2, 3, or 4, but you were always asked to respond to one specific stimulus. Based on previous research, we know that when a tone is sounded you will bind a visual stimulus to that sound. The question we are looking at here is whether you can bind 2 (or more) visual stimuli to the same sound.

Our hypothesis is that you will be able to bind more than one, depending on the conditions that we set out. Previous research (Van der Burg et al., 2013) suggested that one is the maximum, but they used at least 16 circles (twice as many as we did) and 200 ms alternations (the fastest condition that we use). We expect that by removing distractors and slowing down the task, we will be able to increase the number of visual stimuli that you can bind to the auditory stimulus. We also expect that by presenting less sets of dots before the set you need to remember, it will be increase your ability to remember those critical dots.

If you have any questions regarding your participation in this study, or would like to receive information about the results once they are available, feel free to contact Dr. Ben Dyson. We would be happy to provide you with your own data as well as the overall findings of our study.

Finally, if you are interested in taking part and learning more about visual and auditory perception research in the H.E.A.R Lab, feel free to contact Dr. Dyson.

Reference

Van der Burg, E., Awh, E., & Olivers, C. N. (2013). The Capacity of Audiovisual Integration Is Limited to One Item. *Psychological science*, *24*(3), 345-351.

Jonathan Wilbiks
PhD Student
Ryerson University
jwilbiks@psych.ryerson.ca

Dr. Ben Dyson
Professor of Psychology
Ryerson University
ben.dyson@psych.ryerson.ca

Appendix E – Manuscript Review by Erik Van der Burg received 1 May 2015

I did review a previous version of the manuscript for a different journal. The current manuscript improved in many ways, but I am still have some major comments which I address below. My major point is that the increased capacity in the 700 ms condition has nothing to do with multisensory integration. Below I specify my major concerns in more detail (not in order of importance).

Major

On p 6. The authors mention: "In particular, it appears the repeated failure of AV capacity to exceed 1under 200 ms SOA condition is due to the inability of the visual cortex to successfully code the number of changing locations in frames prior to the critical one." I disagree with this, as in my study (Van der Burget al. 2013) we found a capacity of ~4 in the visual condition. Importantly, the temporal properties were the same as in the auditory condition in which we found a capacity of 1.

I find Experiment 2 somewhat trivial. If there are less colour- or polarity changes before the target change, then the target becomes more unique, and thus captures attention. It is well known from the literature that an abrupt colour change does pop out (see some work from Jan Theeuwes). It is important to note that in the pip and pop studies, we added lots of noise to the visual system to camouflage the target colour change. Then a sound with one item can improve search. With regard to the present study, the effects reported are probably not due to more/less integration, but reflect a visual effect as the target colour change does pop out more often when it becomes unique. Perhaps the authors can plot K as a function of the number of frames prior the target frame. I guess that the improved K is due to the condition in which the target was preceded by just one frame.

In Experiment 3, the authors always used the fourth (out of five) frame as the target frame, and showed that this temporal knowledge improves K for the long SOA condition. Again, I don't think this has anything to do with AV integration. Instead participants get more temporal knowledge, and are therefore able to do the task better. In the pip & pop experiments and also in the Van der Burg et al. capcity studywe tried to remove as much temporal information as we were interested in the effects of the sound on search and not on the effects of temporal information on search. The results obtained from Experiment 3 are in my opinion not surprising.

This is my main point. I wonder whether the present finding has anything to do with the multisensory integration in the SOA = 700 ms condition. I think that if the authors increase the SOA more (say one or two seconds), then the performance should approach performance for a pure visual task. In the present case, with an SOA of 700 ms, I think that participants can do the task more or less visually. If the participants can do the task visually, or if the effects are visually driven effects, then the conclusion drawn by the authors is incorrect. I propose to do another

control experiment. In this experiment the authors manipulate the modality of the cue (like in the Van der Burg et al. 2013) study. So, in one block the auditory cue is replaced by a visual cue, and in another block they present an auditory cue (as in the present study). Obviously, they must manipulate the SOA to show that for the same participants the capacity increases as a function of SOA for auditory cues,
but not for visual cues. So, a visual cue with the same temporal properties is not able to improve performance.

Overall, the introduction is not that easy to follow, and I think that some clarification is required. For instance, I disagree with the authors that I manipulated the perceptual load by manipulating the set size. And if anything, there was no set size effect, so indicating that it was equally easy to detect the synchronized discs. Furthermore, I know the Holcombe and Chen study, but I don't see a link between their study and the Van der Burg et al. study (2013). For instance, they used a single or two visual events that moved with a certain speed. In our case, the discs were static, and not moving. The Holcombe and Chen task was very difficult and I don't think that is the case if we only present 2 elements. You can easily track around four elements simultaneously (as the authors mention in the introduction, and as have shown in the Van der Burg et al. study).

I believe that the EEG analyses are problematic as it is difficult to measure a target related ERP to a visual event in clutter. As a result, the ERP will contain lots of residual activity caused by the preceding distractors. This is obviously stronger in the dense condition (200 ms) than in the more static condition in which the preceding event was 700 ms away from the target event. It is therefore also not surprisingly that the N1 is reduced in the 200 ms condition compared to the 700 ms condition. Again, I believe this is a visual effect. Unfortunately, the authors have no visual only baseline condition to check this, or a no stim condition to correct for overlapping activity (see e.g. Van der Burg et al. 2011, and see some other studies by Durk Talsma).

Minor.

1. Please report the sizes etcetera in visual angle instead of centimeters.
2. What was the luminance of the stimuli?
3. I don't understand the discussion on p. 11 (top of the page).
4. The Los and Van der Burg (2013) study is a nice example of temporal preparation effects with integration. This might be of interest.

Signed: Erik van der Burg

References

Abikoff, H., & Gittelman, R. (1985). Hyperactive children treated with stimulants: Is cognitive training a useful adjunct? *Archives of General Psychology, 42*(10), 953.

Adler, R. F., & Benbunan-Fich, R. (2015). The effect of task difficulty and multitasking on performance. *Interacting with Computers, 27*(4), 430-439.

Ahissar, M., & Hochstein, S. (1997). Task difficulty and the specificity of perceptual learning. *Nature, 387*, 401-406.

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology, 14*(3), 257-262.

Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science, 15*(2), 106-111.

Anderson, K. J. (1990). Arousal and the inverted-u hypothesis: A critique of Neiss's "Reconceptualizing Arousal". *Psychological Bulletin, 107*, 96-100.

Arnal, L. H., Doelling, K. B., & Poeppel, D. (in press). Delta-beta coupled oscillations underlie temporal prediction accuracy. *Cerebral Cortex*.

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

Awh, E., Vogel, E. K., Oh, S. H. (2006). Interactions between attention and working memory. *Neuroscience, 139*(1), 201-208.

Baddeley, A. D., & Hitch, G. (1974). Working memory. *Psychology of Learning and Motivation, 8*, 47-89.

Bahrami, B. (2003). Object property encoding and change blindness in multiple object tracking. *Visual Cognition, 10*, 949-963.

Bettencourt, K. C., & Somers, D. C. (2009). Effects of target enhancement and distractor suppression on multiple object tracking capacity. *Journal of Vision, 9*(7), 9-9.

Bierderman, I., Checkosky, S. F. (1970). Processing redundant information. *Journal of Experimental Psychology, 83*(3p1), 486.

Bor, D., & Seth, A. K. (2012). Consciousness and the prefrontal parietal network: insights from attention, working memory, and chunking. *Frontiers in Psychology, 3*, 63.

Brehmer, Y., Westerberg, H., & Backman, L. (2012). Working-memory training in younger and older adults: training gains, transfer, and maintentance. *Frontiers in Human Neuroscience, 6*(63), 1-7.

Burr, D., Banks, M. S., & Morrone, M. C. (2009). Auditory dominance over vision in the perception of interval duration. *Experimental Brain Research, 198*(1), 49-57.

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

Cavanagh, P., & Alvarez, G. A. (2005). Tracking multiple targets with multifocal attention. *Trends in Cognitive Sciences, 9*(7), 349-354.

Chan, Y. M., Pianta, M. J., & McKendrick, A. M. (2014). Older age results in difficulties separating auditory and visual signals in time. *Journal of Vision, 14*(11), 13-13.

Chun, M. M. (2011). Visual working memory as visual attention sustained internally over time. *Neuropsychologia, 49*(6), 1407-1409.

Chun, M. M., Golomb, J. D., & Turk-Browne, N. B. (2011). A taxonomy of external and internal attention. *Annual Review of Psychology, 62*, 73-101.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity [Target article and commentaries]. *Behavioral and Brain Sciences, 24,* 87-185.

Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why? *Current Directions in Psychological Science, 19*(1), 51-57.

Crisinel, A.-S., & Spence, C. (2009). Implicit association between basic tastes and pitch. *Neuroscience Letters, 464*, 39-42.

Crowder, R. G. (1976). *Principles of learning and memory*. Oxford: Erlbaum.

Culham, J. C., Brandt, S. A., Cavanagh, P., Kanwisher, N. G., Dale, A. M., & Tootell, R. B. H. (1998). Cortical fMRI activation produced by attentive tracking of moving objects. *Journal of Neurophysiology, 80*(5), 2657-2670.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience, 18*(1), 193-222.

Doran, M. M., & Hoffman, J. E. (2010). The role of visual attention in multiple object tracking: Evidence from ERPs. *Attention, Perception, & Psychophysics, 72*(1), 33-52.

Drew, T., McCollough, A. W., Horowitz, T. S., & Vogel, E. K. (2009). Attentional enhancement during multiple-object tracking. *Psychonomic Bulletin & Review, 16*(2), 411-417.

Drew, T., Horowitz, T. S. & Vogel, E. K. (2013). Swapping or dropping? Electrophysiological measures of difficulty during multiple object tracking. *Cognition, 126,* 213-223.

Drew, T., & Vogel, E. K. (2008). Neural measures of individual differences in selecting and tracking multiple moving objects. *The Journal of Neuroscience, 28*(16), 4183-4191.

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review, 96*(3), 433.

Drullman, R., & Bronkhorst, A. W. (1999). Multichannel speech intelligibility and talker recognition using monaural, binaural, and three-dimensional auditory presentation. *The Journal of the Acoustical Society of America, 107*(4), 2224-2235.

Dyson, B. J. (2014) Practice, expertise, and aging. *The Encyclopedia of Adulthood and Aging.*

Dyson, B. J., & Quinlan, P. T. (2003). Feature and conjunction processing in the auditory modality. *Perception & Psychophysics, 65*(2), 254-272.

Evans, K. K., & Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features. *Journal of Vision, 10*(1), 1-12.

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.

Fiebelkorn, I. C., Foxe, J. J., & Molholm, S. (2010). Dual mechanisms for the cross-sensory spread of attention: How much do learned associations matter? *Cerebral Cortex, 20*(1), 109-120.

Folstein, J. R., & van Petten, C. (2008). Influence of cognitive control and mismatch on the N2 component of the ERP: a review. *Psychophysiology, 45*(1), 152-170.

Francis, A. L., & Nusbaum, H. C. (2009). Effects of intelligibility on working memory demand for speech perception. *Attention, Perception, and Psychophysics, 71*(6), 1360-1374.

Franconeri, S. L., Jonathan, S. V., & Scimeca, J. M. (2010). Tracking multiple objects is limited only by object spacing, not by speed, time, or capacity. *Psychological Science, 21*(7), 920-925.

Fujisaki, W., Koene, A., Arnold, D., Johnston, A., & Nishida, S. (2006). Visual search for a target changing in synchrony with an auditory signal. *Proceedings of the Royal Society B: Biological Sciences, 273*(1588), 865-874.

Fujisaki, W., Shimojo, S., Kashino, M., & Nishida, S. (2004). Recalibration of audiovisual simultaneity. *Nature Neuroscience, 7*(7), 773-778.

Gallace, A., & Spence, C. (2006). Multisensory synesthetic interactions in the speeded classification of visual size. *Perception & Psychophysics, 68*(7), 1191-1203.

Garner, K. G., Tombu, M. N., & Dux, P. E. (2014). The influence of training on the attentional blink and psychological refractory period. *Attention, Perception, & Psychophysics, 76*(4), 979-999.

Gilbert, A. C., Boucher, V. J., & Jemel, B. (2014). Perceptual chunking and its effect on memory in speech processing: ERP and behavioral evidence. *Frontiers in Psychology, 5.*

Girard, S., Pelland, M., Lepore, F., & Collignon, O. (2013). Impact of the spatial congruence of redundant targets on within-modal and cross-modal integration. *Experimental Brain Research, 224*(2), 275-285.

Gmeindl, L., Walsh, M., & Courtney, S. M. (2011). Binding serial order to representations in working memory: a spatial/verbal dissociation. *Memory & Cognition, 39*(1), 37-46.

Gobet, F., Lane, P. C. R., Croker, S., Cheng, P. C-H., Jones, G., Oliver, I., & Pine, J. M. (2001). Chunking mechanisms in human learning. *Trends in Cognitive Sciences, 5*(6), 236-243.

Gobet, F., & Simon, H. A. (1998). Expert chess memory: Revisiting the chunking hypothesis. *Memory, 6*(3), 225-255.

Hartshorne, J. K. (2008). Visual working memory capacity and proactive interference. *PLoS One, 3*(7), e2716.

Hasegawa, T., Matsuki, K. I., Ueno, T., Maeda, Y., Matsue, Y., Konishi, Y., & Sadato, N. (2004). Learned audio-visual cross-modal associations in observed piano playing activate the left planum temporale. An fMRI study. *Cognitive Brain Research, 20*(3), 510-518.

Heron, J., Roach, N. W., Hanson, J. V., McGraw, P. V., & Whitaker, D. (2012). Audiovisual time perception is spatially specific. *Experimental Brain Research, 218,* 477-485.

Heron, J., Whitaker, D., McGraw, P. V., & Horoshenkov, K. V. (2007). Adaptation minimizes distance-related audiovisual delays. *Journal of Vision, 7*(13), 1-8.

Ho, C., Reed, N., & Spence, C. (2007). Multisensory in-car warning signals for collision avoidance. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 49*(6), 1107-1114.

Holcombe, A. O., & Chen, W. Y. (2013). Splitting attention reduces temporal resolution from 7 Hz for tracking one object to <3 Hz when tracking three. *Journal of Vision, 13*(1), 1-19.

Houghton, G. & Tipper, S. P. (1994) A model of inhibitory mechanisms in selective attention. In*: Inhibitory mechanisms in attention, memory, and language,* ed. D. Dagenbach & T. Carr. Academic Press.

Intriligator, J., & Cavanagh, P. (2001). The spatial resolution of visual attention. *Cognitive Psychology, 43,* 171-216.

Jones, G., Gobet, F., & Pine, J. M. (2007). Linking working memory and long-term memory: a computational model of the learning of new words. *Developmental Science, 10,* 853-873.

Kane, M. J., Bleckley, M. K., Conway, A. R., & Engle, R. W. (2001). A controlled-attention view of working-memory capacity. *Journal of Experimental Psychology: General, 130*(2), 169.

Kane, M. J., & Engle, R. W. (2000). Working-memory capacity, proactive interference, and divided attention: limits on long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(2), 336.

Kanisza, G. (1976). Subjective contours. *Scientific America*, *234*(4), 48-52.

Karatekin, C. (2004). Development of attentional allocation in the dual task paradigm. *International Journal of Psychophysiology, 52*(1), 7-21.

Kawachi, Y., Grove, P. M., & Sakurai, K. (2014). A single auditory tone alters the perception of multiple visual events. *Journal of Vision*, *14*(8), 16-16.

Kim, R. S., Seitz, A. R., & Shams, L. (2008). Benefits of stimulus congruency for multisensory facilitation of visual learning. *PLoS One*, *3*(1), e1532.

Klingberg, T. (2010). Training and plasticity of working memory. *Trends in Cognitive Sciences, 14*(7), 317-324.

Klingberg, T., Forssberg, H., & Westerberg, H. (2002). Increased brain activity in frontal and parietal cortex underlies the development of visuospatial working memory capacity during childhood. *Journal of Cognitive Neuroscience*, *14*(1), 1-10.

Koelewijn, T., Bronkhorst, A., & Theeuwes, J. (2010). Attention and the multiple stages of multisensory integration: A review of audiovisual studies. *Acta Psychologica*, *134*(3), 372-384.

Lakatos, P., Karmos, G., Mehta, A. D., Ulbert, I., & Schroeder, C. E. (2008). Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science, 320*, 110-113.

Large, E. W., & Snyder, J. S. (2009). Pulse and meter as neural resonance. The Neurosciences and Music III – Disorders and Plasticity. *Annals of the New York Academy of Sciences, 1169*, 46-57.

Lavie, N. (2005). Load theory of selective attention and cognitive control. *Trends in Cognitive Science, 9*, 75-82.

Leboe, L. C., & Mondor, T. A. (2007). Item-specific congruency effects in nonverbal auditory Stroop. *Psychological Research, 71*(5), 568-575.

Logie, R. H., Brockmole, J. R., & Jaswal, S. (2011). Feature binding in visual short-term memory is unaffected by task-irrelevant changes of location, shape, and color. *Memory & Cognition, 39*(1), 24-36.

Luck, S. J., Heinze, H. J., Mangun, G. R., & Hillyard, S. A. (1990). Visual event-related potentials index focused attention within bilateral stimulus arrays. II. Functional dissociation of P1 and N1 components. *Electroencephalography and Clinical Neurophysiology, 75*, 528-542.

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

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

Luria, R., Balaban, H., Awh, E., & Vogel, E. K. (2016). The contralateral delay activity as a neural measure of visual working memory. *Neuroscience & Biobehavioral Reviews, 62*, 100-108.

Lustig, C., May, C. P., & Hasher, L. (2001). Working memory span and the role of proactive interference. *Journal of Experimental Psychology: General, 130*(2), 199.

Ma, Z., & Flombaum, J. I. (2013). Off to a bad start: Uncertainty about the number of targets at the onset of multiple object tracking. *Journal of Experimental Psychology: Human Perception and Performance, 39*(5), 1421-1432.

Makeig, S., Westerfield, M., Townsend, J., Jung, T. P., Courchesne, E., & Sejnowski, T. J. (1999). Functionally independent components of early event-related potentials in a visual spatial attention task. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 354*(1387), 1135-1144.

Makovski, T., & Jiang, Y. V. (2008). Proactive interference from items previously stored in working memory. *Memory & Cognition, 36*(1), 43-52.

Mangun, G. R., & Hillyard, S. A. (1991). Modulations of sensory-evoked brain potentials indicate changes in perceptual processing during visual-spatial priming. *Journal of Experimental Psychology: Human Perception and Performance, 17*(4), 1057.

Marks, L. E. (1987). On cross-modal similarity: Auditory-visual interactions in speeded discrimination. *Journal of Experimental Psychology: Human Perception and Performance, 13*, 384-394.

Marks, L. E., Ben-Artzi, E., & Lakatos, S. (2003). Cross-modal interactions in auditory and visual discrimination. *International Journal of Psychophysiology, 50*(1), 125-145.

Marois, R., & Ivanoff, J. (2005). Capacity limits of information processing in the brain. *Trends in Cognitive Science, 9*, 296-305.

Martino, G., & Marks, L. E. (2000). Cross-modal interaction between vision and touch: The role of synesthetic correspondence. *Perception, 29,* 745-754.

Mathy, F., & Feldman, J. (2012). What's magic about magic numbers? Chunking and data compression in short-term memory. *Cognition, 122*(3), 346-362.

Matusz, P. J., & Eimer, M. (2011). Multisensory enhancement of attentional capture in visual search. *Psychonomic Bulletin and Review, 18*, 904-909.

Mazza, V., & Caramazza, A. (2011). Temporal brain dynamics of multiple object processing: the flexibility of individuation. *PLoS One, 6*(2), e17453.

McCabe, D. P., Roediger III, H. L., McDaniel, M. A., Balota, D. A., & Hambrick, D. Z. (2010). The relationship between working memory capacity and executive functioning: Evidence for a common executive attention construct. *Neuropsychology*, *24*(2), 222.

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

Moran, J., & Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science*, *229*(4715), 782-784.

Ngo, M. K., Pierce, R. S., & Spence, C. (2012). Using multisensory cues to facilitate air traffic management. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *54*(6), 1093-1103.

Norman, D. A., & Bobrow, D. G. (1975). On data-limited and resource-limited processes. *Cognitive Psychology*, *7*(1), 44-64.

Norman, D. A., & Bobrow, D. G. (1976). On the analysis of performance operating characteristics. *Psychological Review, 83*(6), 508.

O'Donnell, B. F., Swearer, J. M., Smith, L. T., Hokama, H., & McCarley R. W. (1997). A topographic study of ERPs elicited by visual feature discrimination. *Brain Topography, 10*(2), 133-143.

Oksama, L., & Hyöna, J. (2004). Is multiple object tracking carried out automatically by an early vision mechanism independent of higher-order cognition? An individual difference approach. *Visual Cognition, 11*, 631-671.

Olesen, P. J., Westerberg, H., & Klingberg, T. (2004). Increased prefrontal and parietal activity after training of working memory. *Nature Neuroscience, 7*(1), 75-79.

Parise, C. V., & Spence, C. (2009). 'When birds of a feather flock together': Synesthetic correspondences modulate audiovisual integration in non-synesthetes. *PLoS One 4*(5), e5664.

Pavlovskaya, M., & Hochstein, S. (2011). Perceptual learning transfer between hemispheres and tasks for easy and hard feature search conditions. *Journal of Vision, 11*, 8.

Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: evidence for a parallel tracking system. *Spatial Vision, 3*, 179-197.

Ramachandran, V. S., & Hubbard, E. M. (2001). Synaesthesia – A window into perception, thought and language. *Journal of Consciousness Studies, 8*, 3-34.

Recanzone, G. H., Merzenich, M. M., & Dinse, H. R. (1992). Expansion of the cortical representation of a specific skin field in primary somatosensory cortex by intracortical microstimulation. *Cerebral Cortex, 2*(3), 181-196.

Repp, B. H., & Su, Y-H. (2013). Sensorimotor synchronization: A review of recent research (2006-2012). *Psychonomic Bulletin & Review, 20*, 403-452.

Rusconi, E., Kwan, B., Giordano, B. L., Umilta, C., & Butterworth, B. (2006). Spatial representation of pitch height: the SMARC effect. *Cognition, 99*(2), 113-129.

Sandhu, R., & Dyson, B. J. (2013). Modality and task switching interactions using bi-modal and bivalent stimuli. *Brain and Cognition, 82*, 90-99.

Sargent, J., Dopkins, S., Philbeck, J., & Chichka, D. (2010). Chunking in spatial memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(3), 576.

Sarmiento, B. R., Shore, D. I., Milliken, B., & Sanabria, D. (2012). Audiovisual interactions depend on context of congruency. *Attention, Perception, & Psychophysics, 74*(3), 563-574.

Saults, J. S., & Cowan, N. (2007). A central capacity limit to the simultaneous storage of visual and auditory arrays in working memory. *Journal of Experimental Psychology: General, 136*(4), 663.

Sears, C. R., & Pylyshyn, Z. W. (2000). Multiple object tracking and attentional processing. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Experimentale, 54*(1), 1.

Shin, J. C., & Ivry, R. B. (2002). Concurrent learning of temporal and spatial sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(3), 445.

Sivonen, V. P., & Ellermeier, W. (2006). Directional loudness in an anechoic sound field, head-related transfer functions, and binaural summation. *The Journal of the Acoustical Society of America, 119*(5), 2965-2980.

Slutsky, D. A., & Recanzone, G. H. (2001). Temporal and spatial dependency of the ventriloquism effect. *NeuroReport 12*(1), 7-10.

Sobel, K. V., Gerrie, M. P., Poole, B. J., & Kane, M. J. (2007). Individual differences in working memory capacity and visual search: The roles of top-down and bottom-up processing. *Psychonomic Bulletin & Review, 14*(5), 840-845.

Soto-Faraco, S., & Alsius, A. (2009). Deconstructing the McGurk-MacDonald illusion. *Journal of Experimental Psychology: Human Perception and Performance, 35*(2), 580.

Spence, C. (2010). The color of wine – Part 1. *The World of Fine Wine, 28*, 122-129.

Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics, 73*, 971-995.

Spence, C., & Deroy, O. (2012). Crossmodal correspondences: Innate or learned? *i-Perception, 3*(5), 316-318.

Spence, C., & Gallace, A. (2011). Tasting shapes and words. *Food Quality and Preference, 22*(3), 290-295.

Spence, C., & Santangelo, V. (2009). Capturing spatial attention with multisensory cues: a review. *Hearing Research, 258*, 134-142.

Spence, C., & Squire, S. (2003). Multisensory integration: maintaining the perception of synchrony. *Current Biology, 13*, R519-R521.

Stone, J. V., Hunkin, N. M., Porrill, J., Wood, R., Keeler, V., Beanland, M., Port, M., & Porter, N. R. (2001). When is now? Perception of simultaneity. *Proceedings of the Royal Society of London B: Biological Sciences, 268*(1462), 31-38.

Stumpf, K. (1883). *Tonpsychologie I [Psychology of the tone]*. Leipzig: Hirzel.

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *The Journal of the Acoustical Society of America, 26*(2), 212-215.

Sussman, E., Ritter, W., & Vaughan Jr., H. G. (1998). Predictability of stimulus deviance and the mismatch negativity. *Neuroreport, 9*, 4167-4170.

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

Theeuwes, J., & Van der Burg, E. (2013). Priming makes a stimulus more salient. *Journal of Vision, 13*(3), 1-11.

Thomaschke, R., & Dreisbach, G. (2013). Temporal predictability facilitates action, not perception. *Psychological Science, 24*(7), 1335-1340.

Todd, J. J., & Marois, R. (2004). Capacity limit of visual short-term memory in human posterior parietal cortex. *Nature, 428,* 751-754.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*(1), 97-136.

Van der Burg, E., Alais, D., & Cass, J. (2013). Rapid recalibration to audiovisual asynchrony. *The Journal of Neuroscience, 33*(37), 14633-14637.

Van der Burg, E., Awh, E., & Olivers, C. N. L. (2013). The capacity of audiovisual integration is limited to one item. *Psychological Science, 24*(3), 345-351.

Van der Burg, E., Olivers, C. N. L., Bronkhorst, A. W., & Theewues, J. (2008). Pip and pop: Non-spatial auditory signals improve spatial visual search. *Journal of Experimental Psychology: Human Perception and Performance, 34,* 1053-1065.

Van der Burg, E. Olivers, C. N. L., & Theeuwes, J. (2012). The attentional window modulates capture by audiovisual events. *PLoS ONE, 7*(7), e39137.

Van der Burg, E., Talsma, D., Olivers, C. N. L., Hickey, C., & Theeuwes, J. (2011). Early multisensory interactions affect the competition among multiple visual objects. *Neuroimage, 55*(3), 1208-1218.

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

Van Wassenhove, V., Grant, K. W., Poeppel, D. (2007). Temporal window of integration in audio-visual speech perception. *Neurophysiologica, 45,* 598-607.

Verleger, R., Jaskowski, P., & Wascher, E. (2005). Evidence for an integrative role of P3b in linking reaction to perception. *Journal of Psychophysiology, 19*(3), 165-181.

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

Vogel, E. K., & Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature, 428*(6984), 748-751.

Vogel, E. K., McCollough, A. W., & Machizawa, M. G. (2005). Neural measures reveal individual differences in controlling access to working memory. *Nature, 438,* 500-503.

Vroomen, J., & Keetels, M. (2010). Perception of intersensory synchrony: a tutorial review. *Attention, Perception, & Psychophysics, 72*(4), 871-884.

Vroomen, J., Keetels, M., de Gelder, B., & Bertelson, P. (2004). Recalibration of temporal order perception by exposure to audiovisual asynchrony. *Cognitive Brain Research, 22*, 32-35.

Walker, P., & Smith, S. (1985). Stroop interference based on the multimodal correlates of haptic size and auditory pitch. *Perception, 14,* 729-736.

Walsh, V. (2003). A theory of magnitude: Common cortical metrices of time, space, and quality. *Trends in Cognitive Sciences, 7,* 483-488.

Wasserman, E. A., Chatlosh, D. L., & Neunaber, D. J. (1983). Perception of causal relations in humans. *Learning and Motivation, 14,* 406-432.

Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin, 88*, 638-667.

Wilbiks, J. M. P. & Dyson, B. J. (2013a). Effects of temporal asynchrony and stimulus magnitude on competitive audio-visual binding. *Attention, Perception, & Psychophysics*, *75*(8), 1883-1891.

Wilbiks, J. M. P. & Dyson, B. J. (2013b). The influence of previous environmental history on audio-visual binding occurs during visual-weighted but not auditory-weighted environments. *Multisensory Research, 26*, 561-568.

Yantis, S. (1992). Multielement visual tracking: Attention and perceptual organization. *Cognitive Psychology, 24*, 295-340.

Zampini, M., Shore, D. I., & Spence, C. (2003). Audiovisual temporal order judgments. *Experimental Brain Research, 152*, 198-210.