Facilitating visual target identification using non-visual cues

Mary Kim Ngo

Jesus College, Oxford

Dissertation submitted for the degree of

Doctor of Philosophy in the University of Oxford

Hilary 2012

ACKNOWLEDGMENTS

I would like to express my sincerest gratitude to the people who have motivated, encouraged, and supported me, in mind and in spirit, through this arduous process:

Thomas Strybel, who got me hooked on multisensory perception and taught me how to be a researcher. Thank you for all your encouragement, advice, support, and for keeping me in check whenever I'd go off course; for being selfless and putting my interests and future first, even when it meant getting me the hell out of Long Beach so I could explore other opportunities! I'm so thankful to have you as my mentor. You're still the mostawesomeist!

Kim Vu, who's tough-loving (but warm and bubbly inside), hard-working mentorship style helped me stay focused and inspired me to strive for excellence. Thanks for giving me the opportunity to teach with you and for teaching me how to teach.

John (Jack) Dwyer, for your words of wisdom. I will always remember to "No matter what, be nice". You encouraged me to rise to any challenge, to work on my weaknesses, and to foster my strengths.

A HUGE THANK YOU to my friends and family for all your love, patience, and support. Jeannique and Nicholas, Azmeen and Imti (+ the Bangla Boys), Ilmo and Hayzell, Anne-Sylvie, Georgiana, Yi-Chuan, Vanessa (+ Maya and the boys), Alexandra, Zita, Ian, Robin, Esther, Jin, Kim and Douglas, Penny, Carlos and Kari, and of course, John (for being by my side through it all). I love you all!!!

Finally, Charles…I can't thank you enough for pushing me to write, write, write! I could not have done what I did in the amount of time that I did it without you. =) Thank you for supporting me for the last three years, for never knocking my crazy ideas, for your generosity, and for always presenting me with and looking out for every cool and exciting collaborative opportunity out there! Your culinary expertise, your crazy, awesome obsession with chili, your eclectic taste for music, and your eccentric, coloured trousers will be dearly-missed. THANK YOU!!

PUBLICATIONS AND PRESENTATIONS

PUBLICATIONS

Ngo, M. K., Cadieux, M. L., Sinnett, S., Soto-Faraco, S., & Spence, C. (2011). Reversing the Colavita visual dominance effect. *Experimental Brain Research*, **214**, 607-618.

Ngo, M. K., & Spence, C. (2010a). Auditory, tactile, and multisensory cues facilitate search for dynamic visual stimuli. *Attention, Perception, & Psychophysics*, **72**, 1654-1665.

Ngo, M. K., & Spence, C. (2010b). Crossmodal facilitation of masked visual target identification. *Attention, Perception, & Psychophysics*, **72**, 1938-1947.

Ngo, M. K., & Spence, C. (2010c). Crossmodal facilitation of masked visual target discrimination by informative auditory cuing. *Neuroscience Letters*, **479**, 102-106.

Ngo, M. K., Sinnett, S., Soto-Faraco, S., & Spence, C. (2010). Repetition blindness and the Colavita effect. *Neuroscience Letters*, **480**, 186-190.

Ngo, M. K., Misra, R., & Spence, C. (2011). Assessing the shapes and speech sounds that people associate with chocolate samples varying in cocoa content. *Food Quality and Preference*, **22**, 567-572.

Ngo, M. K., & Spence, C. (2011). Assessing the shapes and speech sounds that people associate with different kinds of chocolate. *Journal of Sensory Studies*, **26**, 421-428.

Ngo, M. K., Piqueras-Fiszman, B., & Spence, C. (2012). On the colour and shape of still and sparkling water: Implications for product packaging. *Food Quality and Preference*, **24**, 260-268.

Ngo, M. K., & Spence, C. (under review). Facilitating masked visual target identification with auditory oddball stimuli. *Experimental Brain Research.*

Gallace, A., Ngo, M. K., Sulaitis, J., & Spence, C. (2011). Multisensory presence in virtual reality: Possibilities & limitations. In G. Ghinea, F. Andres, & S. Gulliver (Eds.), *Multiple sensorial media advances and applications: New developments in mulsemedia* (pp. 1-38). Hershey: IGI Global.

Spence, C., Ngo, M. K., Lee, J.-H., & Tan, H. (2010). Solving the correspondence problem in haptic/multisensory interface design. In M. H. Zadeh (Ed.), *Advances in haptics* (pp. 47-74). InTech. Available at http://www.intechopen.com/articles/show/title/solving-the-correspondence-problem-in-haptic-multisensory-interface-design.

Spence, C., & Ngo, M. K. (in press). Does crossmodal attention or multisensory integration explain the crossmodal facilitation of masked visual target identification in the freezing effect? To appear in B. E. Stein (Ed.), *The new handbook of multisensory processing*. Cambridge, MA: MIT Press.

Spence, C., Ngo, M. K., Percival, B., & Smith, B. (under review). Crossmodal correspondences: Assessing shape symbolism for foods with a complex flavour profile. *Journal of Sensory Studies*.

PRESENTATIONS

Ngo, M. K., & Spence, C. (2011). Reversing the Colavita visual dominance effect. Poster presented at the *Oxford Meeting of the Experiment Psychological Society* (13-15th April), Oxford, UK.

Ngo, M. K., & Spence, C. (2010). Crossmodal facilitation of rapidly-masked visual target identification. Poster presented at the *11th International Multisensory Research Forum* (16-19th June), Liverpool, UK.

Ngo, M. K., & Spence, C. (2009a). Using non-visual cues to enhance the efficiency of visual search. Talk given at the *2nd Year PGR Student Presentations* (5th October), Oxford, UK.

Ngo, M. K., & Spence, C. (2009b). Comparing the effectiveness of auditory & tactile cues in facilitating visual search performance. Poster presented at the *Haptic and Audio Interaction Design 2009 4th International Workshop* (10-11th September), Dresden, Germany.

Ngo, M. K., & Spence, C. (2009c). Evaluating the effectiveness of temporally synchronous and spatially predictive cues in visual search. Poster presented at the *10th International Multisensory Research Forum* (29th June – 2nd July), New York, New York.

Ngo, M. K., & Spence, C. (2009d). Multisensory interfaces and warning signals: Enhancing user experience and maximizing efficiency. Talk given at the *Old Members' Annual General Meeting* (18th April), Oxford, UK.

Ngo, M. K. (2009). Auditory cuing of visual target identification. Talk given at the *Human Factors in Transport Security Academic Forum* (16-17th April), Oxford, UK.

TABLE OF CONTENTS

SHORT ABSTRACT

The research presented in this thesis was designed to investigate whether and how the temporal synchrony and spatial congruence of non-visual cues with visual targets could work together to improve the discrimination and identification of visual targets in neurologically-healthy adult humans. The speed and accuracy of participants' responses were compared following the presence or absence of temporally synchronous and/or spatially congruent or incongruent auditory, vibrotactile, and audiotactile cues in the context of dynamic visual search and rapidly-masked visual target identification. The understanding of the effects of auditory, vibrotactile, and audiotactile cues derived from these laboratory-based tasks was then applied to an air traffic control simulation involving the detection and resolution of potential conflicts (represented as visual targets amidst dynamic and cluttered visual stimuli). The results of the experiments reported in this thesis demonstrate that, in the laboratory-based setting, temporally synchronous and spatially informative non-visual cues both gave rise to significant improvements in participants' performance, and the combination of temporal and spatial cuing gave rise to additional improvements in visual target identification performance. In the real-world setting, however, only the temporally synchronous unimodal auditory and bimodal audiotactile cues gave rise to a consistent facilitation of participants' visual target detection performance. The mechanisms and accounts proposed to explain the effects of spatial and temporal cuing, namely multisensory integration and attention, are examined and discussed with respect to the observed improvements in participants' visual target identification performance.

EXTENDED ABSTRACT

Searching for a visual target amidst an array of visual distractors can be difficult and time-consuming, especially when the target and distractors are similar in one or more features (Duncan & Humphreys, 1989). Two techniques that have consistently been demonstrated to be successful in facilitating the search process are temporal and spatial cuing using either visual or non-visual cues. In studies on temporal cuing, the presentation of the cue is synchronized with some change in the visual target (e.g., motion, onset, orientation, or colour, etc.; Staufenbiel, van der Lubbe, & Talsma, 2011; Van der Burg, Olivers, Bronkhorst, & Theeuwes, 2008a, 2009; Van der Burg, Talsma, Olivers, Hickey, & Theeuwes, 2011; Yantis, 1993; Yantis & Jonides, 1984). In studies of spatial cuing, the cue is either presented in the same relative (e.g., left or right hemi-field) or exact location as the visual target (Perrott, Cisneros, McKinley, & D’Angelo, 1996; Perrott, Saberi, Brown, & Strybel, 1990; Perrott, Sadralodabai, Saberi, & Strybel, 1991). While temporally synchronous and spatially informative visual and non-visual cues have independently been shown to improve participants’ visual target identification and/or discrimination performance, the effects of the combination of these two cue attributes have, surprisingly, not been directly studied to date.

The first goal of the present thesis was therefore to test whether the combination of temporally synchronous and spatially informative cues would give rise to a larger facilitation of participants’ visual search performance than

either temporally synchronous or spatially informative cues when used in isolation. The second goal of this thesis was to determine whether the benefits of various non-visual (auditory, vibrotactile, and audiotactile), spatial and temporal cues observed in controlled laboratory settings could carry over to a real-world setting.

The introductory chapter of this thesis (Chapter 1) therefore provides brief descriptions of the three paradigms comprising the experimental chapters of the thesis. The first two paradigms involve laboratory-based visual target identification and discrimination tasks, namely dynamic visual search and rapid serially-masked visual presentation. Both tasks involve visual target detection and identification/discrimination, but differ in terms of the temporal and/or spatial nature of the tasks. The third and final paradigm is based on a real-world air traffic control simulation which involves the time-sensitive detection of visual target conflicts. Once a brief understanding of the experimental paradigms used in the thesis has been developed, the potential mechanism(s) underlying the temporal and spatial non-visual cuing of visual perception is then introduced.

Chapter 2 explores the use of non-visual cues in a dynamic visual search task, where participants must locate and discriminate a target line segment from amongst a number of distractor line segments in a cluttered search display (Van der Burg et al., 2008a, 2009). The dynamic and cluttered nature of the task allowed for the examination of how temporal and spatial cues might interact in order to guide spatial attention. The results of the experiments reported in this chapter demonstrate that the efficiency of participants' visual search performance could be significantly improved simply by presenting a temporally synchronous non-visual (either auditory, vibrotactile, or audiotactile) cue at the same time as the colour change of the visual target. Importantly, the unimodal auditory, vibrotactile, and bimodal audiotactile cues all led to equivalent performance improvements even though they were completely uninformative with regard to the location or identity (i.e., horizontal or vertical) of the visual target.

Additionally, the results demonstrated that the presentation of temporally synchronous, spatially-informative auditory and vibrotactile cues enhanced the efficiency of participants' visual search performance more than following either spatially uninformative or else spatially invalid cues (cf. Tan, Gray, Spence, Jones, & Rosli, 2009). Importantly, the benefit appeared to be larger following auditory cuing than following tactile cuing; this despite the fact that the spatial information transmitted by the cues was identical in both cases. This result suggests that auditory cues may be more effective in facilitating participants' visual search performance, at least when the cues provide spatial information regarding the likely target side.

Having demonstrated the benefits of temporal and spatial cuing in a task that is spatial in nature, the experiments reported in Chapter 3 next went on to explore the same potential cuing benefits in a task that is more temporal in nature. In the rapid serially-masked visual target discrimination task, participants are presented with a series of displays, each presented briefly and followed immediately by a mask. The participants' task is to identify in which one of the four corners of the visual array the target display (a diamond-shaped arrangement of four dots) appeared (Vroomen & de Gelder, 2000).

The results of the experiments reported in Chapter 3 demonstrated that any oddball stimulus (a single tone, a single vibration, a tone plus vibration, a visual cue, a louder tone, a quieter tone, or even a silent period; i.e., the absence of a stimulus) embedded within an otherwise regular sequence of tones all led to significant improvements in participants' visual target discrimination performance when the stimulus was presented at the same time as the visual target display.

Consistent with the findings of Chapter 2, the results of Chapter 3 demonstrated that making a temporally synchronous auditory cue spatially informative with regard to the likely location of the visual target led to an additional improvement in the accuracy of participants' target discrimination responses compared to either the spatially uninformative or spatially invalid cuing conditions. Here, even though the task was temporally demanding in nature, participants still benefited significantly from the spatially informative nature of the cue.

The laboratory-based dynamic visual search and rapid serially-masked visual presentation tasks used in Chapters 2 and 3, respectively, differed in terms of the temporal and/or spatial nature of the tasks. The relative effectiveness of temporal and spatial cuing might therefore have been expected to vary depending on the task at hand. Nevertheless, the pattern of improvements in participants' visual target identification and discrimination performance appeared to be nearly identical across the two experimental paradigms. Having demonstrated reliable non-visual temporal and spatial cuing effects in these two laboratory-based tasks, the goal of the final experimental chapter of this thesis was to test these cues in a more applied, real-world task.

The task utilized in Chapter 4 was an air traffic control simulation where participants were required to monitor and control various aircraft flying in their airspace. This air traffic simulation combined the spatial demand of the dynamic visual search task and the temporal demand of the rapid serially-masked visual target presentation task: Participants had to monitor various locations on the visual display in order to keep track of each aircraft. Meanwhile, each aircraft updated its location in the airspace at regular intervals and participants had to make sure the aircraft did not come too close together or too close to boundaries, making the task extra time-sensitive.

The benefits of temporal and spatial non-visual cuing that had been reported in Chapters 2 and 3 were only partially replicated in the experiments reported in Chapter 4. In particular, only temporally synchronous cues presented at the onset of conflict situations gave rise to significantly faster response times to the conflicts. More specifically, only multisensory audiotactile and unimodal auditory cues gave rise to improved performance, while unimodal vibrotactile cues failed to facilitate the speed of participants' responses. This may have reflected the high perceptual and cognitive demands of the air traffic simulation task, which contrasts with the relatively lower demands of the laboratory-based tasks used in Chapters 2 and 3 (see Spence, 2010, for a recent review). Additionally, making the temporally synchronous auditory cues also spatially informative with regard to the location of the conflict did not result in any additional performance benefits beyond what was observed following temporally synchronous cuing alone.

Taken together, the results of the three experimental chapters reported in this thesis provide robust evidence that unimodal auditory and multisensory audiotactile cues, when presented at the same time as critical visual target events, give rise to reliable performance benefits across a variety of basic and applied research paradigms. Spatial cuing, on the other hand, only appears to be effective in certain situations. As such, this may reflect the operation of somewhat different mechanisms underlying the effects of temporal and spatial non-visual cuing. This possibility is explored in Chapter 5, where multisensory integration and attention are discussed as two potential accounts for the cuing effects observed.

CHAPTER 1:

MULTISENSORY INTERACTIONS IN HUMAN INFORMATION PROCESSING: A REVIEW

1.0. Introduction

Humans are multisensory beings moving through, perceiving, and interacting with a world that, at times, may bombard them with an overwhelming amount of sensory information. Sights, sounds, smells, tastes, and even tactile stimuli surround us all most of the time. Just imagine being in a cramped underground subway, standing shoulder-to-shoulder with other passengers. The clashing scents of perfume, cologne, and body odour fill the air. The wheels of the train screech against the tracks as the train comes to a halt. You see a group of teenagers board the train and, over their chatter, you catch the muffled announcement *"The next station is Marble Arch"*. We normally don't think about all the sensory information flooding into our brains at every moment, let alone how it is that our brains process and make sense of it all because this all normally happens so effortlessly. Of course, not all of the sensory information surrounding us is processed at all times. Imagine taking an exam and being distracted by every sight, sound, or smell that just so happened to be present in the room. It would be very difficult to focus on get anything done! In order to combat this, it is often the case that only certain modes/modalities of information, quite often sight and sound, are selected for further processing. It has been argued that this strategy of selectively attending to certain stimuli over others can help to minimize potential sensory overload and ensure our seamless and, more importantly, safe interaction with the environment (Luck, Woodman, & Vogel, 2000).

Humans are said to be “visual creatures / visually dominant” as we tend to rely heavily on our sense of sight to help us understand and navigate through the environment (e.g., Posner, Nissen, & Klein, 1976; Spence, Shore, & Klein, 2001). What is more, it has been estimated that over half of our brains are involved in visual processing, much more than any other sense (Felleman & Van Essen, 1991; see also Gallace, Ngo, Sulaitis, & Spence, 2011). Our eyes are like cameras—they can zoom in to focus on tiny details or zoom out to scan and capture entire fields of view. And, like cameras, things presented centrally to our eyes, or in foveal vision, are in sharp focus while things presented peripherally are not only blurred, but also sometimes neglected altogether (Ali & Klyne, 1985; Jacobson, 1951). Thus, the information that we need to process visually should ideally be within our field of view, and the accuracy with which we process and interact with this information is often limited by our visual acuity. Moreover, while our visual systems are able to take in large amounts of information at any given time, there are, nevertheless, limitations in terms of the amount of information that humans can attend to and process efficiently without suffering a noticeable decrement in performance (Coltheart, 1980; Di Lollo, 1977; see also Wickens, 2008). While driving, for example, we should keep our eyes on the road ahead, occasionally check our side and rear view mirrors, frequently check the speedometer to make sure we are driving at the appropriate speed, scan the road for potential hazards, and use our eyes to help perform numerous other driving-related visual tasks. At times like these, when the resources available to process visual information are taxed, information received through other sensory modalities, such as hearing and touch, can provide effective alternative means for communicating critical details of a given event (e.g., Ho & Spence, 2008; Wickens, 1984, 2008).

Auditory signals are omnidirectional (that is, you don’t have to look at a sound in order to hear it; Proctor & Van Zandt, 1994). This makes the auditory modality an ideal alternative in terms of relaying information when the visual system is either overloaded or else insensitive to peripheral stimulation. The omnidirectional nature of auditory signals is an advantage in cases when they are used to simply alert people, but the limitations of human auditory localization become problematic when knowing the source of an auditory signal matters. Humans are fairly good at localizing sounds in the horizontal plane (a minimum audible angle of approximately 1° is not unusual under optimal conditions; e.g., Perrott & Saberi, 1990), but we tend to be poor at localizing sounds vertically and distinguishing sounds coming from the front and back (Wightman & Kistler, 1999). Moreover, auditory stimuli are susceptible to amplitude and frequency masking (Oxenham & Wojtczak, 2010). This means that some auditory cues might be rendered inaudible against background noise and sounds that are either louder than the cue or else share similar frequencies with it. In order to combat potential masking effects and maximize the effectiveness of the auditory signals, designers typically design auditory stimuli to be presented at greater intensities and lower frequencies than the noise present in the environment (e.g., Miller & Beaton, 1994; Patterson, 1982; Proctor & Van Zandt, 1994).

Information presented tactually is relatively impervious to auditory noise whilst being easily localized on the surface of the body (Proctor & Van Zandt, 1994). As such, presenting information tactually is a promising alternative when the visual and auditory systems are either overloaded or otherwise unavailable. In recent years, many mobile phone manufacturers have implemented vibrotactile feedback to compensate for the shortcomings of auditory stimulation in the presence of noise by also vibrating to alert users of incoming calls and messages. When a mobile phone rings at a party, for example, it can be difficult to hear amidst all the noise and chatter. That said, the phone’s vibration can easily be felt, thus allowing its owner to avoid a potentially missed call.

Our skin has widely varying degrees of sensitivity over the surface of the human body. This is directly related to the amount of cortical representation reserved for each part of the body (Proctor & Van Zandt, 1994; see also Weinstein, 1968). Moreover, some receptors that are responsive to certain types of tactile stimulation are only found in certain parts of the skin (e.g., pleasure receptors are only found in hairy skin; Olausson et al., 2008; see also McGlone & Spence, 2010, for a recent review). The effectiveness of tactually-presented stimuli at capturing attention and communicating information is, therefore, influenced by the location of the tactile stimulation. An individual whose cell phone was located at a position away from the body is likely not going to hear or feel the vibration of the phone and would, consequently, probably miss the incoming call anyway!

Understanding the limitations of vision, audition, and touch is vital to the design and development of warning signals which aid humans in the prevention of potential dangers, including vehicular and avionic (aeronautic) collisions, and ensure safe navigation. Single modes of sensory stimulation might provide us with fragmented bits of information regarding the environment, but it is the combined stimulation of our different senses that gives us a more complete and accurate representation of the world around us. It is this multisensory information that humans optimally integrate and interact with (Ernst & Banks, 2002; see also Trommershäuser, Landy, & Körding, 2011). Product designers capitalize on this fact by creating products that can stimulate multiple senses simultaneously. Things don't just look good or sound good—they look good, sound good, and feel good (on occasion, they even smell and/or taste good too!). Aside from the appeal that multisensory products have to consumers and product designers, the benefits of, and need for, multisensory stimulation become even more apparent when applied to the design of warning signals, where time-sensitive information needs to be conveyed/communicated as rapidly as possible. Information provided by multisensory stimulation has the capacity to help interface operators respond more quickly and appropriately to environmental obstacles and/or events (Ho, Reed, & Spence, 2006, 2007; Ho, Tan, & Spence, 2006; Santangelo, Ho, & Spence, 2008). One experimental paradigm, in particular, that has demonstrated the robust benefits of multisensory / crossmodal cuing is the visual search task (Van der Burg et al., 2008a, 2009, 2011).

1.1. Visual Search

In a typical visual search paradigm, participants must locate and identify a pre-specified visual target in amidst a number of visual distractors (see Wolfe, 2001, for a review).[1] When searching for a target defined by a single basic feature amongst uniform/homogenous distractors (feature search), search is quite efficient and the target is said to "pop-out" from amongst the distractors. In fact, adding more distractor stimuli to the search field does not increase the amount of

---

[1] Although the focus of this thesis is solely on visual search, it should be noted that other forms of search have also been studied, including auditory search (Asemi, Sugita, & Suzuki, 2003), tactile search and exploration (Forti & Humphreys, 2005; James et al., 2002; Karnath & Perenin, 1998), and even gustatory search (Delwiche, Lera, & Breslin, 2000; Halpern, 1983).

time taken to find the target. However, when the target shares one or more feature(s) with the distractors (conjunction search), search can become rather difficult, with each added distractor increasing the amount of time taken to find the target. Feature search is said to involve parallel search, where each item in the search field can be processed in parallel. Conjunction search, on the other hand, often involves serial search, where each item in the search field is processed sequentially and either identified as the target or ruled out as a distractor (Thornton & Gilden, 2007; Treisman & Gelade, 1980; Wolfe, 1998)[2]. It should be noted, however, that several studies have shown that certain conjunctions of features can be searched for more efficiently, or in parallel, than others (David & Driver, 1994; Humphreys, Hodsoll, & Riddoch, 2009; Kim & Cave, 1995; McLeod, Driver, & Crisp, 1988; McLeod, Driver, Dienes, & Crisp, 1992; Yamani & McCarley, 2011), leaving many researchers to conclude that the distinction between serial vs. parallel and feature vs. conjunction is inappropriate and unnecessary (Duncan & Humphreys, 1989; Wolfe, 1998).

Visual search tasks have evolved from those early studies involving the detection of simple, stationary targets (such as red or green horizontal and vertical bars) that are either present or absent to involving the detection of moving stimuli, targets and distractors that are dynamically changing in some feature, and complex/ecologically-real targets and distractors (Chen & Zelinsky, 2006; see also Krummenacher, Müller, Deubel, Wolfe, & Humphreys, 2010; Wolfe, 1994, 2001, for reviews). The evolutionary path from basic to complex visual search studied in the laboratory makes sense given the fact that the

---

[2] Rather than a distinction between serial and parallel search, Duncan and Humphreys (1989) suggested that: "*For all search materials, instead, difficulty increases with increased similarity of targets to nontargets and decreased similarity between nontargets, producing a continuum of search efficiency*" (p. 433).

findings from these laboratory-based paradigms have important implications for real-world tasks, such as baggage screening at airports (e.g., McCarley, Kramer, Wickens, Vidoni, & Boot, 2004; Wolfe, Horowitz, & Kenner, 2005), vehicular navigation (e.g., Underwood, Chapman, Bowden, & Crundall, 2002), and air traffic control management (Remington, Johnston, Ruthruff, Gold, & Romera, 2000), where people often must locate and identify visual target events amidst dynamic and cluttered backgrounds. In fact, one of the criticisms of many laboratory-based visual search tasks is that they often miss key aspects of real world visual search, such as the fact that target events do not typically occur as frequently in the real world as they do in laboratory studies (Wolfe & Van Wert, 2010). What is more, while targets are pre-defined in laboratory-based tasks, visual targets in real-world tasks are often undefined, so it may be the case that the participant doesn’t quite know what the target looks like (e.g., improvised explosive devices or weapons) when searching for it (amongst all the various items stuffed into a packed holiday suitcase; e.g., see Wolfe et al., 2005).

As search tasks have evolved to become more complex, it seems that the amount of time required to locate and identify visual target events has also increased. Typical search times for simple (parallel) search tasks average 400-500 ms for set sizes greater than 12, whereas complex search tasks typically average upward of 1000 ms (Taylor & Cutsuridis, 2011; Treisman & Gelade, 1980; Treisman & Sato, 1990; Treisman & Souther, 1985) and upward of 2000 ms when searching for targets in naturalistic scenes (Chen & Zelinsky, 2006; Taylor & Cutsuridis, 2011; Van Wert, Horowitz, & Wolfe, 2009; Wolfe, 1994; Wolfe, Võ, Evans, & Greene, 2011). Real-world visual search typically involves the search for more complex stimuli than simple lines and basic shapes. As such, it is inevitable that visual search is likely going to take longer than the 400-500 ms search times observed in the basic visual search laboratory setting and probably be more likely to reflect the >1000 ms search times typical of complex visual search. While this may be acceptable in a laboratory setting, it might be considered too long (i.e., unacceptable) in a real-world setting where lives and safety often depend on split-second reactions (e.g., to potential threats, road hazards, collisions, etc.; e.g., see Ho & Spence, 2008). Perhaps unsurprisingly, then this has spurred a lot of interest, for good reason, in finding ways to improve visual search performance.

Many researchers have explored the introduction of various cuing techniques/paradigms that might aid in speeding-up visual search under complex viewing conditions. The extant literature on the cuing of visual search is extensive, ranging from unimodal visual cuing through to crossmodal auditory and tactile cuing. These studies have typically focused separately on either spatially informative or temporally synchronous cues. While unimodal visual cues give rise to significant improvements in search performance, crossmodal and multisensory cues have proven to be more effective at not only reducing search times, but also increasing the accuracy of participants' responses, especially under conditions of high perceptual load (Ho, Reed, & Spence, 2006, 2007; Ho, Tan, & Spence, 2006; Santangelo et al., 2008; see also Spence, 2010, for a recent review).

1.2. The Role of Spatial Congruence/Informativeness in the Crossmodal Cuing of Visual Search

Spatially informative non-visual cues, when presented at the same (exact or relative) location as visual targets, have been shown to improve the speed and accuracy of participants' responses when detecting and identifying the visual targets (e.g., Atchley, Kramer, Andersen, & Theeuwes, 1997; Bolia, D'Angelo, & McKinley, 1999; Perrott et al., 1990, 1996; see Spence & Santangelo, 2009, for a review). In applied settings, the presentation of spatially informative cues may offer an effective means not only of reducing the time needed to detect potential threats (e.g., potential collisions, road hazards, weapons hidden in airport luggage, etc.), but also of improving the subsequent discrimination of those threats. The presentation of spatially informative non-visual cues, specifically auditory cues that are spatially coincident with visual targets, has been shown to reduce visual search latencies by several thousand milliseconds (ms) for peripherally located visual targets (i.e., for targets presented at eccentricities exceeding 90° from central fixation; e.g., Perrott et al., 1990, 1991, 1996).

Rudmann and Strybel (1999) investigated whether the presentation of auditory cues that were spatially coincident, displaced by 6°, or else spatially-uninformative with regard to the location of the visual target would facilitate participants' visual search performance. Rudmann and Strybel demonstrated that spatially coincident auditory cues were found to be most effective in terms of reducing participants' visual search latencies. They also demonstrated that the displaced auditory cues were nevertheless still effective in terms of enhancing visual search when compared to performance in the uninformative cue condition (see also Brill, Terrence, Downs, Gilson, Hancock, & Mouloua, 2004).

Meanwhile, Spence and Driver (1997a, b) have reported that the auditory pre-cuing of the side of a visual target can provide an effective means of improving participants' target discrimination performance, even when the side of the auditory cue happens to be non-predictive with regard to the side on which the visual target is likely to occur (see also Dufour, 1999; Spence, 2010; and Spence, McDonald, & Driver, 2004, for reviews). These crossmodal exogenous spatial attentional cuing benefits were found to be maximal when the auditory cue preceded the visual target by approximately 100-300 ms. More recent research, however, has demonstrated that for certain search tasks, cues that are presented simultaneously with the presentation of the visual target event are more effective at reducing search times than those presented before the target (Van der Burg et al., 2008a; Vroomen & de Gelder, 2000). This raises the possibility that this difference in time course may hint at the involvement of different mechanisms underlying the auditory cuing effects.

1.3. The Role of Temporal Synchrony in the Crossmodal Cuing of Visual Search

The presentation of spatially uninformative but temporally synchronous[3] non-visual cues has also been shown to reduce visual search latencies and enhance participants' perception of visual targets (Chan & Chan, 2006; Dalton & Spence, 2007; Staufenbiel et al., 2011; Van der Burg et al., 2008a, 2009; Vroomen & de Gelder, 2000). Recent research by Olivers and Van der Burg (2008) has demonstrated that the presence of a single tone whose onset was temporally synchronous with the target(s) in a rapid visual stimulus presentation

---

[3] It is important to note that even though the stimuli are physically simultaneously presented (i.e., temporally synchronous), this does not necessarily mean that they will arrive at the relevant brain regions synchronously (e.g., see Schroeder & Foxe, 2002, 2004, 2005; Spence & Squire, 2003).

(RSVP) stream led to a 10% improvement in participants' visual target identification performance. Importantly, Olivers and Van der Burg demonstrated that the synchronous tone enabled the second target in the RSVP stream to *"escape the attentional blink"* (p. 191), thus leading to a level of performance that was comparable to the first target in the RSVP stream. Similarly, Dalton and Spence (2007) have also reported that auditory singletons (i.e., auditory stimuli that are unique from other items on a dimension such as stimulus duration), when presented at the same time as visual targets, lead to significantly faster and more accurate identification of the duration of the visual target (when participants had to indicate whether its duration was shorter or longer than that of the non-target stimuli) in a stream of centrally-presented visual and auditory stimuli.

1.3.1. The 'Pip and Pop' Paradigm

Van der Burg et al. (2008a) reported a dynamic visual search task in which participants had to locate and discriminate a target line segment (horizontal vs. vertical) from amongst a number of distractor line segments (tilted at ±22.5°). On a given trial, a number of distractors (or the target) change colour between red and green at random intervals. This task is particularly difficult because of the relatively large, cluttered search field, with the set size varying from 24 to 48 items. Additionally, the target shares a number of features with the target, including colour, shape, and size. Even though the target and distractor line segments differ in orientation, this difference is relatively small (only 22.5°). Moreover, the task is dynamic, rather than static, as the target and distractors change back and forth between green and red at random intervals (see Figure 1.1).

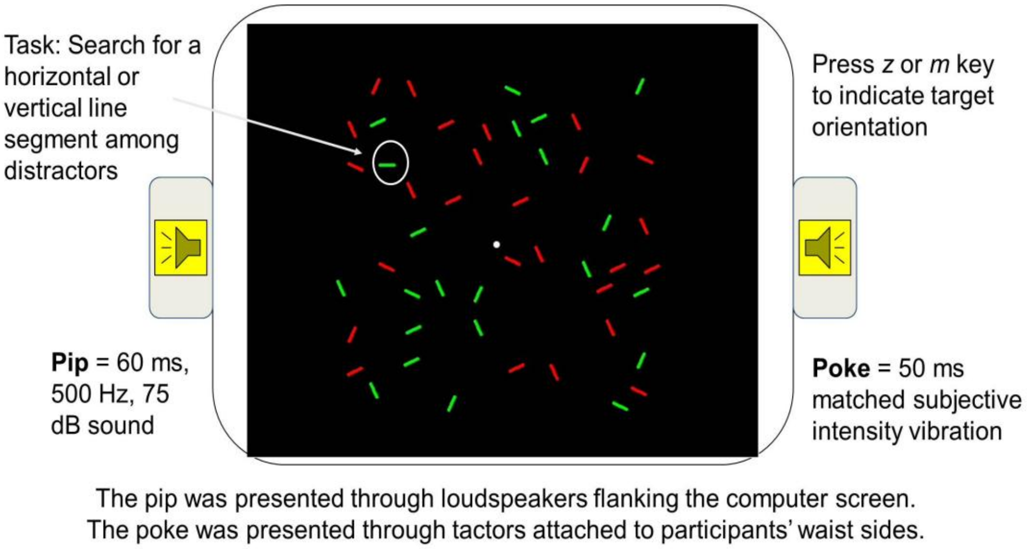


Figure 1.1. Illustration of the dynamic visual search display used in Experiments 2.1-2.6 of this thesis.

Although this visual search task seems to be spatial in nature, in the sense that participants could have scanned all the different locations of the search field in order to find the visual target (if, in fact, it was present), there was a temporal component to the task in terms of the timing of the colour change of the target and distractors. Importantly, when Van der Burg et al. (2008a, 2009) presented either an auditory (“pip”) or vibrotactile (“poke”) cue simultaneously with the colour change of the visual target, participants’ search latencies were reduced by an average of more than 1000 ms! Van der Burg et al. concluded that the temporally synchronous auditory or vibrotactile cue and the synchronously colour-changing visual target were likely being integrated, thus creating a more salient visual target that somehow “popped out” from amongst the distractors.

1.3.2. The ‘Flickering-Dots’ Paradigm

Vroomen and de Gelder (2000) introduced a visual target identification task involving the rapid serially-masked visual presentation of visual displays. In this task, participants are presented with a series of four displays (three distractors and one target), each containing four quasi-randomly positioned white dots. Each display is presented briefly and followed immediately by a mask. One of the displays contains four dots forming a diamond shape (the target display) which could appear in one of the four corners of the visual display. The participants' task is to identify in which quadrant the target diamond appeared (see Figure 1.2).

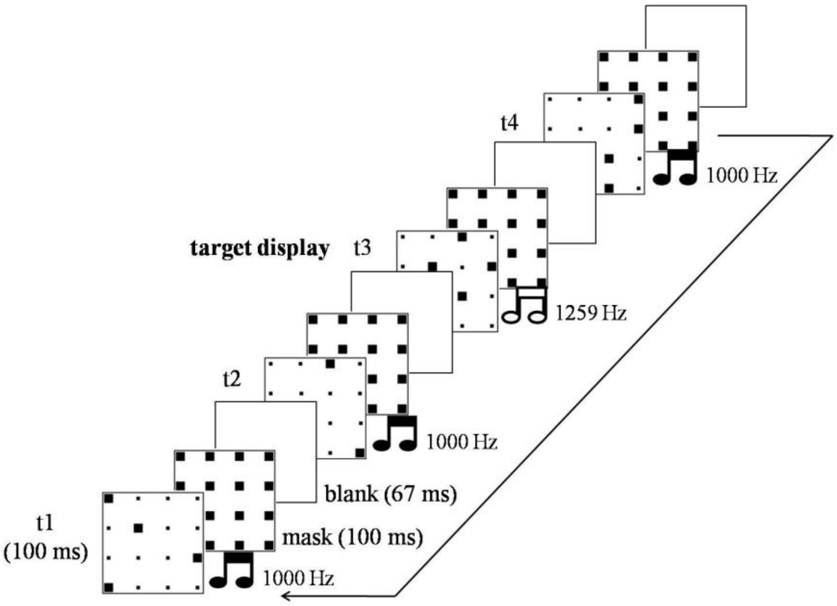


Figure 1.2. Illustration of the rapid serially-masked visual presentation used in Experiments 3.1-3.9 of this thesis.

This task is particularly difficult because each visual display is presented very briefly (<100 ms) and then rapidly-masked. Thus, the task is temporally demanding in nature, and requires participants to focus their attention on brief moments in time in order to discriminate the target from the distractor displays. Importantly, Vroomen and de Gelder (2000) demonstrated that by presenting each distractor display with a low-frequency tone and the target display with a high-frequency tone, this gave rise to significantly better identification of the visual target's location. They concluded that the high-frequency tone was perceptually-grouped and hence integrated[4] with the simultaneously-presented visual target display and that this crossmodal pair of stimuli was segregated from the other low-frequency tone + distractor display pairs. Vroomen and de Gelder claimed that this crossmodal segregation gave rise to a brief, subjective freezing of the visual target display together with a high-frequency tone, which ultimately gave rise to the improved visual target identification performance.

1.3.3. The Air Traffic Scenarios Test (ATST)

In 1996, the Federal Aviation Administration developed a low-fidelity air traffic management simulation task (air traffic scenarios test; ATST), which they used to screen potential air traffic controllers. The task requires operators to monitor and manage air traffic in a terminal airspace/sector, where planes were landing and exiting. In its current form, all of the information in the ATST is presented visually, including the signal used to alert operators when conflicts arise. When participants have to monitor a large number of airplanes visually, this may result in visual overload (see Figure 1.3).

---

[4] Perceptual grouping refers to the subjective organization/parsing of stimulus arrays into subgroups of stimuli (O'Leary & Rhodes, 1984; Spence & Chen, in press; see also Spence & Soto-Faraco, 2010, for a review), whereas multisensory integration refers to the unification of separate/different sensory stimuli into coherent, singular events (see Stein & Stanford, 2008, for a review).

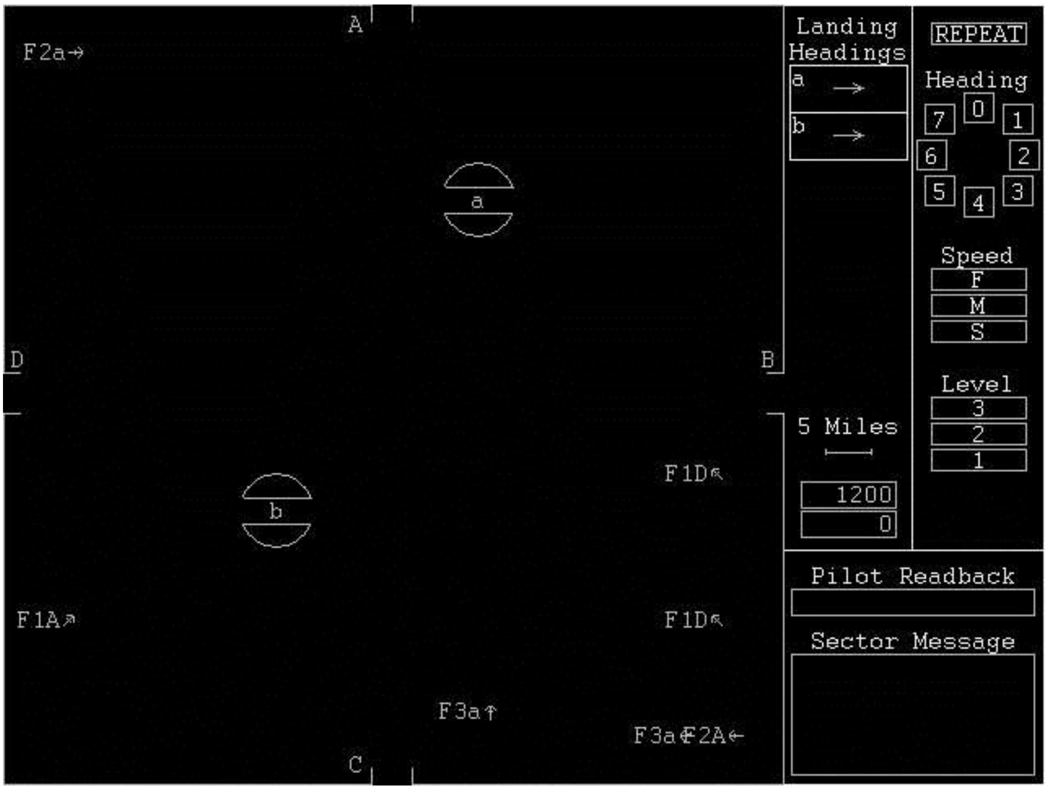


Figure 1.3. A black-and-white screenshot of the ATST used in the experiments reported in Chapter 4.

Aside from the fact that the ATST is completely visual, performing the task is challenging for many reasons. First, the task is dynamic, with each aircraft updating its position in the airspace every seven seconds. Second, there are several rules that participants need to obey when controlling and monitoring the aircraft. These rules include making sure that the aircraft maintain 5 nautical miles of lateral separation with other aircraft and boundaries, landing at a slow speed and low altitude, and exiting at a fast speed and high altitude, to name a few. As such, the task is cognitively demanding.

The visually overloaded environment of the ATST, as well as the real-world air traffic control environment, would seem to be the perfect environment in which to implement and test the potential benefits of non-visual cuing. Yet, the research on non-visual cuing in such applied settings has been surprisingly limited (e.g., Hameed, Jayaraman, Ballard, & Sarter, 2007).

1.4. Potential Mechanisms Underlying Non-Visual Cuing

There are two schools of thought regarding the way in which spatially or temporally informative non-visual cues facilitate visual target identification. According to one account, the non-visual cue and visual target, when presented in close spatial and temporal proximity, are integrated and perceived as a single, bound object through a process known as *multisensory integration* (Lewald & Guski, 2003; Meredith, Nemitz, & Stein, 1987; Meredith & Stein, 1986). This binding of the two sources of information occurs automatically, in a bottom-up manner. Thus, the enhanced perception of the visual target in the presence of a spatially or temporally informative non-visual cue occurs at a perceptual level, prior to any top-down influences on information processing.

The other school of thought is that spatially or temporally informative non-visual cues elicit a shift in participants' *attention*, either in the spatial and/or temporal dimensions. It is well-known that attention can be directed to specific locations in space using spatially informative cues (see Spence & Driver, 2004, for a review). There have also been numerous studies demonstrating the fact that temporal attention can be directed to a specific moment in time by the use of temporally informative cues in order to facilitate the processing of visual targets (Chen & Yeh, 2008; Correa, Lupiáñez, Madrid, & Tudela, 2006; Coull, Frith, Büchel, & Nobre, 2000; Nobre & Coull, 2009). Thus, by shifting attention to a specific location or moment in time, in either a top-down or bottom-up manner, the perception of any object(s) occurring at that very location or moment is enhanced.

1.4.1. Multisensory Integration

Multisensory integration refers to the way in which inputs from different sensory modalities are combined by the nervous system. One potential result of such multisensory integration is that the various sources of sensory information come together to form coherent, unified perceptual experiences. The temporal and spatial alignment of multiple sources of sensory stimuli are two key factors determining whether or not the constituent unisensory stimuli will be integrated to form unified perceptual experiences (Meredith et al., 1987; Meredith & Stein, 1986; Stein & Stanford, 2008). The closer together the constituent unimodal stimuli are presented in time and space, the more likely that multisensory integration will be observed.

When applied to visual search, the presentation of temporally synchronous and spatially co-localized non-visual cues with the visual targets should provide the optimal conditions for multisensory integration to occur. It is this multisensory integration of the auditory (and/or tactile) and visual stimuli that is said to underlie the facilitatory effects observed following the presentation of either temporally synchronous or spatially informative non-visual cues in visual search tasks. In this case, the binding of the crossmodal stimuli ultimately gives rise to the enhanced saliency of the visual target stimulus, which, in turn, results in improved visual target identification performance.

1.4.2. Attention

Attention is broadly defined as the selective processing of relevant sensory information (Milliken & Tipper, 1998; Spence, 2010). Attention can be oriented/shifted to different sensory stimuli in the environment and our cognitive resources can be allocated to the attended sensory stimuli (Posner, 1980). Posner considers the ability to shift our attention and selectively attend to certain sensory stimuli to be a high level, cognitive skill, which often results in the enhanced processing of the attended stimuli.

When applied to visual search, any cue that might provide information about the whereabouts of the visual target in either the temporal, spatial, or temporal and spatial dimension(s) can elicit a shift in a participant's attention. Several studies have demonstrated that attention can be intentionally directed to a specific moment in time by the use of temporally informative cues to facilitate the processing of a visual target (Correa et al., 2006; Coull et al., 2000; Coull & Nobre, 1998). Attention can also be directed to specific locations in space using spatially informative cues (Spence & Driver, 2004). More recently, Lange, Krämer, and Röder (2006) reported that people can use the temporal and spatial features of auditory stimuli in parallel in order to facilitate their processing of the auditory stimuli.

Attention to a specific location or time can also be elicited in a stimulus-driven, bottom-up manner, where the cue need not be temporally or spatially informative (cf. Driver & Spence, 1998). Indeed, numerous studies have shown that certain cues can capture attention in an automatic fashion, and draw participants' attention to the location or timing of the presentation of the cue (Debener, Kranczioch, Herrmann, Engel, 2002; Mazza, Turatto, Rossi, & Umiltà, 2007; McDonald, Teder-Sälejärvi, & Hillyard, 2000).

1.5. Outline of the Experimental Research in this Thesis

The benefits of having non-visual (i.e., auditory and vibrotactile) cues that are either spatially informative (Ho, Santangelo, & Spence, 2009; Ho, Tan, &

Spence, 2006; Perrott et al., 1990, 1991, 1996) or temporally synchronous (e.g., Chan & Chan, 2006; Dalton & Spence, 2007; Van der Burg et al., 2008a, 2009; Vroomen & de Gelder, 2000) have been repeatedly demonstrated in previous research. The potential benefit of the combination of these two cue attributes for participants' visual target identification performance, however, has not been studied in either the dynamic visual search or the rapid serially-masked visual presentation paradigms before. Moreover, whether or not temporally synchronous and spatially informative non-visual cuing can improve visual target detection and identification in real-world settings has not been adequately studied, to date.

The goal of this thesis, therefore, was to link basic to applied research by extending what is found in the controlled experimental settings of the dynamic and cluttered visual search and rapid serially-masked visual presentation paradigms to the realistic, simulated air traffic control setting of the ATST. By carrying out the same experimental manipulations across different paradigms, this should allow for a better comparison of the effects of non-visual cuing under a variety of experimental conditions.

Additionally, the mechanism(s), be it multisensory integration, attention, or a combination of both, underlying the crossmodal effects of spatial and temporal cuing remain unclear (cf. Spence & Ngo, in press; Van der Burg et al., 2008a; Vroomen & de Gelder, 2000). The experiments reported in this thesis were designed to help to shed light on the best explanation for the facilitatory effects of spatial and temporal cuing in the context of dynamic visual search, rapid serially-masked visual presentation, and air traffic control simulation tasks.

The question of interest in all of the experiments reported in **Chapters 2-4** is whether, and under what conditions, crossmodal auditory and vibrotactile cues (be they spatially informative, temporally synchronous, or both) facilitate visual target identification performance. **Chapters 2** and **3** first examine temporal and spatial cuing in a basic, laboratory setting using dynamic visual search and rapid serially-masked visual presentation tasks. Here, comparing many of the same experimental manipulations across different tasks highlights the generality of the cuing effects observed. Once the benefits of temporal and spatial cuing have been established and understood in the setting of basic research paradigms, they are then extended to the applied setting of the ATST in **Chapter 4** in order to determine whether the potential enhancements in participants' visual perception hold in such real-world tasks.

Chapter 2 of this thesis presents a series of six experiments using the "pip and pop" paradigm to further explore the effectiveness of temporally synchronous non-visual cuing. The findings of Van der Burg et al. (2008a, 2009) are replicated and extended by showing that temporally synchronous auditory, vibrotactile, and audiotactile cues all give rise to similar improvements in participants' visual target identification performance. Making the temporally synchronous cues spatially informative with regard to the likely location of the visual target facilitates participants' performance still further. This is perhaps unsurprising given the fact that this type of visual search task might benefit most from spatial cuing as a result of the spatially-sensitive nature of the task (i.e., it involves the potential scanning/searching of the visual field of space). As such, the potential benefits resulting from the same temporal and spatial cues as those used in **Chapter 2** were next explored in a paradigm where the temporal demand of the task is high.

Chapter 3 presents a series of nine experiments using the “flickering-dots” paradigm in order, in the first place, to replicate and extend Vroomen and de Gelder’s (2000) original findings by demonstrating that any cue, when presented at the same time as the visual target display, can give rise to a significant improvement in participants’ visual target identification performance. Additionally, temporally synchronous and spatially informative cues are shown to facilitate participants’ performance even more than temporally synchronous cues alone. Thus, even for temporally-demanding tasks, spatially informative cues appear to be useful for improving visual target detection and subsequent identification. After demonstrating the potential benefits of temporal and spatial cuing in controlled, laboratory settings, **Chapter 4** goes on to test whether the same crossmodal facilitatory effects could be similarly observed in an applied, real-world setting.

Chapter 4 presents a series of four experiments exploring the use of non-visual, temporally synchronous and/or spatially informative cues in the ATST. Here, the benefits that can be observed by providing operators with non-visual cues to signal the advent of a conflict or dangerous situation are clearly demonstrated. Finally, the potential mechanisms responsible for the effects of temporal and spatial cuing are discussed in **Chapter 5**. Specifically, the roles of multisensory integration and attention are discussed in terms of explaining the crossmodal facilitatory effects observed in the presence of temporally synchronous and spatially informative visual and non-visual cues under a range of different experimental paradigms and conditions. The findings from the laboratory-based and real-world tasks used in the present thesis will be extended to inform and make recommendations for the design and implementation of multisensory / non-visual warning signals in a variety of applied domains.

CHAPTER 2:

DYNAMIC VISUAL SEARCH IN A CLUTTERED FIELD

2.0. Introduction

Searching for a target defined by a conjunction of features in a complex and dynamically-changing visual display often requires slow and exhaustive search. That is, each item in the search display has to be examined individually in order to determine whether or not it is a target (e.g., Treisman & Gelade, 1980; Treisman & Sato, 1990; see Quinlan, 2003; Treisman, 1996, for reviews). For interface operators, such as pilots and air traffic controllers, this can pose a major problem, as successful conflict resolution often requires not only the rapid detection of potential threats (or dangers), but also the accurate interpretation of those threats (see Pawlak & Vicente, 1996; Previc, 2000; Vicente & Rasmussen, 1992). Under such conditions, the presentation of spatially informative cues may offer an effective means not only of reducing the time needed to detect potential threats, but also of improving the subsequent discrimination of those threats. The presentation of spatially informative non-visual cues, specifically auditory cues that are co-localized with visual targets, has been shown to reduce visual search latencies by several thousand milliseconds for peripherally located visual targets (i.e., for targets presented at eccentricities exceeding 90° from central fixation; e.g., Perrott et al., 1990, 1991, 1996). The presentation of spatially uninformative auditory cues can also reduce visual search latencies for visual targets presented in the central field by more than 200 ms (e.g., Dufour, 1999; Perrott et al., 1990, 1991, 1996).

Spatially-uninformative auditory and vibrotactile cues can facilitate participants' visual search performance when they are temporally synchronized with a change in the target stimulus (e.g., Bolognini, Frassinetti, Serino, & Ladavas, 2005; Chan & Chan, 2006; Van der Burg et al., 2008a, 2009; Vroomen & de Gelder, 2000). For example, Van der Burg et al. (2008a, 2009) measured search latencies for visual targets presented in a frequently-changing central search field. The participants in these studies had to search for a horizontal or vertical line segment (i.e., for a line segment presented at 0 or 90°) presented amongst distractor line segments oriented at +22.5°. On a given trial, the colour of the target and distractor stimuli changed regularly from red to green or vice versa. The participants were either presented with an auditory tone cue over headphones (Van der Burg et al., 2008a), or else with a vibrotactile cue to their left hand (Van der Burg et al., 2009). The onset of the cue was either synchronized with the colour change of the target stimulus, or else no cue was presented. Importantly, the cue gave participants absolutely no information concerning the orientation of the visual target, hence ruling out a response priming account of any facilitatory effects that were observed (cf. Spence & Driver, 1997a, b).

Nevertheless, Van der Burg and his colleagues (2008a) reported that the presentation of a temporally synchronous cue still resulted in a significant crossmodal facilitation of participants' visual search performance relative to their performance in those trials in which no cue was presented. In fact, the average reduction in search latencies for temporally synchronous cues was in excess of

1000 ms (for displays consisting of 24-48 items). Van der Burg et al. also demonstrated that the search slopes were significantly shallower (indicating more efficient visual search performance) for the auditory cue trials (31 ms/item) as compared to the no cue trials (147 ms/item). Van der Burg et al. concluded that the temporally synchronous auditory or vibrotactile cue and the synchronously colour-changing visual target were likely being integrated, thus creating a more salient visual target that somehow "popped out" from amongst the distractors. Just what the perceptual or psychological consequences of non-visual cuing are is something that will be discussed later in the general discussion of this thesis.

The experiments reported in this chapter were designed to replicate and extend Van der Burg et al.'s (2008a, 2009) findings using the auditory-visual "pip and pop" and vibrotactile-visual "poke and pop" paradigms in order to explore how visual search performance is modulated by cues that are not only temporally synchronous (presented in synchrony with the colour change of the visual target) but also spatially informative with regard to the likely location of the target (in the left or right hemi-field). While the benefits of having auditory or vibrotactile cues that are either temporally synchronous (e.g., Chan & Chan, 2006; Dalton & Spence, 2007; Van der Burg et al., 2008a, 2009; Vroomen & de Gelder, 2000) or spatially informative (Ho, Santangelo, & Spence, 2009; Ho, Tan, & Spence, 2006; Perrott et al., 1990, 1991, 1996) have been repeatedly demonstrated in previous research, the potential benefit of having cues that are both spatially informative and temporally synchronous has not been studied before in the context of a dynamic visual search paradigm.

Experiment 2.1 was designed to replicate Van der Burg et al.'s (2008a, 2009) studies in which they compared visual search performance between a cue-absent condition and a temporally-synchronized cue-present condition. Experiment 2.2 compared the consequences of presenting unimodal auditory and vibrotactile cues to the consequences of presenting multisensory audiotactile cues to check for any benefits of multisensory over unimodal cuing. Having successfully replicated Van der Burg et al.'s basic effects in Experiments 2.1-2.3 went on to explore the combination of temporally synchronous and spatially informative cues in order to ascertain whether or not participants' visual search performance could be improved still further by making the non-visual cue spatially informative with regard to the likely target side. For the first three experiments reported in this chapter, the changes in visual search performance following auditory and tactile cuing were compared in order to determine whether one cue was better than the other in facilitating visual search. Experiment 2.4 examined the influence of cue location (headphones vs. external loudspeaker cones) on participants' visual search performance in order to determine whether the location of the auditory cue relative to the visual display would make a difference in terms of the cuing effects observed. Experiment 2.5 directly compared participants' performance in the presence of temporally synchronous, spatially informative, or both temporally synchronous and spatially informative auditory cues in order to determine which cue would give rise to the greatest improvements in performance. Finally, in Experiment 2.6, the visual search task was conducted on a large screen in order to explore whether the results observed on a small computer screen would hold over a large visual field of view. Given that visual search in the real-world often occurs over a large visual field (e.g., scanning the road while driving, managing air traffic, or searching for a face in a crowd) and not just a small computer screen, this is a particularly important aspect of visual search that merits further exploration. Here, the spatially informative non-visual cues might be expected to be even more useful than on the small screen, given the large visual field in which participants must search for the target.

2.1. Experiment 2.1

While Van der Burg and his colleagues reported that the performance benefits observed following vibrotactile cuing was comparable to that observed following auditory cuing, they did not perform any direct comparison between the results of these two cue conditions (experiments). The goal of the first experiment in this thesis was therefore to replicate the findings of Van der Burg et al.’s (2008a and 2009) recent studies and to extend their findings by directly comparing the performance of participants who were presented with auditory cues to those who were presented with vibrotactile cues.

2.1.1. Methods

Participants. Twenty-two participants from the University of Oxford (12 female) ranging in age from 20-38 years (mean age = 27 years) took part. All of the participants reported normal or corrected-to-normal vision, hearing, and tactile sensitivity for all experiments in this thesis and were right-handed by self-report. The experiment took approximately 45 minutes to complete. The participants received a £5 (UK Sterling) gift voucher in return for taking part in the study.

Apparatus and stimuli. All of the experiments reported in this thesis (with the exception of Experiments 2.6, 4.5, and 4.6) were conducted in a dimly-lit, soundproof chamber. The experiments reported in Chapters 2 and 3 were all conducted on a PC using E-Prime (Psychology Software Tools, Pittsburgh). The participants sat in a chair approximately 80 cm from a 17” CTX PR711F visually-flat CRT computer monitor (CTX, Watford, UK; screen refresh rate = 75 Hz). The auditory stimulus consisted of a 500-Hz tone (44.1 kHz sampling rate, 16-bit) with a 60 ms duration (including a 5-ms fade-in and fade-out to avoid clicks) which was presented from two Dell A215 PC loudspeaker cones (Dell Inc., UK), one placed 16° to either side of the center of the visual search display. The tactile stimulus consisted of a 50 ms duration 200 Hz vibration, presented through two VBW32 (Audiological Engineering Corporation, Somerville, MA, USA) tactors fastened to a waist-belt, with one tactor placed on either side of the participant’s waist. In pre-testing, four participants (who did not take part in the main experiment) subjectively matched the intensity of the vibrotactile stimulus to that of the auditory stimulus. The visual search displays consisted of 24, 36, or 48 red (RGB=255, 0, 0) or green (RGB=0, 255, 0) line segments (length 0.57° visual angle) presented against a black background. The colour of each display item was determined randomly. All of the line segments were randomly placed in an invisible 10 x 10 grid (9.58° x 9.58°, 0.00°-0.34° jitter), centered on a white fixation point, with the sole constraint that the target was never presented in one of the four central positions. The orientation of each line deviated randomly by exactly +22.5° from the horizontal or vertical, except for the target, which was presented in either the horizontal or vertical orientation.

Each of 9 sequentially presented display screens (making up one complete display cycle) were presented for 50, 100, or 150 ms, with each display duration being repeated randomly 3 times within a sequence, during which time a certain number of items (either a certain number of distractors or the target) changed colour: If the set size was 24, 1, 2, or 3 distractors changed colour; If the set size was 36, 1, 3, or 5 distractors changed colour; and if the set size was 48, 1, 4, or 7 distractors changed colour. When the colour of the target changed, it was the only item in the display to do so. If the target changed colour, then the preceding display duration was always 150 ms and the following display duration was always 100 ms. The target only changed colour once per display cycle, so that the average frequency with which the target's colour changed was 1.11 Hz (i.e., once every 900 ms). That said, the target did not change colour during the first 500 ms of the very first display cycle of each trial. Ten different display cycles were generated, and presented one after the other (without any gap) until the participant responded or the 10th display cycle had been presented, at which time the whole sequence (of 10 display cycles) was repeated.

Design and procedure. The modality of the cue was varied on a between-participants basis. The presence versus absence of the cue was varied on an alternating block-by-block basis. Set size was varied on a trial-by-trial basis. During the cue-present blocks, the participants (N = 10) either heard a tone or else felt a vibration (N = 12) that was synchronized with the colour change of the target. The reaction time (RT) and accuracy of participants' responses were measured. RTs were measured from the onset of the search display until the initiation of the participant's response. Each trial began with the presentation of a fixation point for 1,000 ms. The search display was presented until the participant made a response. The participants were instructed to press the z or m key (on the keyboard) as rapidly and accurately as possible in order to indicate the orientation of the target. The assignment of the targets to the response keys was counterbalanced across participants. Each target orientation was randomly determined and occurred equally often per block of 36 trials. There were four cue-absent and four cue-present blocks of trials presented in a counterbalanced, alternating order. These test blocks were preceded by two 36-trials practice blocks. After each block of trials, the participants received feedback concerning their overall mean accuracy and RT. They were given the opportunity to take a break before starting the next block.

2.1.2. Results and Discussion

All of the participants exceeded 85% correct for the practice blocks at the start of the experiment (the same criterion was also met by all of the participants in the subsequent experiments reported in this chapter). The data from the two practice blocks and from those trials in which the participants made an erroneous response (*M*=6.7% of trials) were excluded from the data analysis. Inverse efficiency (IE) scores were computed to correct for any potential speed-accuracy trade-offs that might have been present in the data (see Spence, Kingstone, Shore, & Gazzaniga, 2001; Townsend & Ashby, 1983). IE scores were calculated by dividing each participant's mean RT for each condition by their proportion of correct responses for that condition.

Given that the participants in Experiment 2.1 reported that they felt that the task became easier over time, it seemed appropriate to compare participants' performance on the first half of the experimental session (blocks 1-4) to that on the second half (blocks 5-8) in order to look for any potential practice effects in the data. A mixed univariate analysis of variance (ANOVA) was conducted on the IE data with Practice (First vs. Second half of the experiment), Set size (24, 36, or 48) and Cue presence (Absent or Present) as the within-participants factors and Cue modality (Auditory or Tactile) as the between-participants factor. The Huynh-Feldt correction was applied whenever sphericity was violated (alpha = .05).

Analysis of the IE data revealed significant main effects of Practice, $F(1,20)=19.71$, $p<.001$, $\eta 2=.50$, Cue presence, $F(1,20)=7.90$, $p=.01$, $\eta 2=.28$, and Set size, $F(2,40)=21.51$, $p<.001$, $\eta 2=.52$. There was a significant Cue presence x Set size interaction, $F(2,40)=7.29$, $p=.004$, $\eta 2=.27$. When the set size was 24, there was a marginally significant difference between the Cue present and Cue absent conditions, $p=.09$, with better performance (i.e., lower IE scores) being reported in the presence ($M=2797$ ms) as compared to the absence ($M=3210$ ms) of a cue. The magnitude of these differences increased as the set size increased from 36 ($p=.03$) up to 48 ($p=.005$). When the set size was 36, the average IE was 3609 ms in the presence of a non-visual cue and 4506 ms when no cue was presented. When the set size was 48, the average IE was 4105 ms in the presence of the non-visual cue and 5763 ms when no cue was presented. Thus, the average reduction in the IE score was 413 ms when the set size was 24, 897 ms when the set size was 36, and 1658 ms when the set size was 48 (see Figure 2.1).

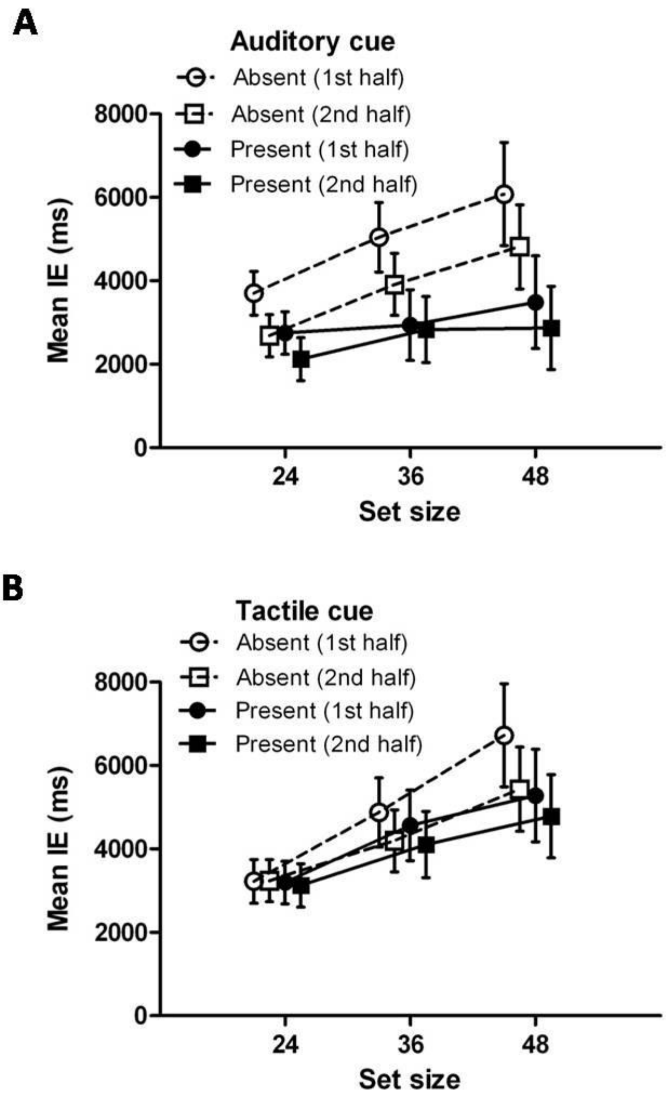


Figure 2.1. Mean IE scores (in ms) as a function of Practice (1^{st} or 2^{nd} half of experimental trial blocks), Set Size (24, 36, or 48), Cue Presence (Absent or Present), and Cue Modality (Auditory or Tactile; Panels A and B, respectively) in Experiment 2.1. Error bars represent the standard errors of the means.

The analysis also revealed a marginally significant Practice x Cue presence interaction, *F*(1,20)=4.29, *p*=.05, η2=.18. While participants' IE scores were significantly lower in the second half of the experimental session than in the first, the reduction in IE was larger for the Cue absent (*M*=895 ms) than for the Cue present (*M*=397 ms) condition. Furthermore, the participants benefited more from the cue in the first half as compared to the second half of the experiment: In fact, the average reduction in IE scores from the Cue absent to Cue present condition was 1238 ms in the first half as compared to 740 ms in the second half of the experiment.

There was no main effect of Cue modality, *F*<1, nor any interaction between Cue modality and Cue presence, *F*(1,20)=2.44, *p*=.13, thus showing that visual search performance was facilitated just as much by the auditory as by the vibrotactile cues in Experiment 2.1. The search slopes for the auditory and vibrotactile Cue present conditions averaged 85 and 121 ms/item, respectively. The search slopes averaged 137 ms/item for both the auditory and vibrotactile Cue absent conditions (note that these conditions were physically identical but completed by different groups of participants). Mean IE scores for the auditory (*M*=4372 ms) and vibrotactile (*M*=4614 ms) Cue absent conditions were also indistinguishable, *t*<1.

It is important to note that the effect of the non-visual cues on participants' visual search performance can manifest as a change in either the slope-intercept or the slope itself. A change in the y-intercept of the search slope is indicative of an overall reduction in RT (IE in this case), which represents an overall facilitation of performance. Such facilitation suggests that participants are simply faster, but not necessarily more efficient at searching for the target amongst the distractors in the presence of the temporally informative non-visual cues. A change in the slope of participants' search functions, on the other hand, is indicative of search efficiency, with more shallow slopes representing performance approaching parallel search. In Experiment 2.1, the presence of the temporally synchronous auditory and vibrotactile cues both led to search slopes that were not significantly different from 0 ms/item (*p*=.30 and .14, respectively), while the absence of any such cues gave rise to search slopes that differed significantly from 0 ms/item.

Thus, the results of Experiment 2.1 therefore revealed that the efficiency of participants' visual search performance was improved significantly simply by presenting a temporally synchronous non-visual (either auditory or vibrotactile) cue at the same time as the colour change of the visual target. Importantly, the non-visual cues used in Experiment 2.1 were completely noninformative with regard to the identity (i.e., horizontal or vertical) of the target. The improvement in participants' performance following non-visual cuing resulted in a reduction of the IE score of nearly 1000 ms on average, with the magnitude of the cuing benefit increasing as the set size increased. Moreover, the search slopes were much shallower (and not significantly different from 0 ms/item) in the presence of either of the non-visual cues.

The comparison of participants' performance in the two halves of the experiment confirmed their subjective reports that the task became easier over time. The non-visual cues seemed to have helped the participants more in the first half of the experiment than in the second half, as the magnitude of the reduction in IE from cue absent to cue present trials was larger in the first than in the second half of the experiment. This result suggests that participants did not require much practice in order to "learn how to use" the auditory and vibrotactile cues. Thus, the temporally synchronous non-visual cues appear to be "intuitive" (Ho, Reed, & Spence, 2006) and effective in producing significant improvements in visual search performance that are immediately tangible.

The magnitude of the performance benefit observed in Experiment 2.1 was numerically very similar to that reported by Van der Burg and colleagues (2008a, 2009). Thus, the results of Experiment 2.1 successfully replicated Van der Burg and colleagues' (see Van der Burg et al., 2008a, Experiment 1; Van der Burg et al., 2009) recent findings using a near-identical experimental set-up. However, in contrast to these earlier studies, the vibrotactile warning signals in the present study were presented from participants' waists rather than from their wrists (as in Van der Burg et al., 2009) and from external loudspeakers rather than over headphones (as in Van der Burg et al., 2008a).

2.2. Experiment 2.2

While a number of previous studies have demonstrated that bimodal cuing can produce performance benefits that are significantly larger than those seen following unimodal cuing (e.g., Ho et al., 2007, 2009, Experiment 2; Santangelo et al., 2008; Spence & Santangelo, 2009), others have reported bimodal cuing to be no better than unimodal cuing, or even, on occasion, worse than unimodal cuing (e.g., Fitch, Kiefer, Hankey, & Kleiner, 2007; Ho et al., 2009, Experiment 1; Lee & Spence, 2009; Lindeman, Yanagida, Sibert, & Lavine, 2003; Santangelo, Van der Lubbe, Olivetti Belardinelli, & Postma, 2006; see also Spence & Ho, 2008). Thus, the answer to the question of whether or not bimodal (or multisensory) cuing leads to improved performance relative to that seen following the presentation of the best of the unimodal cues seems to depend on the particular task and experimental setting under investigation.

Given this uncertainty, Experiment 2.2 was designed to compare bimodal audiotactile cuing to unimodal auditory and tactile cuing using the same experimental set-up as had been used in Experiment 2.1. In particular, the goal of Experiment 2.2 was to determine whether bimodal cues would further improve participants' visual search performance as compared to either unimodal auditory or unimodal vibrotactile cues. Should the addition of a redundant crossmodal cue work to increase the saliency of the visual target (e.g., Matusz & Eimer, 2011; Stein, London, Wilkinson, & Price, 1996; Van der Burg et al., 2008a, 2009) relative to the visual distractors, then one might expect the bimodal cues to facilitate visual search more than the unimodal cues. If, however, the non-visual cues facilitated participants' performance by way of attentional capture, perhaps by providing the participants with some sort of temporal marker as to when the target colour change occurred (see Van der Burg et al., 2008a; Vroomen & de Gelder, 2000), then one might expect that a bimodal cue would not be any more effective than the best of the unimodal cues (since both types of cues would most likely provide equivalent temporal information to the participant).

2.2.1. Methods

Ten participants took part in Experiment 2.2 (8 female; mean age=27 years; age range=21-33 years). The experiment took approximately 60 minutes to complete. The experimental set-up was identical to that used in Experiment 2.1 with the following exceptions: 1) The various cue conditions were now interleaved within each block of experimental trials; 2) Within each block of trials, four Cue types (No cue, Auditory, Vibrotactile, or Audiotactile) were presented equiprobably and in a random order; and 3) There was only one practice block at the beginning of the experiment followed by seven experimental blocks.

2.2.2. Results and Discussion

The IE data are shown in Figure 2.2. The IE data from the practice block and from those trials in which the participants made an erroneous response (M=3.0% of trials) were excluded from the data analysis. The remaining data were subjected to a repeated measures ANOVA with Set size (24, 36, or 48) and Cue type (No Cue, Auditory, Vibrotactile, Audiotactile) as the within-participants factors.

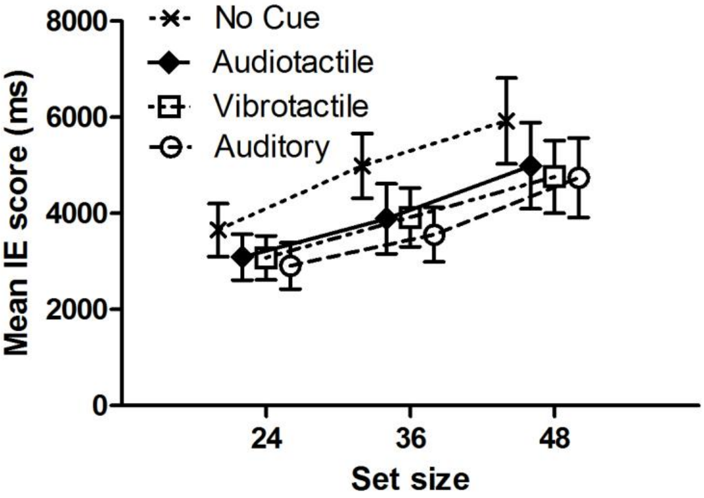


Figure 2.2. Mean IE scores as a function of Set size (24, 36, or 48), Cue type (No cue, Auditory, Vibrotactile, or Audiotactile) in Experiment 2.2. Error bars represent the standard errors of the means.

The analysis revealed a significant main effect of Set size, *F*(2,18)=31.04, *p*<.001, η2=.78, with IE scores increasing as the set size increased. There was also a significant main effect of Cue type, *F*(3,27)=5.60, *p*=.01, η2=.38. All three of the non-visual cue conditions (Auditory, Vibrotactile, and Audiotactile) resulted in performance that was significantly better than in the No cue condition (*p*=.025, *p*=.017, and *p*=.034, respectively), as shown by paired-samples t-test post-hoc comparisons. However, there were no significant differences among the three cued-conditions. IE scores were higher in the No cue condition (*M*=4856 ms) than in the Auditory (*M*=3736 ms), Vibrotactile (*M*=3990 ms), or Audiotactile cue conditions (*M*=3918 ms). The search slope averaged 135 ms/item for the No cue condition, 104 ms/item for the Auditory, 109 ms/item for the Vibrotactile, and 111 ms/item for the Audiotactile cue conditions. There was no interaction between Set size and Cue type, *F*<1.

Experiment 2.2 was designed to test whether bimodal audiotactile cues would lead to a greater facilitation of visual search performance than unimodal auditory or vibrotactile cues. The results did not highlight any added performance benefits associated with the presentation of the bimodal audiotactile cues whose onsets were synchronized with the colour change of the visual target, as compared to the unimodal non-visual cues. While no further improvements in participants' visual search performance were observed with the multisensory (audiotactile) cues in the present study, it is important to note that overall performance was still significantly better than in the uncued condition, and comparable to that seen in the two other unimodal cue conditions. Thus, rather than increasing the saliency of the visual target by way of the redundant crossmodal cue, it would appear as if the non-visual cues used in the present study facilitated participants' visual search performance in more of a shift in participants' attention, possibly by providing the participants with a temporal marker as to when the target colour change was likely to occur (see Coull & Nobre, 1998).

It is important to note, however, that a question has been raised over the temporal marker account of the facilitation of visual search performance by the results of another of Van der Burg et al.'s (2008a) control experiments. In that study (Experiment 2b), Van der Burg and his colleagues replaced the synchronous auditory cue with a peripheral visual halo surrounding the entire visual search display. This visual cue provided the same temporal information regarding the colour change of the visual target as the auditory cue, yet it did not lead to any facilitation of participants' visual search performance. That is, there was no benefit (relative to the No cue condition) from this form of visual cuing. Therefore, the temporal marker hypothesis cannot easily account for the crossmodal facilitation of visual search performance resulting from the presentation of the non-visual cue.

An important difference, however, between the present Experiment 2.2 and the experiments reported by Van der Burg et al. (2008a, 2009) is that the four cue conditions (No cue, Auditory, Vibrotactile, and Audiotactile) were interleaved here within each block of experimental trials. Thus, the presentation of each cue type varied on a trial-by-trial basis, while in Van der Burg et al.'s experiments, each of the cue conditions were presented to the participants in separate blocks of trials. This raises the possibility that participants may have adopted different 'attentional control settings' (Folk, Remington, & Johnston, 1992) in response to the somewhat different task demands required of the participants in Experiment 2.2 of the present study and Van der Burg et al.'s experiments.

2.3. Experiment 2.3

Perrott et al. (1990) reported that the presence of a spatially-informative auditory cue reduced participants' visual target detection and identification latencies by nearly 200 ms, even when the visual targets appeared within 10° of central fixation. The auditory cue in their study consisted of a 10-Hz click train whose onset was simultaneous with that of the visual display and was presented until the participants made a response. The auditory cue was either presented from a stationary loudspeaker positioned directly in front of the participant or from a boom-mounted loudspeaker that was positioned directly behind the visual target on each trial. Thus, the onset of the auditory cue could be both temporally synchronous with that of the visual target and spatially informative regarding its precise location. Perrott et al. reported that target identification latencies were significantly faster following the presentation of the spatially informative auditory cue as compared to performance following the presentation of the spatially uninformative cue. They concluded that the spatial coincidence of the auditory cue and visual target was necessary to guide participants' spatial attention efficiently in the direction of the visual target (see also Dufour, 1999).

Experiment 2.3 was therefore designed to investigate whether the combination of temporally synchronous and spatially informative cuing would further improve participants' visual search performance relative to cues that were temporally synchronous but spatially uninformative with respect to the location of the target in the left or right hemi-field (as in Experiments 2.1 and 2.2). Additionally, Experiment 2.3 was also designed in order to examine whether the presentation of spatially invalid cues would result in significant visual search costs relative to the spatially uninformative or spatially valid cue conditions (cf. Tan et al., 2009).

2.3.1. Methods

Thirty-eight participants took part in Experiment 2.3 (23 female; mean age=26 years; age range=19-44 years), 18 of whom received auditory cues and the remainder received vibrotactile cues. The experimental set-up was identical to that used in Experiment 2.1 with the exception that a temporally synchronous cue was now presented on every trial (that is, the No cue blocks from Experiment 2.1 were no longer presented). In half of the blocks, the cues were spatially uninformative while in the remainder of the blocks, they were spatially informative, with 80% of these trials being spatially valid and the remaining 20% being spatially invalid with regard to the likely location of the visual target in either the left or right hemi-field. For spatially valid cue trials, the auditory or vibrotactile cue was presented from the loudspeaker or tactor corresponding to the side on which the target appeared (left or right hemi-field of the search screen), while the opposite was true for spatially invalid cue trials. On spatially uninformative cue trials, the auditory or vibrotactile cues were presented from both loudspeakers or tactors.

The spatially uninformative and spatially informative cues were presented in separate blocks of experimental trials in order to avoid the possibility that the spatial location of the cues might have influenced the participants' responses even in the spatially uninformative blocks (cf. Zampini, Guest, Shore, & Spence, 2005). The order of presentation of the blocks (spatially uninformative or informative) was counterbalanced and the blocks were presented in alternating order. The experiment took approximately 45 minutes to complete.

2.3.2. Results and Discussion

The IE data are shown in Figure 2.3. The IE data from the two practice blocks and from those trials in which the participants made an erroneous response (*M*=2.1% of trials) were excluded from the data analyses. The remaining data were subjected to a mixed ANOVA with Set size (24, 36, or 48) and Cue type (Spatially uninformative, Spatially valid, or Spatially invalid) as the within-participants factors, and Cue modality (Auditory or Tactile) as the between-participants factor. Once again, the Huynh-Feldt correction was used whenever sphericity was violated (alpha = .05).

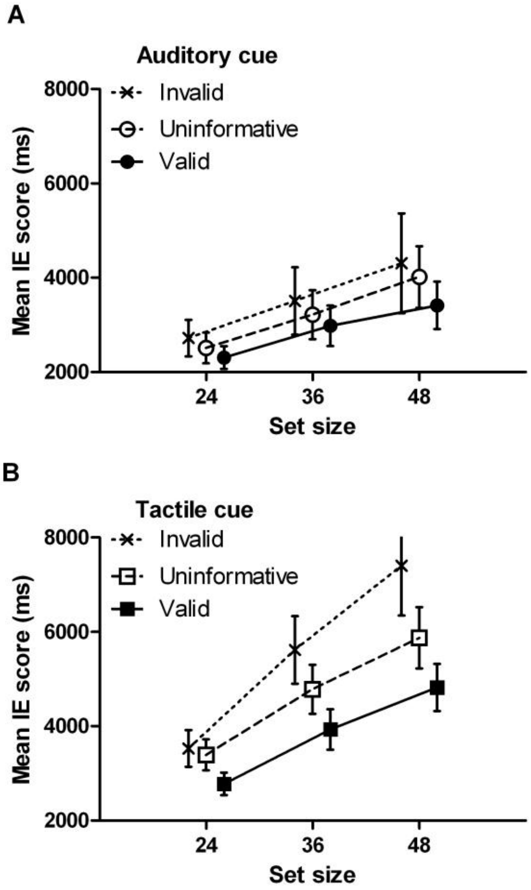


Figure 2.3. Mean IE scores as a function of Set Size (24, 36, or 48), Cue type (Uninformative, Valid, or Invalid), and Cue modality (Auditory or Tactile; Panels A and B, respectively) in Experiment 2.3. Error bars represent the standard errors of the means.

The analysis of the IE data revealed significant main effects of Cue type, *F*(2,72)=7.38, *p*=.005, η2=.17, Set size, *F*(2,72)=43.98, *p*<.001, η2=.55, and Cue modality, *F*(1,36)=4.79, *p*=.035, η2=.12. The Cue type x Set size interaction was marginally significant, *F*(4,144)=2.84, *p*=.06, η2=.07, and the main effects of Set size and Cue modality were qualified by a significant Set size x Cue modality interaction, *F*(2,72)=4.93, *p*=.022, η2=.12. Post-hoc comparisons (independent-samples *t*-tests) revealed a marginally significant difference between the IE scores for auditory (*M*=2514 ms) and vibrotactile (*M*=3234 ms) cues when the set size was 24 (*p*=.07). IE scores were significantly lower for auditory (*M*=3235 ms) as compared to vibrotactile (*M*=4777 ms) cues when the set size was 36 (*p*=.031). They were also significantly lower for auditory (*M*=3913 ms) than for vibrotactile (*M*=6033 ms) cues when the set size reached 48 (*p*=.027). The search slopes averaged 98 ms/item for the spatially invalid, 90 ms/item for the spatially uninformative, and 81 ms/item for the spatially valid auditory cue conditions. The search slopes for the spatially valid auditory cues were significantly shallower (i.e., signifying more efficient search) than for either the spatially uninformative (*p*=.009) or the spatially invalid (*p*=.034) auditory cues. For the vibrotactile cues, the search slopes averaged 153 ms/item, 130 ms/item, and 107 ms/item, respectively. Similarly, the search slopes for the spatially valid vibrotactile cues were significantly shallower than either the spatially uninformative (*p*=.003) or the spatially invalid (*p*=.018) vibrotactile cues.

Experiment 2.3 was designed to investigate whether making the temporally synchronous auditory or vibrotactile cue spatially informative with regards to the likely location (side) of the visual target would further improve participants' visual search performance. To the best of my knowledge, this study is the first to combine temporally synchronous and spatially informative non-visual cues and to measure their effectiveness directly in a visual search task. The results of this experiment clearly show that the presentation of both spatially-informative auditory and vibrotactile cues did indeed enhance the efficiency of participants' visual search performance relative to that seen following either the spatially uninformative or else spatially invalid cues (cf. Tan et al., 2009). However, the crossmodal benefit appeared to be larger following auditory cuing than following vibrotactile cuing, this despite the fact that the spatial information transmitted by the cues was identical in both cases. The search slopes were somewhat shallower following auditory as compared to vibrotactile cuing. It is also important to note that IE scores were overall lower when the auditory (*M*=3220 ms) rather than vibrotactile cues (*M*=4681 ms) were presented. This result therefore suggests that auditory cues may crossmodally facilitate participants' visual search performance more effectively, at least when the cues provide spatial information regarding the likely target side (cf. Ho, Tan, & Spence, 2006).

Significant overall differences between the auditory and vibrotactile cue conditions were observed in Experiment 2.3, whereas no such differences had been detected between the temporally synchronous but spatially uninformative cues in Experiments 2.1 and 2.2 (nor in Van der Burg et al.'s, 2008a, 2009, previous studies). This result points to the possibility that there may be important differences in the nature of the spatial information provided by cues in the two sensory modalities that may influence the relative effectiveness of auditory or vibrotactile cues in providing spatially salient information regarding a target's likely location. Indeed, Ho, Tan, and Spence (2006) have already suggested that presenting a cue that is directionally congruent (i.e., in the front-back or left-right dimension) with respect to a visual target might not be sufficient in-and-of-itself to produce significant spatial attentional effects (i.e., performance benefits). Rather, they argued that the approximate spatial co-localization of the cue and target (within the same functional region of space) was also necessary to elicit the maximal attentional facilitation (cf. Perrott et al., 1990).

In the present study, the auditory cues came from the same functional region of space as the target (in extrapersonal space), whereas the vibrotactile cues were delivered to the body (in peripersonal space; cf. Previc, 1998, 2000). This difference may help to explain the overall performance differences reported between the spatially informative auditory and vibrotactile cue conditions in Experiment 2.3. In order to test this hypothesis, a fourth experiment was conducted in this chapter in which spatially informative auditory cues were now presented from either close to the participant’s body (i.e., over headphones) or from close to the visual display (but far from the participant’s body, just as in Experiment 2.3). If the difference between the auditory and vibrotactile cues observed in Experiment 2.3 primarily reflected differences in the position in space from which the cue had been presented (i.e., rather than differences in cue modality per se) then a significant difference in the magnitude of the crossmodal cuing effects as a function of whether the auditory cue was presented from close to versus far from the participants’ body might be expected. Based on the results of Experiment 2.3, the facilitation of participants’ visual search performance should be larger in the latter case (i.e., when the cue and target originate from the same, distal, region of space). Experiment 2.4 therefore tested whether a difference in the efficiency of participants’ visual search performance would be observed between a condition in which the auditory cues were presented via headphones (far from the visual display) and a condition in which they were played over external loudspeaker cones situated close to the visual display.

2.4. Experiment 2.4

Ho, Tan, and Spence (2006) suggested that presenting a cue that is directionally congruent (i.e., in the left-right dimension) with respect to a visual target might not be sufficient in-and-of-itself to produce significant spatial attentional effects (i.e., performance benefits). Rather, they argued that the approximate spatial co-localization of the cue and target (within the same functional region of space) is also necessary to elicit attentional facilitation (cf. Perrott et al., 1990).

If the spatial co-localization of the cue and target is as important as Ho, Tan, and Spence (2006) asserted, then one might expect to see a significant difference in the magnitude of any crossmodal cuing effect as a function of whether the auditory cue was presented from close to versus far from the participants' body. Based on this line of reasoning, the facilitation of participants' visual search performance should be larger in the latter case (i.e., when the cue and target originate from the same, distal, region of space). If, however, spatial co-localization is not vital in terms of facilitating participants' visual search performance with spatially informative cues, then one would expect to find no significant differences between conditions in which the cue originated from close to versus far from the participants' body.

In order to test these two possibilities, Experiment 2.4 was conducted in which spatially informative auditory cues were now presented from either close to the participant’s body (i.e., over headphones) or from close to the visual display (but far from the participant’s body, just as in Experiment 2.3). This latest experiment tested whether a difference in the efficiency of participants’ visual search performance would be observed between a condition in which the auditory cues were presented via headphones (i.e., far from the visual display) and a condition in which they were played over external loudspeaker cones situated close to the visual display.

2.4.1. Methods

Twenty-eight participants took part in Experiment 2.4 (13 female; mean age=27 years; age range=20-40 years). The experimental set-up was identical to that used in Experiment 2.3 with two exceptions: 1) The spatially informative cues were now always valid with regard to the location of the visual target in the left or right hemi-field (that is, no invalid trials were included, as had been the case in Experiment 2.3); and 2) For half of the participants (N=14), auditory cues were presented over closed ear headphones (Beyer Dynamic DT 531), while for the remainder, the auditory cues were presented by means of external loudspeakers. The intensity of the auditory cues (based on the combined output of the speaker pair) was subjectively matched for both modes of stimulus presentation. In half of the blocks of trials, the cues were spatially uninformative with regard to the likely side of the visual target, while in the remainder of the blocks, they were informative (100% valid) with regard to the target side. The order of presentation of the blocks (spatially uninformative or spatially informative) was counterbalanced and the blocks were presented in alternating order. The experiment took approximately 45 minutes to complete.

2.4.2. Results and Discussion

The IE data are shown in Figure 2.4. The IE data from the two practice blocks and from those trials in which the participants made an erroneous response (M=2.7% of trials) were excluded from the data analyses. The remaining data were subjected to a mixed ANOVA with Set Size (24, 36, or 48) and Cue Type (Spatially Uninformative or Spatially Informative) as the within-participants factors, and Mode of Presentation (Headphone or Loudspeaker) as the between-participants factor.

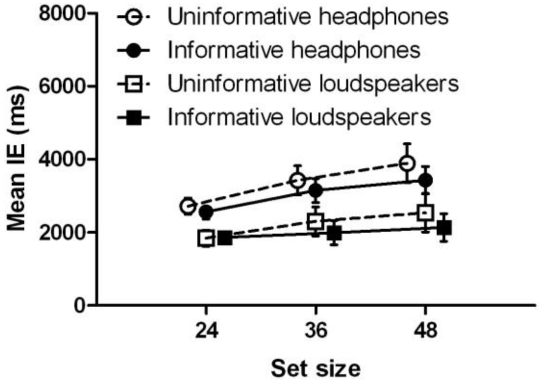


Figure 2.4. Mean IE scores as a function of Set Size (24, 36, or 48), Cue Type (Uninformative or Informative), and Mode of Presentation (Headphones or Loudspeakers) in Experiment 2.4. Error bars represent the standard errors of the means.

Analysis of the IE data revealed a significant main effect of Set Size, $F(2,52)=16.59$, $p<.001$, $\eta^2=.39$, with IE scores increasing as set size increased. There was a marginally significant main effect of Cue Type[5], $F(1,27)=3.72$, $p=.06$, $\eta^2=.12$. Of particular interest, there was a significant main effect of Mode

---

[5] Note that the marginally significant main effect of Cue Type may have been due to a lack of power (i.e., too few participants). Nevertheless, the trend was that participants were faster to find the target in the presence of spatially informative as compared to spatially uninformative auditory cues.

of Presentation on the IE data, $F(1,26)=5.82$, $p=.023$, $\eta^2=.18$, with lower IE scores (i.e., improved performance) reported for the Loudspeaker condition (M=2045 ms) as compared to the Headphone condition (M=3139 ms).

Experiment 2.4 was designed to test whether the mode of presentation of the auditory cues would affect the efficiency of participants' visual search performance. Previous research had suggested that the spatial co-localization of a cue and target event can yield larger performance benefits (e.g., see Bolia et al., 1999; Ho, Tan, & Spence, 2006; Perrott et al., 1990). The results of Experiment 2.4 confirmed this suggestion by demonstrating that participants' visual search performance was significantly enhanced (i.e., IE scores were significantly lower) when the auditory cues were presented from close to the visual display (via external loudspeakers) than when they were presented from close to the participant's body (via headphones). Thus, a greater crossmodal facilitation of visual search performance was observed when the cue and target stimuli originated from the same, distal, region of space. In fact, presenting the auditory cues via loudspeakers led to a 34% improvement in participants' performance relative to when the same auditory information was presented via headphones. This suggests that the difference in the effectiveness of the auditory and vibrotactile cues reported in Experiment 2.3 may have had more to do with differences in cue position (relative to the visual target display) rather than differences in the cue modality itself.

2.5. Experiment 2.5

Having demonstrated that temporally synchronous auditory cues that are made to also be spatially informative with regard to the location of the visual target can give rise to even greater performance improvements than cues that are temporally synchronous but spatially uninformative, Experiment 2.5 went on to compare the advantage of temporal versus spatial cuing directly. Additionally, participants' performance in the presence of these two cuing conditions to the temporally synchronous and spatially informative cuing condition was compared. In light of the findings of Experiment 2.3, temporally synchronous and spatially informative cues were expected to give rise to the greatest facilitation of participants' visual target identification performance.

2.5.1. Methods

Twenty-six participants took part in Experiment 2.5 (19 female; mean age=27 years; age range=19-36 years). The experimental set-up was similar to that of Experiment 2.4 in that the spatially informative auditory cues were 100% valid whenever they were presented and the cues were presented over loudspeakers. Importantly, the spatially informative cue was temporally uninformative with regards to the colour change of the target and consisted of a tone that beeped continuously, at a fast pulse rate (approximately four beeps per second). The presence of the temporally synchronous, spatially informative, or temporally and spatially informative cue was varied between blocks. The order of presentation of the blocks was counterbalanced across participants and presented in alternating order.

2.5.2. Results and Discussion

The IE data are shown in Figure 2.5. The IE data from the practice block and from those trials in which the participants made an erroneous response (*M*=2.7% of trials) were excluded from the data analysis. The remaining data were subjected to a repeated measures ANOVA with Set Size (24, 36, or 48) and Cue Type (Temporal, Spatial, Temporal-Spatial) as the within-participants factors.

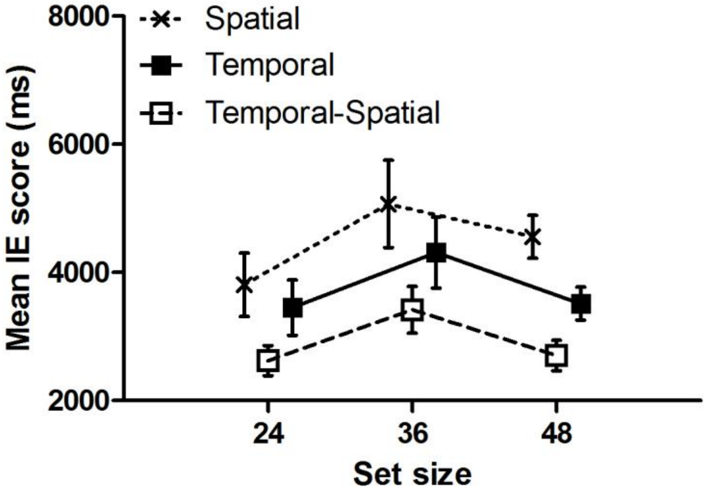


Figure 2.5. Mean IE scores as a function of Set Size (24, 36, or 48) and Cue Type (Temporal, Spatial, Temporal-Spatial) in Experiment 2.5. Error bars represent the standard errors of the means.

The analysis of the IE data revealed significant main effects of Cue type, $F(2,50)=33.57$, $p<.001$, $\eta^2=.57$, and Set size, $F(2,50)=3.41$, $p=.049$, $\eta^2=.12$. Post-hoc comparisons (independent-samples t-tests) revealed that the presence of the temporal-spatial auditory cue (M=2914 ms) gave rise to significantly better visual target identification performance than either the temporally synchronous (M=3759 ms) or the spatially informative auditory cue (M=4479 ms), $p<.001$ for both comparisons. Interestingly, participants' performance following the presentation of temporal cuing was also significantly better than following spatial cuing, $p<.001$.

As mentioned previously, a change in the search slope reflects a change in search efficiency, whereas a change in the slope-intercept reflects overall facilitation of search performance. The results of Experiment 5 appear to reflect both efficient search performance and overall facilitation. The search slopes for all three cue types appeared to be flat in Experiment 2.5. This is indicative of participants' performance approaching levels of parallel search, where an increase in the number of distractors in the search field does not results in any corresponding increase in search time. This finding may be taken to suggest that all three cue types gave rise to more efficient visual search performance. What differentiates the three cue types from one another are their respective slope-intercepts. The temporally-synchronous and spatially-informative cues gave rise to the fastest search times across all set sizes, while the spatially-informative and temporally-uninformative cues gave rise to the slowest search times. Taken together, the results of Experiment 2.5 therefore suggest that in the context of dynamic visual search in small, cluttered visual fields, temporal cuing is more effective than spatial cuing at facilitating visual target identification. Still, having a cue that is not only temporally synchronous with the target but also spatially informative with regards to its location proves to be the most effective means of improving visual target identification.

It is possible that the relative weakness of the spatially informative compared to temporally synchronous auditory cue might have been due to the fact that the screen size was relatively small, such that any additional spatial information would have resulted in a minimal amount of facilitation given that most of the visual stimuli were within participants' central field of view (Andersen & Dyre, 1989; Stiles, 1949). In this case, the information provided by the temporal cue provided a greater reduction in uncertainty (Proctor & Van Zandt, 1994) and was, thus, more effective than the spatial cue. Experiment 2.6

therefore went on to test whether the same pattern of results would emerge from a setting in which the visual field was much larger, presumably rendering the spatial information provided by the cue even more important/informative than when presented in the context of a small visual search field.

2.6. Experiment 2.6

2.6.1. Methods

Twenty-four participants took part in Experiment 2.6 (15 female; mean age=25 years; age range=18-34 years). The experimental design and procedure was identical to that of Experiment 2.5 with the exception that the stimuli were now projected onto a large (approximately 90 cm x 70 cm) screen made of a white audio-transparent cloth and the loudspeakers were placed directly behind the screen.

2.6.2. Results and Discussion

The IE data are shown in Figure 2.6. The IE data from the practice block and from those trials in which the participants made an erroneous response (*M*=2.1% of trials) were excluded from the data analysis. The remaining data were subjected to a repeated measures ANOVA with Set Size (24, 36, or 48) and Cue Type (Temporal, Spatial, Temporal-Spatial) as the within-participants factors.

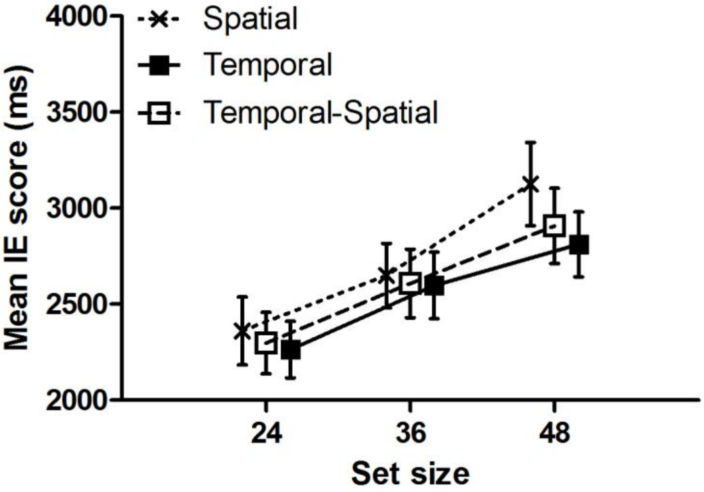


Figure 2.6. Mean Inverse Efficiency scores (IE) as a function of Set Size (24, 36, or 48) and Cue Type (Temporal, Spatial, Temporal-Spatial) in Experiment 2.6. Error bars represent the standard errors of the means.

The analysis only revealed a significant main effect of Set Size, $F(2,46)=64.61$, $p<.001$, $\eta^2=.74$, with mean IE scores increasing with set size. There was, however, no significant main effect of Cue Type, $F<1$. The results of Experiment 2.6 demonstrate that there was no difference among participants' performance in the presence of temporally synchronous, spatially informative, or temporally synchronous and spatially informative cues. This result suggests that all three cues gave rise to similar improvements in visual target identification performance, at least in the context of a large visual search field. Importantly, participants' performance was similar in the presence of either the temporally synchronous or spatially informative auditory cue. This result could be taken to suggest that in a larger search field, the spatial information provided by the cue is just as important as the temporal information. Interestingly, the combination of temporally synchronous and spatially informative auditory cuing did not result in a crossmodal facilitation effect that was significantly better than that observed following either temporally synchronous or spatially informative cuing alone.

2.7. General Discussion

The goal of the six experiments reported in Chapter 2 of this thesis was to extend Van der Burg et al.'s (2008a, 2009) recent studies using the auditory-visual "pip and pop" and tactile-visual "poke and pop" paradigms. Van der Burg et al. investigated the facilitation of participants' visual search performance when non-visual cues were temporally synchronized with the colour change of a visual target whose orientation (horizontal vs. vertical) their participants had to discriminate. Van der Burg et al. reported that the presence of either a temporally synchronous auditory or vibrotactile (but interestingly, not visual) cue resulted in significantly faster search latencies relative to trials in which no cue was presented.

It is important to note here that rather than presenting the auditory cues via headphones (as in Van der Burg et al., 2008a) and presenting the vibrotactile cues to the participants' hand (as in Van der Burg et al., 2009), the auditory cues in the experiments reported here were presented via loudspeakers while the vibrotactile cues were presented to the participants' waists. Even though the position of the cues differed between the present study and those reported by Van der Burg et al. (2008a, 2009), the results of Experiment 2.1 nevertheless converge in showing that the presence of auditory or vibrotactile cues can give rise to a substantial crossmodal facilitation of participants' visual search performance when compared to the performance seen when no cue is present. Importantly, the results of Experiments 2.1, 2.2, and 2.3 also demonstrated that both the auditory and vibrotactile cues were equally effective in terms of enhancing participants' visual search performance when the cues were presented at the same time as the target colour change. The effectiveness of unimodal auditory and vibrotactile cues were compared to bimodal audiotactile cues in Experiment 2.2. The results demonstrated that the presentation of the combined audiotactile cues resulted in a performance enhancement that was similar to, but importantly no bigger than that observed following unimodal (either auditory or vibrotactile) cuing.

The results of Experiments 2.3 through 2.5 provide evidence that the presentation of auditory (and vibrotactile) cues that are not only temporally synchronous, but also spatially informative with respect to the likely location of a visual target hidden in a cluttered, dynamic visual search display can lead to an even larger performance benefit. These experiments constitute the only ones to date to have combined both temporally synchronous and spatially informative cues and to have shown that the combination of these two cue features can further improve participants' visual search performance as compared to cues that are only temporally synchronized (but spatially-uninformative) with respect to the dynamic colour change of the visual target. The spatially informative, non-visual, cues improved participants' visual search performance still further (on the valid trials) relative to the no cue baseline trials, invalid spatial cuing gave rise to significant search costs (see also Tan et al., 2009).

Several studies have highlighted the importance of the spatial correspondence between auditory and visual stimuli in modulating participants' performance in tasks involving visual targets. For example, an earlier study by Bolia et al. (1999) demonstrated that the presentation of free-field auditory spatial cues in a visual search task led to rapid search latencies that are characteristic of parallel search, whereas similar virtual auditory cues, while also resulting in significant reductions in search latencies as compared to when no sound was presented, nevertheless led to performance that was most consistent with serial search (cf. Roberts, Summerfield, & Hall, 2009). An intriguing study of crossmodal temporal adaptation conducted by Di Luca, Machulla, and Ernst (2009) reported that when repeatedly-presented, asynchronous auditory and visual stimuli were spatially co-localized (with the auditory stimuli presented via loudspeakers placed behind the light source), significant changes were observed in the perceptual latency of the visual stimuli. However, when the asychronously-presented auditory and visual stimuli were presented from different locations (i.e., when the auditory stimuli were presented via headphones while the visual stimuli were still presented from in front of the participant) this resulted in significant changes in the perceptual latency of the auditory stimuli instead. Zampini et al. (2005) also found that when required to judge the simultaneity of auditory and visual stimuli, participants were more likely to report them as having been presented simultaneously when the stimuli were presented from the same spatial position as compared to when they originated from different spatial positions.

Previous research by Ho, Tan, and Spence (2006) suggested that spatial cuing effects may differ as a function of the cue modality (i.e., auditory or vibrotactile). In their study, participants performed a visual discrimination task (making speeded discrimination responses concerning the colour change of a car’s numberplate) following the presentation of a spatially predictive auditory or vibrotactile cue from either the front or rear. Ho and her colleagues observed a significant facilitatory effect following auditory, but not vibrotactile, cuing. They suggested that non-visual cues that are directionally congruent with respect to a visual target may primarily give rise to response priming benefits, whereas cues that share (i.e., come from) the same functional region of space as the target can give rise to attentional facilitation as well. Ho et al. concluded that in order for a spatial cue to be maximally effective, it should be co-localized with the target event.

In the experiments reported in this chapter, it was found that the spatially informative auditory and vibrotactile cues were both effective in facilitating participants' visual search performance, even though the cues came from different functional regions of space (i.e., the auditory spatial cues came from loudspeakers positioned close to the visual display, whereas the vibrotactile cues were presented close to the body and away from the visual display). Still, it is important to note that the spatially informative auditory cues were better than the vibrotactile cues in facilitating participants' search performance. This suggests that in order for a crossmodal cue to be maximally effective, it should be in close spatial alignment with the target.

The spatial uncertainty with respect to the location of the visual target may have been limited by the fact that the participants only had to search a small visual field in the present study. The temporal uncertainty with regard to the colour change of the visual target and distractors, however, was a major contributor to the difficulty of the task. Thus, the participants' use of the temporal or spatial information provided by the cue may have been based on the demands of the task. In this case, participants may have strategically used the cue attribute/feature (i.e., temporal synchrony) that was most informative in helping them correctly identify the visual target. The results of Experiment 2.6 indeed corroborated this notion by showing that when trying to locate and identify a visual target in a large (approximately 90 cm x 70 cm) search field where, presumably, the task becomes more spatially demanding, the benefit of having either a temporally synchronous or spatially informative cue gave rise to an equivalent magnitude of facilitation. Moreover, the facilitation in participants' visual target identification performance following the combination of temporal and spatial cuing was no better than having either temporal or spatial cues alone.

It is important to note that the results of Experiment 2.4 revealed that overall performance was 34% better (in terms of search latencies) when auditory cues were presented via loudspeakers situated to either side of the visual display rather than via headphones (i.e., from close to the participant), thus supporting the conclusions of the many previous studies insisting on the spatial co-localization of the cue and target (e.g., Bolia et al., 1999; Di Luca et al., 2009; Ho, Tan, & Spence, 2006). Moreover, although the spatial co-localization of the cue and target proved not to be vital in facilitating participants' visual search performance in Experiment 2.4, it may have contributed to the lack of any observed improvements in participants' performance following the presentation of the bimodal, as compared to the unimodal, cues in Experiment 2.2, as the auditory and vibrotactile cues in the bimodal cue condition originated from different locations.

The fact that the vibrotactile and auditory cues were not spatially co-localized with the visual target may have required additional cognitive processing from the participants, in the sense that the participants may have had to interpret the fact that the vibrotactile cue located on the body (in their personal space) corresponded with the visual target located in front of the participant (in their peripersonal space; cf. Ho, Tan, & Spence, 2006; Làdavas & Farnè, 2004). Furthermore, while it may have been possible for auditory cues to capture participants' attention, causing them to automatically shift their attention toward the location of the auditory stimulus, the position of the vibrotactile cue may have impeded this type of automatic orienting due to the fact that the tactile stimulus was presented to the participants' torsos. If participants' attention were to be automatically directed to the location of the tactile stimulation, as might have been the case for the auditory stimuli, this would mean that their attention would have been directed to their torsos and not the computer monitor. It would be interesting to see whether positioning the auditory and vibrotactile cues in the same functional region of space would result in significantly better performance in the bimodal compared as compared to the unimodal cuing condition.

Taken together, the results of the experiments reported in Chapter 2 therefore demonstrate the robust benefits in participants' visual search performance that can be observed following the presentation of temporally synchronous auditory and vibrotactile cues. Importantly, even when the cues were spatially uninformative with regards to the likely location of the visual target, the temporally synchronous cues effectively guided participants' attention to the spatial location of the target. When the cues were made to be spatially informative, participants' visual target identification performance was improved still further.

In the context of the dynamic visual search task, where the visual target is positioned in some location amidst a cluttered search field, it is perhaps not so surprising that spatially informative cues would give rise to some facilitation of participants' performance. The potential benefits of temporal and spatial cuing will be further explored in the experiments reported in Chapter 3, where participants now had to discriminate the location of briefly-presented and masked visual targets embedded in a rapid serial visual presentation stream. In contrast with the dynamic visual search task where the spatial nature of the task might contribute to the effectiveness of the spatially informative cues, the demanding temporal nature of the rapid serially-masked visual discrimination task used in Chapter 3 might lead one to predict a smaller, perhaps non-significant, improvement in performance resulting from spatial cuing. However, that is not what the results of Chapter 3 showed.

CHAPTER 3:

RAPID SERIALLY-MASKED VISUAL SEARCH

3.0. Introduction

Inspired by the pioneering crossmodal research of O'Leary and Rhodes (1984), Vroomen and de Gelder (2000) reported a study in which they explored whether the perceptual organization of an auditory stimulus stream would modulate the perceptual organization of simultaneously-presented visual information. Vroomen and de Gelder were particularly interested in the question of whether the organization of a sequence of tones could facilitate the identification of a visual target in a rapidly-presented sequence of visual displays. The visual target in their experiment was presented in one of the four corners of a visual display. The participants had to indicate in which corner the target appeared. In one condition, the sequence of visual displays was accompanied by a sequence of four low tones (LLLL), while, in another condition, the visual displays were accompanied by a sequence of three low tones and a high tone (LLHL). Importantly, the third out of every four visual displays in the sequence always contained a target (see Figure 3.1), and was thus either synchronized with a low tone in the LLLL tone sequence or else with a high tone in the LLHL tone sequence (see Vroomen & De Gelder, 2000, Experiment 1).

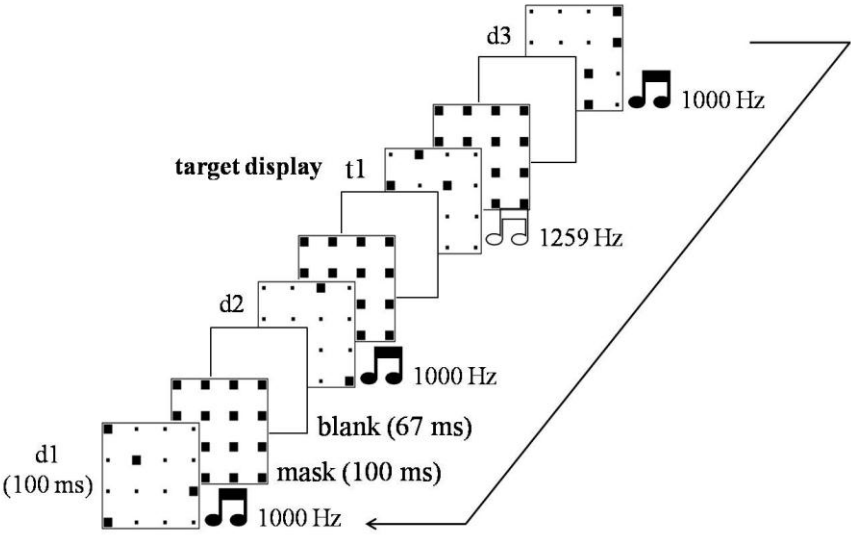


Figure 3.1. Schematic representation of a sequence of stimulus displays for the LLHL condition in Experiment 3.1. The figure shows a sequence of four masked visual displays (d1, d2, t1, and d3), with each display accompanied by a tone. The three distractor displays (d1-d3) are each accompanied by a low (1000-Hz) tone while the target display (t1) is accompanied by a high (1259-Hz) tone.

Vroomen and de Gelder (2000) reported that their participants responded significantly more rapidly and accurately when identifying the visual target if the displays were accompanied by the LLHL sequence as compared to the LLLL sequence. They attributed this crossmodal facilitation of participants' masked visual target identification performance following the presentation of a unique sound to the segregation taking place in the auditory stream which occurred for the LLHL tone sequence, but not for the LLLL tone sequence. Vroomen and de Gelder argued that the participants could clearly differentiate between the low and high frequency tones in the LLHL sequence and that this auditory stream segregation carried over to influence their perception of stimuli in the visual modality (cf. O'Leary & Rhodes, 1984; see also Spence & Chen, in press; Spence, Sanabria, & Soto-Faraco, 2007; Spence & Soto-Faraco, 2010). Vroomen and de Gelder stated that "*perceptual evaluation in one modality may then have* *consequences in other modalities so that coherence is maintained. A sound that segregates in the auditory modality may, for that reason, provoke segregation in the visual modality*" (p. 1590). In the context of crossmodal perceptual organization, Vroomen and de Gelder argued that the segregation of the auditory cue and the visual target display from the distractor displays allowed the participants to access the information contained within the visual target display for longer than in the LLLL sequence trials, thus resulting in enhanced visual target identification performance.

Vroomen and de Gelder (2000, Experiment 5) directly tested the crossmodal perceptual organization account of their results by manipulating the likelihood that the high tone would segregate from the sequence of low tones. In particular, they introduced a LMHL (low-middle-high-low) tone sequence condition with the rationale that the high tone in this sequence would be less abrupt (and noticeable) when it was preceded by the presentation of the middle tone (two semitones higher than the low tone) than the same high tone embedded in the LLHL sequence. Given that abrupt sounds are more likely to segregate than less abrupt sounds (Bregman, 1990), they predicted that the high tone would be less likely to segregate from the LMHL sequence as compared to the LLHL sequence, which in turn would lead to worse performance when the target display was synchronized with the high tone from the LMHL sequence than from the LLHL sequence. This is indeed what was observed, thus supporting Vroomen and de Gelder's crossmodal perceptual organization account of their data.

The importance of a sound's abruptness in the context of auditory stream segregation can be taken as evidence in support of Vroomen and de Gelder's (2000) crossmodal perceptual organization account. It can also be taken, however, to suggest that any single abrupt cue (i.e., visual, auditory, or vibrotactile) may, in-and-of-itself, be sufficient to enhance participants' visual target identification performance. This may not necessarily be because it allows for better auditory stream segregation (and hence, crossmodal perceptual organization), but perhaps because the abruptness of the auditory cue somehow increases the saliency of the synchronous visual target (cf. Anderson & Mamassian, 2008; Fiebelkorn, Foxe, Butler, & Molholm, 2011; Matusz & Eimer, 2011; Noesselt, Bergmann, Hake, Heinze, & Fendrich, 2008; Van der Burg, Cass, Olivers, Theeuwes, & Alais, 2010). Vroomen and de Gelder did not test this possibility, however. Instead, they used an auditory sequence in all 6 of the experiments that they report with the high tone always embedded in a sequence of lower tones. Thus, whether the effect of the tone's abruptness worked to modulate the crossmodal perceptual organization of the auditory and visual information or whether instead it affected some other mechanism responsible for the observed performance benefits could not be dissociated. Moreover, the "abruptness" of the sound, in the sense that the term was used by Vroomen and de Gelder, is ambiguous. In the context of the stream of auditory tones used in their study, the abruptness of the high tone referred to the magnitude of the frequency change (or difference) between the synchronized high tone and the distracting low tones. The abruptness of a cue, in its traditional sense, though is used to refer to the rise/decay time of the cue. Because there appears to be a discrepancy in the traditional use of the term abrupt to describe the sudden onsets of cues and that used by Vroomen and de Gelder to describe the change in frequency between high and low tones, it is important to note that, in this chapter, any single onset cue is referred to as abrupt and the high tone within a sequence of other low tones as a frequency change.

In another experiment, Vroomen and de Gelder (2000; Experiment 4) demonstrated that even when the temporal structure or order of the LLHL tone sequence was disrupted by varying the number of low tones (i.e., distractors) preceding the presentation of the high tone (i.e., target) so that the occurrence of the high tone was made unpredictable, they still found a significant facilitation of participants' visual target identification performance when the high tone was presented at the same time as the visual target. Vroomen and de Gelder went on to conclude that the synchrony of the auditory cue with the visual target was a key factor in eliciting the crossmodal performance benefits that were observed. Similarly, Van der Burg et al. (2008a, 2009) demonstrated that the presentation of temporally synchronous auditory and vibrotactile cues gave rise to a significant reduction in search latencies (in excess of 1000 ms) for visual targets, as described and reviewed in the preceding chapter. While the visual search task used by Van der Burg and his colleagues is comparable to the rapid serially-masked visual target identification task used by Vroomen and de Gelder, it is important to note that the former involves the presentation of a static, permanent target, while the latter involves the transient presentation of the target. As such, the tasks may be fundamentally different, thereby revealing the operation of potentially different underlying mechanisms to account for the temporally synchronous cuing effects observed.

Nevertheless, based on the results of Vroomen and de Gelder's (2000) experiments and those reported by Van der Burg et al. (2008a, 2009, 2010), it appears as though the necessary conditions for achieving a significant improvement in participants' visual target identification involve the presentation of cues that are abrupt and temporally synchronous with the onset of the visual target. If the abruptness and temporal synchrony of the cue are at stake, it seems logical to predict that a single, temporally synchronous cue (be it auditory, vibrotactile, or even visual, for that matter) should suffice to produce the same, if not larger, improvements in performance as a high tone embedded in a sequence of low tones. Vroomen and de Gelder did not test this possibility, presumably because their study was framed in the context of crossmodal perceptual organization.

In fact, to date, no studies have directly attempted to replicate Vroomen and de Gelder's (2000) important findings with regards to crossmodal perceptual organization or tested the possibility that a single cue, when presented in synchrony with the target display, could lead to similar performance improvements to those observed in the presence of the stream segregated high tone embedded in the sequence of low tones. If a single, temporally synchronous cue (auditory, vibrotactile, or visual) were found to facilitate participants' visual target identification performance, this would suggest that one does not necessarily need to involve the notion of auditory stream segregation, or crossmodal perceptual organization, in order to explain the reported performance benefits. Perhaps, by itself, a single cue might suffice to produce the freezing effect and hence enhance the saliency of the neural representation of the visual target. However, if a single cue were not to give rise to any facilitation of participants' visual target identification performance, this would then provide support for the idea that it is the perceptual organization of the auditory and visual streams that critically underlies the crossmodal facilitatory effects that were observed in Vroomen and de Gelder's study. The first goal of the experiments reported in this chapter was therefore to test these two possibilities using an experimental set-up identical to that of Vroomen and de Gelder (Experiment 1).

3.1. Experiment 3.1

3.1.1. Methods

Participants. Eighteen participants (15 female; age range = 18-36 years; mean age = 26 years) took part in this experiment. The experiment took approximately 20 minutes to complete. The participants received a £5 (UK Sterling) gift voucher in return for taking part in the study.

Apparatus and stimuli. The visual display consisted of a 4 x 4 array of quasi-randomly flickering white dots (each dot was 4 x 4 pixels in size) and the entire array measured 4.2 x 4.2 cm in total. The flickering display was created by presenting a series of visual stimuli in rapid succession. Each visual display contained four uniquely-positioned white dots, which made up the full matrix of 16 dots when the various four-dot displays were overlaid.

The third stimulus in each quartet of displays always contained four dots positioned so as to form a diamond. This diamond constituted the target stimulus that participants had to detect and whose location they had to indicate. The target display always appeared in one of the four corners of the matrix: Top-left, top-right, bottom-left, or bottom-right. The participants made their responses using the numerical keypad of a standard QWERTY keyboard by pressing "9" if the target appeared in the top-right, "7" if the target appeared in the top-left, "3" if the target appeared in the bottom-right, or "1" if the target appeared in the bottom-left corner of the display. Note that the third (target) display in each 4-display sequence was the only one to contain dots that were arranged so as to form a diamond (i.e., the dots never formed a diamond in any of the distractor displays).

Each four-dot display (including the target display) was presented for 100 ms (i.e., for 6 screen refreshes) and was immediately followed by a mask containing all 16 dots (i.e., the entire matrix), which was also presented for 100 ms. Immediately after the presentation of the mask, a blank screen was presented for 67 ms. This cycle was repeated continuously with no interruption until either the participant made a response or until 10 full cycles had been presented. If the participants had not made a response by the end of the 10th cycle, the next trial began automatically. Each trial began when the participant pressed the space bar and ended when either the participant made a response or else when the 10th display cycle had been presented.

The auditory stimuli in Experiment 3.1 consisted of sequences of either four 1000-Hz low tones (denoted as LLLL) or else three low tones and a single, 1259-Hz high tone (denoted as LLHL, with the high tone always presented in the third position of each four tones and in temporal synchrony with the target display; 44.1 kHz sampling rate, 16-bit). Note that these tones were presented at exactly the same frequencies as those used in Vroomen and de Gelder’s (2000) original study. The tones were 100 ms long (including a 5-ms fade-in and fade-out to avoid clicks) and were presented from two loudspeaker cones, one placed 16° to either side of the center of the visual display.

Design and procedure. The tone sequence (LLLL or LLHL) was varied randomly on a trial-by-trial basis, with each tone sequence occurring equiprobably within the experimental session. The target was presented 20 times in each of the four corners of the display with both the LLLL and the LLHL sequences, giving rise to a total of 160 pseudorandomly ordered trials in each participant's experimental session.

Vroomen and de Gelder (2000) noted that auditory stream segregation takes time to build up (e.g., see Bregman, 1990) and, therefore, had their participants listen to a series of LLLL and LLHL tone sequences before starting the main experimental session. In the present experiment, participants were also presented with tone sequences alternating between four and eight repetitions (with no gaps between the successive sequences) giving rise to a total of four alternations. Thus, the participants heard the LLLL sequence four times in a row, followed by the LLHL sequence four times in a row, and then the LLLL sequence eight times in a row, followed by the LLHL sequence eight times in a row. The entire sequence was repeated four times with no gaps between any of the tone sequences. During this pre-experimental session, the participants listened to the tones, which were presented in synchrony with masks (i.e., displays containing all 16 dots) and were told not to make any response. Each mask was presented for 200 ms, followed by a blank display for 67 ms. Importantly, auditory stream segregation was not built-up just before each experimental trial, as had been the case in Vroomen and de Gelder's original study. The participants were simply given 16 practice trials before beginning the main experimental session, with half of the trials presented with the LLLL sequence and the remainder presented with the LLHL sequence. The participants were instructed to identify the location of the target diamond as rapidly and accurately as possible. They were also told that the target would always be presented in the third visual display out of each four-display quartet.

3.1.2. Results and Discussion

The RT and accuracy of participants' responses were measured and analyzed[6]. RTs were measured from the onset of the first target display until the participant responded. RTs and percentages of correct responses were subjected to an ANOVA with Tone Sequence (LLLL vs. LLHL) as the within-participant factor (the RT and error data for each condition and experiment are also reported for completeness in Table 3.1). The ANOVA revealed a significant main effect on the percentage of correct responses, $F(1,17)=6.50$, $p=.02$, $\eta 2=.28$. That is, the participants responded more accurately when the visual displays were accompanied by the LLHL tone sequence ($M=81\%$ correct) than when they were presented with the LLLL tone sequence ($M=75\%$ correct; see Figure 3.2). The participants also responded significantly more rapidly when the displays were presented simultaneously with the LLHL tone sequence ($M=3181$ ms) than when they were presented with the LLLL tone sequence ($M=3613$ ms), $F(1,17)=9.57$, $p=.007$, $\eta 2=.36$.

---

[6] Rather than computing IE scores, RTs and percentages of correct responses were measured and analyzed separately for each experiment reported in Chapter 3. This was done in order to allow for direct comparisons between the results reported in Vroomen and de Gelder's (2000) original study and those reported in this chapter. The other rationale for using RTs and accuracy instead of IE scores was that, in contrast with the dynamic visual search task used in Chapter 2 where participants' accuracy was high (approximately 95% correct responses on average) across experimental conditions, participants' accuracy was considerably lower (approximately 75% correct responses on average) and more varied across conditions in the rapid serially-masked visual target discrimination task. Moreover, the various cuing effects observed in the experiments reported in Chapter 3 gave rise to different patterns of facilitation that were either manifest in the RT data only, accuracy data only, or in both RT and accuracy. These varied patterns of facilitation might be taken to reflect the operation of different mechanisms underlying the cuing effects, a point that will be returned to in the General Discussion of this chapter.

Table 3.1. Mean RTs, their Standard Errors (SEM), Percentage of Errors, and IEs for the Various Cue Conditions of Experiments 3.1 through 3.9.

Experiment	Cue condition	RT (ms)	SEM	Error (%)	IE (ms)
3.1	LLLL	3613	322	25	5329
	LLHL	3181	239	19	4272
3.2	Tone absent	3036	276	30	4630
	Tone present	2114	268	19	2917
3.3	No cue	4273	304	30	7003
	Auditory	3664	322	25	5695
	Vibrotactile	3724	286	27	6003
	Audiotactile	3673	290	29	6069
3.4	Visual absent	2767	265	28	4327
	Visual present	2684	325	19	3776
3.5	No cue	3017	385	28	5090
	Auditory	2265	434	13	3606
	Visual	2495	348	16	2878
3.6	MMMM	3746	357	47	6749
	MMLM	3383	344	38	5378
3.7	MMMM	2653	325	28	4248
	MMnoM	2457	339	24	3568
3.8	MMMM	3630	359	30	6022
	MMnoM	3602	381	20	5143
	MMQM	3389	367	24	5161
	MMLM	3209	400	23	5079
3.9	Uninformative	2425	351	19	3453
	Invalid	3191	387	24	4976
	Valid	2455	300	14	3062

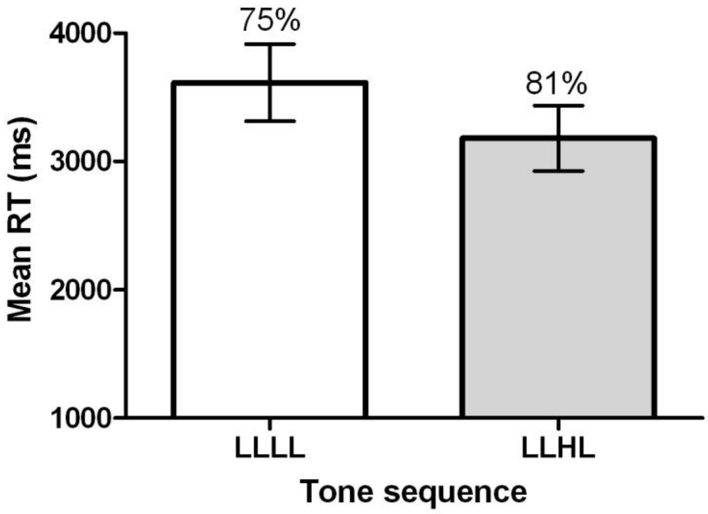


Figure 3.2. Mean RT (in ms) for LLLL and LLHL conditions of Experiment 3.1. The error bars represent the standard errors of the means. Mean percentages of correct responses for each tone sequence are shown above the corresponding error bar.

The results of Experiment 3.1 demonstrate that participants identified the visual targets both significantly faster and more accurately when they were accompanied by a temporally synchronous high tone in the LLHL tone sequence as compared to when accompanied by a low tone in the LLLL tone sequence. Thus, Vroomen and de Gelder’s (2000, Experiment 1) basic findings were replicated in terms of significant effects being reported in both the speed and accuracy of participants’ target identification responses in an experimental set-up that was similar to theirs, with the exception that a build-up of auditory stream segregation was not induced prior to each experimental trial. The fact that participants’ performance was significantly improved in the presence of the synchronous high tone despite the fact that auditory stream segregation was not induced as it had been in Vroomen and de Gelder’s original study suggests that crossmodal perceptual organization may not be the mechanism underlying the facilitation observed in Experiment 3.1. Instead, it is likely due to the abruptness and saliency of the frequency change coinciding with the presentation of the target display. As such, Experiment 3.2 next went on to examine whether or not a single tone presented in synchrony with the visual target would lead to a similar performance benefit to that observed with the LLHL tone sequence.

3.2. Experiment 3.2

3.2.1. Methods

Eighteen participants from the University of Oxford (12 female) ranging in age from 18-39 years (mean age = 26 years) took part in this experiment. The experimental set-up for Experiment 3.2 was essentially the same as that of Experiment 3.1, with the following exceptions: First, instead of having tones presented on each trial, only half of the trials were Tone-present trials whereas the remainders were Tone-absent trials. Second, instead of having four synchronous tones presented with each of the four, four-dot displays (as in Experiment 3.1), only a single 1259-Hz tone was presented simultaneously with the target display on the Tone-present trials. The presence versus absence of the tone was varied on a trial-by-trial basis.

3.2.2. Results and Discussion

RTs and percentages of correct responses were once again subjected to separate ANOVAs with Tone presence (Absent vs. Present) as the within-participant factor. The ANOVA on the mean RT data revealed a significant main effect, $F(1,17)=32.87$, $p<.001$, $\eta 2=.66$. The participants responded significantly more rapidly on the Tone-present trials ($M=2114$ ms) than on the Tone-absent trials ($M=3036$ ms). There was also a significant main effect of Tone presence on the percentages of correct responses, $F(1,17)=12.29$, $p=.003$, $\eta 2=.42$, with participants responding more accurately on the Tone-present trials (*M*=81%) than on the Tone-absent trials (*M*=70%; see Figure 3.3).

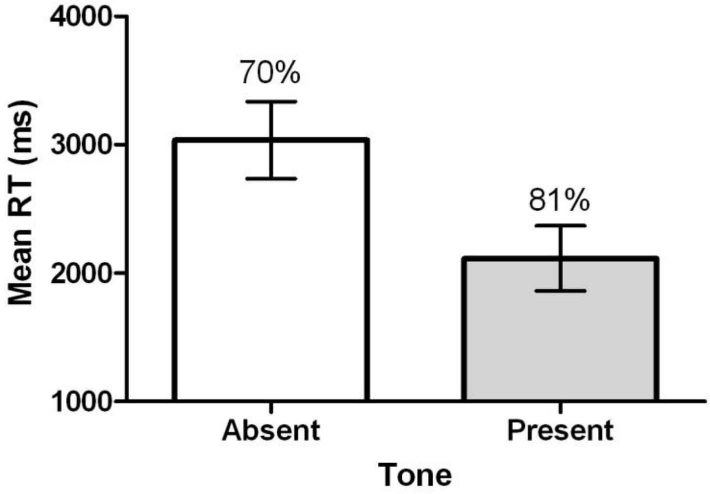


Figure 3.3. Mean RT (in ms) for Tone-absent and Tone-present conditions of Experiment 3.2. The error bars represent the standard errors of the means. Mean percentages of correct responses for each condition are shown above the corresponding error bar.

The results of Experiment 3.2 revealed that a single tone, when presented in synchrony with a visual target, led to significantly faster and more accurate visual target identification performance. Interestingly, the effect sizes when a single tone was presented in Experiment 3.2 were larger for both the RT and accuracy data than those for the LLHL tone sequence in Experiment 3.1. Moreover, the average reduction in RTs was larger (*M*=922 ms) when the single tone was presented than when the tone was embedded in the LLHL sequence (*M*=432 ms). These results therefore support the prediction that a single, temporally synchronous cue would at least lead to an equivalent (and in this case, a much larger) improvement in participants' visual target identification performance as those reported in Vroomen and de Gelder's (2000) previous study. It is possible, then, that rather than crossmodal perceptual organization, another mechanism may underlie the crossmodal facilitation effect observed in the present study and that of Vroomen and de Gelder.

The results of Experiment 3.2 provide initial evidence to support the notion that the abruptness and temporal synchrony of the cue were vital in enhancing participants' visual target identification performance. Several recent studies have shown that vibrotactile cues produce comparable improvements in participants' visual target identification performance to those observed with auditory cues (Van der Burg et al., 2008a, 2009; see also Experiment 2.2. of the present thesis). Experiment 3.3 therefore went on to investigate whether similar visual target performance benefits might be seen using vibrotactile rather than auditory cues in the masked visual target identification paradigm introduced by Vroomen and de Gelder (2000). At the same time, Experiment 3.3 was designed to investigate whether multisensory (audiotactile) cues would give rise to performance benefits that were any larger than those seen following unimodal (auditory or vibrotactile) cuing (cf. Santangelo et al., 2008; Spence & Santangelo, 2009). The results reported in the previous chapter (Experiment 2.2) demonstrated a significant improvement in participants' visual target identification performance following temporally synchronous, multisensory audiotactile cuing, but this improvement was no greater in magnitude than that observed following unimodal auditory and vibrotactile cuing. It possible, therefore, that a similar pattern of equivalent bimodal and unimodal cuing will be observed in Experiment 3.3.

3.3. Experiment 3.3

3.3.1. Methods

Sixteen participants from the University of Oxford (9 female) ranging in age from 18-36 years (mean age = 25 years) took part in this experiment. The experimental set-up was similar to that used in Experiment 3.2 in that it included a tone absent and tone present condition. Additionally, however, a unimodal vibrotactile and a bimodal audiotactile cue condition were also included in the design. The tactile cue consisted of a 50 ms vibration presented by means of a single tactor attached to the back of the participant's left hand. The participants placed their left hand in the middle in front of the space bar of the keyboard, which was centered in front of the computer screen. The intensity of the vibrotactile stimulus was subjectively matched to that of the auditory tone prior to conducting the actual experiment by four volunteers, who did not take part in the main experiment. The four cue conditions (No cue, Auditory, Vibrotactile, or Audiotactile) were interleaved randomly within each block of trials. Each of the four blocks of trials consisted of 80 trials, with each of the four cue conditions and target positions occurring equiprobably and in random order within each block of trials.

3.3.2. Results and Discussion

RTs and percentages of correct responses were once again subjected to separate ANOVAs with Cue type (No cue, Auditory, Vibrotactile, or Audiotactile) as the within-participant factor. The ANOVA revealed a significant main effect on the mean RT data, $F(3,45)=6.88$, $p=.003$, η2=.31. All three cue types (Auditory, Vibrotactile, and Audiotactile) led to faster responses ($M=3664$, 3724, and 3673 ms, respectively) than the No cue condition (*M*=4273 ms), *p*=.012, *p*=.003, and *p*=.008, respectively (see Figure 3.4). The differences in accuracy between the various cue conditions failed to reach significance, *F*(3,45)=1.9, *p*=.14.

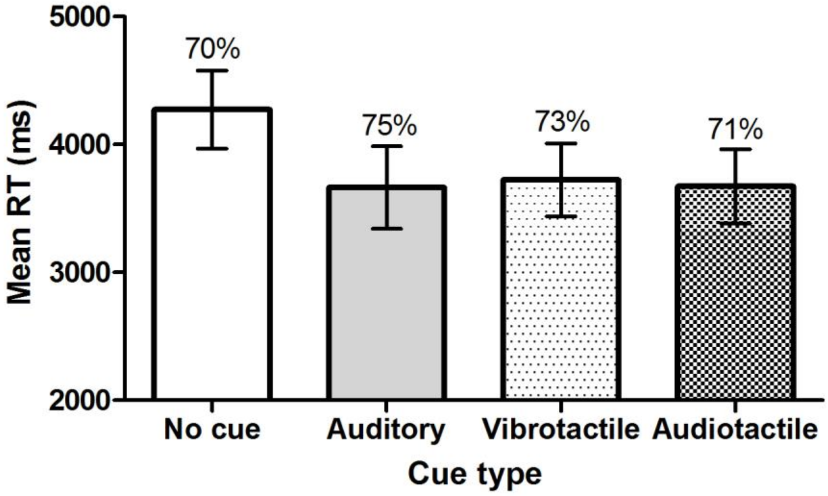


Figure 3.4. Mean RT (in ms) for the No cue, Auditory, Vibrotactile, and Audiotactile cue conditions of Experiment 3.3. The error bars represent the standard errors of the means. Mean percentages of correct responses for each condition are shown above the corresponding error bar.

The results of Experiment 3.3 demonstrated that, while the presentation of a non-visual cue synchronized with the visual target once again led to faster responses by the participants (thus replicating the results of Experiment 3.2), there were no differences among the three cue conditions. Thus, auditory, vibrotactile, and audiotactile cues all led to similar performance benefits in terms of improved visual target identification performance relative to that seen in the uncued baseline condition. This finding suggests that the tactile modality can provide an alternative cue modality that may be just as effective as the more typically used auditory cue modality in terms of crossmodally facilitating participants' performance on visual tasks. What is more, bimodal audiotactile cues did not lead to any further improvements in participants' performance over that seen in response to the unimodal cues. In fact, of the three cued-conditions, the average accuracy for the audiotactile cue condition was numerically somewhat lower (71%) than in either the auditory (75%) or vibrotactile (73%) cue conditions. Mean RTs were also numerically slightly faster for the auditory (3664 ms) than for the audiotactile (3673 ms) cue condition. Thus, there was no evidence of any speed-accuracy trade-offs in the data. This pattern of results is consistent with that observed in Experiment 2.2 (in Chapter 2 of the present thesis) where the temporally synchronous unimodal auditory and vibrotactile cues were just as effective as bimodal audiotactile cues in (significantly) reducing search latencies for visual targets presented amidst the dynamic and cluttered visual search display (cf. Spence & Santangelo, 2009; Van der Burg et al., 2008a, 2009).

3.4. Experiment 3.4

The results of Experiment 3.1-3.3 demonstrated that unimodal and multisensory non-visual cues, when presented simultaneously with the visual target, resulted in improved visual target identification in terms of both the speed and accuracy of participants' responding. Several recent studies have shown that crossmodal cuing is better than unimodal cuing (see Spence & Santangelo, 2009, for a recent review). Van der Burg et al. (2008a; Experiment 2b) even found that the presentation of a visual cue did not give rise to any significant improvements in participants' visual target identification performance (see also Noesselt et al., 2008). Experiment 3.4 examined whether a visual cue would give rise to a significant visual target performance benefit using Vroomen and de Gelder’s (2000) rapid serially-masked visual presentation paradigm. If the visual cue did not lead to any improvements in participants’ visual target identification performance, this would suggest that crossmodal cues are necessary for facilitating performance on this visual target identification task. If, on the other hand, the presentation of a visual cue was to significantly facilitate participants’ visual target identification performance, this would instead suggest that the benefit of presenting a temporally synchronous cue need not be crossmodal. This might suggest instead that the cue simply provides a temporal marker for the visual target and hence, its modality is irrelevant (Watson, Humphreys, & Olivers, 2003).

3.4.1. Methods

Sixteen participants from the University of Oxford (11 female) ranging in age from 20-34 years (mean age = 24 years) took part in Experiment 3.4. The experimental set-up was nearly identical to that of Experiment 3.2, with the sole exception that rather than a tone, on half of the trials, the target was presented together with a light gray outline box surrounding the 4.2 x 4.2 cm grid in which the target appeared (i.e., the visual cue outlined the entire visual display; see Figure 3.5). The presence versus absence of this visual cue was varied on a trial-by-trial basis, just as for the tone presented in Experiment 3.2.

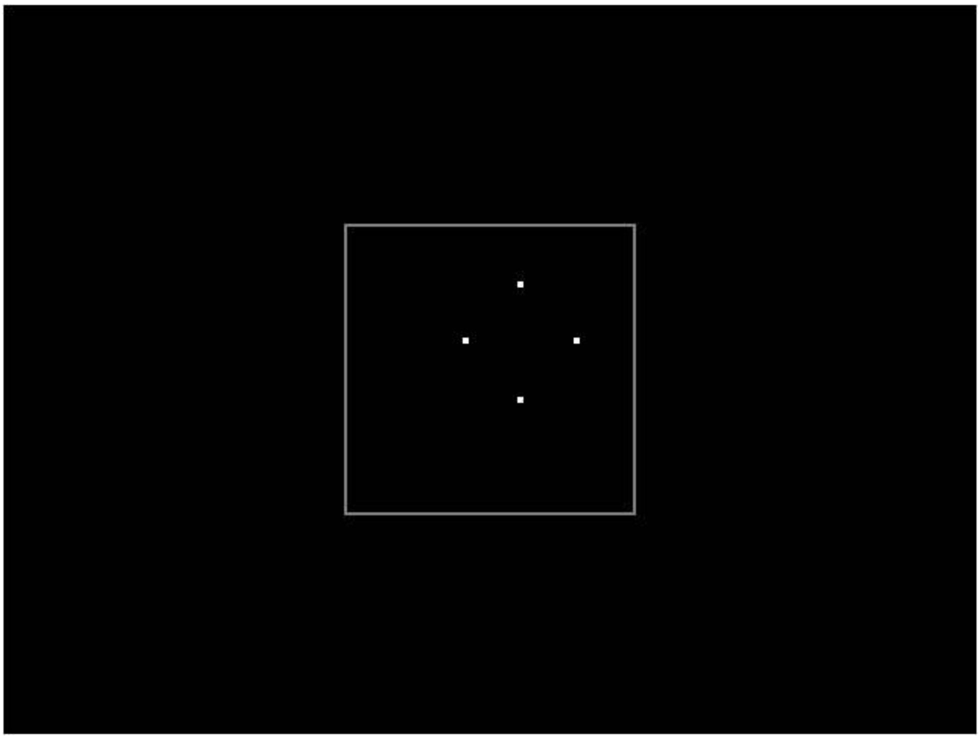


Figure 3.5. Illustration of the light gray outline of a box (visual cue) surrounding the visual target display presented on the computer screen in Experiment 3.4.

3.4.2. Results and Discussion

RTs and percentages of correct responses were subjected to separate ANOVAs with visual cue presence (Absent vs. Present) as the within-participant factor (the RT and error data for each condition are also reported for completeness in Table 3.1). The ANOVA revealed a significant main effect of the presence of the visual cue on the accuracy of participants' responses, $F(1,15)=6.83$, $p=.02$, $\eta2=.31$, with more accurate responding being observed when it was present (M=81%) than when it was absent (M=72%). There was, however, no significant effect on mean RTs, $F<1$ (see Figure 3.6).

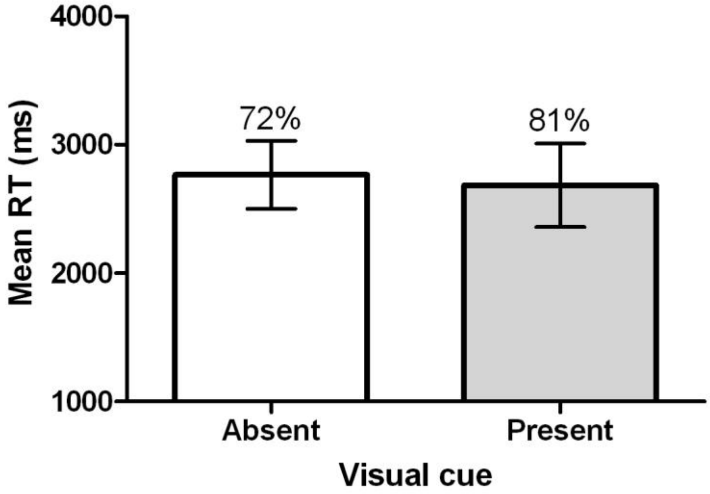


Figure 3.6. Mean RT (in ms) for the visual cue Absent and Present conditions of Experiment 3.4. The error bars represent the standard errors of the means. Mean percentages of correct responses for each condition are shown above the corresponding error bar.

While participants were not significantly faster at identifying the visual target in the presence of the visual cue as compared to when it was absent, their target identification responses were significantly more accurate. Thus, visual cues, when presented in synchrony with the visual target, led to an improvement in participants' visual target identification performance, at least in terms of the accuracy of their responding. A closer look at the effect sizes from this experiment and those reported in Experiment 3.2 also revealed that the effect size for the presence of the single tone in Experiment 3.2 was larger than the effect size observed here in Experiment 3.4. That is, the auditory cue in Experiment 3.2 appeared to be more effective at facilitating participants' visual target identification performance than the visual cue in Experiment 3.4. In fact, the improvement in participants' performance in the presence of the auditory cue in Experiment 3.2 was a 922 ms reduction in average RTs and an 11% increase in the accuracy of participants' responses, whereas the presence of the visual cue in

Experiment 3.4 led to only an 83 ms reduction in average RTs and a 9% increase in accuracy.

The results of Experiment 3.4 suggest that temporally synchronous visual cues can lead to some improvement in participants' visual target identification performance, at least in terms of accuracy, in the Vroomen and de Gelder (2000) task. However, the optimal cue may need to be non-visual (i.e., crossmodal) in order to maximize any potential facilitatory effects observed for visual target perception and identification (cf. Watanabe & Shimojo, 2001). Although there may be an additional benefit for participants' visual target identification performance resulting from the presentation of the auditory (i.e., crossmodal) as compared to the visual (i.e., intramodal) cue, it is important to note the between-participants nature of the comparison leading to this conclusion. It is also important to note that the mean RT from the No cue conditions of Experiments 3.2 and 3.4 (i.e., a condition that was exactly the same in the two studies) differed by nearly 300 ms, pointing to the fact that there may have been individual differences between the participants tested in the two experiments. Therefore, in order to be certain about the relative improvements in participants' visual target identification performance following the presentation of either the auditory or visual cue, a within-participants design, including both cue types, was used in Experiment 3.5. In this way, participants' performance on the visual target identification task when they were exposed to both the auditory and visual cue conditions could be compared.

3.5. Experiment 3.5

3.5.1. Methods

Twelve participants from the University of Oxford (5 female; age range = 19-32 years; mean age = 25 years) took part in this experiment. The experimental set-up was similar to that of Experiment 3.4 with the exception that the participants completed one block of 80 trials in which no cue was presented with the visual target and a second block of 160 trials in which the visual target was presented with the synchronized 1259-Hz tone on half of the trials and the synchronous visual cue (the light gray outline of a box) on the remaining half of the trials. The auditory and visual cues were interleaved randomly within the block of trials and the order of presentation of the blocks was counterbalanced across participants.

3.5.2. Results and Discussion

RTs and percentages of correct responses were subjected to separate ANOVAs with Cue type (No cue, Auditory, or Visual) as the within-participants factor. The ANOVA revealed a significant main effect of Cue type on the accuracy of participants' responses, $F(2,22)=18.49$, $p<.001$, $\eta2=.63$. Participants responded significantly more accurately when the visual target was presented with either the auditory (M=87%) or visual cue (M=84%) than when no cue (M=72%) was presented, $p<.001$ and $p=.001$, respectively (see Figure 3.7). In this experiment, there was no significant difference in accuracy between the Auditory and Visual cue conditions, $p=.28$. There was a significant main effect of Cue type on RTs, $F(2,22)=7.40$, $p=.003$, $\eta2=.40$, with participants responding more rapidly following the presentation of either the tone (M=2265 ms) or the box cue

(*M*=2495 ms) than in the No cue (*M*=3017 ms) condition, *p*=.003 and *p*=.034, respectively. There was no significant difference in RTs between the Auditory and Visual cue conditions, *p*=.24.

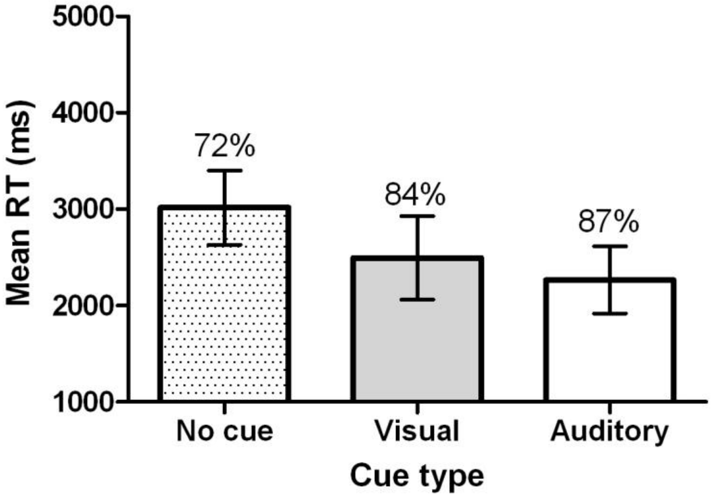


Figure 3.7. Mean RT (in ms) for the No cue, Visual cue, and Auditory cue conditions of Experiment 3.5. The error bars represent the standard errors of the means. Mean percentages of correct responses for each condition are shown above the corresponding error bar.

The results of Experiment 3.5 revealed that the visual cue, just like the auditory cue, was also effective in facilitating participants' performance on the masked visual target identification task (when compared to performance in a condition in which no cue was presented). This result supports the view that a cue need not be crossmodal in order to elicit a significant improvement in the identification of a visual target. Alternatively, however, the fact that the presence of the synchronous auditory and visual cues were interleaved within one block of trials, while the cue absent trials were presented in a separate block, may have led to the participants adopting a search strategy, or attentional set, which might have allowed them to capitalize on the temporally informative nature of the synchronous cue, regardless of its sensory modality (see Leber, Kawahara, & Gabari, 2009). The participants in Experiment 3.4, on the other hand, may not have fully adopted an attentional set that was responsive to the temporal information provided by the visual cue as both cue present and absent trials were interleaved within each block of trials, thus possibly resulting in a partial performance benefit manifesting in an improvement in the accuracy (but crucially not the speed) of their responding.

Nevertheless, the fact that all three cue modalities (auditory, vibrotactile, and visual) gave rise to a significant improvement in participants' visual target identification performance when compared to a condition in which no cue was presented with the target has two important implications: First, the fact that single cues facilitated participants' visual target identification performance suggests that the freezing effect reported by Vroomen and de Gelder (2000) does not necessarily rely on auditory stream segregation. Second, while the freezing effect can be induced crossmodally, that does not necessarily mean that it is a fundamentally crossmodal phenomenon (or that it results from multisensory integration, as originally suggested by Vroomen and de Gelder). The fact that intramodal visual cues facilitated participants' visual target discrimination performance to the same extent as auditory cues argues against the multisensory integration account of the freezing effect. Therefore, on the basis of these results, one might predict that any attention-capturing stimulus (no matter what its modality) would lead to similar facilitatory effects on participants' performance in the freezing effect paradigm.

In Experiments 3.6 through 3.8, various other oddball stimuli, namely the presence of a louder sound (Experiment 3.6), the absence of a sound (Experiment

3.7), or the presence of a quieter sound (Experiment 3.8), in an otherwise regular stream of auditory distractors were explored. The goal of these experiments was to determine whether any oddball stimuli would also give rise to a significant facilitation of participants' visual target discrimination performance. Additionally, these experiments were designed to provide further insight into the role that attentional capture versus multisensory integration plays in the enhancement of visual perception seen in the freezing effect.

3.6. Experiment 3.6

3.6.1. Methods

Sixteen participants (14 of whom were female), ranging in age from 19 to 44 years (mean age = 23 years), took part in Experiment 3.6. The experimental set-up was nearly identical to that used in Vroomen and de Gelder's (2000) original study, where the auditory stimuli consisted of sequences of either four 1000-Hz tones (denoted as MMMM). The four tones were played at an amplitude of 70 dB, or else three 70 dB tones and a single louder tone (denoted as MMLM) played at 75 dB accompanied the visual displays. The louder tone was always presented in the third position of each sequence and in temporal synchrony with the target display. The tone sequence (MMMM or MMLM) was varied randomly on a trial-by-trial basis, with each tone sequence occurring equiprobably within the experimental session.

3.6.2. Results and Discussion

RTs and percentages of correct responses were subjected to separate ANOVAs with Tone sequence (MMMM vs. MMLM) and target location as the within-participants factor. The data from the practice session and from those trials in which the participants failed to make a response (*M*=1.4% of trials) were excluded from the data analysis.

The analysis revealed no significant main effect of target location and no significant interaction between target location and tone sequence for either the RT or the accuracy data. Participants' visual target identification performance was significantly faster when the visual target was presented together with a louder tone in the MMLM sequence (*M*=3383 ms) than when paired with the MMMM sequence (*M*=3746 ms), *F*(1,15)=6.23, *p*=.025, η2=.29. Participants also responded significantly more accurately in the presence of the MMLM sequence (*M*=62%) as compared to the MMMM sequence (*M*=53%), *F*(1,15)=11.35, *p*=.004, η2=.43 (see Figure 3.8).

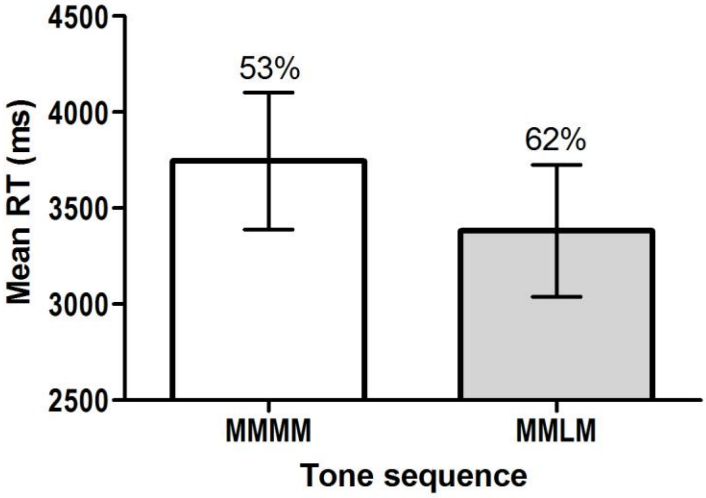


Figure 3.8. Mean RT (in ms) for the MMMM and MMLM conditions tested in Experiment 3.6. The error bars represent the standard errors of the means. Mean percentages of correct responses for each condition are shown above the corresponding error bar.

The results of Experiment 3.6 therefore demonstrate that when a louder tone in an otherwise identical sequence of tones was presented at the same time as the visual target display, significantly better target identification performance was observed, in terms of both speed and accuracy, as compared to a condition in which the tones accompanying the display were all equally loud.

In Vroomen and de Gelder's (2000) original experiment, the pitch of the tone accompanying the target display distinguished the target from the distractors. Vroomen and de Gelder argued that this consequently gave rise to the segregation of the two streams of low-pitched vs. high-pitched tones along with the corresponding crossmodal segregation of the distractor and target displays. Experiment 3.6 evaluated the consequences of varying a different auditory feature, namely loudness, and revealed that when the louder tone accompanied the target display, this also gave rise to a significant performance benefit.

It is well-known that loudness and pitch can mutually influence each other in perception, such that a louder tone will be perceived as being higher in pitch than the same-frequency tone played at quieter amplitude (Neuhoff, Wayand, & Kramer, 2002; Parker & Schneider, 1974). As such, it is possible that the louder tone accompanying the target display in Experiment 3.6 was perceived to be higher in pitch than the tones accompanying the distractor displays. Thus, it may well be the case that the mechanism underlying the facilitation observed in the presence of the MMLM tone sequence in the present study was similar to that observed in the presence of the higher-pitched tone in the LLHL sequence in Vroomen and de Gelder's (2000) original study. In this case, multisensory integration might still account for the facilitation observed in Experiment 3.6. In order to rule out this possibility, therefore, Experiment 3.7 was designed to test whether removing the tone accompanying the visual target display altogether would give rise to any performance benefits.

Should the absence of a tone in the sequence of otherwise identical tones be found to improve participants' visual target identification performance, this would suggest the attentional capture account to be a more appropriate interpretation of the freezing effect than the multisensory integration account, as there is simply no sound for the visual target to be integrated with in this case.

3.7. Experiment 3.7

3.7.1. Methods

Fourteen participants (7 female; age range = 22-36 years; mean age = 27 years) took part in Experiment 3.7. The experimental set-up was the same as that in Experiment 3.6, except that rather than presenting a louder tone in synchrony with the target display, the tone was replaced with a silent period instead. The presence (vs. absence) of a tone with the target display was once again varied on a trial-by-trial basis.

3.7.2. Results and Discussion

RTs and percentages of correct responses were subjected to separate ANOVAs with Tone sequence (MMMM vs. MMnoM) and target location as the within-participants factor. The data from the practice session and from those trials in which the participants failed to make a response (*M*=2.1% of trials) were excluded from the data analysis. There was no significant main effect of target location and no significant interaction between target location and tone sequence. Participants' visual target identification performance was significantly faster when the target was presented together with a louder tone in the MMnoM sequence (*M*=2457 ms) than when paired with the MMMM sequence (*M*=2653 ms), *F*(1,13)=13.52, *p*=.003, η2=.51. Even though the accuracy of participants' responses were numerically greater in the presence of the MMnoM sequence (*M*=76%) as compared to the MMMM sequence (*M*=72%), this difference failed to reach significance, *p*=.14 (see Figure 3.9).

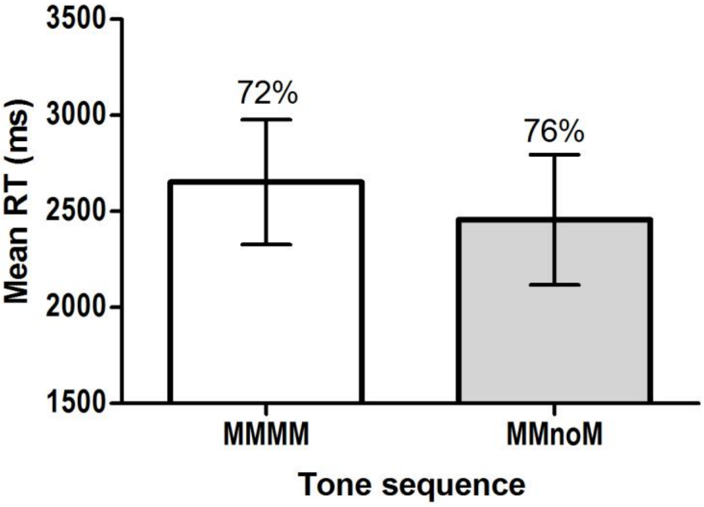


Figure 3.9. Mean RT (in ms) for the MMMM and MMnoM conditions tested in Experiment 3.7. The error bars represent the standard errors of the means. Mean percentages of correct responses for each condition are shown above the corresponding error bar.

The results of Experiment 3.7 demonstrate that the absence of a sound in an otherwise-regular sequence of tones can also give rise to a significant facilitation of participants' visual target identification performance. More specifically, participants' performance was significantly faster, but importantly, not more accurate, when the tone was omitted from an otherwise regular sequence of low tones presented in synchrony with the distractor visual displays (the MMnoM condition) than when a low tone was presented with the target (the MMMM condition). The results of Experiments 3.6 and 3.7 might therefore be taken to suggest that any oddball stimulus can give rise to the same crossmodal facilitation of participants' visual target identification performance. Vroomen and de Gelder (2000) suggested that the crossmodal facilitation effect that they observed when a high tone, which was embedded in a sequence of low tones (LLHL), accompanied the visual target was attributable to multisensory integration. Given that there was no auditory stimulus to integrate with the visual target in the MMnoM condition of Experiment 3.7, it is unlikely that the facilitation observed here could have been due to multisensory integration.

It is important to note that the magnitude of the facilitation effect reported in Experiment 3.7 was numerically smaller (mean cuing effect = 196 ms) than that seen in Experiment 3.6 (mean cuing effect = 363 ms), which might be taken to suggest that the absence of the tone served as a weaker cue as compared to the presence of the louder tone in facilitating participants' visual target discrimination. However, it should also be said that the difference between these cuing effects failed to reach significant, p=.18. Moreover, Watson and Humphreys (1995) demonstrated that both the abrupt onset and offset of stimuli are equally effective at capturing attention in visual search (see also Pfefferbaum, Buchsbaum, & Gips, 1971). Thus, it is perhaps unsurprising that there was no significant difference in the cuing effects observed between the two conditions. Still, it remains possible that there might be a part of the effect that Vroomen and de Gelder (2000) reported that is attributable to multisensory integration and another part that is attributable to attentional capture. Evaluating this possibility, however, is made more difficult on the basis of the results of Experiments 3.6 and 3.7 given the large differences in baseline performance between the two experiments. This makes it difficult to make a meaningful comparison of the magnitude of the facilitatory effect across the two experiments.

Now while the most likely explanation for such differences is simply in terms of individual differences in the motivation/age of the participants, it seemed sensible to compare the effectiveness of the two (MMLM and MMnoM) oddball conditions within the same experiment and within the same group of participants. Additionally, Experiment 3.8 was designed to explore whether an attenuated sound (MMQM) would give rise to an improvement in participants' visual target identification performance that was similar in magnitude to that observed in the presence of the amplified sound in Experiment 3.6.

Research on the effectiveness of receding tones (that is, tones that are falling in amplitude or getting quieter) at capturing attention has been mixed, with some demonstrating that receding auditory signals can capture attention (Abrams & Christ, 2005), some failing to show any attentional capture from falling amplitude tones (Franconeri & Simons, 2003; von Mühlenen & Lleras, 2003), and yet others demonstrating asymmetrical effects of looming versus receding sounds (Schouten, Troje, Vroomen, & Verfaillie, 2011). Based on the inverse effectiveness rule of multisensory integration (Meredith & Stein, 1983), one might expect that an attenuated sound would be more effectively integrated with the simultaneously-presented visual target display than an amplified sound (Spence & Driver, 1999). However, the literature on looming versus receding tones would appear to suggest that any effect of an attenuated sound, if present, is likely to be much weaker as compared to an amplified sound. In Experiment 3.8, therefore, participants' performance in the presence of three oddball types (loud, quiet, or absent tone) was directly compared in order to determine whether the same pattern of facilitation would hold across the three.

3.8. Experiment 3.8

3.8.1. Methods

Eleven participants (6 female; age range = 19-35 years; mean age = 25 years) took part in Experiment 3.8. The experimental set-up for this experiment was the same as that in Experiment 3.7, except that all three oddball tone sequences (MMLM, MMQM, and MMnoM) were included and compared to the baseline regular tone sequence (MMMM). The tone sequences were varied randomly on a trial-by-trial basis, with each tone sequence occurring equiprobably. This gave rise to a total of 320 trials within the experimental session.

The attenuated (quieter) tone was presented at 60 dB (compared to distractor tones still presented at 70 dB), which was a 10 dB decrease in amplitude as compared to a 5 dB increase in amplitude in the amplified (louder) oddball condition. This differential change in amplitude was chosen to compensate for the perceptual asymmetries that exist for rising and falling tones, whereby people reliably overestimate the change in intensity of rising amplitude tones relative to equivalent falling amplitude tones (Neuhoff, 1998, 2001; Neuhoff et al., 2002).

3.8.2. Results and Discussion

RTs and percentages of correct responses were subjected to separate ANOVAs with Tone sequence (MMMM, MMLM, MMQM, and MMnoM) and target location as the within-participants factor. The data from the practice session and from those trials in which the participants failed to make a response (*M*=2.2% of trials) were excluded from the data analysis.

The analyses revealed no significant main effect of target location and no significant interaction between target location and tone sequence for either the RT or the accuracy data. There was a significant main effect of Tone sequence, *F*(3,33)=3.05, *p*=.044, η2=.23 on participants' RT data. Post-hoc comparisons (*t*-tests) revealed that participants' visual target identification performance was significantly faster when the target was presented together with a louder tone in the MMLM sequence (*M*=3209 ms) than when paired with either the MMnoM sequence (*M*=3602 ms; *p*=.046) or the MMMM sequence (*M*=3630 ms; *p*=.035). RTs were also significantly faster in the presence of the MMQM (*M*=3389 ms) as compared to the MMMM sequence, *p*=.038 (see Figure 3.10). Neither the differences between participants' RTs in the presence of the MMLM and MMQM sequences nor between the MMQM and MMnoM were significant, *p*=.29 and *p*=.27, respectively.

Analysis of participants' accuracy data revealed a significant main effect of Tone sequence, *F*(3,33)=6.27, *p*=.002, η2=.39. Participants responded significantly more accurately in the presence of the MMLM sequence (*M*=77%) as compared to the MMMM sequence (*M*=70%), *p*=.029, but there were no differences in the accuracy of participants' responding between the MMLM and either the MMnoM (*M*=80%; *p*=.13) or the MMQM (*M*=76%; *p*=.65) sequences. Participants also responded significantly more accurately in the presence of the MMnoM sequence as compared to the MMMM sequence (*p*=.002) and the MMQM sequence (*p*=.035). Finally, participants also responded significantly more accurately in the presence of the MMQM sequence than the MMMM sequence, *p*=.049.

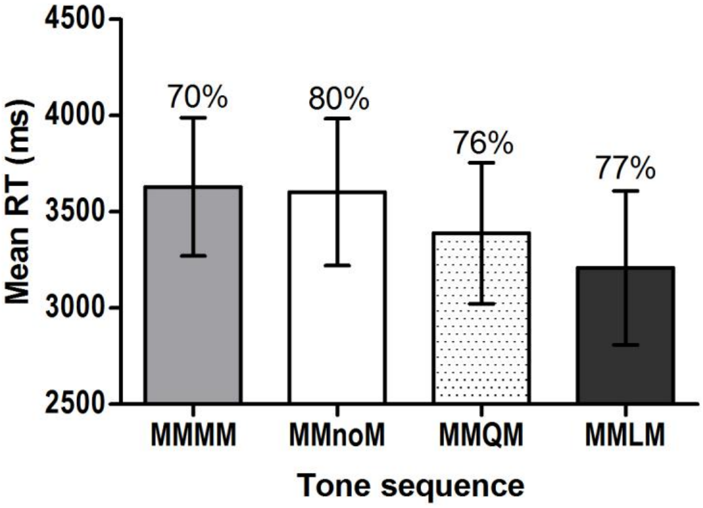


Figure 3.10. Mean RT (in ms) for the MMMM, MMnoM, MMQM, and MMLM conditions tested in Experiment 3.8. The error bars represent the standard errors of the means. Mean percentages of correct responses for each condition are shown above the corresponding error bar.

The results of Experiment 3.8 demonstrate reliable visual target identification performance benefits resulting from oddballs in the form of a salient louder tone, quieter tone, and absence of a tone (in an otherwise regular sequence of tones) presented at the same time as the target display. A closer look at the RT and accuracy data revealed different patterns of facilitation for the various oddball cues. While both the presence of the louder tone in the MMLM sequence and the quieter tone in the MMQM sequence gave rise to benefits in terms of both the speed and accuracy of participants' responses, the absence of the tone in the MMnoM sequence gave rise to enhanced performance only in terms of accuracy. (Note that the improved accuracy was, statistically, even greater than that observed in the presence of the quiet tone oddball, and, numerically, greater than the loud tone oddball). Taken together, it would appear that the performance benefit resulting from the absence of a tone as an oddball was less reliable than either the louder tone or the quieter tone.

On the one hand, the fact that the absence of a crossmodal cue at the same time as the visual target gave rise to any performance benefit at all suggests that multisensory integration does not adequately account for the behavioural benefits of oddball cuing observed in Experiment 3.8, given that there would have been no crossmodal cue present for the visual target to integrate with in this case. Moreover, the fact that all three oddball cues gave rise to some level of improvement in participants' performance (even though this cuing effect was unreliable in the case of the attenuated oddball tone) suggests that an attentional capture account may be more parsimonious and better able to account for the range of behavioural findings from the series of studies exploring the 'freezing phenomenon' that have been published to date.

On the other hand, the fact that the absence of a tone did not result in a significant reduction in search time suggests that the crossmodal cues were overall better at facilitating participants' visual target identification performance, thereby offering some support for the multisensory integration account of the facilitatory effects observed in Experiment 3.8.

It should be noted that there were some apparent speed-accuracy tradeoffs present in the data (see Table 3.1 and Figure 3.10) that might explain the lack of facilitation, in terms of RT, resulting from the absence of the tone (in the otherwise regular sequence of tones) as an oddball in Experiment 3.8. In order to correct for these speed-accuracy trade-offs, participants' RT and accuracy data were therefore converted to IE scores. Participants' IE scores were subjected to a repeated measures ANOVA with Tone sequence (MMLM, MMQM, MMnoM, and MMMM) and target location as the within-participants variables. The analysis revealed a significant main effect of Tone sequence, $F(3,30)=2.99$, $p=.046$. The main effect of target location and interaction between target location and tone sequence were not significant.

Participants' visual target identification performance was significantly better when the target was presented with either the louder tone in the MMLM sequence ($M=5079$ ms; $SE=1100$ ms) or silence in the MMnoM sequence ($M=5143$ ms; $SE=896$ ms) than when paired with a regular tone in the MMMM sequence ($M=4248$ ms; $SE=558$ ms), $p=.007$ and $p=.015$, respectively. The difference between IE scores for the MMQM ($M=5161$ ms; $SE=893$ ms) and MMMM conditions was marginally significant, $p=.086$, with somewhat better performance being reported in the MMQM condition. Importantly, however, none of the oddball cuing conditions gave rise to cuing effects that were significantly different from each other.

The results of this IE scores analysis demonstrated a different pattern of performance facilitation as compared to the separate RT and accuracy data analyses. Specifically, reliable visual target identification performance benefits resulted from oddballs in the form of the salient loud tone or the absence of a tone (in an otherwise regular sequence of tones) presented at the same time as the target display. In fact, the loud tone and the absence of sound both gave rise to performance that were statistically indistinguishable (p=.87) and numerically very similar (M=5079 and 5143 ms, respectively), thus arguing against the multisensory integration account of the facilitatory effects observed in Experiment 3.8.

The presence of the quiet tone as an oddball, however, gave rise to a borderline significant effect, even though the difference among the three oddball conditions was not significant. It would therefore appear that the performance benefit resulting from the quiet tone as an oddball was less reliable than either the loud tone or the absence of a tone. This is perhaps unsurprising, given the mixed results with regards to the differential effectiveness of looming versus receding sounds at capturing attention in previous research (e.g., Abrams & Christ, 2005; Franconeri & Simons, 2003; von Mühlenen & Lleras, 2003; Schouten et al., 2011).

3.9. Experiment 3.9

In all of the preceding experiments reported in this chapter, the cues, though temporally synchronous, were spatially uninformative with regard to the likely location of the visual target. In light of the findings of the previous experiments demonstrating the benefits of spatial cuing in the dynamic and cluttered visual search task reported in Chapter 2 of this thesis, the final experiment to be reported in the present chapter explored whether the same benefits could be observed when carried over to the rapid serially-masked visual presentation paradigm.

Having demonstrated that any oddball cue can give rise to improvements in participants' visual target identification performance, as long as the cue is salient and temporally synchronous with the visual target display, Experiment 3.9 next went on to explore whether any additional performance benefits could be seen by making the oddball cue spatially informative with regard to the location of the target.

3.9.1. Methods

Twelve participants from the University of Oxford (7 female; age range=21-30 years; mean age=25 years) took part in Experiment 3.9. The visual target display was synchronized with a single tone which was either uninformative with regard to the side (left or right) on which the target was presented, spatially valid, or else spatially invalid. For the central (uninformative) trials, the tones were presented through both loudspeakers at the same time, giving rise to a sound that was subjectively localized in the middle of the display. For spatially valid trials, the tone was presented from the loudspeaker in the hemi-field in which the target appeared, while for invalid trials, the tone was presented from the loudspeaker on the opposite side. Uninformative and spatially informative cues were presented in separate blocks of experimental trials with the order of presentation counterbalanced across participants. The uninformative cue block consisted of 80 trials. The spatially informative cue block consisted of 100 trials; i.e., 80 valid and 20 invalid), giving rise to a total of 180 trials completed by each participant.

3.9.2. Results and Discussion

RTs and percentages of correct responses were subjected to separate ANOVAs with Cue type (Uninformative, Valid, and Invalid) as the within-participants factor. The ANOVA on the mean RT data revealed a significant main effect, $F(2,22)=6.77$, $p=.005$, $\eta^2=.38$. Participants responded significantly more slowly on the Invalid cue trials ($M=3191$ ms) than on either the Valid ($M=2455$ ms) or Uninformative ($M=2425$ ms) cue trials, $p=.002$ and $p=.026$, respectively. There was no significant difference in RT between the Valid and Uninformative cue trials, *p*=.89 (see Figure 3.11). There was also a significant main effect of Cue type on the accuracy data, *F*(2,22)=6.19, *p*=.007, η^2=.36. Participants responded significantly more accurately when a Valid (*M*=86%) cue was presented as compared to when either an Uninformative (*M*=81%) or Invalid cue (*M*=76%) was presented, *p*=.02 and *p*=.014, respectively. The difference between the latter two conditions failed to reach statistical significance, *p*=.13.

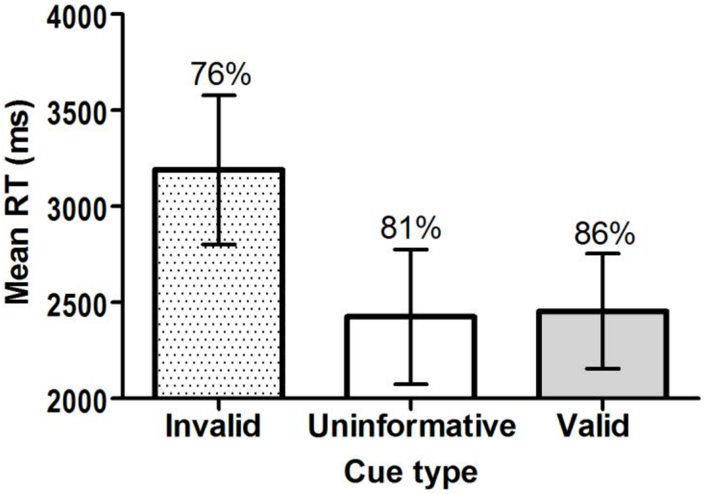


Figure 3.11. Mean RT (in ms) for the spatially Invalid, Uninformative, and Valid cue conditions tested in Experiment 3.9. The error bars represent the standard errors of the means. The mean percentages of correct responses for each tone sequence are shown above the corresponding error bar.

Two key findings emerged from the analysis of the results from Experiment 3.9. First, the crossmodal facilitation of participants' performance can be observed using a speeded elevation discrimination version of the 'flickering-dots' task (i.e., rather than in the 4-choice localization response utilized by the previous experiments in this chapter). Second, when the temporally synchronous auditory cue was made spatially informative with regard to the likely location of the visual target, an additional improvement in the accuracy of participants' target discrimination responses was observed on the validly cued trials as compared to either the spatially uninformative or invalid cuing trials. These results are thus consistent with those reported in Chapter 2 using the “pip and pop” visual search task. At this point, it is important to note that although valid spatial cuing led to significantly better overall performance than either spatially uninformative or invalid cuing in the present study, the results with respect to the RT and accuracy data were different (a point I will return to later in the General Discussion).

3.10. General Discussion

The results of Experiment 3.1 revealed that in the presence of an auditory sequence of tones a frequency change of sufficient magnitude, when synchronized with the onset of the visual target, can lead to a significant improvement in visual target identification. The results of Experiments 3.2-3.5 demonstrated that any abruptly-onset cue that is synchronous with the presentation of the visual target also significantly improves participants’ visual target identification performance. Specifically, the results of Experiment 3.3 showed that unimodal auditory, vibrotactile, and bimodal audiotactile cues were all equally effective in terms of improving participants’ target identification performance. Based on the results of Experiments 3.1-3.3, and Vroomen and de Gelder’s (2000) previous research, one might have assumed that the facilitation of participants’ visual target identification performance was crossmodal in nature, given that the cues used were always non-visual. The results of Experiments 3.4 and 3.5, however, demonstrated that the presentation of a synchronous visual cue (the outline of a box surrounding the visual display) also led to a significant improvement in participants’ performance on the visual target identification task.

The results of Experiment 3.5 revealed that both auditory and visual cues effectively facilitated participants' visual target identification performance, although the auditory cues appeared to be more effective than the visual cues at doing so. Previous studies have also revealed the effectiveness of intramodal visual cuing on participants' visual target identification (e.g., Jonides, 1981; Klein & Dick, 2002; Posner & Cohen, 1984) and visual perception (Watanabe & Shimojo, 1998). For example, Klein and Dick demonstrated that presenting a visual cue at the same time and location as the first target presented in an RSVP stream led to the more accurate visual target identification. It is important to note that studies such as Klein and Dick's have typically used cues that were spatially informative with regard to the likely location of the target. The results of Experiment 3.5, however, demonstrated that the benefits observed following the presentation of the synchronous cue need not be spatially informative nor necessarily crossmodal. Nevertheless, the reduction in RTs following the presentation of the temporally synchronous auditory cues was 752 ms whereas the reduction following the visual cues was only 522 ms, suggesting perhaps that auditory (i.e., crossmodal) cues may still be somewhat more effective than visual (i.e., intramodal) cues in facilitating participants' visual target identification performance (see also Noesselt et al., 2008).

The results of Experiments 3.6 through 3.8 demonstrate that any salient oddball can give rise to a significant improvement in participants' visual target identification performance. Specifically, analysis of participants' IE scores revealed that the louder tone in Experiment 3.6 (partial $\eta 2=.36$) and the complete absence of a tone in Experiment 3.7 (partial $\eta 2=.38$) were found to be equally effective in terms of facilitating participants' performance when these oddball stimuli were synchronized with the presentation of the visual target display. When the benefits of these two oddball stimuli were compared directly in Experiment 3.8, both gave rise to similar levels of performance improvements; the improvement in performance resulting from the quiet tone as an oddball stimulus as compared to the regular tone sequence was, however, only marginally significant.

Finally, the presentation of the temporally synchronous auditory cue in Experiment 3.9 demonstrated that participants' visual target identification performance could be enhanced still further by making the cue spatially informative. The benefit arising from the allocation of voluntary attention to a valid, spatially informative peripheral cue typically manifests itself in an improvement in the accuracy of participants' responses through a process which Prinzmetal, McCool, and Park (2005) have labelled *channel enhancement*. Because the spatially informative cue in the present study provided information about the side on which the target was likely to be presented, the perceptual representation of the target should have been enhanced given that the participants were aware that the cues validly informed them about the likely location of the target and hence would have had an incentive to attend to the cued location. Consistent with Prinzmetal et al.'s account, participants' visual target identification responses were indeed significantly more accurate when the cue provided valid spatial information as compared to when the cues were either spatially uninformative or else spatially invalid.

Taken together, the results of the experiments reported in Chapter 3 therefore build upon those reported in the previous chapter by demonstrating that across two distinct visual target discrimination and identification paradigms, the benefits of temporal and spatial cuing were comparable. Here, the temporal demand of the rapid serially-masked visual target discrimination task was relatively high whereas the spatial demand of the dynamic visual search task was relatively high. Nevertheless, both tasks proved to benefit from the presentation of temporally synchronous and spatially informative non-visual cuing.

Having demonstrated the effectiveness of temporal and spatial non-visual cuing in two distinct laboratory-based paradigms, the goal of the final experimental chapter reported in the main thesis was to implement and test such cues in a real-world air traffic control simulation task, which, in many ways, combines the difficulties of the two laboratory-based studies. In the air traffic control task used in Chapter 4, the visual target is not only dynamic and the environment cluttered, but also the target is less clearly defined and less frequently occurring. Moreover, the critical visual target events in the simulation are time-sensitive, requiring participants' speedy detection and response. Should the benefits of temporal and spatial cuing observed in the laboratory settings be similarly observed in this real-world setting, this would not only highlight the generality of the non-visual cuing effects, but it would also speak to the importance of the potential design and implementation of warning signals in current-day air traffic management systems.

CHAPTER 4:

DYNAMIC VISUAL MONITORING OF VISUAL TARGETS

4.0. Introduction

It has been well-documented that auditory and, more recently, tactile cues can help people to find and identify visual targets more efficiently than when no cue is presented. The benefits seem to be maximal when the cues are presented at the same time and from the same relative location (or direction) as the visual target (e.g., Ho, Tan, & Spence, 2006; Santangelo et al., 2008; Van der Burg et al., 2008, 2009; Vroomen & de Gelder, 2000; see also Chapters 2 and 3 of the present thesis). These studies have demonstrated that not only can crossmodal cues enhance the perceived saliency of target visual events, but also that the cues can guide a person's attention toward the location of the target visual event, even when the cues provide no information with regards to the target's location.

The paradigms used to study the effectiveness of crossmodal auditory and tactile cuing have typically involved basic, low-level, visual perceptual tasks, such as visual search, simple detection, and visual target discrimination, where the variables and stimuli can be systematically controlled. In visual search tasks, for example, researchers typically vary the set size (the number of stimuli presented on the screen at any one time), the presence of the target, and the similarity between distractors and the target in order to examine how these factors might affect participants' visual search performance. On a given trial, the target and distractors typically remain in static positions and the participant simply has to search for and identify the target in the display.

The findings from such laboratory-based paradigms have been questioned due to their lack of ecological validity (Wolfe, 1994; Wolfe et al., 2005, 2011). First, visual search in the real-world often involves complex/ecologically-real targets and distractors (e.g., keys, faces, weapons, etc.; Chen & Zelinsky, 2006; McCarley et al., 2004; Van Wert et al., 2009; Wolfe, 2001; Wolfe et al., 2005; 2011; Wolfe & Van Wert, 2010). Second, observers are not always aware of the exact identity of the target and, to make matters worse, clear view of the target may be obstructed / occluded by overlapping objects. Third, target events such as bombs and air traffic conflicts, fortunately for us, occur very infrequently.

The understanding of the potential benefits of multisensory/crossmodal cuing observed in laboratory-based research have only recently been applied to settings such as air traffic control and vehicular operation (Brill et al., 2004; Fitch et al., 2007; Ho & Spence, 2008). Still, the focus for much of this applied research seems to be on the utilization of multisensory warning signals for driving simulation tasks, while research on the potential benefits of implementing non-visual warning signals in air traffic management environments seems to have fallen somewhat by the wayside. This is perhaps not so surprising given the fact that for air traffic simulation tasks, the visual target is harder to define, control, and predict, as compared to the situation in a driving simulator where the visual target might be a clearly-defined collision (front-end or rear-end) or specific visual target (e.g., road sign) in the simulated visual scene. Moreover, in a driving simulator, the operator is only responsible for the car s/he is driving, whereas in an air traffic management simulation, the operator is responsible for a number of aircraft and required to monitor each of them in parallel.

That said, there has been some promising crossmodal research in the area of air traffic control. For instance, Hameed, Jayaraman, Ballard, and Sarter (2007) introduced tactile cues to an air traffic control simulation task and found that participants found it significantly easier to detect target visual events (e.g., requests for altitude change, potential conflict, or handoff) and responded more rapidly to these events. While the results of Hameed et al.'s study are certainly promising, there are several points worth noting about the conditions of the study. First, the visual target events occurred quite frequently, once every 15 seconds, which is much more frequently than would be expected during realistic air traffic control situations (where such events are exceedingly rare; visit http://www.airsafe.com/ten_faq.htm for a recent review of incidents). Second, there may inevitably be a general cry-wolf situation where operators simply start to ignore the warning signals altogether if the warnings occur too frequently (Breznitz, 1983; though see also Wickens, Rice, Keller, Hutchins, Hughes, & Clayton, 2009, for more recent counter evidence). In this case, the benefit of the warning signal, when actually needed, may no longer be apparent.

The goal of the experiments reported in Chapter 4 was therefore to examine whether the benefits of non-visual, auditory and tactile cuing could be exploited in an air traffic control simulation, where participants were required to monitor and control a number of aircraft that are dynamically-changing and being updated regularly. Importantly, the auditory and vibrotactile cues, in the cue present conditions, were only presented with a conflict situation (e.g., loss of separation with either a boundary or another aircraft), rather than with every visual event[7], given the fact that conflict situations are potentially more dangerous than aircraft coming into or exiting an operator's airspace.

4.1. Experiment 4.1

4.1.1. Methods

Participants. Ten participants from the University of Oxford, ranging in age from 21 to 32 years of age (8 female; mean age=27 years), took part in this experiment. The experiment took approximately one hour and fifteen minutes to complete. The participants received £10 in gift vouchers as compensation for their participation.

Air Traffic Scenarios Test (ATST). The ATST is a low-fidelity, simplified air traffic control simulation developed by the Federal Aviation Administration (www.faa.gov) in 1996. It was originally designed to measure a candidate's aptitude for air traffic management by focusing on the issues of safety, procedure, and delay. The participant's task in the ATST is to guide icons representing aircraft to one of two airports (*a* and *b*) or four sector gates (*A, B, C,* and *D*; see Figure 4.1). The participants are given instructions concerning how to achieve the goals of the simulation and obey various rules of flight by moving the aircraft through the airspace, maintaining safe separation, and ensuring that the aircraft landed or exited at the correct altitude, speed, and heading for its destination. Participants choose from three altitudes (1 for low, 2 for medium, and 3 for high), three speeds (F for fast, M for medium, and S for slow), and eight headings (0 = north, 1 = northeast, 2 = east, 3 = southeast, 4 = south, 5 = southwest, 6 = west, and 7 = northwest). They were instructed to obey the following flight rules: a 5-nautical miles (nm) lateral separation and 1-nm altitude-level vertical separation

---

[7] Conflict situations comprised approximately 50% of all visual events.

must be maintained between aircraft; aircraft must maintain a 5-nm lateral distance from all boundaries; aircraft must land at slow speed and low altitude; aircraft must land at airports from a heading of left-to-right; aircraft must exit at fast speed and high altitude.

Participants controlled the aircraft by clicking on a target aircraft and then on a command button (for heading, speed, and altitude) on the right side of the display. The target aircraft turned yellow once it was clicked, indicating that a change was now ready to be made. The target aircraft remained yellow until a change (even if it was to maintain the same heading, speed, and altitude) was made, at which point the aircraft returned to its original green colour. Aircraft responded immediately to any change and with 100% accuracy. Aircraft positions were updated every 7 seconds to simulate radar sweeps. Scenarios started with seven aircraft already in motion, and a new aircraft appeared in grey at the periphery of the airspace every 30 seconds. Aircraft in grey did not move and were considered to be awaiting handoff. When participants clicked on an aircraft's icon to take the handoff, it turned green and began moving after the next "radar sweep".

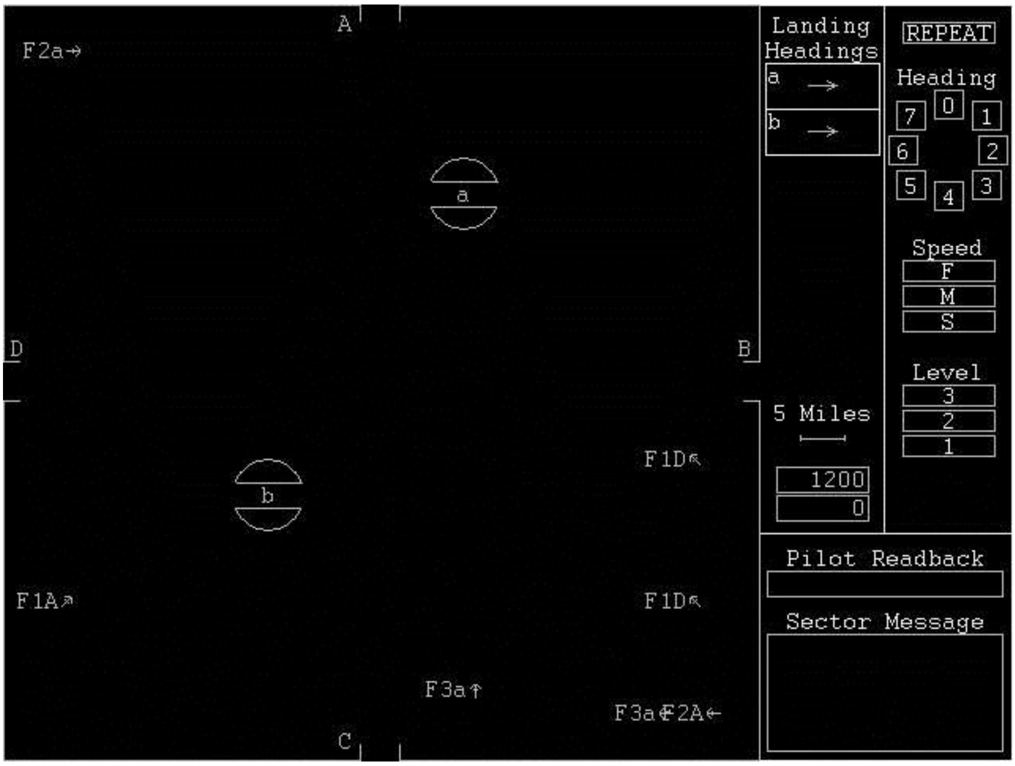


Figure 4.1. A black-and-white screenshot of the ATST used in Experiments 4.1-4.4 in this thesis. Sector gates are marked with capital letters (*A*, *B*, *C*, or *D*) and airports are marked with lowercase letters (*a* or *b*). A data tag with the aircraft's speed (F for Fast, M for Medium, or S for Slow), altitude (3, 2, or 1), and destination (to a sector gate or airport) was provided next to the arrow representing each aircraft. Command buttons for altering the Heading, Speed, and Altitude of the aircraft were placed on the top-right side of the screen. A 5-mile distance indicator and a time-indicator were also displayed to the bottom-left of the command buttons.

The ATST measures four aspects of performance (handoff delay, en route delay, simulation errors, and conflict resolution delay). At the start of each radar sweep, each aircraft in grey awaiting its initial selection accrued 7 seconds worth of *handoff delay*. *En route delay* was calculated as the sum of the differences between the actual flight durations and optimal flight durations of each correctly handled aircraft. Optimal flight durations were determined by the simulator software as the time it would take for an aircraft to fly in a straight line from the point of origin to its destination at maximum speed. *Simulation errors* were tallied whenever an aircraft lost separation, arrived at an incorrect destination, arrived improperly (i.e., at the wrong speed, altitude, or heading), or intersected a boundary, another aircraft, or airport. *Conflict resolution delay*, or RT, was measured from the onset of the loss of separation between an aircraft and either a boundary or another aircraft until the participant initiated a response (i.e., clicking on the aircraft in conflict in an attempt to resolve the conflict).

Design. The experiment was conducted on a desktop computer using Microsoft Visual Studio 2010. Four air traffic scenarios (two practice and two test scenarios) were created and designed to be equivalent in their level of difficulty. Each scenario started with seven aircraft already in motion, with a new aircraft appearing every 30 seconds. Participants were only required to monitor and control the seven aircraft in motion, but they were also encouraged to try to take on more traffic if they felt that they could handle it. Whenever an aircraft crashed into a boundary or another aircraft, a red cross appeared in place of the aircraft(s) at the location of the crash. Whenever an aircraft violated separation requirements (with a boundary or with another aircraft) the aircraft(s) was highlighted in red in order to indicate that a loss of separation had occurred. In the experimental condition of interest, a tone (500 Hz, 60 ms) was played at the same time that the conflict occurred in addition to the red highlighting. The presence or absence of the tone was varied on a block-by-block basis and counterbalanced across participants. The experiment consisted of two 10-minute practice blocks and two 20-minute test blocks.

Procedure. Participants were first given instructions concerning how to achieve the goals of the simulation. The participants were also informed about the rules of flight that they were meant to follow. Before engaging in practice, the experimenter provided participants with a brief demonstration and introduction to all the elements of the simulation. The participants were instructed on how to make changes to the aircraft, how to move the aircraft through the airspace, how to maintain safe separation, and how to appropriately land or exit the aircraft. The participants were also shown what a crash and loss of separation conflict would look (and sound) like. Once they confirmed that they understood the task, they began their first practice block, followed by the second practice and two test blocks; they were allowed to take breaks in-between blocks if they needed it[8].

4.1.2. Results and Discussion

RTs were measured instead of computed IE scores due to the fact that accuracy would be impossible to determine for each measure in the context of the ATST simulation. Separate paired-samples *t*-tests were conducted in order to evaluate the differences between the cue present and cue absent conditions for each measure of task performance (handoff delay, en route delay, simulation errors, and conflict resolution delay). The mean values for each measure of task performance are also reported for completeness in Table 4.1 for each condition and experiment. The analyses revealed that participants responded significantly more rapidly to a conflict when an auditory cue was presented at the time of conflict (in addition to the visual cue; M=2111 ms; SE=389 ms) as compared to when only the visual cue was presented (M=3151 ms; SE=735 ms), $t(9)=2.55$, p=.03 (see Figure 4.2). None of the other measures of task performance differed significantly between cue present and absent conditions, $t<1$ for all comparisons.

---

[8] Note that in the real-world situation, potential air traffic control candidates must undergo an 8-hour aptitude screening test with the ATST, or similar, software (visit http://www.oig.dot.gov/sites/dot/files/Controller%20training%204.1.10.pdf for more details).

Table 4.1. Mean Conflict Resolution, Handoff, and En-route Delays, and Simulation Errors for the Various Cue Conditions of Experiments 4.1 through 4.4.

Experiment	Cue condition	Conflict resolution delay (ms)	Handoff delay (ms)	En-route delay (ms)	Simulation errors
4.1	Auditory absent	3151	12855	6565	11
	Auditory present	2111	12433	6675	11
4.2	Tactile absent	2898	9445	6513	16
	Tactile present	2606	9082	6342	13
4.3	No cue	4104	5767	4822	17
	Auditory	2956	6249	4670	16
	Vibrotactile	3564	5416	4816	21
	Audiotactile	2282	3664	4986	16
4.4	No cue	3277	10369	6775	17
	Uninformative	2694	10808	6744	16
	Informative	2400	10584	6730	13

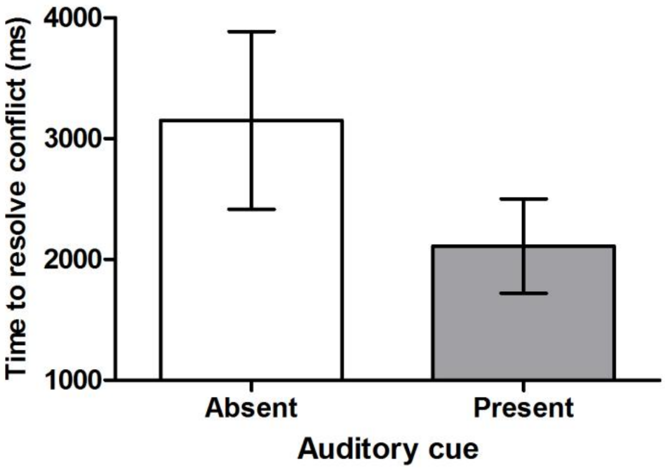


Figure 4.2. Mean conflict resolution delay (in ms) for auditory cue present and absent conditions in Experiment 4.1. The error bars represent the standard errors of the means.

The results of the Experiment 4.1 clearly demonstrate that the presentation of an auditory cue, in addition to the existing visual cues used to signal conflict situations, gave rise to a significant reduction in the amount of time it took participants to respond to impending potential collisions. These results therefore support and extend the extant literature on visual search in the laboratory to an applied setting and demonstrate the value of implementing crossmodal, auditory warning signals in air traffic control environments, not only during training, but more importantly, in real-time.

While the success of temporal and spatial auditory cuing at facilitating performance has been demonstrated consistently in a variety of tasks involving visual target detection and identification (e.g., Haas, Pillalamarri, Stachowiak, & Lattin, 2005; Noesselt et al., 2008; Santangelo & Spence, 2007; Van der Burg et al., 2008a, Vroomen & de Gelder, 2000), the effectiveness of tactile cuing has been shown less consistently (see Ho, Tan, & Spence, 2006; Santangelo et al., 2008; Van der Burg et al., 2008a; see also Spence & Santangelo, 2009, for a review). The goal of Experiment 4.2 therefore was to test whether the effectiveness of temporally synchronous tactile cuing demonstrated in Chapters 2 and 3 of the present thesis could be similarly demonstrated in the ATST.

4.2. Experiment 4.2

4.2.1. Methods

Eleven new participants from the University of Oxford (7 female; age range = 19-32 years; mean age = 26 years) took part in Experiment 4.2. The experimental design and procedure were identical to those used in Experiment 4.1. Instead of an auditory cue, however, participants were presented with a vibrotactile cue at the exact moment at which a conflict occurred. White noise was presented through closed-cup headphones in order to mask any potential sounds coming from vibrotactors fastened to a waist-belt, with one tactor placed on either side of the participant's waist (as had been the case in previous experiments reported in this thesis).

4.2.2. Results and Discussion

Separate paired-samples *t*-tests conducted on participants' handoff delays, en route delays, simulation errors, and conflict resolution delays revealed no significant differences between the vibrotactile cue present and absent conditions ($t<1$ for all comparisons, except for the comparison of simulation errors, $t(10)=1.66$, $p=.13$).

The results of Experiment 4.2 therefore demonstrate that temporally synchronous vibrotactile cues were ineffective at facilitating participants' dynamic visual monitoring performance. Specifically, participants were no faster to detect and respond to conflicts when the vibrotactile cue was presented along with the visual cue as compared to when the visual cue was presented alone. This finding is consistent with previous research by Ho, Tan, and Spence (2006) showing that vibrotactile cues did not give rise to any facilitation of visual target discrimination performance for their participants, while auditory cues did[9].

The fact that the auditory cues in Experiment 4.1 gave rise to a clear improvement in participants' conflict resolution performance, while the vibrotactile cues in Experiment 4.2 did not, might lead one to conclude that tactile cuing is ineffective when compared to auditory cuing. In order to draw such a conclusion, however, it would be necessary to directly compare participants' performance in the presence of either auditory or vibrotactile cues. The goal of Experiment 4.3 was therefore to directly compare participants'

---

[9] Although between-experiments comparisons of the four performance measures might have clarified the differential effects of the auditory cue in Experiment 4.1 and the vibrotactile cue in Experiment 4.2, it should be noted that the large individual differences in participants' performance between the two experiments made such comparisons difficult.

performance on the ATST following either an auditory cue or a vibrotactile cue to a baseline condition in which no crossmodal cue was presented. Additionally, Experiment 4.3 explored the use of multisensory audiotactile cuing in order to determine whether the multisensory cues might lead to a greater magnitude of facilitation than either unimodal auditory or vibrotactile cue alone.

The results obtained in the laboratory-based settings of Experiments 2.2 and 3.3 of this thesis demonstrated no additional benefit resulting from presenting participants with multisensory audiotactile cues as compared to unimodal auditory and vibrotactile cues. This finding might lead one to expect the benefits following multisensory and unimodal cuing to be equal. However, previous research by Ho et al. (2007) examined the use of unimodal auditory and vibrotactile and bimodal audiotactile cues in an applied, real-world driving simulation task. Ho et al. observed that participants initiated braking responses significantly more rapidly when an audiotactile cue was presented simultaneously with the onset of an impending front-end collision than when either a unimodal auditory or vibrotactile cue was presented.

Similarly, research by Santangelo and Spence (2007) and Santangelo et al. (2008) has demonstrated that bimodal audiotactile cues captured spatial attention more effectively than unimodal cues in a laboratory-based elevation discrimination task under conditions of high perceptual load. In light of these findings, it is possible that the pattern of multisensory versus unimodal cuing will be different from that observed in the previous experiments reported in this thesis (i.e., Experiments 2.2 and 3.3) given that the task here in Chapter 4 involves a more complex, real-world simulation with higher perceptual and cognitive load than the laboratory-based dynamic visual search and rapid serially-masked visual presentation tasks.

4.3. Experiment 4.3

4.3.1. Methods

Fifteen participants from the University of Oxford, ranging in age from 18 to 32 years of age (7 female; mean age=24 years), took part in Experiment 4.3. The experimental set-up was identical to that used in Experiment 4.2. In addition to the vibrotactile present and absent conditions, however, Experiment 4.3 included additional auditory and audiotactile cuing conditions. The four cuing conditions (Absent, Auditory, Vibrotactile, and Audiotactile) were varied on a block-by-block basis and randomized across participants[10]. The experiment consisted of two 10-minute practice blocks and four 20-minute test blocks and took approximately two hours to complete.

4.3.2. Results and Discussion

Separate repeated measures ANOVAs were conducted for each measure of task performance (handoff delay, en route delay, simulation errors, and conflict resolution delay) with Cuing condition (Absent, Auditory, Vibrotactile, and Audiotactile) as the within-participants factor. The ANOVA on the conflict resolution delay revealed a significant effect of Cuing condition, $F(3,42)=10.19$, $p<.001$. Post-hoc *t*-tests revealed that participants responded significantly more rapidly to a conflict following the presentation of the audiotactile cue ($M=2282$ ms) as compared to the auditory cue ($M=2956$ ms; $p=.05$), vibrotactile cue ($M=3485$ ms; $p=.001$), or when the visual cue was presented alone ($M=4104$ ms,

---

[10] It is possible that with time-on-task (i.e., more practice) the benefit of the cues might be expected to decline as participants get better at the task. However, closer inspection of the data revealed no pattern of decline or increase in the effect of the crossmodal cues.

p<.001) (see Figure 4.3A). Participants also responded to conflicts more rapidly following the presentation of the unimodal auditory cue as compared to the absence of any crossmodal cue, *p*=.01. There was no significant difference between participants' conflict resolution delays in the auditory and tactile cuing conditions, *p*=.16, and between the vibrotactile and crossmodal cue absent conditions, *p*=.13.

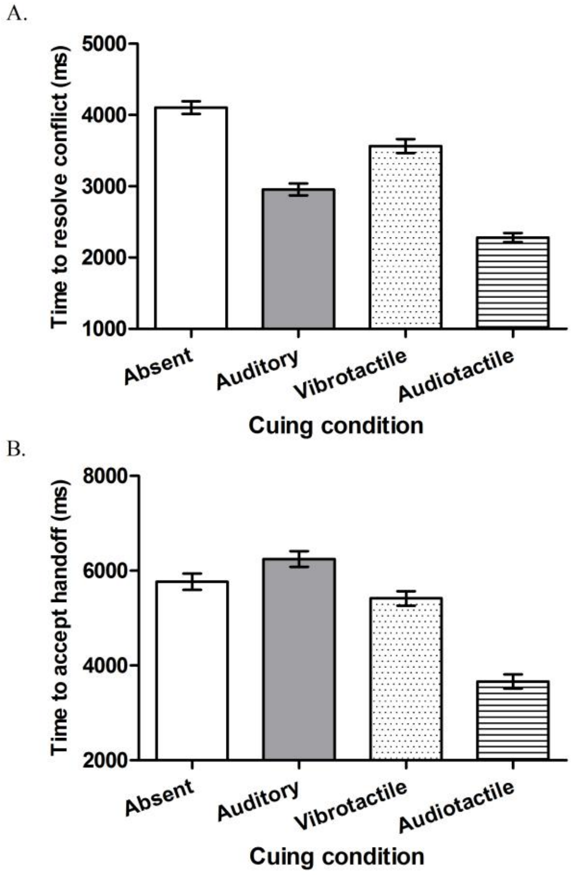


Figure 4.3. A) Mean conflict resolution delay (in ms) for the four cuing conditions (Absent, Auditory, Vibrotactile, and Audiotactile) in Experiment 4.3. B) Mean handoff delay (in ms) for the four cuing conditions. The error bars represent the standard errors of the means.

There were no significant differences among the cuing conditions in either the mean number of simulation errors, *p*=.35, or en route delays, *F*<1. There was, however, a significant effect of cuing condition on participants' handoff delays, *F*(3,42)=3.74, *p*=.018. Post-hoc *t*-tests revealed that participants accepted handoffs significantly more rapidly following the presentation of the audiotactile cue (M=3664 ms) as compared to the auditory cue (M=6249 ms; p=.017), vibrotactile cue (M=5416 ms; p=.041), or when the visual cue was presented alone (M=5767 ms, p=.037) (see Figure 4.3B).

The results of Experiment 4.3 suggest that only the unimodal auditory and multisensory audiotactile cues gave rise to a statistically significant and reliable improvement in the speed of participants' responses to conflicts, while the unimodal vibrotactile cue did not give rise to any such performance improvements. Thus, consistent with the results of Experiment 4.2, the unimodal vibrotactile cue proved to be ineffective at facilitating participants' performance in the present experiment. Interestingly, the presence of the multisensory/bimodal audiotactile cue also gave rise to more efficient performance in terms of participants' handoff delays. That is, participants accepted handoffs (i.e., accepted control of new aircraft entering their airspace) more rapidly in the bimodal audiotactile cuing condition than compared to any of the other unimodal or crossmodal cue absent conditions, even though the audiotactile cues, or any of the other cues for that matter, provided no information about the onset of any incoming aircraft.

The results of Experiment 4.3 are consistent with previous research by Ho et al. (2007) and Santangelo and Spence (2007) comparing the relative effectiveness of unimodal auditory and vibrotactile cues to bimodal/multisensory audiotactile cues in visual tasks. In these studies, and in the experiments reported in the present thesis, multisensory audiotactile cues appear to be most consistently effective, especially when compared to the unimodal vibrotactile, at facilitating visual target detection and identification performance across experimental settings, including those carried out in laboratory-based and real-world paradigms.

4.4. Experiment 4.4

Experiments 4.1 through to 4.3 all utilized cues that were temporally synchronous with the onset of conflict situations but spatially uninformative with regards to their locations. Having demonstrated the potential added benefit of having temporally synchronous cues that are also spatially informative in the experiments reported in Chapters 2 and 3, the goal of Experiment 4.4 was to test whether similar benefits could be observed here in the applied, real-world ATST environment.

4.4.1. Methods

Twenty-six participants from the University of Oxford, ranging in age from 18 to 35 years of age (19 female; mean age=22 years), took part in Experiment 4.4. The experimental set-up was identical to that used in Experiment 4.1. However, in addition to the two test blocks, participants completed a third test block in which the auditory cue was spatially informative with regard to the location of the conflict. The three cuing conditions (Absent, Spatially Uninformative, and Spatially Informative) were varied on a block-by-block basis and counterbalanced across participants. In the Uninformative cuing block, the auditory cue was presented from both loudspeakers. In the Spatially informative cuing block, the auditory cue was played through the left loudspeaker when a conflict occurred in the left hemi-field of the visual display or played through the right speaker when the conflict was in the right hemi-field. The experiment consisted of two 10-minute practice blocks and three 20-minute test blocks. The experiment took approximately 1.5 hours to complete.

4.4.2. Results and Discussion

Separate repeated measures ANOVAs were conducted for each measure of task performance (handoff delay, en route delay, simulation errors, and conflict resolution delay) with Auditory cuing condition (Absent, Spatially-uninformative, and Spatially-informative) as the within-participants factor.

The ANOVA on the conflict resolution delay was the only analysis to reach statistical significance, $F(2,50)=4.31$, $p=.019$. Post-hoc *t*-tests revealed that participants responded significantly more rapidly to a conflict following the presentation of both spatially uninformative ($M=2694$ ms) and spatially informative cues ($M=2400$ ms) as compared to when only the visual cue was presented ($M=3277$ ms), $p=.046$ and $p=.016$, respectively (see Figure 4.4). There was no significant difference between the two crossmodal cuing conditions, $p=.32$. None of the other measures of task performance differed significantly among the cuing conditions ($p=.31$ for handoff delays, $F<1$ for en route delays and simulation errors).

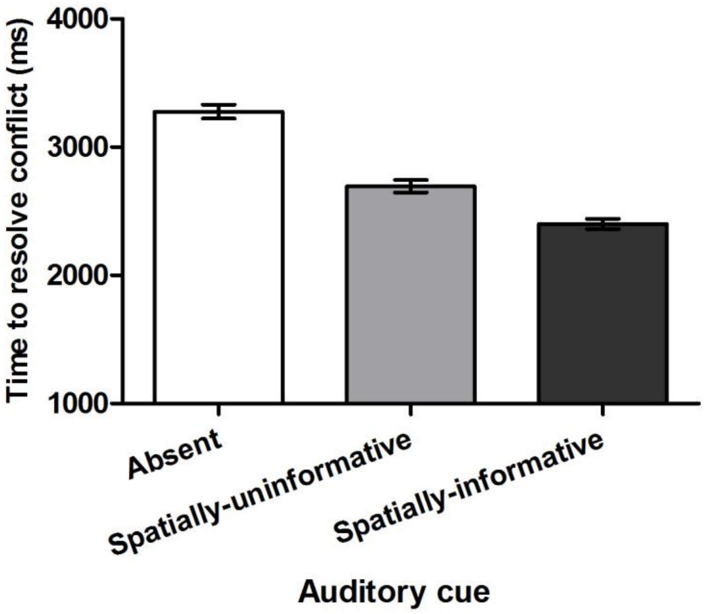


Figure 4.4. Mean conflict resolution delay (in ms) for the three auditory cuing conditions (Absent, Spatially-uninformative, and Spatially-informative) in Experiment 4.4. The error bars represent the standard errors of the means.

The results of Experiment 4.4 demonstrate that spatially uninformative and spatially informative auditory cues, when presented at the same time as the onset of a conflict, both gave rise to significantly faster response times than when only a visual cue signaled the onset of the conflict. Although participants were numerically faster (nearly 300 ms on average) when responding to conflicts following the presentation of the spatially informative as compared to the uninformative auditory cue, the difference between these two conditions failed to reach statistical significance[11]. This suggests that both cues are equally effective at signaling conflict situations and ensuring speeded responses to them in a real-world air traffic management environment.

---

[11] The fact that such large numerical differences failed to reach significance in this chapter (as compared to previous chapters where much smaller differences did) is probably due to the fact that there was more variance in, and overall larger, RTs and other performance measures in this chapter.

4.5. General Discussion

The results of the four experiments reported in this chapter highlight a number of key points. First, the results of Experiments 4.1, 4.3, and 4.4 all demonstrated that an auditory cue that was presented at the same time as the onset of a conflict (in addition to the standard visual cue) resulted in faster responses to the conflicts than when only the visual cue was presented. Second, the results of Experiments 4.2 and 4.3 both suggest that vibrotactile cues are ineffective cues for signaling the occurrence of a conflict. Third, the results of Experiment 4.3 also suggest that multisensory/crossmodal audiotactile cues are more effective and give rise to a more consistent benefit in participants' response times than either of the unimodal auditory or vibrotactile cues. Finally, the results of Experiment 4.4 suggest that temporally synchronous cues are sufficient to significantly facilitate conflict detection compared to the visual cue baseline, while making the cues additionally spatially informative gave rise to a small, but in this case non-significant improvement in participants' response times.

When comparing the findings reported in the present chapter to those of the previous two experimental chapters, it is worth noting that the patterns of crossmodal cuing effects are somewhat different. For one thing, vibrotactile cues appeared to be consistently ineffective at facilitating performance in the real-world ATST, whereas in the laboratory-based visual search and visual target discrimination tasks of Chapters 2 and 3, respectively, the same vibrotactile cues were just as effective as the auditory and even the bimodal audiotactile cues. That said, in the laboratory-based tasks, both of the unimodal cues were just as effective as the bimodal cue at facilitating participants' visual target identification performance, whereas in the ATST, the bimodal cue gave rise to significantly better performance than either of the unimodal cues. Moreover, the performance benefit resulting from the presentation of the bimodal cue was even greater than that observed following the presentation of the unimodal auditory cue.

Additionally, participants' performance in terms of handoff delays also appeared to be facilitated by the presence of the bimodal audiotactile cues, despite the fact that the cues provided no information about any of the incoming aircraft. It is possible that the participants may have adopted an attentional set that was responsive to the most salient cue, which in this case, happened to be the bimodal audiotactile cue (cf. Folk et al., 1992; Leber et al., 2009; Yantis, 1993; see also Theeuwes, 1994). In turn, this attentional set for the bimodal cue may have evoked an overall state of "*attentional readiness*" (Yantis, 1993; p. 676) in the bimodal cuing block, which might explain why participants' improved handoff delays were also observed in this condition.

Previous research by Santangelo et al. (2008) and Santangelo and Spence (2007) have demonstrated that under conditions of high perceptual load, only bimodal audiotactile cues effectively captured participants' visuospatial attention, while unimodal auditory and vibrotactile cues failed to do so. The unimodal cues appeared to be effective only under conditions of low perceptual load. The fact that the bimodal cue gave rise to a more reliable improvement in the speed of participants' responses to conflicts (and to handoffs) than either of the unimodal cues may similarly reflect high perceptual and cognitive load in the ATST (Linnell & Caparos, 2011; Santangelo et al., 2008; Santangelo & Spence, 2007). Here, participants were required to monitor a number of aircraft, all moving dynamically through their small and cluttered airspace, and control the speed, altitude, and heading of the aircraft. The task was also time-sensitive as participants had to respond as quickly as possible to conflicts in order to ensure that the aircraft did not crash and that they arrived or exited efficiently to and from airports and exit gates. Therefore the ATST was considerably higher in both perceptual and cognitive load as compared to either of the laboratory-based tasks utilized in Chapters 2 and 3. As such, and consistent with previous research, it is likely that the high cognitive load of the ATST is what rendered the unimodal vibrotactile cue ineffective (and the unimodal auditory cue less effective) in comparison to the multisensory audiotactile cue, which was reliably effective at facilitating participants' performance.

Another key difference between the cuing effects observed in the ATST and those in the laboratory-based paradigms is that the spatially informative cues failed to facilitate performance any more than did the temporally synchronous cues alone in the ATST. Linnell and Caparos (2011) have recently demonstrated that the focus of spatial attention can be influenced by both perceptual and cognitive load: "*Increasing perceptual load caused the spatial profile to focus when cognitive load was low (Caparos & Linnell, 2009, 2010) but not when it was high, compatible with the idea that perceptual load exerts its focusing effect only with the involvement of cognitive resources, hence nonautomatically*" (p. 5) Although Linnell and Caparos refer to visuospatial attention in their study, it is possible that auditory spatial attention is similarly affected by perceptual and cognitive load. In this case, because perceptual and cognitive load were high in the ATST, it is possible that participants' auditory spatial attention became diffuse, making the informative nature of the spatial cue less useful in terms of facilitating participants' performance crossmodally.

Finally, it should be noted that there were large variations in participants' measures of task performance (e.g., handoff delays, en route delays, and simulation errors) between the four experiments reported in Chapter 4. This may be one of the problems associated with using realistic stimuli / set-ups, where the environment and stimuli are typically more complex, less controlled / predictable, and participants' performance is more varied. In future research, it may be worthwhile to increase training / practice (rather than just two 10-minute practice blocks as used in Experiments 4.1-4.4) in order for participants to reach a plateau in performance[12] (Pierce, submitted) and then test the effects of non-visual cuing thereafter. This would not only potentially minimize the variance across experiments and participants, but it would also ensure that any changes in cuing effects are not simply attributable to time-on-task.

Taken together, the results of the four experiments reported in Chapter 4 suggest that the best cues for improving the speed of participants' responses to conflicts in an air traffic management environment appear to be multisensory audiotactile cues that are temporally synchronous with the onset of the conflict. These findings are consistent with those of Ho et al. (2007) who demonstrated that when participants simulated a car-following scenario, they initiated their braking responses significantly more rapidly following the presentation of multisensory audiotactile warning signals than following the presentation of either unimodal auditory or vibrotactile warning signals. Thus, in real-world tasks, temporally synchronous multisensory audiotactile cues appear to have the greatest potential for significantly improving operator performance and reducing

---

[12] Although, just what a "plateau in performance" (i.e., performance criterion) is exactly might be difficult to define in real-world tasks, such as air traffic management. Moreover, the plateau is likely to vary from person to person.

the time it takes to detect and respond to potential visual target (e.g., threats, dangers, conflicts, etc.) events.

CHAPTER 5:
GENERAL DISCUSSION

5.0. Summary of Key Findings

The results from the three experimental chapters reported in this thesis converge in demonstrating that temporally synchronous and spatially informative non-visual cues give rise to significant improvements in visual target detection and identification, in both basic and applied experimental settings. Specifically, the experiments reported in **Chapter 2** demonstrate that temporally synchronous auditory, vibrotactile, and audiotactile cues all give rise to similar improvements in participants' visual target identification performance in dynamic and cluttered search displays. Here, the hypothesis was that, due to the spatially relevant nature of the task, making these cues also spatially informative would give rise to further improvements in participants' visual target identification performance. This hypothesis was indeed confirmed in Experiments 2.3 and 2.5, which demonstrated that temporally synchronous and spatially informative auditory and vibrotactile cues did indeed enhance participants' visual search performance relative to that seen following temporally synchronous cues.

Similarly, in **Chapter 3**, temporally synchronous and spatially informative cues facilitated participants' performance to a greater extent than temporally synchronous cues alone, even when the task was defined by its temporally-demanding nature. Additionally, the experiments reported in **Chapter 3** demonstrate that any oddball cue (auditory, vibrotactile, or audiotactile), including a visual cue and the absence of a tone in an otherwise regular sequence of tones, when presented at the same time as the visual target in a rapidly-masked stream of visual stimuli, can give rise to a significant improvement in participants' visual target identification performance in terms of speed and accuracy of their responding. Having demonstrated the potential benefits of crossmodal temporal and spatial cuing in two different laboratory-based experimental paradigms, the experiments reported in **Chapter 4** explored whether these benefits would hold in a real-world paradigm, in the context of air traffic control.

In **Chapter 4**, some of the same facilitatory effects observed in the two basic, laboratory settings were indeed shown to carry over to the applied, real-world experimental setting of the ATST. Namely, by providing participants with either unimodal auditory or bimodal audiotactile cues to signal the advent of a conflict or threatening situation, a significant reduction in the time it took participants to detect and respond to the conflict was observed. Importantly, and in contrast with the results of the two basic, laboratory-based paradigms, unimodal vibrotactile cues gave rise to no such facilitation effect, while bimodal audiotactile cues gave rise to an even greater facilitation than either of the unimodal cues. Additionally, when the auditory cues were made to be spatially informative with regard to the location of the conflict in the ATST, there appeared to be no further significant reductions in participants' response times. Thus, the findings from the basic, laboratory paradigms were at least partially replicated in the applied, more real-world paradigm, at least in terms of the benefits of temporal auditory cuing.

5.1. Mechanisms Underlying Non-Visual Cuing

Van der Burg et al. (2008a, 2009) reported that their participants described the visual target as “popping out” from amongst the distractors. Vroomen and de Gelder (2000), on the other hand, stated that their participants reported the visual target to “freeze” for a short period of time in the presence of the segregated high tone. Although the descriptions of the visual target as “freezing” or “popping-out” seem to be phenomenally different, with the former reflecting an enhancement in the duration and the latter reflecting an enhancement in the saliency of the visual target, it is possible that the effect that a temporally synchronous non-visual cue has on a visual target in each case may be the same phenomenally, but the use of either description (e.g., by participants and/or experimenter) may simply depend on the task, stimuli, and experimental paradigm used.

For example, in Van der Burg et al.’s (2008a, 2009) studies and in the experiments reported in Chapter 2 of the present thesis, participants searched for the visual target amongst a dynamic and cluttered field of distractors. In this case, the participants would have benefited most from having a synchronous cue that could make the visual target more salient than the distractors (i.e., making the visual target “pop-out” from amongst the surrounding distractor stimuli). As such, it is possible that the synchronous cue somehow resulted in the enhancement of the saliency of the visual target (Chen & Yeh, 2008; Fiebelkorn et al., 2011; Jaekl & Soto-Faraco, 2010; Lippert, Logothetis, & Kayser, 2007; Matusz & Eimer, 2011; Stein et al., 1996; Van der Burg et al., 2008a, 2008b, 2009; though see Odgaard et al., 2003).

In Vroomen and de Gelder's (2000) study, and in the experiments reported in Chapter 3 of the present thesis, on the other hand, the visual target and distractors were rapidly presented and then masked. The very brief presentation of the visual target is what, presumably, made the participants' task so difficult. It would therefore probably have benefited participants most if the synchronous tone were somehow able to "freeze" the frame of the target, allowing them to continue to access the information contained within the target frame for a short, but extended period of time.

At present, it is impossible to say whether any such enhancement of the target representation primarily affected the perceived duration and/or saliency of the target (note that this distinction might map on to the subjectively reported distinction between the "freezing" phenomenon and the visual target "popping-out"). Rather than reflecting the operation of different mechanisms, however, it may very well be the case that whether participants report the visual target as "freezing" or "popping out" when presented in synchrony with an abrupt cue depends on the descriptors that are most appropriate or beneficial for the task and stimuli at hand.

Nevertheless, it seems likely that the effects observed in the presence of the temporally synchronous and spatially informative cues of the present thesis did indeed lead to an enhanced representation (be it in terms of visible persistence or saliency) of the visual target (Chen & Yeh, 2008, 2009; Lippert et al., 2007; Noesselt et al., 2008; Pascucci, Megna, Panichi, & Baldassi, 2011; Prinzmetal et al., 2005; Stein et al., 1996; Van der Burg et al., 2008a, 2008b, 2009). Two putative mechanisms underlying the enhanced representation and, as a result, improved performance observed in the presence of temporally synchronous and spatially informative cues have been put forth and tested in the present thesis: One is *multisensory integration* and the other is *attention* (spatial and/or temporal).

5.1.1. The Role of Multisensory Integration in Temporal and Spatial Cuing

The *multisensory integration* account suggests that any facilitation observed in the presence of a crossmodal cue would be based on the relative temporal and spatial alignment of the crossmodal stimuli at stake. Once aligned in either time, space, or both, the stimuli are automatically integrated and perceived as a single, bound object. In the case of visual target identification, the enhanced perception of the visual target in the presence of a spatially or temporally informative non-visual cue occurs at a perceptual level, prior to any top-down influences on information processing.

5.1.1.1. Dynamic Visual Search

Van der Burg et al. (2008a, 2009) suggested that the performance benefits observed following the presentation of synchronous auditory or vibrotactile cues in their study were likely due to the crossmodal enhancement of the saliency of the visual targets. Van der Burg and his colleagues ruled out the possibility that the cues merely had a crossmodal alerting effect (McDonald et al., 2000; Spence & Driver, 1994, 1997a, 1997b; Vroomen & de Gelder, 2000), by showing that auditory cues did not give rise to a visual search benefit if they were presented 150 ms prior to the change of the colour of the target (see Van der Burg et al., 2008a, Experiment 3). Instead, the greatest improvements in participants' visual search performance were reported when the auditory cue and target colour change were presented simultaneously, or within 50 ms of each other (see Van der Burg et al., 2008a, Experiment 3). Van der Burg et al. therefore concluded that rather than simply alerting participants to the colour change of the target, the synchronous presentation of the cue and target somehow resulted in the visual target "popping-out" from amongst the distractor line segments. They suggested that the non-visual cue *"boosts the saliency of the concurrently presented visual event, resulting in a salient emergent feature that pops out from the cluttered visual environment, and guides attention to the relevant location*" (Van der Burg et al., 2009, p. 63), hence facilitating participants' visual search performance.

The two cue attributes exploited in the present thesis, spatial correspondence and temporal synchrony, happen to be the key factors relevant to enhancing the likelihood of multisensory integration. Consistent with a multisensory integration account, the finding that invalid spatial cuing gave rise to poorer visual target identification performance as compared to spatially valid and spatially uninformative cuing points to the fact that the spatial correspondence between crossmodal stimuli does (at least sometimes) play a role in whether or not information from the different sensory modalities will be integrated efficiently and, in turn, enhance participants' performance.

Van der Burg et al. (2008a, 2009) suggested that the "pip and pop" and "poke and pop" phenomena occur via an automatic, low-level, process. Their idea was that the "pop-out" effect observed in the presence of the auditory or vibrotactile cues was consistent with the results of previous research by Stein et al. (1996) in which weak (i.e., low intensity) visual stimuli (flashes of light) were rated as being significantly brighter when they were accompanied by pulses of white noise than when no sound was present. Stein and his colleagues argued that this increase in perceived brightness was due to the crossmodal enhancement of the visual stimulus by the simultaneously-presented auditory stimulus (one might think of it in terms of ‘superadditivity’, Stein & Meredith, 1993; Stein & Stanford, 2008). Rather than visual stimulus enhancement, however, Odgaard et al. (2003) argued that Stein et al.’s results may simply have reflected response bias, which is considered to occur at a later (i.e., post-perceptual) decisional level of information processing.

It is also important to note that Van der Burg et al. (2008a) also demonstrated that a temporally synchronous visual cue gave rise to significantly improved visual target identification as compared to a No cue baseline condition. In this case, there would have been no crossmodal cue with which the visual target could integrate, thereby making the multisensory integration account unlikely. It therefore appears that the visual target stimulus enhancement via multisensory integration explanation cannot fully account for the visual search performance benefits observed in the dynamic visual search task used, in either the experiments reported in Chapter 2 of the present thesis or those reported by Van der Burg and his colleagues (2008a, 2009; though see Jaekl & Soto-Faraco, 2010; Van der Burg, Talsma, Olivers, Hickey, & Theeuwes, 2011, for recent neural evidence of the early multisensory interactions elicited by synchronous audio-visual events).

5.1.1.2. Rapid Serially-Masked Visual Search

Similar to Van der Burg et al. (2008a, 2009), Vroomen and de Gelder (2000) showed that when the visual target was synchronized with the presentation of a deviant tone, participants identified the visual target more rapidly (and more accurately). Vroomen and de Gelder proposed that the auditory and visual stimuli became bound together and integrated into a single, coherent multisensory event. Specifically, their suggestion is that the auditory, tactile, and audiotactile cue and visual target may become crossmodally, perceptually grouped in a manner that aids in the segregation of the target from the surrounding distractors, which ultimately leads to its improved identification (see Spence & Chen, in press).

In their original study, Vroomen and de Gelder (2000) argued that one of the principal reasons why the LLHL sequence (as compared to the LLLL sequence) improved participants' visual target identification was because the segregation of the high tone from the low tones gave rise to crossmodal visual stream segregation. They demonstrated that two key factors in determining the strength of the auditory stream segregation, and therefore the strength of the improvement in the identification of the synchronous visual target, were the abruptness and the temporal synchrony of the auditory tone with respect to the onset of the visual target display (although note that they use "abruptness" to describe the magnitude of the frequency change from the low to high tone rather than to describe the abrupt onset of the auditory cue).

Consistent with this notion of abruptness and temporal synchrony, research by Olivers and Van der Burg (2008) has shown that the presence of a single tone whose "abrupt" onset was temporally synchronous with the target(s) in a rapid visual stimulus presentation (RSVP) stream led to a 10% improvement in participants' visual target identification performance. Importantly, Olivers and Van der Burg showed that the synchronous tone helped the second target in the RSVP stream to "*escape the attentional blink*" (p. 191), leading to a level of performance that was actually comparable to the first target in the RSVP stream. Similarly, the results of Experiments 3.2 through 3.5 of the present thesis, where a single, abrupt, and temporally synchronous cue was synchronized with the presentation of the visual target, suggest that perhaps the two cuing conditions (LLHL or single high tone) reflect the operation of somewhat different mechanisms.

Vroomen and de Gelder's (2000) proposed crossmodal perceptual organization and multisensory integration account may help to explain part of the performance improvements observed when comparing LLLL to LLHL tone sequence conditions (in Experiment 3.1, and even in Experiment 3.6, where the louder tone might be perceived as a higher-pitched tone in the sequence of tones). However, it is likely that the facilitation that occurs when a single abrupt onset cue is presented is different from the facilitation that occurs when a stream of auditory stimuli is presented. Furthermore, multisensory integration cannot fully account for the facilitatory effects observed in the presence of deviant/oddball stimuli, in light of the findings from Experiments 3.4, 3.5, 3.7, and 3.8. The results of Experiments 3.4 and 3.5 demonstrated that a visual cue was effective at improving participants' visual target identification performance (perhaps to a lesser extent than an auditory cue, but significant nonetheless). The results of Experiments 3.7 and 3.8 also demonstrated that the absence of a tone in an otherwise regular sequence of tones gave rise to a significant facilitation of visual target identification performance that was comparable to that observed in the presence of a deviant louder tone with the visual target display. The results of these four experiments all converge in showing that even when there is no crossmodal cue to be integrated with the visual target, significant performance enhancements could still be observed, thus providing evidence against the multisensory integration account and suggesting that a different account of the facilitatory effects observed may be in order.

5.1.2. The Role of Attention in Temporal and Spatial Cuing

The *attention* account suggests that, rather than resulting from the binding of the crossmodal stimuli, any cue (unimodal or crossmodal) that is either spatially informative, temporally informative, or both, can elicit a shift in participants' attention either to a specific location (Andersen, Heinke, & Humphreys, 2010, 2011; see also Spence & Driver, 2004, for a review), a specific time (Chen & Yeh, 2008; Correa et al., 2006; Coull et al., 2000), or both (Coull & Nobre, 1998; Nobre & Coull, 2009). This direction of attention can be driven by the cue in a top-down or bottom-up manner, giving rise to the enhanced perception of any stimulus that happens to be presented at the same location or time as the cue.

5.1.2.1. Dynamic Visual Search

Van der Burg et al. (2008a, 2009) themselves noted that the automatic, bottom-up processes involved in multisensory integration could not fully account for the crossmodal cuing effects observed in the dynamic visual search task (p. 1063):

"*Although we believe the results demonstrate a strong automatic component to the pip and pop effect, some of the results suggest that this is not as strong as other, previously reported automatic attentional capture effects (e.g., for color, Theeuwes, 1992; or abrupt onset, Yantis & Jonides, 1984). As we have already pointed out, even with synchronized sounds, not only were overall RTs quite high (for good reasons), but search slopes never quite reached the values* *typical for parallel search…Therefore, we cannot (nor do we wish to) exclude some top-down influences on the pip and pop effect…This would be consistent with other evidence that auditory–visual integration requires at least some attention (see, e.g., Alsius, Navarra, Campbell, & Soto-Faraco, 2005; Talsma et al., 2007).*"

Indeed, there has been convincing neurophysiological (Lakatos, O'Connell, Barczak, Mills, Javitt, & Schroeder, 2009; Talsma, Doty, & Woldorff, 2007) as well as behavioral evidence (Alsius et al., 2005) demonstrating that attention is a prerequisite for multisensory integration to occur. Lakatos et al. argue that the top-down influences of attention "*play an important role in most forms of multisensory integration*" and that "*attention can control neurophysiological context, thus determining the representation of specific sensory content in primary sensory cortex*" (p. 12)

Lippert et al. (2007) demonstrated that improved performance was observed only under those conditions in which the participants were aware that the cue provided reliable information concerning the timing of the visual target. Lippert et al. also suggested that the participants may have used their "knowledge" of this temporal relationship in order to focus their attention on the processing of the stimuli presented at "*the particular instant of time highlighted by the sound*" (p. 104). That is, the auditory cue may have facilitated participants' performance by helping them to endogenously focus their temporal attention (Coull et al., 2000). Lippert et al. would argue that rather than improving the identification of the visual target through multisensory integration, the observed benefit following the presentation of the temporally synchronous and spatially informative cues might reflect the influence of top-down, cognitive processes of attention.

Whenever the target changed colour, it was the only item in the search display to do so and the participants were explicitly told that this was the case. The temporal information provided by the target alone may not have been sufficient to make the target stand out amongst the distractors due to the possible overload in visual information. However, the temporal information provided by the auditory and vibrotactile cue (and visual cue, in Van der Burg et al.'s, 2008a, Experiment 2) may have allowed the participants to inhibit the selection of the items that changed colour at different times from the target (i.e., the distractors). Additionally, the spatially informative auditory and vibrotactile cues used in Chapter 2 of the present thesis provided further reliable information to help guide participants' attention to the correct relative location (left or right hemi-field) of the target. By shifting their attention to one, or other, side of the visual search screen, the spatially informative cues effectively worked to narrow the search space, making it so that participants only needed to search half of the visual field for the target (see also Brill et al., 2004).

Rather than enhancing the saliency of the visual target via multisensory integration, then, the temporally synchronous and spatially informative cues are likely to have drawn participants' attention to the exact moment and relative location of the target's colour change in the dynamic visual search task employed in this thesis (Andersen et al., 2010, 2011; Coull & Nobre, 1998; Watson et al., 2003).

5.1.2.2. Rapid Serially-Masked Visual Search

Similar to the effects observed in the presence of temporally and spatially informative cues from the dynamic search task in Chapter 2 of this thesis, it is possible that the visual or non-visual cues from the rapid serially-masked visual presentation task used in Chapter 3 merely acted as a temporal marker, signifying the exact moment in time at which the target appeared in the rapidly-changing displays (Watson et al., 2003). In this case, the synchronous cue, regardless of its sensory modality of occurrence, may have helped participants become more attuned to the timing of the presentation of the displays. Consequently, the participants may have directed their attention to the time at which the target appeared, thus leading to the speeded identification of the visual target in the presence of the synchronized cue (cf. Coull & Nobre, 1998; Kingstone, 1992).

In Lippert et al.'s (2007) study, participants indicated whether a visual target (presented simultaneously with a sound) was presented above or below a fixation dot. A significant improvement in the accuracy of participants' performance was reported when a sound was presented at the same time as the visual target as compared to a condition in which no sound was presented. At first glance, it might be concluded that the occurrence of the simultaneously-presented sound in Lippert et al.'s study (and by extension in the present thesis) led to a significant enhancement in the saliency of the visual target, as has been proposed by several researchers (e.g., Fiebelkorn et al., 2011; Frassinetti, Bolognini, & Ladavas, 2002; Jaekl & Soto-Faraco, 2010; Matusz & Eimer, 2011; Noesselt et al., 2008, Stein et al., 1996; Van der Burg et al., 2008a, 2009, 2011; Vroomen & de Gelder, 2000). In a follow-up experiment, however, Lippert et al. went on to show that this enhancement of visual target identification performance disappeared if the stimulus onset asynchrony between the visual target and the auditory cue (-400 – 400 ms) was varied randomly on a trial-by-trial basis (cf. Odgaard et al., 2003).

It is likely that the results reported in the present thesis may, at least in part, reflect similar cognitive effects. Similar to those tested in Chapter 2, the participants tested in the experiments reported in Chapter 3 of this thesis certainly had prior knowledge about the temporal and spatial relationship between the various cues and the visual target. Thus, they could have used both the temporal and spatial information provided by the cues in order to focus their visual attention on the exact time and side indicated by the sound in the informative cue blocks.

Analogous to the "freezing" effect reported by Vroomen and de Gelder (2000), the "time's subjective expansion" (TSE) effect put forward by Tse, Intrilligator, Rivest, and Cavanagh (2004) describes the fact that low-probability, "oddball" stimuli (stimuli that differ from the high probability stimuli in one or more features) tend to be perceived as having been presented for longer when presented amidst a series of other high-probability stimuli. In Chapter 3 of the present thesis, the synchronous presentation of the cue (be it crossmodal or intramodal) with the visual target may effectively have made the target display into an oddball amongst the other unimodal visual distractor displays.

Assuming that the perceived duration of the oddball target display was increased, this might, then, help to explain the improved identification of the visual target when it was presented at the same time as a cue as compared to when no cue was presented. Thus, it is possible that the effect of the synchronous cue reflected some kind of enhancement of visible persistence (Chen & Yeh, 2009; Coltheart, 1980; Kanai, Paffen, Hogendoorn, & Verstratem, 2006) of the target which, in turn, may have led to an enhancement of the saliency of the visual target (Chen & Yeh, 2008; Lippert et al., 2007; Stein et al., 1996; Van der Burg et al., 2008a, 2008b, 2009; though see Odgaard et al., 2003). This is consistent with the notion that attetion can modulate multisensory integration processes (Lakatos et al., 2009; Talsma et al., 2007). In this case, attention to the oddball stimulus is what, presumably, gives rise to the subsequent enhancement of the oddball stimulus.

This oddball account may explain why there was a reported perceptual “freezing” effect, or TSE, in which the participants perceived the duration of the target display to be longer than the distractor displays. It may also help to explain the reported “pop-out” effect of the target amongst the distractors in the dynamic visual search task, which, in turn, may have manifested in the observed improvements in both the speed and accuracy of participants’ visual target identification performance.

The presentation of the valid spatial cues gave rise to significantly faster and more accurate visual target discrimination. Here, the effect of the valid spatial cue may have reflected the operation of both the multisensory integration of the auditory cue and visual target as well as the allocation of participants’ temporal and spatial attention to the timing and the location of the presentation of the multisensory information. It is, however, important to note that the valid spatial cues led to significantly more accurate, but importantly, not faster, responding than the spatially uninformative cues. Many previous studies have shown that multisensory integration can occur in the absence of the spatial correspondence between the constituent sensory information (e.g., Van der Burg et al., 2008a; Vroomen & Keetels, 2006). Thus, while the spatially valid and uninformative cues may both involve multisensory integration, only the valid spatial cues can elicit spatial attention. It would appear then that the additional information provided by the spatially valid cue reflected the operation of spatial attention, which gave rise to the enhanced perception (i.e., increased accuracy) of participants' responses when the visual target was presented at the cued location (consistent with the *channel enhancement* account put forth by Prinzmetal et al., 2005).

5.1.3. Where Multisensory Integration and Attention Meet

At this point, it could be argued that temporal and spatial attentional capture provides the most parsimonious account of the available data on the freezing effect. The evidence provided in the present thesis suggests that this form of temporal and spatial attentional capture is driven exogenously (though some modulation by endogenous factors should not be ruled out) by any salient event, be it intramodal or crossmodal. Attentional capture appears to give rise to a significant facilitation of participants' visual discrimination and identification performance by means of enhancing the perception of the attended visual stimulus (by either prolonging the perceived duration or the salience) that happens to be presented at the same time and in the same place as the salient cue.

While Experiments 3.4, 3.5, 3.7, and 3.8 of the present thesis provided initial evidence that multisensory integration could not be responsible for the facilitatory effects observed, the remaining experiments could not rule out a multisensory integration account. That is, it is still plausible that where the temporally synchronous or spatially informative crossmodal auditory or tactile cues facilitated visual target discrimination and identification performance, the operation of multisensory integration was also at play.

In order to tease the attentional capture and multisensory integration accounts apart, however, it would be necessary to isolate one or the other, by making it impossible for participants to either attend to the target event, or to integrate any crossmodal events. While it is difficult to ensure that multisensory integration does not occur in the presence of crossmodal stimuli, given that multisensory integration has been observed despite wide spatial misalignments (see Fiebelkorn et al., 2011), it is possible to control for attention. This could be done by employing a dual-tasking paradigm in which participants' visual attention is divided between a central/foveal task (e.g., tracking) and a peripheral visual task (e.g., orientation discrimination). Here, one could compare participants' event-related potentials (ERPs) to the peripheral visual task on crossmodally-cued versus uncued trials under varying levels of attentional allocation (e.g., 100% for central task, 100% for peripheral task, or 50% central and 50% peripheral task). If attention does indeed play a role in enhancing the perception of the crossmodal stimuli, then one might expect to observe a modulation of participants' evoked potentials to the crossmodally-cued visual targets based on the attentional allocation to the peripheral task.

5.2. Application of the Present Findings

Wolfe (1994) cautioned that the implications of research on visual search in the laboratory should be taken with a grain of salt, stating that "*At the very least, the usual laboratory visual search task has greatly simplified the task of determining the set of items to be searched. At worst, it is possible that the task* *has become so artificial that the body of sophisticated theory mustered to explain laboratory search tasks will have nothing to say about real world tasks*" (p. 1188).

Indeed, the results of the four experiments reported in Chapter 4 confirm that the cuing effects observed in the laboratory-based tasks of Chapters 2 and 3 were only partially replicated in the real-world ATST simulation. The benefit of presenting the visual target with temporally synchronous bimodal audiotactile cues proved to hold across both laboratory-based and applied settings. In particular, multisensory (or bimodal) audiotactile cues proved to give rise to the reliable facilitation of participants' responses to conflicts in the ATST when presented at the same time as the onset of the conflicts. Temporally synchronous unimodal auditory cues, by comparison, were less consistent in facilitating participants' performance; vibrotactile cues just didn't seem to work at all, a finding that is inconsistent with those of Chapters 2 and 3 showing equal improvements in participants' visual target identification and discrimination performance following temporally synchronous unimodal auditory and vibrotactile cues. In contrast to the results of the laboratory-based experiments was also the finding that when the auditory cues were made to also be spatially informative with regards to the location of the conflict (in the left or right visual field), no additional improvements in conflict resolution delays were observed.

Taken together, it would therefore appear that Wolfe (1994) was correct in his warning some years ago about the application of laboratory-based visual search findings to applied settings such as, in this case, air traffic control, where the task is much more cognitively demanding than the dynamic visual search and rapid serially-masked target discrimination tasks. Nevertheless, the findings reported in this thesis have important implications for the design and implementation of non-visual and multisensory warning signals in the automotive and air traffic management industries. In such real-world settings, the timing of the warning signal proves to be more crucial than its spatial alignment in improving the rapid detection and identification of potential threats (e.g., in air traffic management, pilot operations, vehicular navigation, collision avoidance, and military operations). Additionally, the sensory modality of the cue/warning signal also plays an important role in determining just how effective the cue will be at facilitating performance. Based on the findings of the present thesis, temporally synchronous multisensory cues appear to offer the greatest magnitude and most reliable facilitation compared to either unimodal auditory or unimodal vibrotactile cues.

5.3. Future Directions and Concluding Remarks

The results from the real-world ATST simulation in Chapter 4 of this thesis suggested a non-significant role of spatial congruence in the facilitation of participants' visual target detection performance following the presentation of a temporally synchronous auditory cue. That is, no further performance benefits were observed when the temporally synchronous cue provided additional information about the location of the visual target in the left or right hemifield of the visual display. It is, nevertheless, possible that the spatial coincidence / co-localization of the warning signal and target is a necessary prerequisite in order to bring about any significant improvements in participants' visual target identification performance beyond what can already be observed with temporally synchronous cues (Ferris & Sarter, 2008; Jones, Gray, Spence, & Tan, 2008; Sarter, 2000; Spence & Ho, 2008). It would therefore be interesting to explore whether spatially coincident auditory cues might be more effective than merely spatially congruent cues in facilitating participants' performance. It would also be interesting to explore how spatially co-localized tactile cues might be implemented, given that the tactile stimulation would need to be on the body in order for it to be felt.

In conclusion, the findings of the present thesis have hopefully shed light on the potential application of temporally synchronous and spatially informative crossmodal cues for the enhancement of visual target detection, discrimination, and identification in both basic, laboratory-based and applied, real-world tasks. The findings of the present thesis have hopefully helped to also shed light on which of the previously proposed accounts, multisensory integration or attention, can better explain the temporal and spatial cuing effects observed here and elsewhere in the visual search literature. Initially, it seemed a rather difficult feat to tease apart the two mechanisms of, on the one hand, multisensory integration and/or, on the other, attention. Taken together, it appears that attention wins out over multisensory integration in accounting for the range of results and various cuing paradigms used in this thesis (see also Lakatos et al., 2009; Talsma et al., 2007, for recent neurophysiological evidence supporting this claim). Both temporally synchronous and spatially informative cues (crossmodal and intramodal) appear to have elicited a shift in participants' attention in time and space, respectively. Through the direction of attention to specific instances and locations, temporally synchronous, and, potentially, spatially informative, cues can aid in the improved detection and identification of visual targets in both basic and applied settings.

REFERENCES

Abrams, R. A., & Christ, S. E. (2005). The onset of receding motion captures attention: Comment on Franconeri and Simons (2003). *Perception & Psychophysics*, **67**, 219-223.

Ali, M. A., & Klyne, M. A. (1985). *Vision in vertebrates*. New York: Plenum Press.

Alsius, A., Navarra, J., Campbell, R., & Soto-Faraco, S. (2005). Audiovisual integration of speech falters under attention demands. *Current Biology*, **15**, 839-843.

Andersen, G. J., & Dyre, B. P. (1989). Spatial orientation from optic flow in the central visual field. *Perception & Psychophysics*, **45**, 453-458.

Anderson, G. M., Heinke, D., & Humphreys, G. W. (2010). Featural guidance in conjunction search: The contrast between orientation and color. *Journal of Experimental Psychology: Human Perception and Performance*, **36**, 1108-1127.

Anderson, G. M., Heinke, D., & Humphreys, G. W. (2011). Differential time course of implicit and explicit cueing by colour and orientation in visual search. *Visual Cognition*, **19**, 258-288.

Anderson, T. S., & Mamassian, P. (2008). Audiovisual integration of stimulus transients. *Vision Research*, **48**, 2537-2544.

Asemi, N., Sugita, Y., Suzuki, Y. (2003). Auditory search asymmetry between pure tone and temporal fluctuating sounds distributed on the frontal-horizontal plane. *Acta Acustica united with Acustica*, **89**, 346-354.

Atchley, P., Kramer, A. F., Andersen, G. J., & Theeuwes, J. (1997). Spatial cuing in a stereoscopic display: Evidence for a "depth-aware" attentional focus. *Psychonomic Bulletin & Review*, **4**, 524-529.

Bakeman, R., & McArthur, D. (1996). Picturing repeated measures: Comments on Loftus, Morrison, and others. *Behavior Research Methods, Instruments, & Computers*, **28**, 584-589.

Begault, D. R. (1993). Head-up auditory displays for traffic collision avoidance system advisories: A preliminary investigation. *Human Factors*, **35**, 707-717.

Bolia, R. S., D'Angelo, W. R., & McKinley, R. L. (1999). Aurally aided visual search in three-dimensional space. *Human Factors*, **41**, 664-669.

Bolognini, N., Frassinetti, F., Serino, A., & Ladavas, E. (2005). "Acoustical vision" of below threshold stimuli: Interaction among spatially converging audiovisual inputs. *Experimental Brain Research*, **160**, 273-282.

Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, **26**, 211-252.

Bregman, A. S. (1990). *Auditory scene analysis*. Cambridge: MIT Press.

Breznitz, S. (1983). *Cry-wolf: The psychology of false alarms*. Hillsdale: Lawrence Erlbaum.

Brill, J. C., Terrence, P. I., Downs, J. L., Gilson, R. D., Hancock, P. A., & Mouloua, M. (2004). Search space reduction via multi-sensory directional cueing. In *Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting* (pp. 2134-2136). New Orleans: Human Factors and Ergonomics Society.

Chan, A. H. S., & Chan, K. W. L. (2006). Synchronous and asynchronous presentations of auditory and visual signals: Implications for control console design. *Applied Ergonomics*, **37**, 131-140.

Chen, K.-M., & Yeh, S.-L. (2009). Asymmetric cross-modal effects in time perception. *Acta Psychologica*, **130**, 225-234.

Chen, X., & Zelinksy, G. (2006). Real-world visual search is dominated by top-down guidance. *Vision Research*, **46**, 4118-4133.

Chen, Y.-C., & Yeh, S.-L. (2008). Visual events modulated by sound in repetition blindness. *Psychonomic Bulletin & Review*, **15**, 404-408.

Coltheart, M. (1980). Iconic memory and visible persistence. *Perception & Psychophysics*, **27**, 183-228.

Coull, J. T., Frith, C. D., Büchel, C., & Nobre, A. C. (2000). Orienting attention in time: behavioural and neuroanatomical distinction between exogenous and endogenous shifts. *Neuropsychologia*, **38**, 808-819.

Coull, J. T., & Nobre, A. C. (1998). Where and when to pay attention: The neural systems for directing attention to spatial locations and to time intervals as revealed by both PET and fMRI. *Journal of Neuroscience*, **18**, 7426-7435.

Dalton, P., & Spence, C. (2007). Attentional capture in serial audiovisual search tasks. *Perception & Psychophysics*, **69**, 422-438.

Davis, G., & Driver, J. (1994). Parallel detection of Kanizsa subjective figures in the human visual system. *Nature*, **371**, 791-793.

Debener, S., Kranczioch, C., Herrmann, C. S., & Engel, A. K. (2002). Auditory novelty oddball allows reliable distinction of top–down and bottom–up processes of attention. *International Journal of Psychophysiology*, **46**, 77-84.

Di Luca, M., Machulla, T.-K., & Ernst, M. O. (2009). Recalibration of multisensory simultaneity: Cross-modal transfer coincides with a change in perceptual latency. *Journal of Vision*, **9**, 1-16.

Driver, J., & Spence, C. (1998). Cross-modal links in spatial attention. Philosophical Transactions of the Royal Society B: Biological Sciences, **353**, 1319-1331.

Dufour, A. (1999). Importance of attentional mechanisms in audiovisual links. *Experimental Brain Research*, **126**, 215-222.

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, **96**, 433-458.

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

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in primate cerebral cortex. *Cerebral Cortex*, **1**, 1-47.

Ferris, T. K., & Sarter, N. B. (2008). Cross-modal links among vision, audition, and touch in complex environments. *Human Factors*, **50**, 17-26.

Fiebelkorn, I. C., Foxe, J. J., Butler, J. S., & Molholm, S. (2011). Auditory facilitation of visual-target detection persists regardless of retinal eccentricity and despite wide audiovisual misalignments. *Experimental Brain Research*, **213**, 167-174.

Fitch, G. M., Keifer, R. J., Hankey, J. M., & Kleiner, B. M. (2007). Toward developing an approach for alerting drivers to the direction of a crash threat. *Human Factors*, **49**, 710-720.

Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance*, **18**, 1030-1044.

Forti, S., & Humphreys, G. W. (2005). Cross-modal visuo-tactile matching in a patient with a semantic disorder. *Neuropsychologia*, **43**, 1568-1579.

Franconeri, S. L., & Simons, D. J. (2003). Moving and looming stimuli capture attention. *Perception & Psychophysics*, **65**, 999-1010.

Franconeri, S. L., & Simons, D. J. (2005). The dynamic events that capture visual attention: A reply to Abrams and Christ (2005), *Perception & Psychophysics*, **67**, 962-966.

Fujawa, G. E., & Strybel, T. Z. (1997). The effects of cue informativeness and signal amplitude on auditory spatial facilitation of visual performance. In *Proceedings of the Human Factors and Ergonomics Society 41st Annual*

Meeting (pp. 556-560). Santa Monica: Human Factors and Ergonomics Society.

Gallace, A., Ngo, M. K., Sulaitis, J., & Spence, C. (2011). Multisensory presence in virtual reality: Possibilities and limitations. In G. Ghinea, F. Andres, & S. Gulliver (Eds.), *Multiple sensorial media advances and applications: New developments in mulsemedia* (pp. 1-38). Hershey: IGI Global.

Haas, E. C., Pillalamarri, R. S., Stachowiak, C. C., & Lattin, M. A. (2005). Audio cues to assist visual search in robotic system operator control unit displays. *Army Research Laboratory*. Aberdeen Proving Ground, MD.

Halpern, B. P. (1983). Tasting and smelling as active, exploratory sensory processes. *American Journal of Otolaryngology*, **4**, 246-249.

Hameed, S., Jayaraman, S., Ballard, M., & Sarter, N. (2007). Guiding visual attention by exploring crossmodal spatial links: An application in air traffic control. In *Proceedings of the Human Factors and Ergonomics Society 51st Annual Meeting* (pp. 220-224). Santa Monica: Human Factors and Ergonomics Society.

Harrar, V., Winter, R., & Harris, L. R. (2008). Visuotactile apparent motion. *Attention, Perception, & Psychophysics*, **70**, 807-817.

Ho, C., Reed, N., & Spence, C. (2006). Assessing the effectiveness of “intuitive” vibrotactile warning signals in preventing front-to-rear-end collisions in a driving simulator. *Accident Analysis and Prevention*, **38**, 988-996.

Ho, C., Reed, N., & Spence, C. (2007). Multisensory in-car warning signals for collision avoidance. *Human Factors*, **49**, 1107-1114.

Ho, C., Santangelo, V., & Spence, C. (2009). Multisensory warning signals: When spatial correspondence matters. *Experimental Brain Research*, **195**, 261-272.

Ho, C., & Spence, C. (2005). Assessing the effectiveness of various auditory cues in capturing a driver’s visual attention. *Journal of Experimental Psychology: Applied*, **11**, 157-174.

Ho, C., & Spence, C. (2006). Verbal interface design: Do verbal directional cues automatically orient visual spatial attention? *Computers in Human Behavior*, **22**, 733-748.

Ho, C. & Spence, C. (2008). *The multisensory driver*. Implications for ergonomic car interface design. Aldershot: Ashgate.

Ho, C., Tan, H. Z., & Spence, C. (2006). The differential effect of vibrotactile and auditory cues on visual spatial attention. *Ergonomics*, **49**, 724-738.

Humphreys, G. W., Hodsoll, J., & Riddoch, M. J. (2009). Fractionating the binding process: Neuropsychological evidence from reversed search efficiencies. *Journal of Experimental Psychology: Human Perception and Performance*, **35**, 627-647.

Jacobson, H. (1951). The informational capacity of the human eye. *Science*, **16**, 292-293.

Jaekl, P. M., & Soto-Faraco, S. (2010). Audiovisual contrast enhancement is articulated primarily via the M-pathway. *Brain Research*, **1366**, 85-92.

James, T. W., Humphrey, G. K., Gati, J. S., Servos, P., Menon, R. S., & Goodale, M. A. (2002). Haptic study of three-dimensional objects activates extrastriate visual areas. *Neuropsychologia*, **40**, 1706-1714.

Johnson, D. N., McGrath, A., & McNeil, C. (2002). Cuing interacts with perceptual load in visual search. *Psychological Science*, **13**, 284-287.

Jones, C. M., Gray, R., Spence, C., & Tan, H. Z. (2008). Directing visual attention with spatially informative and spatially noninformative tactile cues. *Experimental Brain Research*, **186**, 659-669.

Jonides, J. (1981). Voluntary versus automatic control over the mind's eye's movement. In J. Long & A. Baddeley (Eds.), *Attention and performance* (Vol. 9, pp. 187-203). Hillsdale: Erlbaum.

Kanai, R., Paffen, C. L., Hogendoorn, H., & Verstraten, F. A. J. (2006). Time dilation in dynamic visual display. *Journal of Vision*, **6**, 1421-1430.

Kingstone, A. (1992). Combining expectancies. *Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, **44(A)**, 69-104.

Kleffner, D. A., & Ramachandran, V. S. (1992). On the perception of shape from shading. *Perception & Psychophysics*, **52**, 18-36.

Klein, R. M., & Dick, B. (2002). Temporal dynamics of reflexive attention shifts: A dual-stream rapid serial visual presentation exploration. *Psychological Science*, **13**, 176-179.

Làdavas, E., & Farnè, A. (2004). Neuropsychological evidence for multimodal representations of space near specific body parts. In C. Spence & J. Driver (Eds.), *Crossmodal space and crossmodal attention* (pp. 69-98). Oxford: Oxford University Press.

Leber, A. B., Kawahara, J.-I., & Gabari, Y. (2009). Long-term abstract learning of attentional set. *Journal of Experimental Psychology: Human Perception and Performance*, **35**, 1385-1397.

Lee, J.-H., & Spence, C. (2009). Feeling what you hear: Task irrelevant sounds modulate tactile perception delivered via a touch screen. *Journal of Multimodal User Interfaces*, **2**, 145-156.

Lewald, J., & Guski, R. (2003). Cross-modal perceptual integration of spatially and temporally disparate auditory and visual stimuli. *Cognitive Brain Research*, **16**, 468-478.

Lindeman, R. W., Yanagida, Y., Sibert, J. L., & Lavine, R. (2003). Effective vibrotactile cuing in a visual search task. In M. Rauterburg, M. Menozzi, & J. Wesson (Eds.), *Proceedings of the ninth IFIP TC13 international conference on human-computer interaction* (INTERACT 2003) (Sept. 1-5, pp. 89-98). Zurich: IOS Press.

Linnell, K. J., & Caparos, S. (2011). Perceptual and cognitive load interact to control the spatial focus of attention. *Journal of Experimental Psychology: Human Perception and Performance*. Advance online publication. doi: 10.1037/a0024669.

Lippert, M., Logothetis, N. K., & Kayser, C. (2007). Improvement of visual contrast detection by a simultaneous sound. *Brain Research*, **1173**, 102-109.

Luck, S. J., Woodman, G. F., & Vogel, E. K. (2000). Event-related potential studies of attention. *Trends in Cognitive Sciences*, **4**, 432-440.

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

McCarley, J. S., Kramer, A. F., Wickens, C. D., Vidoni, E. D., & Boot, W. R. (2004). Visual skills in airport-security screening. *Psychological Science*, **15**, 302-306.

McDonald, J. J., Teder-Sälejärvi, W. A., & Hillyard, S. A. (2000). Involuntary orienting to sound improves visual perception. *Nature*, **407**, 906-908.

McLeod, P., Driver, J., & Crisp, J. (1988). Visual search for a conjunction of movement and form is parallel. *Nature*, **332**, 154-155.

McLeod, P., Driver, J., Dienes, Z., & Crisp, J. (1991). Filtering by movement in visual search. *Journal of Experimental Psychology: Human Perception and Performance*, **17**, 55-64.

Meredith, M. A., Nemitz, J. W., & Stein, B. E. (1987). Determinants of multisensory integration in superior colliculus neurons: I. Temporal factors. *Journal of Neuroscience*, **7**, 3213-3229.

Meredith, M. A., & Stein, B. E. (1986). Spatial factors determine the activity of multisensory neurons in cat superior colliculus. *Brain Research*, **365**, 350-354.

Miller, M. E., & Beaton, R. J. (1994). The alarming sounds of silence. *Ergonomics in Design*, **January**, 21-23.

Milliken, B., & Tipper, S. P. (1998). Attention and inhibition. In H. Pashler (Ed.), *Attention* (pp. 191-222). East Sussex: Psychology Press.

Neuhoff, J. G. (1998). Perceptual bias for rising tones. *Nature*, **395**, 123-124.

Neuhoff, J. G. (2001). An adaptive bias in the perception of looming auditory motion. *Ecological Psychology*, **13**, 87-110.

Neuhoff, J. G., Kramer, G., & Wayand, J. (2002). Pitch and loudness interact in auditory displays: Can the data get lost in the map? *Journal of Experimental Psychology: Applied*, **8**, 17-25.

Nobre, K., & Coull, J. (Eds.) (2009). *Attention and time*. Oxford: Oxford University Press.

Noesselt, T., Bergmann, D., Hake, M., Heinze, H.-J., & Fendrich, R. (2008). Sound increases the saliency of visual events. *Brain Research*, **1220**, 157-163.

Odgaard, E. C., Arieh, Y., & Marks, L. E. (2003). Cross-modal enhancement of perceived brightness: Sensory interaction versus response bias. *Perception & Psychophysics*, **65**, 123-132.

Olausson, H., Cole, J., Rylander, K., McGlone, F., Lamarre, Y., Wallin, B. G., Krämer, H., Wessberg, J., Elam, M., Bushnell, M. C., & Vallbo, A. (2008). Functional role of unmyelinated tactile afferents in human hairy skin: Sympathetic response and perceptual localization. *Experimental Brain Research*, **184**, 135-140.

O'Leary, A., & Rhodes, G. (1984). Cross-modal effects on visual and auditory object perception. *Perception & Psychophysics*, **35**, 565-569.

Olivers, C. N. L., & Van der Burg, E. (2008). Bleeping you out of the blink: Sound saves vision from oblivion. *Brain Research*, **1242**, 191-199.

Olivers, C. N. L., Van der Burg, E., Talsma, D., Bronkhorst, A. W., & Theeuwes, J. (2009). Sound increases visual saliency: Evidence from EEG. *Journal of Vision*, **9**:84. doi:10.1167/9.8.84

Osborn, W. C., Sheldon, R. W., and Baker, R. A. (1963). Vigilance performance under conditions of redundant and nonredundant signal presentation. *Journal of Applied Psychology*, **47**, 130-134.

Oxenham, A. J., & Wojtczak, M. (2010). Frequency selectivity and masking. In C. J. Plack (Ed.), *The Oxford handbook of auditory science: Hearing* (pp. 5-44). Oxford: Oxford University Press.

Pascucci, D., Megna, N., Panichi, M., Baldassi, S. (2011). Acoustic cues to visual detection: A classification image study. *Journal of Vision*, **11**, 1-11.

Patterson, R. D. (1982). Guidelines for auditory warning systems on civil aircraft: The learning and retention of warnings. *MRC Applied Psychology Unit, Civil Aviation Authority Contract* 7D/S/0142.

Pawlak, W. S., & Vicente, K. J. (1996). Inducing effective operator control through ecological interface design. *International Journal of Human-Computer Studies*, **44**, 653-688.

Perrott, D. R., Cisneros, J., McKinley, R. L., & D'Angelo, W. R. (1996). Aurally aided visual search under virtual and free-field listening conditions. *Human Factors*, **38**, 702-715.

Perrott, D. R., & Saberi, K. (1990). Minimal audible angle thresholds for sources varying in both elevation and azimuth. *Journal of the Acoustical Society of America*, **87**, 1728-1731.

Perrott, D. R., Saberi, K., Brown, K., & Strybel, T. Z. (1990). Auditory psychomotor coordination and visual search performance. *Perception & Psychophysics*, **48**, 214-226.

Perrott, D. R., Sadralodabai, T., Saberi, K., & Strybel, T. Z. (1991). Aurally aided visual search in the central visual field: Effects of visual load and visual enhancement of the target. *Human Factors*, **33**, 389-400.

Pfefferbaum, A., Buchsbaum, M., & Gips, J. (1971). Enhancement of the average evoked response to tone onset and cessation. *Psychophysiology*, **8**, 332-339.

Pierce, R. S. (submitted). The effect of SPAM administration during a dynamic simulation. *Human Factors*.

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, **32**, 3-25.

Posner, M. I., & Cohen, Y. (1984). Components of performance. In H. Bouma & D. Bowhuis (Eds.), *Attention and performance X* (pp. 531-556). Hillsdale: Erlbaum.

Previc, F. H. (1998). The neuropsychology of 3-D space. *Psychological Bulletin*, **124**, 123-164.

Previc, F. H. (2000). Neuropsychological guidelines for aircraft control stations. *IEEE Engineering in Medicine and Biology Magazine*, **19**, 81-88.

Prinzmetal, W., McCool, C., & Park, S. (2005). Attention: Reaction time and accuracy reveal different mechanisms. *Journal of Experimental Psychology: General*, **134**, 73-92.

Proctor, R. W., & Van Zandt, T. (1994). *Human factors in simple and complex systems*. Needham Heights: Allyn & Bacon.

Quinlan, P. T. (2003). Visual feature integration theory: Past, present, and future. *Psychological Bulletin*, **129**, 643-673.

Remington, R. W., Johnston, J. C., Ruthruff, E., Gold, M., & Romera, M. (2000). Visual search in complex displays: factors affecting conflict detection by air traffic controllers. *Human Factors*, **42**, 349-366.

Roberts, K. L., Summerfield, A. Q., & Hall, D. A. (2009). Covert auditory spatial orienting: An evaluation of the spatial relevance hypothesis. *Journal of Experimental Psychology: Human Perception and Performance*, **35**, 1178-1191.

Rudmann, D. S., & Strybel, T. Z. (1999). Auditory spatial facilitation of visual search performance: Effect of cue precision and distractor density. *Human Factors*, **41**, 146-160.

Santangelo, V., & Spence, C. (2007). Multisensory cues capture spatial attention regardless of perceptual load. *Journal of Experimental Psychology: Human Perception and Performance*, **33**, 1311-1321.

Santangelo, V., Ho, C., & Spence, C. (2008). Capturing spatial attention with multisensory cues. *Psychonomic Bulletin & Review*, **15**, 398-403.

Santangelo, V., Van der Lubbe, R. H. J., Olivetti Belardinelli, M., & Postma, A. (2006). Spatial attention triggered by unimodal, crossmodal, and bimodal exogenous cues: A comparison of reflexive orienting mechanisms. *Experimental Brain Research*, **173**, 40-48.

Sarter, N. B. (2000). The need for multisensory interfaces in support of effective attention allocation in highly dynamic event-driven domains: The case of cockpit automation. *International Journal of Aviation Psychology*, **10**, 231-245.

Sarter, N. B. (2001). Multimodal communication in support of coordinative functions in human-machine teams. *Journal of Human Performance in Extreme Environments*, **5**, 50-54.

Schouten, B., Troje, N. F., Vroomen, J., & Verfaillie, K. (2011). The effect of looming and receding sounds on the perceived in-depth orientation of depth-ambiguous biological motion figures. *PLoS ONE*, **6**, 1-8.

Schroeder, C. E, & Foxe, J. J. (2002). The timing and laminar profile of converging inputs to multisensory areas of the macaque neocortex. *Cognitive Brain Research*, **14**, 187-198.

Schroeder, C. E., & Foxe, J. J. (2004). Multisensory convergence in early cortical processing. In G. A. Calvert, C. Spence, & B. E. Stein (Eds.), *The handbook of multisensory processes* (pp. 295-309). Cambridge, MA: MIT Press.

Schroeder, C. E, & Foxe, J. (2005). Multisensory contributions to low-level, ‘unisensory’ processing. *Current Opinion in Neurobiology*, **15**, 454-458.

Spence, C. (2010). Crossmodal attention. *Scholarpedia*, **5(5)**: 6309.

Spence, C., & Chen, Y.-C. (in press). Intramodal and crossmodal perceptual grouping. To appear in B. E. Stein (Ed.), *The new handbook of multisensory processing*. Cambridge: MIT Press.

Spence, C., & Driver, J. (1997a). Audiovisual links in exogenous covert spatial orienting. *Perception & Psychophysics*, **59**, 1-22.

Spence, C., & Driver, J. (1997b). On measuring selective attention to a specific sensory modality. *Perception & Psychophysics*, **59**, 389-403.

Spence, C., & Driver, J. (1999). Multiple resources and multimodal interface design. In D. Harris (Ed.), *Engineering psychology and cognitive ergonomics, Vol. 3: Transportation systems, medical ergonomics and training* (pp. 305-312). Hampshire: Ashgate Publishing.

Spence, C., & Driver, J. (Eds.). (2004). *Crossmodal space and crossmodal attention*. Oxford: Oxford University Press.

Spence, C., & Ho, C. (2008). Tactile and multisensory spatial warning signals for drivers. *IEEE Transactions on Haptics,* **1**, 121-129.

Spence, C. [J.], & Driver, J. (1994). Covert spatial orienting in audition: Exogenous and endogenous mechanisms facilitate sound localization. *Journal of Experimental Psychology: Human Perception & Performance*, **20**, 555-574.

Spence, C., Kingstone, A., Shore, D. I., & Gazzaniga, M. S. (2001). Representation of visuotactile space in the split brain. *Psychological Science*, **12**, 90-93.

Spence, C., McDonald, J., & Driver, J. (2004). Exogenous spatial cuing studies of human crossmodal attention and multisensory integration. In C. Spence & J. Driver (Eds.), *Crossmodal space and crossmodal attention* (pp. 277-320). Oxford: Oxford University Press.

Spence, C., & Ngo, M. K. (in press). Does attention or multisensory integration explain the crossmodal facilitation of masked visual target identification? To appear in B. E. Stein (Ed.), *The new handbook of multisensory processing*. Cambridge: MIT Press.

Spence, C., Sanabria, D., & Soto-Faraco, S. (2007). Intersensory Gestalten and crossmodal scene perception. In K. Noguchi (Ed.), *Psychology of beauty and Kansei: New horizons of Gestalt perception* (pp. 519-579). Tokyo: Fuzanbo International.

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

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

Spence, C., & Soto-Faraco, S. (2010). Auditory perception: Interactions with vision. In C. J. Plack (Ed.), *The Oxford handbook of auditory science: Hearing* (pp.271-296). Oxford: Oxford University Press.

Staufenbiel, S. M., van der Lubbe, R. H. J., & Talsma, D. (2011). Spatially uninformative sounds increase sensitivity for visual motion change. *Experimental Brain Research*, **213**, 457-464.

Stein, B. E., London, N., Wilkinson, L. K., & Price, D. P. (1996). Enhancement of perceived visual intensity by auditory stimuli: A psychophysical analysis. *Journal of Cognitive Neuroscience*, **8**, 497-506.

Stein, B. E., & Meredith, M. A. (1993). *The merging of the senses*. Cambridge: MIT Press.

Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: Current issues from the perspective of the single neuron. *Nature Reviews Neuroscience*, **9**, 255-267.

Stiles, W. S. (1949). Increment thresholds & the mechanisms of colour vision. *Documenta Ophthalmologica*, **3**, 138-165.

Talsma, D., Doty, T. J., & Woldorff, M. G. (2007). Selective attention and audiovisual integration: is attending to both modalities a prerequisite for early integration? *Cerebral Cortex*, **17**, 691-701.

Tan, H. Z., Gray, R., Spence, C., Jones, C. M., & Rosli, R. M. (2009). The haptic cuing of visual spatial attention: Evidence of a spotlight effect. In B. E. Rogowitz & T. N. Pappas (Eds.), *Proceedings of SPIE-IS&T Electronic Imaging, Human Vision and Electronic Imaging XIV* (Vol. 7240, pp. 12). San Jose: SPIE-IS&T.

Theeuwes, J. (1992). Perceptual selectivity for color and form. *Perception & Psychophysics*, **51**, 599-606.

Theeuwes, J. (1994). Stimulus-driven capture and attentional set: Selective search for color and visual abrupt onsets. *Journal of Experimental Psychology: Human Perception and Performance*, **20**, 799-806.

Theeuwes, J., & Van der Burg, E. (2007). The role of spatial and nonspatial information in visual selection. *Journal of Experimental Psychology: Human Perception and Performance*, **33**, 1335-1351.

Townsend, J. T., & Ashby, F. G. (1983). *Stochastic modeling of elementary psychological processes*. New York: Cambridge University Press.

Treisman, A. (1996). The binding problem. *Current Opinion in Neurobiology*, **6**, 171-178.

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

Treisman, A. M., & Sato, S. (1990). Conjunction search revisited. *Journal of Experimental Psychology: Human Perception and Performance,* **16**, 459-478.

Trommershäuser, J., Landy, M.S., & Körding, K.P. (Eds.). (2011). *Sensory cue integration*. New York: Oxford University Press.

Tse, P. E., Intriligator, J., Rivest, J., & Cavanagh, P. (2004). Attention and the subjective expansion of time. *Perception & Psychophysics*, **66**, 1171-1189.

Underwood, G., Chapman, P., Bowden, K., & Crundall, D. (2002). Visual search while driving: Skill and awareness during inspection of the scene. *Transportation Research Part F*, **5**, 87-97.

Van der Burg, E., Cass, J., Olivers, C. N. L., Theeuwes, J., & Alais, D. (2010). Efficient visual search from synchronized auditory signals requires transient audiovisual events. *PLoS ONE*, **5**: e10664. doi:10.1371/journal.pone.001066

Van der Burg, E., Olivers, C. N. L., Bronkhorst, A. W., & Theeuwes, J. (2008a). Pip and pop: Nonspatial 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., Bronkhorst, A. W., & Theeuwes, J. (2008b). Audiovisual events capture attention: Evidence from temporal order judgments. *Journal of Vision*, **8(5):2**, 1-10.

Van der Burg, E., Olivers, C. N. L., Bronkhorst, A. W., & Theeuwes, J. (2009). Poke and pop: Tactile-visual synchrony increases visual saliency. *Neuroscience Letters*, **450**, 60-64.

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**, 1208-1218.

Van Wert, M. J., Horowitz, T. S., & Wolfe, J. M. (2009). Even in correctable search, some types of rare targets are frequently missed. *Attention, Perception, & Psychophysics*, **71**, 541-553.

Vicente, K. J., & Rasmussen, J. (1992). Ecological interface design: Theoretical foundations. *IEEE Transactions of Systems, Man, and Cybernetics*, **22**, 589-606.

von Grünau, M., & Anston, C. (1995). The detection of gaze direction: A stare-in-the-crowd effect. *Perception*, **24**, 1297-1313.

von Mühlenen, A., & Lleras, A. (2007). No-onset looming motion guides spatial attention. *Journal of Experimental Psychology: Human Perception and Performance*, **33**, 1297-1310.

Vroomen, J., & de Gelder, B. (2000). Sound enhances visual perception: Cross-modal effects of auditory organization on vision. *Journal of Experimental Psychology: Human Perception and Performance*, **26**, 1583-1590.

Watanabe, K., & Shimojo, S. (1998). Attentional modulation in perception of visual motion events. *Perception*, **27**, 1041-1054.

Watanabe, K., & Shimojo, S. (2001). When sound affects vision: Effects of auditory grouping on visual motion perception. *Psychological Science*, **12**, 109-116.

Watson, D. G., & Humphreys, G. W. (1995). Attentional capture by contour onsets and offsets: No special role for onsets. *Perception & Psychophysics*, **57**, 583-597.

Watson, D. G., & Humphreys, G. W. (1997). Visual marking: Prioritizing selection for new objects by top-down attentional inhibition of old objects. *Psychological Review*, **104**, 90-122.

Watson, D. G., Humphreys, G. W., & Olivers, C. N. L. (2003). Visual marking: Using time in visual selection. *Trends in Cognitive Sciences*, **7**, 180-186.

Weinstein, S. (1968). Intensive and extensive aspects of tactile sensitivity as a function of body part, sex and laterality. In D. R. Kenshalo (Ed.), *The skin senses* (pp. 195-218). Springfield: Thomas.

Wickens, C. D. (1984). Processing resources in attention. In R. Parasuraman & R. Davies (Eds.), *Varieties of attention* (pp. 63-101). New York: Academic Press.

Wickens, C. D. (2008). Multiple resources and mental workload. *Human Factors*, **50**, 449-455.

Wickens, C. D., Rice, S., Keller, D., Hutchins, S., Hughes, J., & Clayton, K. (2009). False alerts in air traffic control conflict alerting system: Is there a “cry-wolf” effect? *Human Factors*, **51**, 446-462.

Wightman, F. L., & Kistler, D. J. (1999). Resolution of front-back ambiguity in spatial hearing by listener and source movement. *Journal of the Acoustical Society of America*, **105**, 2841-2853.

Wolfe, J. M. (1994). Visual search in continuous, naturalistic stimuli. *Vision Research*, **34**, 1187-1195.

Wolfe, J. M. (1998). What can 1 million trials tell us about visual search? *Psychological Science*, **9**, 33-39.

Wolfe, J. M., Horowitz, T. S., & Kenner, N. M. (2005). Rare items often missed in visual searches. *Nature*, **435**, 439-440.

Wolfe, J. M., & Van Wert, M. J. (2010). Varying target prevalence reveals two dissociable decision criteria in visual search. *Current Biology*, **20**, 121-124.

Wolfe, J. M., Võ, M. L.-H., Evans, K. K., & Greene, M. R. (2011). Visual search in scenes involves selective and nonselective pathways. *Trends in Cognitive Sciences*, **15**, 77-84.

Zampini, M., Guest, S., Shore, D. I., & Spence, C. (2005). Audio-visual simultaneity judgments. *Perception & Psychophysics*, **67**, 531-544.

Yamani, Y., & McCarley, J. S. (2011). Visual search asymmetries in heavy clutter: Implications for display design. *Human Factors*, **53**, 299-307.

Yantis, S. (1993). Stimulus-driven attentional capture and attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance*, **19**, 676-681.

Yantis, S., & Jonides, J. (1984). Abrupt visual onsets and visual attention: Evidence from visual search. *Journal of Experimental Psychology: Human Perception and Performance*, **10**, 601-621.

APPENDIX A:
REVERSING THE COLAVITA VISUAL DOMINANCE EFFECT

A.0. Introduction

When auditory and visual events occur simultaneously in a detection task people tend to fail to respond to the auditory stimulus. This effect, known as the Colavita visual dominance effect (Colavita, 1974; Colavita & Weisberg, 1979; Osborn, Sheldon, & Baker, 1963), occurs even when unimodal auditory and visual stimuli presented in isolation are responded to with equivalent speed and accuracy. The original experiments reporting the Colavita visual dominance effect required participants to press one key whenever they detected the appearance of a visual stimulus (a light flash), another key when they detected an auditory stimulus (a beep), and a third key (or in some cases, pressing both unimodal response keys) when the visual and auditory stimuli were presented simultaneously (Colavita, 1974; Colavita & Weisberg, 1979; Egeth & Sager, 1977; Koppen & Spence, 2007a; Spence, 2009). The tasks and stimuli utilized to explore the Colavita effect have become more sophisticated in recent years (see Spence, 2009; Spence, Chen, & Parise, 2011, for a review). For instance, rather than detecting the occurrence of every flash, beep, or flash plus beep, participants have often been required to respond selectively to pre-specified visual (e.g., the outline drawing of a traffic light), auditory (e.g., the sound of a cat meowing), or bimodal targets (e.g., the traffic light and the cat) embedded into a rapidly-presented stream of irrelevant visual and auditory stimuli (Koppen, Alsius, &

Spence, 2008; Sinnett, Soto-Faraco, & Spence, 2008; Sinnett, Spence, & Soto-Faraco, 2007).

The Colavita effect observed in bimodal stimulus presentations stands in stark contrast to the results of numerous other studies in which it has been shown that the concurrent presentation of visual and auditory stimuli actually leads to multisensory facilitation, as evidenced by faster reaction times (RTs; e.g., Forster, Cavina-Pratesi, Aglioti, & Berlucchi, 2002; Nickerson, 1973; see also the results of the experiments reported in Chapters 2-4 of the present thesis) greater accuracy (e.g., Fiebelkorn, Foxe, Butler, & Molholm, 2011; Giard, & Peronnet, 1999; Vroomen & de Gelder, 2000; see also the results of the experiments reported in Chapter 3 of the present thesis), and stronger neuronal responses (e.g., Molholm, Ritter, Javitt, & Foxe, 2004; Stein & Meredith, 1993) as compared to the unimodal presentations of either visual or auditory stimuli. It is important to bear in mind, however, that whether or not multisensory facilitation or inhibition is observed under conditions of bimodal audiovisual stimulation strongly depends on the task demands and on the response(s) required of participants (see Sinnett et al., 2008, on this point).

Sinnett et al. (2008) recently demonstrated that the nature of participants' response (and task) can modulate whether multisensory facilitation or inhibition (and by extension, the Colavita visual dominance effect) will be observed. For example, when the task was simply to detect the presence of any target (visual, auditory, or bimodal) by pressing a single response key (i.e., regardless of the sensory modality of the target), multisensory facilitation in the form of a redundant target effect (RTE) was observed, with participants responding more rapidly to bimodal than to unimodal targets. On the other hand, when the participants were instructed to make separate responses to the targets presented in each sensory modality, multisensory inhibition (i.e., the Colavita visual dominance effect) was observed instead. Importantly, Sinnett et al. further demonstrated that multisensory facilitation and inhibition could be elicited using exactly the same task and stimuli simply by changing the sensory modality to which participants were instructed to respond. Specifically, when responding only to the visual modality, responses to visual targets were facilitated (i.e., speeded-up) in the context of bimodal stimulation (auditory and visual) with respect to unimodal presentations. Conversely, when responding to auditory targets, responses were inhibited (i.e., slowed down) in the same bimodal simultaneous presentations with respect to unimodal auditory presentations. Regardless of whether facilitation or inhibition is observed, there appears to be a unidirectional advantage in favor of the visual modality, at least in the context of a simple detection task.

It is possible that the task demands used in the basic Colavita paradigm do not provide the optimal circumstances for superior responding to auditory over visual targets, which is why few studies have been able to effectively eliminate the Colavita visual dominance effect. In fact, no study has, to date, been able to reverse the effect (i.e., to show auditory dominance; Colavita & Weisberg, 1979; Koppen & Spence, 2007d; Sinnett et al., 2007). Welch and Warren's (1980, 1986) *modality appropriateness* hypothesis may offer a potential way to explore sensory dominance, and, in effect, to eliminate or reverse the Colavita effect. According to the *modality appropriateness* hypothesis, the sensory modality eliciting superior processing of a specific dimension/stimulus attribute, such as space or time, ought to be the modality that will dominate over others. O'Connor and Hermelin (1972) argued that while vision is best suited for the processing of spatial information (see also Shimojo & Shams, 2001), audition is most effective at processing temporal information. The visual capture of auditory apparent motion (e.g., Soto-Faraco, Spence, & Kingstone, 2004), and the ventriloquist effect (Bertelson, Vroomen, de Gelder, & Driver, 2000; Howard & Templeton, 1966) exemplify how vision can strongly influence the perception of auditory stimuli when spatial judgments are involved.

There are also many examples of visual processing being influenced by auditory stimuli in the temporal domain (e.g., Gebhard & Mowbray, 1959; Morein-Zamir, Soto-Faraco, & Kingstone, 2003; Shams, Kamitani, & Shimojo, 2000; Shipley, 1964; Walker & Scott, 1981; Welch, Duttonhurt, & Warren, 1986). Moreover, there are temporal processing limitations in the visual modality (see McKeeff, Remus, & Tong, 2007) that are more severe than those seen in the auditory modality (Chen & Yeh, 2009; Welch & Warren, 1980). Soto-Faraco and Spence (2002) adapted the repetition blindness (Kanwisher, 1987) paradigm to written (visual) or spoken (auditory) streams of letters and, consistent with the idea of superior auditory temporal processing, they demonstrated repetition blindness to visual repetitions whereas no such repetition deficit (*repetition deafness*) was observed for auditory repeats at the same presentation rate.

The goal of Experiment A.1 was therefore to explore whether using a repetition detection task, in which participants' auditory performance is typically better than visual performance, would lead to the elimination or even to the reversal of the Colavita effect (i.e., to auditory dominance). In Experiment A.1, participants monitored a stream of simultaneously-presented visual and auditory stimuli and responded as soon as they saw the immediate repetition of a picture, heard the immediate repetition of a sound, or saw and heard the immediate repetition of a picture and sound. Using this repetition detection task also provides a different context for studying the Colavita effect, one that is not based on simple detection tasks (as in the traditional Colavita paradigm), but rather relies on a more abstract, rule-based level of representation. Here, participants must not only hold the representation of an object in their working memory, but also make a decision about whether or not that object was a repetition of a previously-presented object.

Based on the *modality appropriateness* hypothesis (e.g., Welch & Warren, 1980, 1986) and given the fact that audition typically dominates in tasks involving temporal processing, one might expect a more accurate detection rate for unimodal auditory than visual repetitions. Critical to the question addressed in Experiment A.1, if the auditory advantage observed in detecting repetitions in rapidly-presented streams of stimuli (Soto-Faraco & Spence, 2002) can be extrapolated to the recent paradigms used to investigate the Colavita effect, then a differential advantage for visual stimuli would no longer be expected. Instead, participants would be expected to be just as, if not more, likely to make auditory-only than visual-only responses to audiovisual targets. If, however, the traditional Colavita effect were again to be observed on the bimodal repetition trials, it would indicate that the basic mechanisms responsible for visual dominance can extend from a sensory level (i.e., detect the appearance of a specific visual, auditory, or audiovisual stimulus) to a more abstract rule-level (i.e., detect a stimulus repetition), regardless of the appropriateness of the modality in the task at hand.

A.1. Experiment A.1

A.1.1. Methods

Participants. Twenty-four participants from the University of Oxford (6 female) ranging in age from 19-40 years (mean age of 28 years) took part in Experiment A.1. All of the participants who took part in the experiments reported in Appendix A were right-handed by self-report.

Apparatus and materials. Fifty line drawings of common objects chosen from the Snodgrass and Vanderwart (1980) database were used as visual stimuli (see Figure A.1). The drawings were randomly rotated 30 degrees to the left or right so as to ensure task difficulty (see Rees, Russell, Frith, & Driver, 1999). Fifty sounds were selected from a database of 103 sounds (downloaded from www.a1freesoundeffects.com, 01/02/2003) on the basis of their clarity and familiarity, as rated by three judges. Loudspeaker cones placed directly beside the computer monitor were used to present the sounds at 70 dB (A), as measured from the participant’s ear position, producing sounds that were subjectively localized to the center of the monitor. All of the sounds were adjusted to have similar lengths (between 400-500 ms; average of 479 ms), with the exception of the sound of a raindrop that was shorter. The pictures were presented for the same duration as the corresponding sounds. The pictures and sounds were presented synchronously (with the same onset and duration) and randomized using the DMDX software (created by J. Forster, see http://www.u.arizona.edu/~jforster/dmdx.htm). The picture-sound stimuli were separated by a 250 ms silent/blank interval (giving rise to a stimulus onset asynchrony, SOA, of between 650 and 750 ms). Care was taken to ensure that none of the sounds or pictures had the same or related semantic meaning when they were presented at the same time.

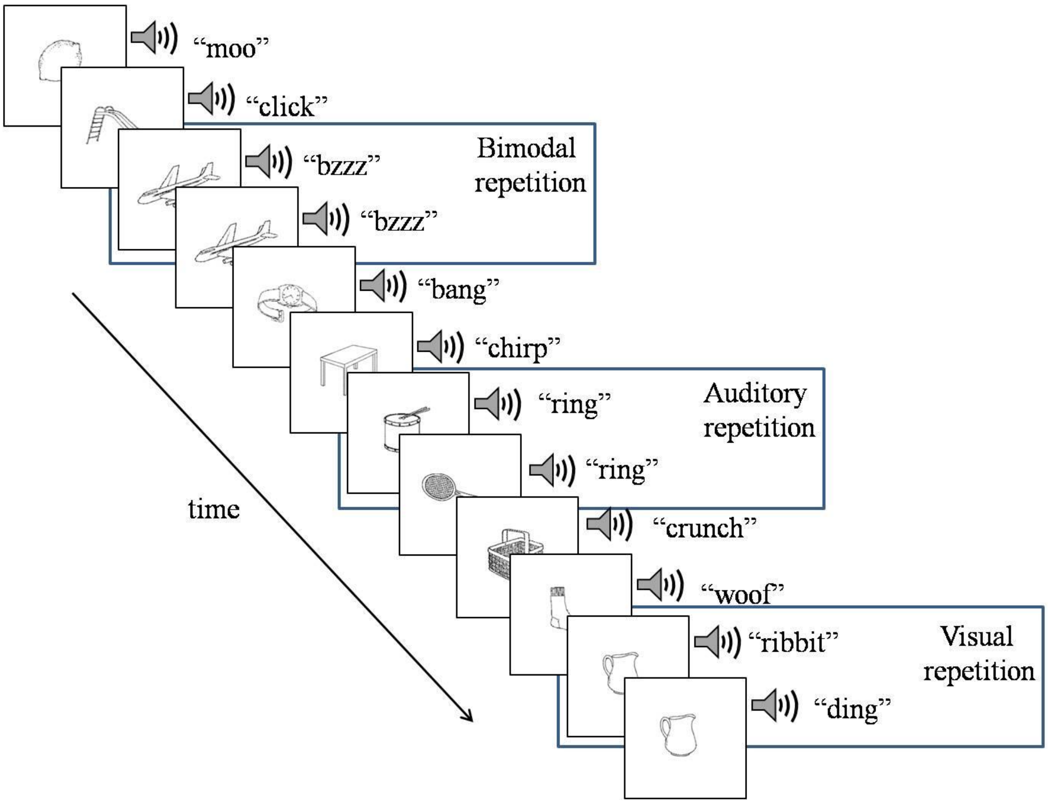


Figure A.1. A schematic illustration of the synchronous presentation of auditory and visual stimuli in Experiment A.1, including an example of a bimodal repetition trial, a unimodal auditory repetition trial, and a unimodal visual repetition trial. Each stimulus pair was presented for 400-500 ms, and separated by a 250 ms blank screen and silence making for an SOA of approximately 650-750 ms.

Procedure. The participants sat approximately 60 cm from a computer screen in a dimly-lit, sound-attenuated room. They were instructed to monitor the streams of auditory and visual stimuli and to respond as rapidly as possible whenever they saw or heard the immediate repetition of a picture, sound, or both by using their right hand to press the "J", "K", or "L" key on a standard keyboard. These three response keys were counterbalanced across participants for each response type: visual, auditory, or bimodal.

There were a total of 1000 trials (presented over 10 blocks), of which 250 trials involved the immediate repetition of a target. Overall, visual and auditory target repetitions were equiprobable (100 of each type), whereas bimodal repetitions were presented 50 times. Thus, of the 250 repetition trials, 40% were visual, 40% were auditory, and 20% were bimodal repetitions (similar proportions have been used in the majority of previous studies of the Colavita effect (Koppen & Spence, 2007a, c, d). In each block of 100 trials, every picture and sound was presented twice, with 25 repeated target trials per block (10 visual, 10 auditory and 5 bimodal). As there were only 100 trials in each block, repeated target trials could appear in succession in each block (with a maximum of 5 being presented in a row). All bimodal stimuli were constructed so that a picture and sound were never combined in the same way. That is, the picture of a chair only appeared with the sound of a cat once, unless, of course, the item constituted a bimodal target repetition (5% of all trials). Note also that the auditory and visual targets on the bimodal target trials were not semantically congruent. The participants were given training with the task using a stimulus sequence containing three repetitions — one visual, one auditory, and one bimodal — before starting the main experimental session. The training session was repeated until the participant felt comfortable with the task. The experiment lasted for approximately 20 minutes.

A.1.2. Results and Discussion

The percentage of correct responses, the percentage of missed repetitions (trials on which participants failed to make a response to a target within the allotted 1250 ms), the distribution of errors to the bimodal targets (where a

Colavita effect might be observed), and the reaction times (RT) of correct responses were calculated for each participant and condition. The errors that participants made in response to bimodal targets were classified into one of three categories: a unimodal visual response (i.e., when the participant failed to respond to the auditory component of the target), a unimodal auditory response (when the participant failed to respond to the visual component of the target), or a miss (when the participant failed to make any response at all). Analyses of the RT data for all of the experiments reported in Appendix A were non-significant ($F<1$ for all experiments except Experiment A.1 and A.3, where $F(2,46)=1.41$, $p=.26$ and $F(2,32)=1.49$, $p=.24$, respectively). These data are presented for completeness in Table A.1 below.

Table A.1. Mean RTs and SEMs (in ms) in response to visual, auditory, and audiovisual targets in Experiments A.1 through A.5.

	Target		
Experiment	Visual	Auditory	Audiovisual
A.1	857 (15)	880 (14)	866 (15)
A.2	905 (17)	914 (18)	913 (22)
A.3	884 (28)	860 (18)	898 (30)
A.4	855 (30)	859 (30)	863 (24)
A.5	866 (21)	854 (18)	863 (16)

A repeated-measures analysis of variance (ANOVA) on the percentage of correct responses with Repetition type (unimodal visual, unimodal auditory, or bimodal audiovisual) as the within-participants factor revealed a significant main effect, $F(2,46)=3.66$, $p=.03$. Separate pairwise comparisons indicated that participants responded less accurately to bimodal repeats (M=81.5%, SE=2.4%)

than to either of the unimodal repeats (auditory, *M*=87.4%, *SE*=2.0%, *p*=.03; visual, *M*=85.5%, *SE*=1.5%, *p*=.06), without significant differences between unimodal trials (*p*=.40, respectively; see Figure A.2A). It is therefore interesting to find out how the different categories of error were distributed.

An ANOVA conducted on the percentage of misses (i.e., where no response was made to a repeated target) revealed a significant main effect of target modality, *F*(2,46)=7.12, *p*=.002. The participants missed significantly more visual repeats (*M*=12.0%, *SE*=1.5%) than auditory (*M*=8.5%, *SE*=1.1%) or bimodal (*M*=5.6%, *SE*=1.7%) repeats, *p*=.04 and *p*=.001, respectively, thereby suggesting that repetition blindness effects are strongest in the visual modality. The difference between the percentage of auditory and bimodal misses failed to reach statistical significance, *p*=.10.

Of interest to the question being addressed in this section of the thesis, an analysis of the erroneous responses restricted to just the bimodal trials revealed that participants made nearly three times as many unimodal visual response errors (*M*=9.6%, *SE*=1.2%) than unimodal auditory response errors (*M*=3.3%, *SE*=0.7%), *t*(23)=5.18, *p*<.001, see Figure A.2B). These results therefore provide convincing evidence of a typical Colavita visual dominance effect in participants' responses to targets consisting of bimodal repeats.

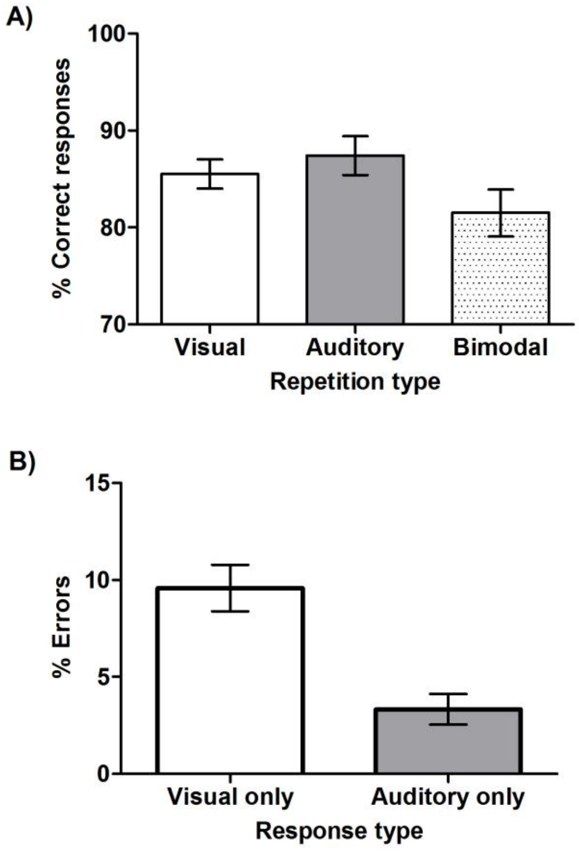


Figure A.2. A) Mean percentage of correct responses for the visual, auditory, and bimodal repetition trials in Experiment A.1. B) Mean percentage of visual versus auditory only responses (i.e., errors) made on bimodal repetition trials.

The results of Experiment A.1 demonstrate that the Colavita visual dominance effect persists under conditions where the participants' task involves the detection of stimulus repetitions, so that the target was defined by a rule requiring continuous updating of the relevant object after every item in the stream had been presented. The Colavita effect has traditionally been demonstrated in tasks in which the targets were defined on the basis of their low-level, sensory attributes involving the simple detection of a unimodal visual, unimodal auditory, or bimodal audiovisual stimulus. The findings of Experiment A.1 therefore provide a first demonstration that the Colavita visual dominance effect also affects higher levels of perceptual processing (cf. Hein, Parr, & Duncan, 2006).

An important aspect of using a repetition detection task in order to demonstrate visual dominance is the putative superiority of the auditory system at this task. That is, previous findings have reported robust repetition blindness deficits in the visual modality but not in the auditory modality using targets embedded in streams of distractor events similar to those that were used here (e.g., Soto-Faraco & Spence, 2002). Accordingly, the Colavita effect was expected to be eliminated, if not reversed, under such favorable conditions for the auditory system. That is not to say, however, that the results of Experiment A.1 contradict the findings of previous research demonstrating auditory dominance in temporally-related tasks (Chen & Yeh, 2009; Kanwisher, 1987; Soto-Faraco & Spence, 2002; Walker & Scott, 1981). In fact, consistent with the findings of Soto-Faraco and Spence’s earlier research, an examination of the percentages of missed visual, auditory, and audiovisual repetitions in the present study revealed that participants missed significantly more visual than auditory repetitions, while the percentage of missed audiovisual repetitions did not differ significantly from the percentage of missed auditory repetitions (despite the additional signal present in the target).

Remarkably, despite this significant unimodal visual repetition deficit, participants made significantly more (erroneous) visual-only responses than auditory-only responses to bimodal trials (a mean difference of 7%), which is consistent with the traditional Colavita visual dominance effect (see Koppen et al., 2008; Koppen & Spence, 2007a, b, c, d, for examples with similar magnitudes of visual dominance in the range of 4-10%). This finding suggests that when participants have to respond to both sensory modalities, vision still dominates even in the context of a repetition, a task which has historically shown advantages for the auditory modality (see Welch et al., 1986; Welch & Warren, 1980, 1986). This points to either a disadvantage for auditory processing or to an advantage for visual processing (or perhaps both; see Sinnett et al., 2008) in the presentation of bimodal stimuli that occurs regardless of the specific parameters of the task, be it temporal or spatial. The results of Experiment A.1 therefore provide further evidence in support of Sinnett et al.'s (2008) suggestion that multisensory stimulation can simultaneously work to enhance the perception of the objects presented in one sensory modality (i.e., visual) while inhibiting the perception of the stimuli presented in another modality (i.e., auditory).

Moro and Steeves (2010) recently reported that unilaterally enucleated patients (i.e., those individuals who had had early unilateral surgical eye removal due to cancer of the retina), who presumably rely less on vision than other normally-sighted individuals, made just as many unimodal auditory responses as visual responses to bimodal targets in a task requiring the participants to respond to pre-specified targets, again demonstrating an elimination of the Colavita visual dominance effect. Assuming that such patients would rely less on vision and more on audition (Lessard, Pare, Lepore, & Lassonde, 1998), Moro and Steeves concluded that in order to eliminate the Colavita effect there must be some form of degradation in the reliability of the visual information presented to participants. Under such conditions, the enucleated patients may have biased their attention, and in turn their responding, toward the auditory modality. That is, while the percentage of auditory-only responses was approximately equal between both groups, the enucleated patients made significantly fewer visual-only responses to bimodal targets as compared to the normally-sighted participants.

Despite the fact that unimodal auditory dominance was observed on unimodal target detection trials, a strong visual dominance effect for bimodal targets, with participants making more visual-only responses (i.e., errors) than auditory responses on these bimodal trials, was nevertheless still observed in Experiment A.1. It would therefore appear that the temporal demands of the repetition detection task used in this experiment were insufficient to bring about a bias toward responding preferentially to stimuli presented in the auditory modality.

One known advantage of auditory processing over visual processing is that short-term auditory (echoic) memory is longer lasting than short-term visual (iconic) memory. According to some estimates, echoic memory lasts upwards of 2000 ms (see Cowan, 1984), whereas iconic lasts for only up to a maximum of 1000 ms (Sperling, 1960), with some studies suggesting conservative estimates of only 200-300 ms (Di Lollo, 1977; Long, 1980). Note that Experiment A.1 used an immediate repetition task in which a potential target repetition required participants to retain the visual information for only about 250 ms. It is possible then that the visual dominance effect observed in this experiment may have been related to the persistence of the visual stimulus in participants' iconic memory store. This is likely given that the average response latency to repeated targets in this immediate repetition task was still well-within the 1000 ms upper limit of iconic memory.

Rather than detecting immediate repetitions, then, participants in Experiment A.2 were required to detect non-adjacent (n-1) repetitions instead.

This potentially induced a sufficient degradation of the internal memory of the visual signal by way of the rapidly decaying iconic memory trace (Di Lollo, 1977; Long, 1980; Sperling, 1960). As the onset of the repeated n-1 target started at least 979 ms after the first item of the repeated pair, it is unlikely that any decision would be based on a simple iconic memory trace. Additionally, introducing an intervening stimulus between target n-1 repetitions may also have given rise to potential masking effects (see Enns & Di Lollo, 2000; Intraub, 1984; Loftus & Ginn, 1984; Potter, 1976, 1999). Specifically, an intervening stimulus might mask potential n-1 targets by competing with and making the first part of the target difficult to access from working memory.

The participants monitored a stream of simultaneously-presented visual and auditory stimuli and responded as soon as they saw or heard the n-1 repetition of a picture (i.e., separated by an interleaved different picture and sound), a sound, or both. If either the temporal decay of iconic memory, or the masking effects of the intervening stimulus, imposes a limit on visual dominance, then by extending the timing of stimulus presentations beyond this temporal limit, an elimination of the Colavita effect, or perhaps even a reversal (i.e., auditory dominance), would be expected.

A.2. Experiment A.2

A.2.1. Methods

Twenty participants from the University of Oxford (12 female) ranging in age from 20-55 years (mean age of 28 years) took part in this experiment. The experimental set-up and procedure was identical to that of Experiment A.1. However, participants were now instructed to monitor the streams of auditory and visual stimuli and to respond as rapidly as possible whenever they saw or heard the repetition of a picture, sound, or both separated by an interleaved different picture and sound (i.e., an n-1 repetition task; see Figure A.3). It is important to note that while the target trials could be presented consecutively, they were never interleaved. That is, the first half of a repetition never preceded the second half of the previous repetition.

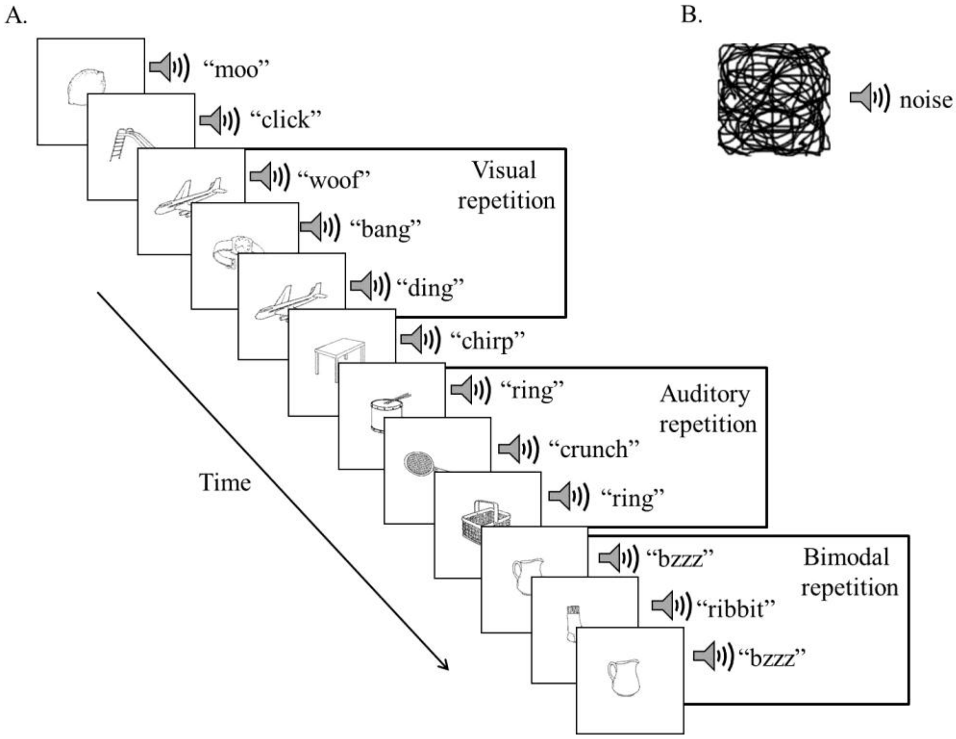


Figure A.3. A) Schematic illustration of the synchronous presentation of visual and auditory stimuli in Experiment A.2. An example of a visual, an auditory, and a bimodal n-1 target are included. Each stimulus pair was presented for 400-500 ms, and separated by a 250 ms blank screen and silence making for an SOA of 650-750 ms. B) Image of visual pattern mask and noise burst between target stimuli used in Experiment A.4.

A.2.2. Results and Discussion

A repeated-measures ANOVA with target type (visual, auditory, or bimodal) as the within-participants factor was conducted on the percentage of correct responses. This analysis highlighted significant differences among the three target types, *F*(2,38)=17.42, *p*<.001 (see Figure A.4A). Separate pairwise comparisons revealed that participants responded more accurately to unimodal auditory targets (*M*=56.0%, *SE*=4.0%) than to either unimodal visual (*M*=35.4%, *SE*=2.8%, *t*(19)=-5.88, *p*<.001) or to bimodal (*M*=34.4%, *SE*=4.7%, *t*(19)=4.64, *p*<.001) targets. The difference between the latter two conditions was not statistically significant, *p*=.81.

An ANOVA conducted on the percentage of misses (i.e., where no response was made to a target) revealed a significant main effect of target type, *F*(2,38)=88.94, *p*<.001 (see Figure A.4B). The participants missed significantly more unimodal visual targets (*M*=57.0%, *SE*=2.7%) than unimodal auditory (*M*=38.9%, *SE*=3.9%) or bimodal (*M*=21.9%, *SE*=2.8%) targets, *p*<.001 for both comparisons. This result, together with the previous analysis of correct responses, strongly suggests that participants found it significantly more difficult to detect n-1 visual repetitions than to detect auditory repetitions. The participants also missed significantly more unimodal auditory than bimodal target trials, *p*<.001, suggesting that the bimodal targets were the easiest to detect. This reduced number of misses in bimodal targets may have been due to a redundancy effect as the participants could have responded to any one of its constituent components (see also Spence et al., 2011).

As mentioned previously, participants could have committed one of three types of errors on the bimodal target trials: A miss, a unimodal visual response, or a unimodal auditory response. Of the errors made on the bimodal trials, 21.9% of those were misses, and the remaining trials were either unimodal visual or auditory responses. A closer look at the pattern of bimodal response errors revealed that participants made significantly more auditory based (*M*=27.2%, *SE*=2.7%) than visual based unimodal responses (*M*=16.0%, *SE*=2.2%), *t*(19)=-3.48, *p*=.002 (see Figure A.4C), thus demonstrating a clear reversal of the Colavita visual dominance effect!

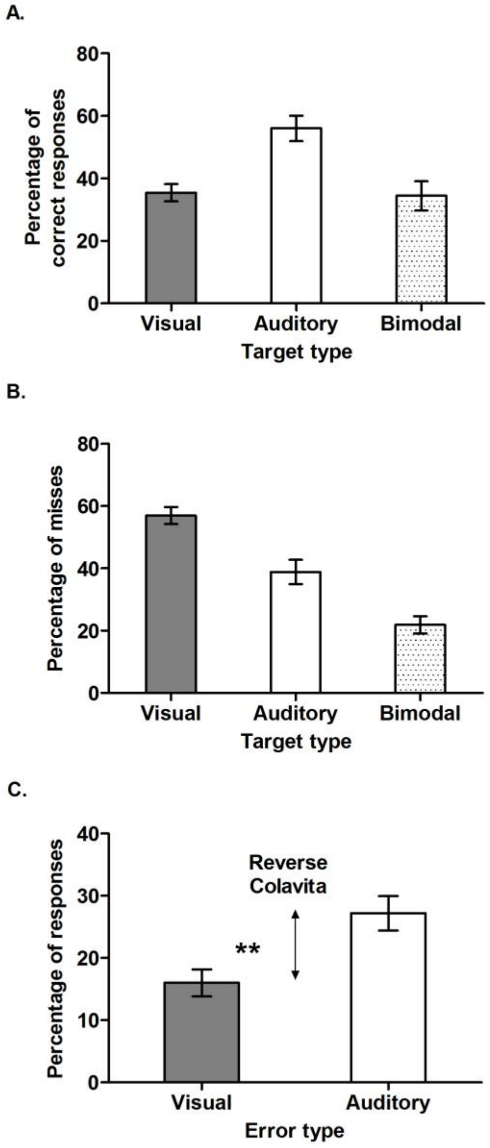


Figure A.4. A) Mean percentage of correct responses for the visual, auditory, and bimodal repetition n-1 trials in Experiment A.2. B) Mean percentage of visual, auditory, and bimodal misses. C) Mean percentage of visual versus auditory only responses (i.e., errors) made on bimodal repetition trials.

The results of Experiment A.2 provide the first clear evidence for a reversal of the Colavita visual dominance effect when participants have to respond to targets consisting of non-adjacent repetitions. One possible account for this finding is couched in terms of the duration of iconic memory, which, as mentioned previously, typically only lasts for up to one second (Sperling, 1960; Di Lollo, 1977; Long, 1980), whereas echoic memory typically lasts upwards of two seconds (Cowan, 1984). Here, the participants would have had to wait at least 979 ms (250 ms blank interval + 479 ms intervening stimulus presentation + 250 ms blank) before being able to determine whether an n-1 repetition had occurred.

Given the known limits of iconic and echoic memory, it is likely that the target (the first instance of the repetition) will have been presented just at, if not beyond, the upper limit of iconic but not echoic memory. Thus, it is possible that the visual information contained in the first stimulus presentation may have decayed prior to the presentation of the potential n-1 repeat (i.e., the third stimulus presented approximately 979 ms after the first stimulus), while the auditory information would have remained relatively less degraded. This would have rendered the visual information relatively “unreliable” as compared to the auditory information in the stream.

Moro and Steeves (2010) previously suggested that degradation in the reliability of the visual information is necessary to eliminate the Colavita visual dominance effect. It would appear then that the n-1 repetition detection task brought about a sufficient degradation in vision not only to eliminate, but apparently to reverse, the Colavita effect observed in Experiment A.2. It is unclear, however, whether the visual degradation was due to the decay of the iconic memory trace or the interference from the intervening stimulus, or some combination of the two.

The goal of Experiment A.3 was therefore to determine the role of the temporal limit of iconic memory in reversing the Colavita visual dominance effect. Here, the duration of the both the stimulus presentations and the blank interstimulus interval were shortened. The amount of time required to detect a visual n-1 target repetition would now fall well within the limits of iconic memory. Any potential masking effects should remain intact, however, as the intervening stimuli in the n-1 task could still have competed with the target stimuli for processing resources. Should the Colavita visual dominance effect be observed once again, and auditory dominance be eliminated, on the bimodal target trials, this would suggest that the limits on sensory memory played an important role in reversing the Colavita effect in Experiment A.2. However, if auditory dominance were to be observed in the next experiment this would, once again, suggest that the masking effects from the intervening stimulus presentations were responsible for the reversal of the Colavita effect.

A.3. Experiment A.3

A.3.1. Methods

Eighteen participants (13 female; age range = 18-36 years; mean age of 27 years) took part in Experiment A.3. The experimental set-up was identical to that used in Experiment A.2, except that the visual and auditory stimulus pairs were now presented for 200 ms and the blank interval was now 150 ms. Thus, the elapsed time between n-1 repetitions in this experiment was 500 ms, which is well within the upper limit of iconic memory.

A.3.2. Results and Discussion

An ANOVA on the percentage of correct responses revealed a significant main effect of target type, *F*(2,34)=35.8, *p*<.001 (see Figure A.5A). Separate pairwise comparisons indicated that participants responded more accurately to unimodal auditory targets (*M*=67.6%, *SE*=3.9%) than to either unimodal visual (*M*=46.6%, *SE*=3.1%, *p*<.001) or bimodal (*M*=34.8%, *SE*=4.8%, *p*<.001) targets. Participants also responded more accurately to unimodal visual than to bimodal targets, *p*=.012.

An ANOVA conducted on the percentage of misses (i.e., where no response was made to a target) also revealed a significant main effect of target type, *F*(2,34)=15.66, *p*<.001 (see Figure A.5B). The participants missed significantly more unimodal visual targets (*M*=44.0%, *SE*=3.5%) than unimodal auditory (*M*=26.7%, *SE*=3.8%), or bimodal (*M*=20.8%, *SE*=3.7%) targets, *p*<.001, for both comparisons. This result, together with the previous correct response analysis, suggests that the visual n-1 repetition detection task was more difficult than the auditory task. The numerical difference between the percentages of missed auditory and bimodal targets failed to reach significance, *p*=.22.

Most importantly, a paired-samples t-test revealed that participants also made significantly more erroneous auditory-only responses (*M*=28.0%, *SE*=2.4%) than erroneous visual-only responses (*M*=20.3%, *SE*=3.0%), *t*(17)=-2.29, *p*=.035 (see Figure A.5C) on the bimodal target trials. Thus, a reversal of the Colavita effect (i.e., auditory dominance) was still observed in the present experiment.

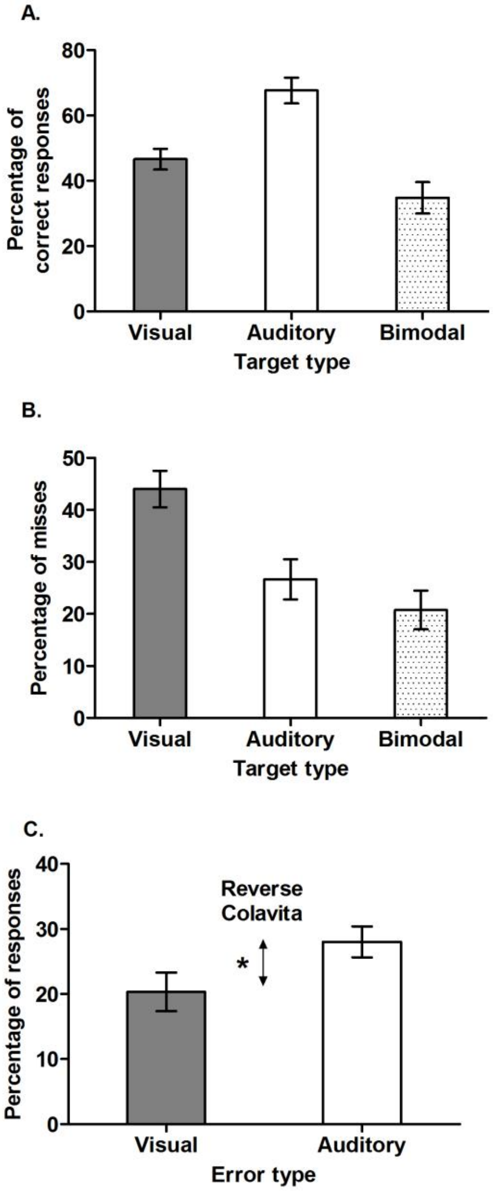


Figure A.5. A) Mean percentage of correct responses for the visual, auditory, and bimodal n-1 repetition trials in Experiment A.3. B) Mean percentage of visual, auditory, and bimodal misses. C) Mean percentage of visual versus auditory only responses (i.e., errors) made on bimodal repetition trials.

The results of Experiment A.3 demonstrate two important points: First, when the minimum time in which an n-1 target repetition could be detected was shortened to be within the limits of both iconic and echoic memory, the task became slightly easier, as evidenced by an overall increase in accuracy and decrease in the percentage of missed targets as compared to the results of

Experiment A.2[13]. Nevertheless, the overall accuracy of participants' responses still fell well below ceiling. Second, and more importantly, a reversal of the Colavita effect, whereby more auditory rather than visual responses were made to bimodal targets, was again observed in Experiment A.3. In fact, for both Experiments A.2 and A.3, participants made approximately 8% more auditory than visual responses to bimodal targets, even though the overall percentage of incorrect responses to bimodal targets was lower in Experiment A.3. Taken together, it would appear then that the temporal limit of visual sensory memory was not responsible for, or at least played only a small role in, the auditory dominance elicited by the n-1 repetition detection task.

Rather than occurring as a result of some intrinsic temporal limitation of sensory memory, it is likely that the reversal of the Colavita visual dominance effect may instead have resulted from interference from the intervening stimulus in the n-1 target repetition. Specifically, each intervening stimulus may have functioned as a conceptual mask (see Loftus & Ginn, 1984), which influenced performance by interrupting and competing for higher-level, semantic processing (Enns & Di Lollo, 2000; Intraub, 1984; Potter, 1976, 1999). Potter (1999) argued for the existence of a short-term memory store in which conceptual information is processed rapidly, but is highly unstable, and decays within a few hundred milliseconds. Unless this conceptual information is selected for further processing, it is quickly forgotten. Similarly, Enns and Di Lollo suggested that

---

[13] Between-experiments comparisons (independent-samples *t*-tests) were conducted on the percentages of correct responses and missed targets data from Experiments A.2 and A.3. These comparisons revealed that participants responded significantly more accurately to visual targets (*mean difference*=11.2%, *p*=.01) and auditory targets (*mean difference*=11.6%, *p*=.05) in Experiment A.3 as compared to Experiment A.2. There was no difference in the percentage of correct responses for bimodal targets (*mean difference*=0.4%) between Experiments A.2 and A.3, *p*=.95. The participants also missed significantly fewer visual targets (*mean difference*=13.0%; *p*=.006), auditory targets (*mean difference*=12.2%, *p*=.03), and bimodal targets (*mean difference*=12.7, *p*=.03) in Experiment A.3 as compared to Experiment A.2.

conceptual masks do not terminate target processing, but rather they become the new focus of attention. Each stimulus presentation in Experiment A.3 carried semantic content and could have been a part of a repetition. Thus, the nature of this n-1 task required participants to constantly update and match previously-presented items to current items based on their semantic content. Consistent with the previous masking literature, the intervening stimulus could easily have become the new focus of attention and replaced the preceding stimulus (the first part of the target) for further processing. It is important to note that the intervening stimulus in this experiment appears to have differentially affected information processing / retention of the target stimuli in the visual and auditory modalities. This is evidenced by the fact that the accuracy of participants' responses was higher for auditory than for visual targets.

The goal of Experiment A.4 was therefore to test whether instead of having an intervening meaningful stimulus separating potential n-1 repetition targets, the presentation of an irrelevant, meaningless visual stimulus (a pattern mask) paired with a burst of white noise, would serve to extinguish any potential conceptual masking effects. Should visual dominance rather than auditory dominance be observed, as had been observed in Experiments A.2 and A.3, on the bimodal target trials, this would suggest that the Colavita effect was indeed due to the conceptual masking of the intervening meaningful stimuli having a differential effect on vision than on audition. If the Colavita effect is still absent in Experiment A.4, however, then it is possible that there is still some processing of the pattern mask giving rise to interference effects with the target stimuli.

A.4. Experiment A.4

A.4.1. Methods

Sixteen participants (12 female; age range = 18-37 years; mean age of 24 years) took part in Experiment A.4. The set-up for this experiment was similar to that used in Experiment A.2 except that between the presentation of each visual and auditory stimulus pair, a pattern mask was presented in the center of the screen along with a 500 ms white noise burst (see Figure A.3B). The pattern mask consisted of a pattern of meaningless, overlapping squiggly lines that did not form a recognizable shape.

A.4.2. Results and Discussion

An ANOVA on the percentage of correct responses revealed a significant main effect of target type, $F(2,30)=13.08$, $p<.001$ (see Figure A.6A). Separate pairwise comparisons indicated that participants responded more accurately to unimodal auditory targets (M=84.3%, SE=4.3%) than to either unimodal visual (M=75.4%, SE=4.3%, $p<.001$) or bimodal (M=73.3%, SE=3.3%, $p=.001$) targets. There were no significant differences between the accuracy of participants' responses to unimodal visual and bimodal targets, $p=.38$.

An ANOVA conducted on the percentage of misses (i.e., where no response was made to a target) also revealed a significant main effect of target type, $F(2,30)=11.16$, $p<.001$ (see Figure A.6B). The participants missed significantly more unimodal visual targets (M=21.4%, SE=4.3%) than unimodal auditory (M=13.4%, SE=4.4%, $p<.001$) or bimodal (M=14.0%, SE=3.2%, $p=.003$) targets. The difference between the percentages of missed auditory and bimodal targets was not significant, $p=.77$. These results, together with the previous correct response analysis, suggest that the visual n-1 repetition detection task was still more difficult than the auditory task.

A paired-samples t-test revealed no significant difference between the percentage of incorrect auditory-only (*M*=3.9%, *SE*=0.9%) and visual-only responses (*M*=5.8%, *SE*=0.9%)**,** *t*(15)=1.15, *p*=.27 (see Figure A.6C) on bimodal target trials. Thus, the participants made similar percentages of visual and auditory based errors. This result demonstrates that the removal of the semantically-meaningful intervening stimulus within the n-1 target stream resulted in an elimination of the auditory dominance that had previously been observed in Experiments A.2 and A.3. Importantly, the Colavita visual dominance effect was not observed in the present experiment either.

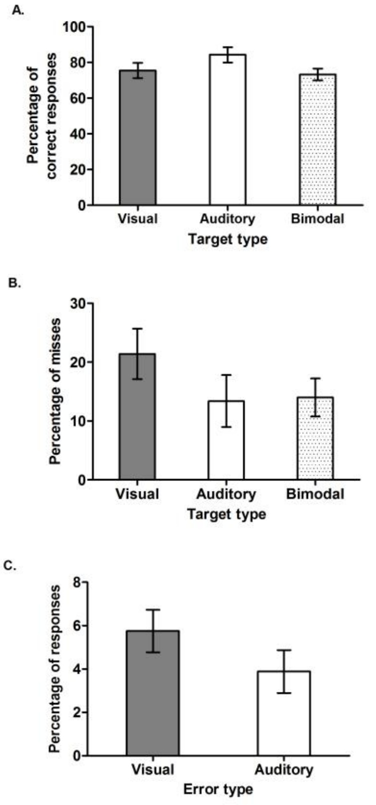


Figure A.6. A) Mean percentage of correct responses for the visual, auditory, and bimodal n-1 repetition trials in Experiment A.4. B) Mean percentage of visual, auditory, and bimodal misses. C) Mean percentage of visual versus auditory only responses (i.e., errors) made on bimodal repetition trials.

The results of Experiment A.4 suggest that it was indeed the interference caused by the processing of the semantically-meaningful intervening stimulus within an n-1 target trial that played a crucial role in determining whether or not the Colavita visual dominance effect would be reversed. Here, the participants had to detect a repetition that was separated by stimuli that were not related to the task (i.e., did not constitute any part of any target repetition). The SOA between target repetitions was the same as in Experiment A.2. Thus, a sensory memory explanation of the effects could be ruled out, where the decay of the visual memory trace due to the passage of time might have influenced the reliability of the visual information. Rather than making more auditory than visual responses to bimodal target repetitions, participants now made similar percentages of visual and auditory responses, thus, demonstrating an elimination (but, importantly, not a reversal) of the Colavita effect. Note that, numerically, participants made more visual-only than auditory-only responses to bimodal targets.

It appears that the pattern mask gave rise to some masking effects (Enns & Di Lollo, 2000; Loftus & Ginn, 1984) and, consequently, some reduction in visual dominance, given that the difference between visual-only and auditory-only responses to bimodal targets was not significant (as had been the case in Experiment A.1, where no intervening stimuli were presented between repetitions). In contrast with the results of Experiment A.3, it appears that the reduction in visual dominance resulting from the pattern mask was not as great as that observed when a semantically meaningful intervening stimulus was presented. Thus, although auditory dominance was no longer observed in Experiment A.4 (as it had been in Experiment A.2. and A.3), the relatively weaker masking effect of the pattern mask (as compared to the semantic intervening stimuli) might explain why the Colavita visual dominance effect was not completely reinstated either. The final experiment to be reported in this Appendix therefore went on to test whether eliminating an intervening stimulus altogether, thereby requiring participants to detect immediate (just like in Experiment A.1) rather than n-1 repetitions, would again give rise to a Colavita visual dominance effect on the bimodal trials. Here, the long SOA between targets (1000 ms) was also maintained in order to minimize the likelihood of a sensory memory explanation of the results.

A.5. Experiment A.5

A.5.1. Methods

Twenty-four participants (20 female; age range = 18-37 years; mean age of 25 years) took part in Experiment A.5. All but one of the participants were right-handed by self-report. The experimental setup was similar to that used in Experiment A.2. However, instead of detecting an n-1 repetition, participants now had to detect any immediate repetition of a picture, sound, or both. Additionally, the blank interstimulus interval was now extended from 250 to 1000 ms.

A.5.2. Results and Discussion

An ANOVA on the percentage of correct responses revealed a significant main effect of target type, *F*(2,46)=11.61, *p*<.001 (see Figure A.7A). Separate pairwise comparisons indicated that participants responded more accurately to unimodal auditory targets (*M*=86.2%, *SE*=2.7%) than to either unimodal visual (*M*=80.5%, *SE*=3.6%, *p*=.001) or bimodal (*M*=77.7%, *SE*=2.8%, *p*<.001) targets. There were no significant differences between the accuracy of participants' responses to the unimodal visual and bimodal targets, *p*=.17.

An ANOVA conducted on the percentage of misses (i.e., where no response was made to a target) also revealed a significant main effect of target type, *F*(2,46)=15.32, *p*<.001 (see Figure A.7B). The participants missed significantly more unimodal visual targets (*M*=17.0%, *SE*=4.3%) than unimodal auditory (*M*=11.3%, *SE*=2.7%, *p*=.001) or bimodal (*M*=9.3%, *SE*=2.3%, *p*<.001) targets. The difference between the percentages of missed auditory and bimodal targets was marginally significant, *p*=.086. This, together with the previous correct response analysis, suggests that detecting auditory repetitions was easier than detecting visual repetitions.

A paired-samples t-test revealed that participants made significantly more visual responses (*M*=8.5%, *SE*=1.0%) than auditory responses (*M*=4.5%, *SE*=1.0%)**,** *t*(23)=2.96, *p*=.007 (see Figure A.7C) on bimodal target trials, a clear demonstration of the Colavita effect.

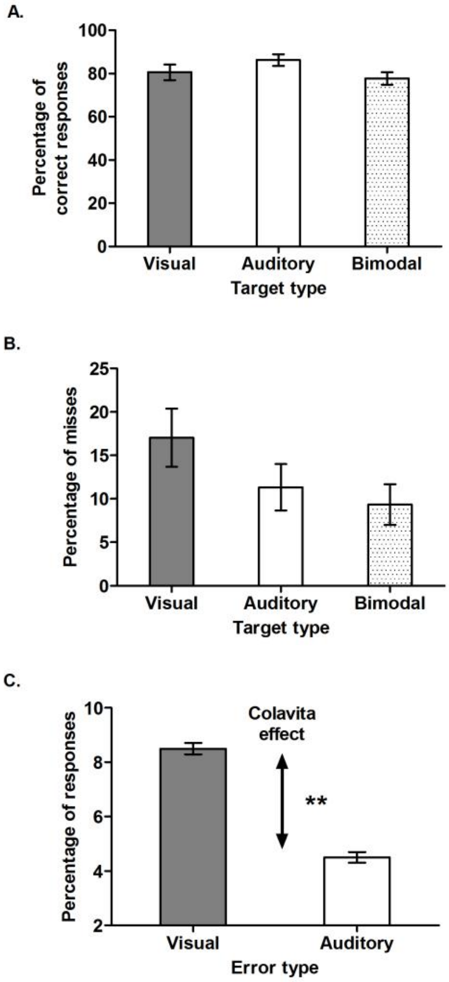


Figure A.7. A) Mean percentage of correct responses for the visual, auditory, and bimodal repetition trials in Experiment A.5. B) Mean percentage of visual, auditory, and bimodal misses. C) Mean percentage of visual versus auditory only responses (i.e., errors) made on bimodal repetition trials.

The results of Experiment A.5 demonstrate that when detecting immediate repetitions of bimodal targets, participants made significantly more visual-only than auditory-only responses. Thus, despite the SOA between targets being as long as 1000 ms, the Colavita visual dominance effect was nevertheless demonstrated. This further suggests that the presentation of an intervening stimulus, be it semantically meaningful or not, rather than temporal processing limits, is the key factor in determining whether or not visual dominance will be observed.

A.6. General Discussion

The results of the experiments reported in Appendix A suggest that the basic mechanisms underlying the Colavita visual dominance effect may not be limited to the detection of targets defined by their low-level sensory attributes/features, but rather they extend to higher-level representations which result from processing based on abstract task rules. Moreover, the results of these experiments demonstrate that under the appropriate experimental conditions and task demands, the Colavita visual dominance effect can be reversed.

In all of the experiments reported here, the participants detected auditory repetitions (both immediate and non-adjacent) more accurately than visual or bimodal repetitions. Importantly, however, the participants in Experiments A.2 and A.3 also made significantly more unimodal auditory than unimodal visual responses to bimodal targets, thus demonstrating a clear reversal of the Colavita visual dominance effect in favour of auditory dominance. While few studies of the Colavita visual dominance effect have successfully eliminated the effect (Koppen & Spence, 2007b; Moro & Steeves, 2010; Sinnett et al., 2007), no study up until now has been able to reverse the effect and show auditory dominance instead. In fact, as far as I am aware, this is the first reported reversal of the Colavita visual dominance effect in more than three decades since the phenomenon was reported in the scientific literature!

Based on accepted constraints of the duration of iconic memory (which lasts for less than 1000 ms; see Sperling, 1960) and echoic memory (which has been estimated to last for upwards of to 2000 ms; see Cowan, 1984), it was predicted that participants would have been biased toward responding to the auditory rather than the visual targets as a default. Such a pattern of responding was thought to arise from the fact that the amount of time needed for the comparison between potential visual n-1 targets in Experiment A.2 exceeded the temporal limit of the iconic memory store. While the results of Experiment A.2 corroborated this prediction, with participants making significantly more auditory than visual responses to bimodal targets, the results of Experiments A.3 and A.4 suggest that the temporal limitations of sensory memory were not crucial in terms of modulating the Colavita effect.

The results of Experiment A.3 demonstrated that shortening the SOA between target repetitions, so that detecting a repetition would be well-within the bounds of iconic memory, not only failed to reinstate the Colavita visual dominance effect, but it reproduced the auditory dominance first demonstrated in Experiment A.2. Thus, it was suspected that the pattern of results might depend on the interference by the meaningful, intervening event separating target repetitions. Because each stimulus presentation in Experiment A.3 could have been a potential n-1 repetition target, it was suspected that visual masking may have played a role, such that the presentation of any semantically meaningful intervening stimulus would have given rise to competition for the higher-level processing resources that are recruited during object recognition (Enns & Di Lollo, 2000; Loftus & Ginn, 1984).

The results of Experiment A.4 demonstrated that presenting semantically meaningless instead of meaningful intervening stimuli gave rise to similar percentages of visual and auditory responses to bimodal targets. Thus, the Colavita visual dominance effect was successfully eliminated, but not reinstated (as it had been in Experiment A.1). Importantly, auditory dominance was no longer observed in Experiment A.4 (as it had been in Experiments A.2 and A.3), suggesting that changing the semantically meaningful intervening stimulus to one that carried no semantic meaning did give rise to the elimination of auditory dominance that was expected to be observed. Still, it appeared that some interference/backward masking from the pattern mask, even though it carried no semantic meaning in either the visual or auditory modality, was present, given that the visual dominance effect was not reinstated as had been predicted.

Loftus and Ginn (1984) demonstrated that while semantically meaningful (e.g., picture) masks produced strong masking effects, pattern masks gave rise to significant (albeit weaker) masking effects as well. In their study, the participants had to recall as many details from a previously-presented naturalistic picture as possible. In one condition, the target picture was masked by either a different naturalistic picture or a random noise pattern mask after a delay of 300 ms. Both the presentation of the picture and pattern mask led to participants recalling significantly fewer details from the target (M=1.86 and M=2.21 details recalled, respectively) as compared to the no-mask condition (M=2.35). Importantly, the picture mask produced a greater magnitude of interference than the pattern mask. The results of Experiments A.3 and A.4 are therefore consistent with Loftus and Ginn's findings in demonstrating a stronger masking effect in the presence of meaningful intervening stimuli as compared to the meaningless pattern mask. When removing the intervening stimulus completely in Experiment A.5, effectively making it an immediate rather than n-1 repetition detection task (just like Experiment A.1, except with a longer SOA), participants made more visual rather than auditory responses to bimodal targets. Thus, the Colavita visual dominance effect was reinstated despite the long SOA between target repetitions.

Taken together, the results of the experiments reported in this Appendix suggest that the Colavita effect can be modulated by the potential masking effects that may differentially affect the retention of visual and auditory information. Introducing an intervening stimulus to the repetition detection task seems to have effectively made the task more difficult overall. More importantly, the interference from the intervening stimulus seems to have had a more pronounced effect on vision than on audition. Not only were participants better at detecting unimodal auditory as compared to unimodal visual targets, but they also made more auditory only rather than visual only responses when it came to bimodal n-1 targets. Thus, the memory of the visual n-1 stimulus seems to have been wiped out by the intervening visual stimulus, much more than the extent to which memory of the auditory n-1 target seems to have been by the intervening auditory stimulus. In other words, the new (intervening) visual stimulus was dominant enough to wipe out the previous visual stimulus, but this effect did not occur at the same proportions in the auditory domain – yet another demonstration of just how dominant the visual system really is!

I suspect that the differential masking effects of the intervening stimuli on vision and audition may be due to the fact that masking has a different time course in vision (see Enns & Di Lollo, 2000, for a review) and audition (Elliot, 1971; Oxenham & Moore, 1994; Oxenham & Wojtczak, 2010). For both vision and audition, forward masking is much more sensitive to the introduction of any temporal delay between the target and mask as compared to backward masking. Forward masking seems to occur in both vision and audition when the mask is presented within 50 ms of the target’s onset. In vision, the disruptive effects of backward masking on the low-level iconic representation of a target are present up to 50 ms after the offset of the target (Loftus & Ginn, 1984). After 300 ms of the offset of the target, Loftus and Ginn suggest that conceptual masking occurs, whereby identification of the mask *"interrupts the higher level processing that is required for long-term storage of the information corresponding to the picture [the target repetition, in our case]"* (p. 435). In audition, low-level backward masking appears to be effective within 200 ms of target offset (Massaro, Cohen, & Idson, 1976). Unlike in vision, however, Massaro and Burke (1991) suggest that auditory masks do not work retroactively as conceptual masks. That is, they do not disrupt the processing of information that might have been obtained before the presentation of the mask.

It would appear then that visual masks are more effective than auditory masks at producing substantial conceptual interference effects, especially at longer SOAs. This notion seems to hold true in light of the findings of the present experiments. Here, not only was detecting n-1 repetitions in the visual modality more difficult than in the auditory modality, but also auditory-only responses dominated over visual-only responses on bimodal trials. The masking effects appear to have given rise to a sufficient degradation in the reliability of the information presented in the visual but not the auditory modality, which has been shown to be necessary for eliminating the Colavita visual dominance effect (Moro & Steeves, 2010).

The take home message, based on the results of the five experiments reported in this Appendix, is that the pattern of visual or auditory dominance can be modulated by the presence of a mask (conceptual, pattern, or absent) in the repetition detection task. That is, in the presence of a conceptual mask, auditory dominance prevails; in the presence of a pattern mask, auditory dominance disappears and vision begins to take the lead; in the absence of any mask, visual dominance is restored on bimodal target trials. Most notably, the fact that the visual system had to be pushed to its upper limit to reverse the Colavita effect and that, with just a little less favouring of the auditory modality, visual dominance was still seen is quite impressive. Moreover, even though unimodal visual target detection was worse than unimodal auditory target detection, vision still dominated over audition in bimodal target trials. These results go to show just how resilient and robust the Colavita visual dominance effect really is!

REFERENCES

Bertelson, P., Vroomen, J., De Gelder, B., & Driver, J. (2000). The ventriloquist effect does not depend on the direction of deliberate visual attention. *Perception & Psychophysics*, **62**, 321-332.

Chen, K.-M., & Yeh, S.-L. (2009). Asymmetric cross-modal effects in time perception. *Acta Psychologica*, **130**, 225-234.

Colavita, F. B. (1974). Human sensory dominance. *Perception & Psychophysics,* **16**, 409-412.

Cowan, N. (1984). On short and long auditory stores. *Psychological Bulletin*, **96**, 341-370.

Di Lollo, V. (1977). Temporal characteristics of iconic memory. *Nature*, **267**, 241-243.

Elliot, L. L. (1971). Backward and forward masking. *Audiology,* **10**, 65-76.

Enns, J. T., & Di Lollo, V. (2000). What's new in visual masking? *Trends in Cognitive Sciences*, **4**, 345-352.

Fiebelkorn, I. C., Foxe, J. J., Butler, J. S., & Molholm, S. (2011). Auditory facilitation of visual-target detection persists regardless of retinal eccentricity and despite wide audiovisual misalignments. *Experimental Brain Research*, **213**, 167-174.

Forster, B., Cavina-Pratesi, C., Aglioti, S. M., & Berlucchi, G. (2002). Redundant target effect and intersensory facilitation from visual-tactile interactions in simple reaction time. *Experimental Brain Research,* **143**, 480-487.

Gebhard, J. W., & Mowbray, G. H. (1959). On discriminating the rate of visual flicker and auditory flutter. *American Journal of Psychology,* **72**, 521-528.

Giard, M.-H., & Peronnet, F. (1999). Auditory-visual integration during multimodal object recognition in humans: A behavioral and electrophysiological study. *Journal of Cognitive Neuroscience*, **11**, 473-490.

Hein, G., Parr, A., & Duncan, J. (2006). Within-modality and cross-modality attentional blinks in a simple discrimination task. *Perception & Psychophysics*, **68**, 54-61.

Howard, I. P., & Templeton, W. B. (1966). *Human spatial orientation*. London: Wiley.

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

Kanwisher, N. G. (1987). Repetition blindness: Type recognition without token individuation. *Cognition*, **27**, 117-143.

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

Koppen, C., Levitan, C., & Spence, C. (2009). A signal detection study of the Colavita effect. *Experimental Brain Research*, **196**, 353-360.

Koppen, C., & Spence, C. (2007a). Assessing the role of stimulus probability on the Colavita visual dominance effect. *Neuroscience Letters*, **418**, 266-271.

Koppen, C., & Spence, C. (2007b). Audiovisual asynchrony modulates the Colavita visual dominance effect. *Brain Research*, **1186**, 224-232.

Koppen, C., & Spence, C. (2007c). Seeing the light: Exploring the Colavita visual dominance effect. *Experimental Brain Research*, **180**, 737-754.

Koppen, C., & Spence, C. (2007d). Spatial coincidence modulates the Colavita visual dominance effect. *Neuroscience Letters*, **417**, 107-111.

Lessard, N., Pare, M., Lepore, F., & Lassonde, M. (1998). Early-blind human subjects localize sound sources better than sighted subjects. *Nature*, **395**, 278-280.

Loftus, G. R., & Ginn, M. (1984). Perceptual and conceptual masking of pictures. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **10**, 435-441.

Long, G. M. (1980). Iconic memory: A review and critique of the study of short-term visual storage. *Psychological Bulletin*, **88**, 785-820.

Massaro, D. W., & Burke, D. (1991). Perceptual development and auditory backward recognition masking. *Developmental Psychology*, **27**, 85-96.

Massaro, D. W., Cohen, M. M., & Idson, W. L. (1976). Recognition masking of auditory lateralization and pitch judgments. *Journal of the Acoustical Society of America*, **59**, 434-441.

McKeeff, T. J., Remus, D. A., & Tong, F. (2007). Temporal limitations in object processing across the human ventral visual pathway. *Journal of Neurophysiology*, **98**, 382-393.

Molholm, S., Ritter, W., Javitt, D. C., & Foxe, J. J. (2004). Multisensory visual-auditory object recognition in humans: A high-density electrical mapping study. *Cerebral Cortex*, **14**, 452-465.

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

Moro, S. S., & Steeves, J. K. E. (2010, June). An absence of the Colavita effect: Enhanced multisensory processing in monocular blindness. Poster presented at the *11th International Multisensory Research Forum*, 16-19th June, Liverpool, UK.

Nickerson, R. S. (1973). Intersensory facilitation of reaction time: Energy summation or preparation enhancement. *Psychological Review*, **80**, 489-509.

O'Connor, N., & Hermelin, B. (1972). Seeing and hearing in space and time. *Perception & Psychophysics*, **11**, 46-48.

Oxenham, A. J., & Moore, B. C. J. (1994). Modeling the additivity of nonsimultaneous masking. *Journal of the Acoustical Society of America*, **98**, 1921-1934.

Oxenham, A. J., & Wojtczak, M. (2010). Frequency selectivity and masking. In C. J. Plack (Ed.), *The Oxford handbook of auditory science* (pp. 5-44). Oxford: Oxford University Press.

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

Potter, M. C. (1999). Understanding sentences and scenes: The role of conceptual short-term memory. In V. Coltheart (Ed.), *Fleeting memories: Cognition of brief visual stimuli* (pp. 13-46). Cambridge, MA: MIT Press.

Rees, G., Frith, C. D., & Lavie, N. (2001). Processing of irrelevant visual motion during performance of an auditory task. *Neuropsychologia*, **39**, 937-949.

Shams, L., Kamitani, Y., & Shimojo, S. (2000). What you see is what you hear. *Nature*, **408**, 788.

Shimojo, S., & Shams, L. (2001). Sensory modalities are not separate modalities: Plasticity and interactions. *Current Opinion in Neurobiology*, **11**, 505-509.

Shipley, T. (1964). Auditory flutter-driving of visual flicker. *Science*, **145**, 1328-1330.

Sinnett, S., Soto-Faraco, S., & Spence, S. (2008). The co-occurrence of multisensory competition and facilitation. *Acta Psychologica*, **128**, 153-161.

Sinnett, S., Spence, C., & Soto-Faraco, S. (2007). Visual dominance and attention: The Colavita effect revisited. *Perception & Psychophysics*, **69**, 673-686.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, **6**, 174-215.

Soto-Faraco, S., & Spence, C. (2002). Modality-specific auditory and visual temporal processing deficits. *Quarterly Journal of Experimental Psychology*, **55A**, 23-40.

Soto-Faraco, S., Spence, C., & Kingstone, A. (2004). Cross-modal dynamic capture: Congruency effects in the perception of motion across sensory modalities. *Journal of Experimental Psychology: Human Perception and Performance*, **30**, 330-345.

Spence, C. (2009). Explaining the Colavita visual dominance effect. *Progress in Brain Research*, **176**, 245-258.

Spence, C., Parise, C., & Chen, Y.-C. (2011). The Colavita visual dominance effect. In M. M. Murray & M. Wallace (Eds.), *Frontiers in the neural bases of multisensory processes* (pp. 523-550). Boca Raton, FL: CRC Press.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied*, **74**, 1-29.

Stein, B. E., & Meredith, M. A. (1993). *The merging of the senses.* Cambridge, MA: MIT Press.

Vroomen, J., & de Gelder, B. (2000). Sound enhances visual perception: Cross-modal effects of auditory organization on vision. *Journal of Experimental Psychology: Human Perception and Performance*, **26**, 1583-1590.

Walker, J. T., & Scott, K. J. (1981). Auditory-visual conflicts in the perceived duration of lights, tones, and gaps. *Journal of Experimental Psychology: Human Perception and Performance*, **7**, 1327-1339.

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

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

Welch, R. B., & Warren, D. H. (1986). Intersensory interactions. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance, Vol. I. Sensory processes in perception* (pp. 1-36). New York: Wiley.

APPENDIX B:

CROSSMODAL TASTE/FLAVOUR ASSOCIATIONS

B.0. Introduction

Gallace, Boschin, and Spence (2011) recently demonstrated that people are more likely to match certain words to the tastes, textures, and flavours of various food and drink items. Their research was motivated by the extensive literature on the topic of sound (or phonetic) symbolism (see Köhler, 1929; Sapir, 1929, for early research on this topic, and Spence, 2011, for a recent review), defined as "the direct linkage between sound and meaning" (Hinton, Nicholls, & Ohala, 1994). Numerous studies have highlighted the fact that people spontaneously associate certain speech sounds with specific visually-presented shapes (e.g., Boyle & Tarte, 1980; Ramachandran & Hubbard, 2001). So, for example, nonsense words such as 'takete' and 'kiki' tend to be associated with angular shapes while nonsense words such as 'maluma' or 'bouba' tend to be associated with rounded 'cloudlike' shapes instead[14]. Interestingly, people all over the world appear to exhibit the same crossmodal correspondences (see Hinton et al., 1994, for a review). Such crossmodal correspondences appear to emerge very early in

---

[14] Here, the nonsense words containing plosive stops or stop consonants, such as 'k', 'p', and 't', which involve the complete closure of the teeth, tongue, and/or lips, followed by a sudden release of air from the mouth are described as 'sharp' speech sounds. By contrast, the term 'rounded' is used to describe those nonsense words containing nasal stops, such as 'm' or 'n', that involve the release of air from the nose rather than the mouth (Klink, 2000; Spence, 2011). In addition to the distinction between consonant sounds, vowel sounds can be distinguished by their frontal or back manner of articulation (Yorkston & Menon, 2004). The 'e' in 'takete' elicits a frontal vowel sound, where the tongue is in the front of the mouth when the word is pronounced. In contrast, the 'u' in 'maluma' elicits a back vowel sound, where the tongue is in the back of the mouth instead. Note that much more research needs to be conducted in order to fine-tune exactly which elements of these nonsense words (i.e., the consonants, the vowels, or their combination) is key in determining the crossmodal correspondences that are chosen for certain foodstuffs.

human development (i.e., within a few months of birth; e.g., Maurer, Pathman, & Mondloch, 2006; Walker et al., 2010).

The majority of the sound symbolism research that has been published to date has focused on the nature of the crossmodal correspondences that exist between speech sounds and the attributes of objects presented visually (see Hinton et al., 1994, p. 4). That said, an anecdotal report by Fónagy (1963, 2001) nearly half a century ago suggested that there might be a crossmodal correspondence between foods on the bitter-sweet continuum and front/back vowel sounds (e.g., an example of a frontal vowel sound is the ‘i’ sound in the word ‘hit’, whereas a back vowel sound would be the ‘o’ in ‘home’; Ladefoged, 1993). Suggestive evidence that the phenomenon of sound symbolism might extend to the case of crossmodal associations between sounds and tastes/flavours comes from marketing research conducted by Yorkston and Menon (2004). They demonstrated that people are more likely to believe that an ice cream will have a creamy taste if it is called ‘Frosch’ than if it is called ‘Frisch’ (see also Klink, 2000). It is important to note, though, that the participants in Yorkston and Menon’s study never actually tasted anything (they only read a short textual description about a fictional product). Thus, to date, there has been virtually no research that has specifically attempted to investigate whether certain phonological stimuli (i.e., speech sounds) also bear a non-arbitrary relationship to non-visual stimulus attributes, be they modality-nonspecific (i.e., amodal), such as an object’s shape, or modality-specific (i.e., modal), such as a food or drink’s bitterness.

One recent exception to this generalization comes from a study by Gallace et al. (2011). They demonstrated that people reliably associate nonsense words (e.g., ‘maluma’ and ‘takete’, as originally popularized by Köhler, 1947; and ‘bouba’ and ‘kiki’, as popularized by Ramachandran & Hubbard, 2001) with specific tastes/flavours/foodstuffs. The participants in their study were given a range of up to ten different foods, including jam, chocolate, crisps, and yoghurt. They had to rate each foodstuff on 24 different visual analogue linear scales with a pair of words such as good/bad, salty/sweet, etc, at the end-points. Amongst these scales (presented via computer) were three scales designed to test whether the phenomenon of sound symbolism extends to the flavour and/or oral-somatosensory (i.e., textural) attributes of foodstuffs, namely: bouba/kiki, takete/maluma, decter/bobolo. The results showed that people consistently rated salt and vinegar crisps (potato chips) as much more kiki (or takete) than cheddar cheese, yoghurt, or blueberry jam. Meanwhile, chocolate with mint chips and crisps (potato chips) were rated as significantly more kiki/takete than regular chocolate.

The goal of Experiment B.1 was to test whether certain meaningless speech sounds and shapes are associated with the taste/flavour/oral-somatosensory attributes of chocolate samples varying primarily in their cocoa content. Here, a number of simple demonstrations highlighting the fact that people really do reliably link certain attributes of shapes and nonsense words to chocolate samples of varying cocoa content are reported. Furthermore, it is demonstrated that these crossmodal correspondences can be observed in relatively small groups of participants using nothing more than a quick and easy-to-administer pencil-and-paper task.

B.1. Experiment B.1

In Experiment B.1, participants evaluated three pieces of chocolate: One piece of 30% cocoa milk chocolate and two pieces of dark chocolate (one with a cocoa content of 70% and the other made with 90% cocoa). The question to be addressed in this first experiment was whether there is a systematic crossmodal correspondence between the cocoa content of chocolate and the angularity of shapes and the sharpness of words.

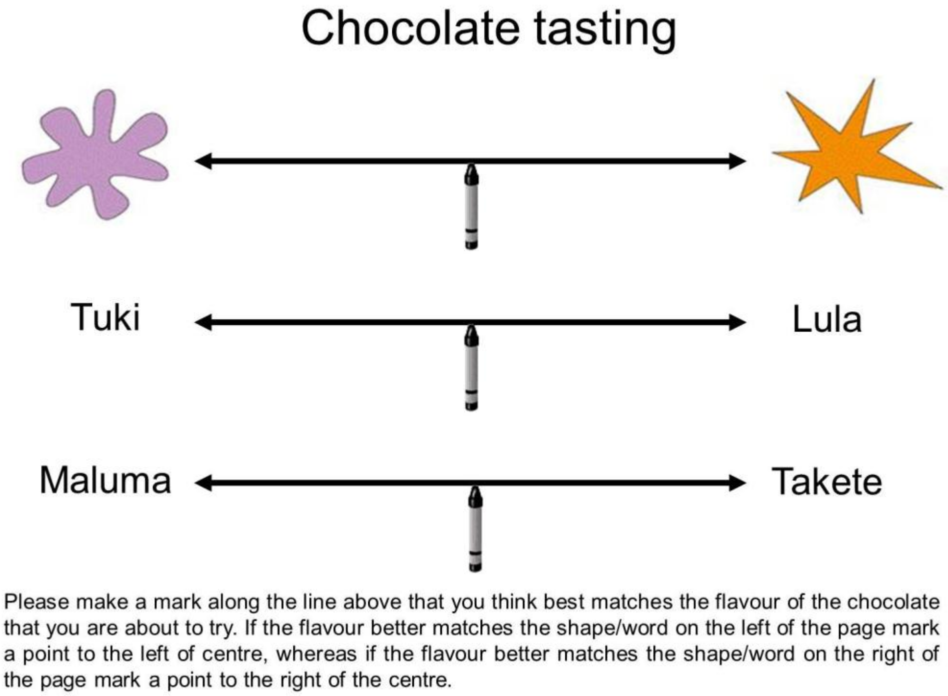
B.1.1. Methods

Twenty-two participants (ranging in age from 16-52 years; mean age of 36 years) were given paper-based analogue labelled line scales on which to score each of the chocolate samples (see Figure B.1 for the scales and labels used). The first scale was anchored with a rounded shape (on the left) and an angular shape (on the right). The second scale was anchored with the words “Tuki” and “Lula”, while the third scale was anchored with the words “Maluma” and “Takete”.[15] Each line scale was 13.5 cm long with an image of a crayon centered on its mid-point. The participants were given a set of three response sheets, each with the same three scales printed on them. They were instructed to close their eyes while a small piece of chocolate (1/4 of a chocolate square, approximately 0.11 ounces) was placed in their hand. The participants tasted the sample and then rated it by marking a line on each of the three response scales on the first response sheet in order to indicate what they thought best matched the flavour. This procedure was repeated for the other two chocolate samples. Thus, in total, each participant tasted three pieces of chocolate (all Lindt Excellence brand), consisting of one

---

[15] Note that the first and third scales were both anchored with the rounded shape / word on the left and the sharper, more angular shape / word on the right. This arrangement was reversed for the second scale, so as to prevent the participants from simply associating one side of the scales (or response sheet) with a particular quality/attribute.

piece of extra creamy milk (30% cocoa), one piece of 70% cocoa, and one piece of 90% cocoa chocolate. The order in which the chocolate samples were tasted was counterbalanced across participants.



Please make a mark along the line above that you think best matches the flavour of the chocolate that you are about to try. If the flavour better matches the shape/word on the left of the page mark a point to the left of centre, whereas if the flavour better matches the shape/word on the right of the page mark a point to the right of the centre.

Figure B.1. The response sheet (including instructions) given to participants. The three line scales highlighted in the figure were used in Experiments B.1 and B.2. In Experiment B.3, the left anchor 'Tuki' was replaced by the word 'Koko' (the name of the chocolate product that participants actually happened to be rating).

B.1.2. Results and Discussion

The participants' responses were measured using a ruler. The mid-point on each scale, shown by the crayon in Figure B.1, was assigned a value of 0. For the first and third scales, participants' responses to the left of the mid-point were assigned negative values, while responses to the right of the crayon were given positive values. The coding of participants' responses was reversed for the second scale. Consequently, negative values indicated more rounded / softer responses, regardless of the specific scale used. Participants' responses for each chocolate sample on each of the three response scales (see Figure B.2) were compared using a 3 (chocolate) x 3 (scale) within-participants ANOVA. The analysis revealed significant main effects of chocolate, *F*(2,42)=18.07, *p*<.001, and scale, *F*(2,42)=3.71, *p*=.03. Participants' ratings of the extra creamy milk chocolate (*M*=-2.89 cm) were significantly different from their ratings of the 70% (*M*=0.98 cm) and 90% dark chocolates (*M*=2.18 cm), *p*<.001 for both comparisons. However, their ratings did not differ significantly between the 70% and 90% dark chocolate samples, *p*=.21. Participants' overall ratings differed significantly between the round vs. angular shape (*M*=-.35 cm) and Lula vs. Tuki scales (*M*=.79 cm), *p*=.04, and between the Lula vs. Tuki and Maluma vs. Takete scales (*M*=-.17 cm), *p*=.007. The interaction between chocolate and scale was not significant, *F*<1 (see Figure B.2).

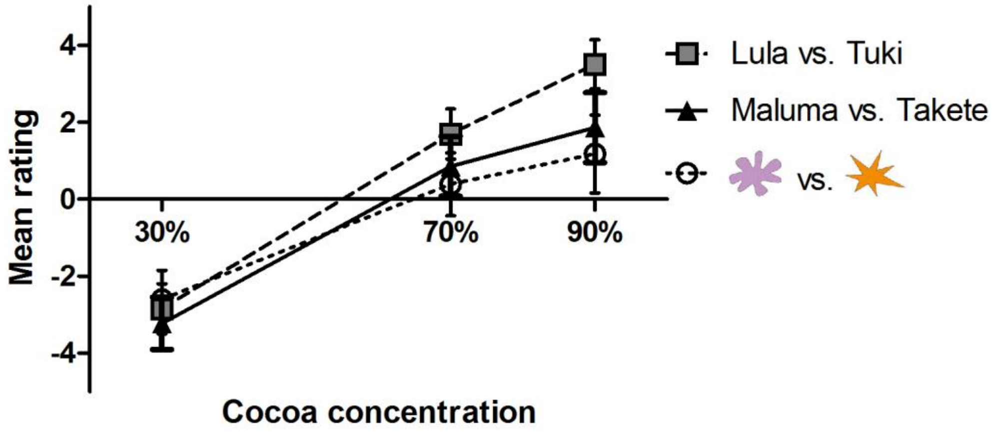


Figure B.2. Participants' mean ratings (in cm) for each chocolate and scale type in Experiment B.1. Negative values correspond to the round shape and words (Lula and Maluma) while positive values correspond to the angular/sharp shape and words (Tuki and Takete). Error bars represent the standard errors of the means.

The results of Experiment B.1 highlight the existence of a systematic crossmodal correspondence between the cocoa content of chocolate and roundedness/angularity, as assessed by participants' responses when given seemingly meaningless visual response scales anchored by either angular versus rounded shapes or by pairs of nonsense words varying in the sharpness of their inflections. Participants matched the 30% cocoa milk chocolate sample with the rounder shape and words (i.e., Lula and Maluma), whereas the 70% and 90% cocoa dark chocolates were matched with the more angular shape and sharper speech sounds (i.e., Tuki and Takete) instead. The perceived bitterness of the chocolate samples used in the present experiment was dependent on the cocoa concentration. This is presumably the key sensory attribute of the chocolate on which participants were basing the correspondence between the chocolate taste and shape/word scale — specifically, the darker the chocolate, the more angular its sound/visual symbolism.

Visual inspection of participants' mean ratings (look again at Figure B.2) might lead one to suggest that the second scale (Tuki vs. Lula) was somewhat more sensitive to any variation in the cocoa content of the chocolates as compared to the other two scales. However, the fact that there was no significant interaction between the chocolate sample and scale factors suggests that all three scales were similarly sensitive to the varying cocoa concentrations present in the sample tested. Having demonstrated a reliable crossmodal correspondence between the angularity of shapes/words and the three types of solid chocolates, Experiment B.2 went on to explore how milk chocolate truffles would be rated using exactly the same set of three response scales.

B.2. Experiment B.2

B.2.1. Methods

In Experiment B.2, participants' ratings of a milk chocolate truffle (Cadbury's Koko brand) were measured using the same procedure as in Experiment B.1. Importantly, the participants were not given any information about the chocolate that they were going to taste (such as, for example, its name). The wrapping encasing each chocolate truffle was also removed prior to presenting it to the participants. Nineteen of the 22 participants who had taken part in Experiment B.1 also took part in this experiment.

B.2.2. Results and Discussion

Just as in Experiment B.1, negative values indicated more rounded / softer responses regardless of the specific response scale used. Participants' ratings for each scale (see Figure B.3A) were subjected to a one-way ANOVA. In contrast to the results of Experiment B.1, this analysis revealed no significant effect of response scale on participants' ratings, $F<1$. Separate one-sample t-tests with the test value set to 0 (corresponding to the mid-point on each scale) were conducted for each scale. Participants' mean ratings for each of the three scales (-3.10 cm for Round vs. Angular, -3.07 cm for Lula vs. Tuki, and -3.04 cm for Maluma vs. Takete) were all significantly different from 0, $t(18)=-4.24$, $p<.001$, $t(18)=-4.19$, $p=.001$, $t(18)=-4.56$, $p<.001$, respectively.

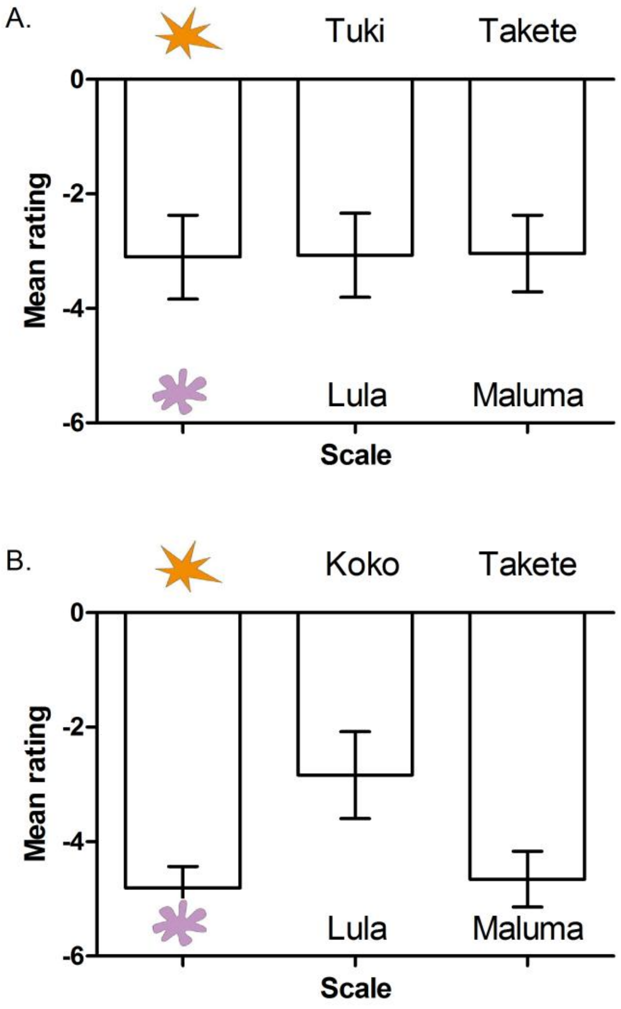


Figure B.3. A) Participants' mean ratings (in cm) for Cadbury's Koko truffles on the Round vs. Angular shape, Lula vs. Tuki, and Maluma vs. Takete scales in Experiment B.2; B) Participants' mean ratings (in cm) for Cadbury's Koko truffles on the Round vs. Angular shape, Lula vs. Koko, and Maluma vs. Takete scales in Experiment B.3. Error bars represent the standard errors of the means.

The correlation between each of the response scales was tested using Pearson's correlation coefficient, *r*. This analysis revealed significant positive correlations among all three scales. There was a significant correlation between the Shape and Lula scales, r=.56, p=.01, between the Shape and Maluma scales, r=.86, p<.001, and between the Lula and Maluma scales, r=.68, p<.001. Given that all three scales were highly correlated with each other, this suggests that they all may be tapping into the same underlying crossmodal association.

The results of Experiment B.2 demonstrated that participants clearly matched the Koko milk chocolate truffles to the round shape and rounded sounds (of Lula and Maluma) rather than to the angular shape and sharp sounds (of Tuki and Takete). In Experiment B.3, the goal was to investigate whether or not people would identify the truffle with the sound of its brand name under conditions where they were unaware that one of the anchors on the scales was actually the name of the product being tested. To this end, the Tuki anchor in the second scale was replaced by the word “Koko”.

B.3. Experiment B.3

B.3.1. Methods

In Experiment B.3, a new group of 21 naïve participants (age range = 21-37 years; mean age of 28 years; none of whom had taken part in either of the two preceding experiments) rated a Cadbury’s Koko truffle using nearly identical scales and exactly the same procedure as in the previous experiment. The only change to the scales was that the word “Tuki” was replaced with the word “Koko”. Once again, the participants were unaware of the fact that the anchor “Koko” was actually the name of the chocolate that they were tasting. Once the experiment had been completed, participants were told that the actual brand name of the chocolate was in fact “Koko” and verbal debriefing by the experimenter confirmed that the participants had indeed been naïve to the truffle’s brand name when rating the chocolate.

B.3.2. Results and Discussion

Participants' ratings for each scale were subjected to a one-way ANOVA. This analysis revealed a significant main effect of scale, *F*(2,40)=5.34, *p*=.03. Participants' ratings differed significantly between the Shape (*M*=-4.81 cm) and Lula (*M*=-2.84 cm) scales, *p*=.02, and between the Lula and Maluma (*M*=-4.66 cm) scales, *p*=.04. Once again, participants' responses for each scale (see Figure B.3B) were subjected to separate one-sample t-tests, with the test value set to 0. These analyses revealed that the mean responses of participants for all three scales were significantly different from 0 (-4.81 cm for Round vs. Angular, *t*(20)=-12.73, *p*<.001; -2.84 cm for Koko vs. Lula, *t*(20)=-3.73, *p*=.001; and -4.66 cm for Maluma vs. Takete, *t*(20)=-9.55, *p*<.001). The correlation between each of the three response scales was again tested using Pearson's correlation coefficient, r. These analyses highlighted a strong, significant positive correlation between the Shape and Maluma scales, *r* =-.85, *p*<.001. In contrast to the results of the preceding experiment, however, no other correlations between scales were observed.

The results of Experiment B.3 demonstrate that participants matched the Koko brand milk chocolate truffle with the round shape and round-sounding words rather than with the angular shape and sharply-inflected sounds when they were naïve as to the name of the product that they were tasting. Importantly, even though the word "Koko" (the actual name of the chocolate) was included as an anchor for one of the scales (vs. Lula as the other anchor), participants still rated the truffle as being closer to the “Lula” end of the scale[16]. This result could be taken to imply that the choice of the brand name by Cadbury’s may have been inappropriate. At the very least, it is inconsistent with the sound symbolic crossmodal association that people tend to exhibit for milk chocolate truffles. It could, of course, be countered that many brand names incorporate a semantic component (see Klink, 2001), and to the extent that the term ‘Koko’ reminds a consumer of cocoa, at least in the context where they are expecting to purchase/eat a chocolate then the sound symbolic aspects of the brand name may be less important than in the present study, where the name was presented in a context where the symbolic content of the sound was all-important. Further research will be needed in order to determine whether any semantic aspects of a brand name modulate the lower level sound symbolic effects.

B.4. Experiment B.4

While the taste/flavour of the chocolates sampled in Experiments B.1 through B.3 presumably drove participants’ shape and sound symbolic correspondences, it is possible that the texture of the chocolate samples played an important role in driving the crossmodal correspondences as well. The goal of the Experiment B.4 was therefore to further examine the contribution of taste and, more specifically, oral-somatosensory texture to sound and shape symbolism. In particular, four types of chocolates were tested: one milk, one dark, and two mint chocolates. Of particular relevance in terms of the goals of Experiment B.4, the two mint chocolates were

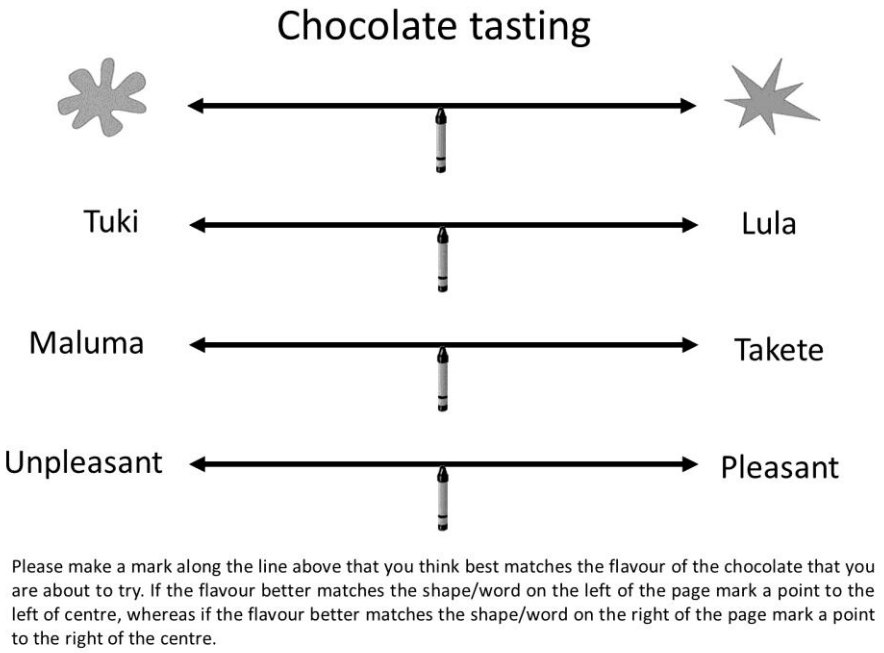
---

[16] A between-experiments (Experiment B.2 vs. B.3) t-test was conducted on participants’ ratings using the second scale revealing no significant difference between the two ratings scales (*M*=3.07 cm for Tuki in Experiment B.2 vs. 2.84 cm for Koko in Experiment B.3; *t*(38)=-.22, *p*=.83.

identical in their cocoa content and taste/flavour, but differed in the texture of the mint (solid vs. fondant). The question to be addressed in this experiment was whether there are any systematic crossmodal correspondences between the taste of chocolate (milk, dark, or mint), the texture of the chocolate (solid or fondant), and the angularity of shapes and the sharpness of words.

B.4.1. Methods

Forty-six participants (17 female; ranging in age from 26-36 years) from the Säid Business School in Oxford were given paper-based analogue labelled line scales on which to score each of the chocolate samples (see Figure B.4 for the scales and labels used). The first scale was anchored with a rounded shape (on the left) and an angular shape (on the right); the second scale was anchored with the words “Tuki” and “Lula”; the third scale was anchored with the words “Maluma” and “Takete”; the fourth and final scale was anchored with the words “Unpleasant” and “Pleasant”. The design and procedure was nearly identical to the of Experiment B.1, except that now participants tasted and evaluated four pieces of chocolate: One piece of Green & Black’s 34% cocoa milk chocolate, one piece of Green & Black’s 70% cocoa dark chocolate, one piece of Bendick’s 50% cocoa mint fondant chocolate, and one piece of Bendick’s 50% cocoa solid mint chocolate. Participants first tasted the mint fondant, followed by the dark, the mint solid, and then the milk chocolate.



Chocolate tasting

Tuki — Lula

Maluma — Takete

Unpleasant — Pleasant

Please make a mark along the line above that you think best matches the flavour of the chocolate that you are about to try. If the flavour better matches the shape/word on the left of the page mark a point to the left of centre, whereas if the flavour better matches the shape/word on the right of the page mark a point to the right of the centre.

Figure B.4. The response sheet (including instructions) given to participants. The four line scales shown in the figure were used in Experiments B.4 and B.5.

B.4.2. Results and Discussion

Participants' responses for each chocolate sample on each of the four response scales (see Figure B.4) were compared using a 4 (chocolate) x 4 (scale) within-participants ANOVA. The analysis revealed significant main effects of scale, $F(3,135)=4.80$, $p=.007$, and chocolate, $F(1,135)=17.99$, $p<.001$. The interaction between scale and chocolate was not significant, $F(9,405)=1.15$, $p=.33$ (see Figure B.5). Participants' ratings on the Unpleasant-Pleasant scale ($M=-1.40$ cm) differed significantly from the Round-Angular shape scale ($M=-.44$ cm, $p=.005$), the Lula-Tuki scale ($M=-.52$ cm, $p=.021$), and the Maluma-Takete scale ($M=-.54$ cm, $p=.014$). Participants' ratings of the milk chocolate ($M=-3.35$ cm) differed significantly from the mint fondant ($M=-.52$ cm, $p<.001$), dark ($M=.39$ cm, $p<.001$), and solid mint chocolates ($M=.58$ cm, $p<.001$). Importantly, participants' ratings also differed significantly between the mint fondant and solid mint chocolates, *p*=.049.

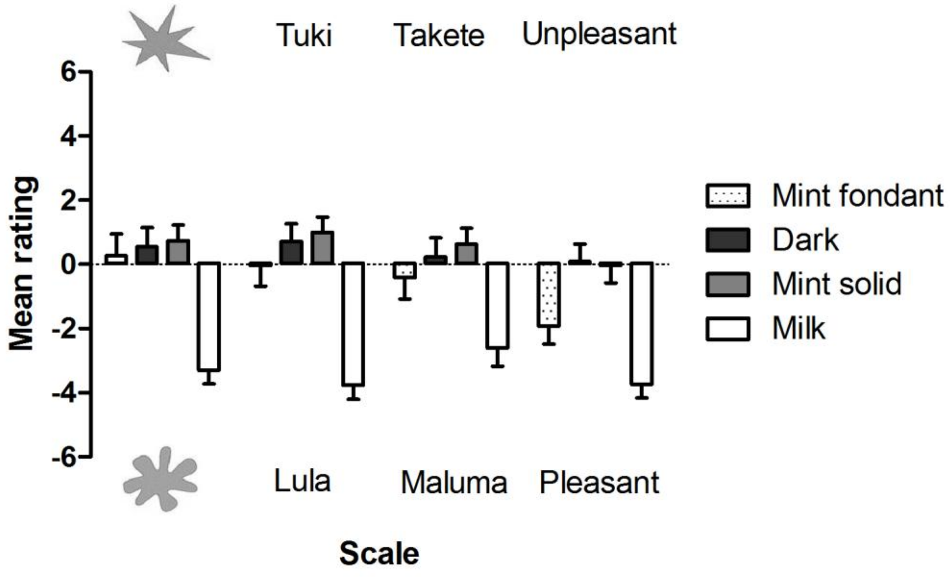


Figure B.5. Participants' mean ratings (in cm) for each chocolate and scale type in Experiment B.4. Negative values correspond to the rounded shape and words (Lula and Maluma), and Pleasant rating, while positive values correspond to the angular shape, sharp words (Tuki and Takete), and Unpleasant rating. Error bars represent the standard errors of the means.

One-sample t-tests were also conducted on each chocolate and scale in order to test whether participants' ratings differed significantly from the mid-point value of 0. The analyses revealed that participants rated the milk chocolate as being significantly round-shaped, *t*(45)=-7.60, *p*<.001, Maluma, *t*(45)=-4.58, *p*<.001, Lula, *t*(45)=-8.28, *p*<.001, and Pleasant, t(45)=-8.68, p<.001. Participants also rated the solid mint chocolate as being significantly Tuki, *t*(45)=2.04, *p*=.047, and the mint fondant as being significantly Pleasant, *t*(45)=-3.42, *p*=.001.

The results of Experiment B.4 demonstrate that milk chocolate was clearly associated with the round shape and rounder-sounding speech sounds (Lula and Maluma), and participants rated it as Pleasant tasting. Participants also rated the mint fondant chocolate as Pleasant, while the dark and solid mint chocolates were rated as neither Pleasant nor Unpleasant. It would appear, therefore, that for milk and mint fondant chocolates, most people would agree that they have a pleasant taste. On the other hand, there seems to be a lack of consensus regarding the pleasantness of the dark and solid mint chocolates (cf. Crisinel & Spence, submitted). Interestingly, while neither the mint fondant nor the dark chocolates were consistently matched with the angular shape or words, the solid mint was consistently matched with the nonsense word Tuki. Given that the main difference between the mint fondant and solid mint chocolate was in terms of the texture, this suggests that texture plays a key role in determining the crossmodal correspondences that the participants had for the chocolate samples used in Experiment B.4.

B.5. Experiment B.5

Having demonstrated the crossmodal correspondences among the chocolates varying in their texture and cocoa content, shapes, and speech sounds in a well-controlled setting, Experiment B.5 went on to investigate whether similar results could be obtained using an internet-based questionnaire. Here, respondents imagined the taste of the chocolates, and made their ratings online, without actually tasting them.

This approach was motivated by the fact that online data can reach larger populations that are more geographically and demographically diverse. Moreover, it is typically cheaper, faster, and easier to administer than laboratory-based studies (Lefever, Dal, & Matthiasdottir, 2007;

Piqueras-Fiszman, Ares, Alcaide, & Diego, 2011; Riva, Teruzzi, & Anolli, 2003). Nevertheless, there are also limitations associated with the use of online data collection that relate to the quality of the data that is acquired (e.g., Denscombe, 2009; Riva et al., 2003).

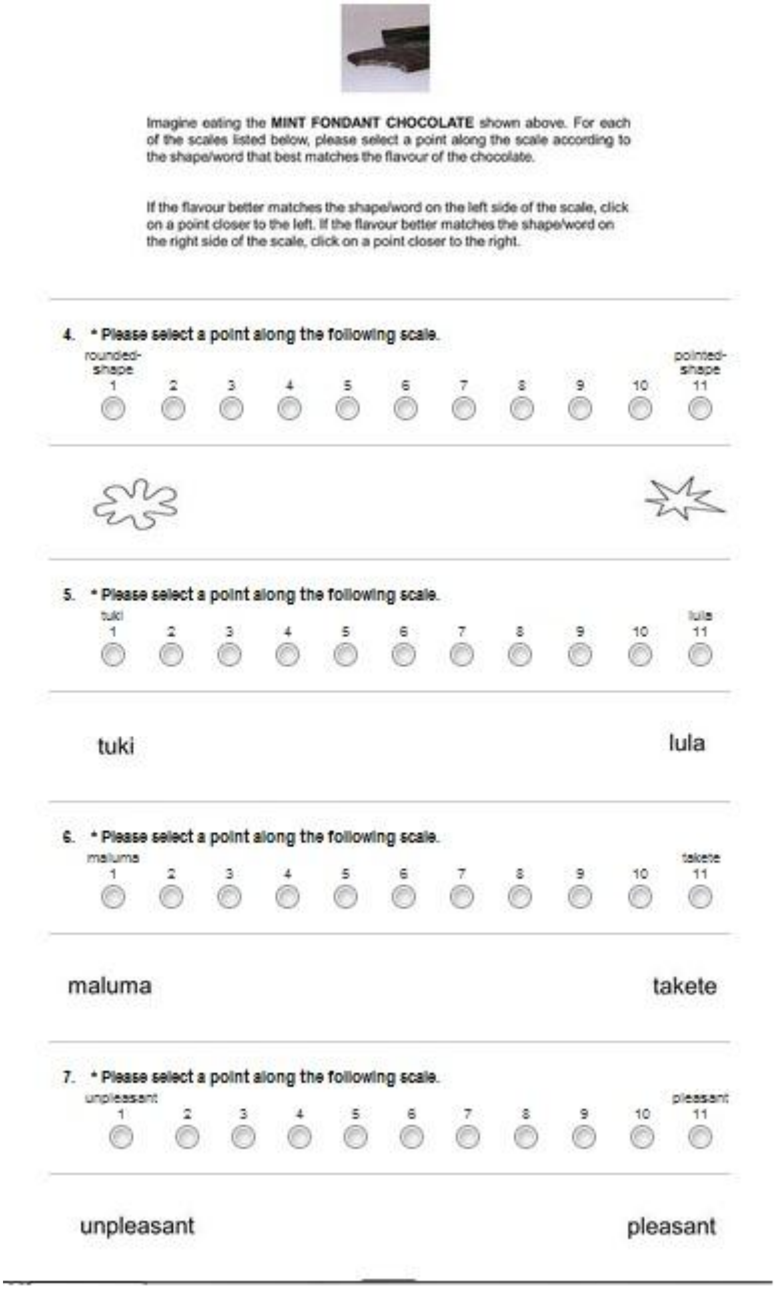
One of the goals of Experiment B.5 was therefore to exploit the advantages of online data collection in order to assess the nature of the crossmodal correspondences between shape and sound on the one hand, and the taste, flavour, and/or oral-somatosensory attributes of chocolates on the other. At the same time, the second goal was to check on the validity of this internet-based approach and test whether the correspondences documented in Experiment B.4, conducted in the laboratory, would similarly carry over to the online version in Experiment B.5.

B.5.1. Methods

Forty-five respondents (6 male; age range = 18-51 years; mean age = 26 years) completed the study questionnaire online. The respondents were invited to complete a brief 5-minute questionnaire posted online using a convenience sampling method. A link to the questionnaire was included in an e-mail recruitment stating that participation was voluntary and that no compensation would be provided.

The online questionnaire was similar to the paper-and-pencil version used in Experiment B.4 in that the four scales per chocolate were presented per screen. Importantly, however, rather than actually tasting the chocolates as the participants had done in Experiment B.4, the respondents were now simply instructed to imagine tasting the chocolates. The presentation of the scales followed the same layout and order (i.e., Round-

Angular shape, Tuki-Lula, Maluma-Takete, and Unpleasant-Pleasant) and the presentation of the chocolates followed the same order as in Experiment B.4. Instead of a line scale with a crayon as the mid-point, however, eleven horizontally aligned radio buttons appeared above each scale; the first and eleventh radio buttons were aligned with the left and right anchors of the scale, respectively, and the sixth radio button served as the mid-point for each scale. The respondents were instructed to click on one of the radio buttons in order to indicate their rating of the chocolate, which was pictured at the top of the screen (see Figure B.5). If the respondent attempted to advance to the next question without selecting a radio button, a prompt appeared indicating that a response was required. Once they had made their ratings on each of the four scales, respondents clicked on a submit button at the bottom of the page in order to advance to the next screen. Once all four chocolate images had been presented and rated, a "thank you" screen appeared indicating that the questionnaire was complete and that the respondent's answers had been submitted.



Imagine eating the **MINT FONDANT CHOCOLATE** shown above. For each of the scales listed below, please select a point along the scale according to the shape/word that best matches the flavour of the chocolate.

If the flavour better matches the shape/word on the left side of the scale, click on a point closer to the left. If the flavour better matches the shape/word on the right side of the scale, click on a point closer to the right.

4. * Please select a point along the following scale.

rounded-shape 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9 ○ 10 ○ 11 ○ pointed-shape

5. * Please select a point along the following scale.

tuki 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9 ○ 10 ○ 11 ○ lula

tuki lula

6. * Please select a point along the following scale.

maluma 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9 ○ 10 ○ 11 ○ takete

maluma takete

7. * Please select a point along the following scale.

unpleasant 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9 ○ 10 ○ 11 ○ pleasant

unpleasant pleasant

Figure B.5. A minimized screenshot of one page of the online questionnaire used in Experiment B.5. Instructions appear at the top of the screen, along with a picture of the chocolate that respondents were supposed to imagine eating, and, thereafter, rate. Above each line were eleven radio buttons (1-11) which respondents clicked on in order to indicate the point along the scale corresponding to the image that best matched their experience when eating the chocolate.

B.5.2. Results and Discussion

Just as in Experiment B.4, the ratings for the second and fourth scales were adjusted so that values to the left of the scale (numbers smaller than 6) indicated more rounded / softer and pleasant ratings regardless of the specific response scale used. Respondents' ratings for each chocolate sample on each of the four response scales were compared using a 4 (chocolate) x 4 (scale) within-participants ANOVA. The analysis revealed significant main effects of scale, *F*(3,132)=41.75, *p*<.001, and chocolate, *F*(3,132)=8.77, *p*<.001, and a significant interaction between scale and chocolate, *F*(9,396)=20.31, *p*<.001 (see Figure B.6).

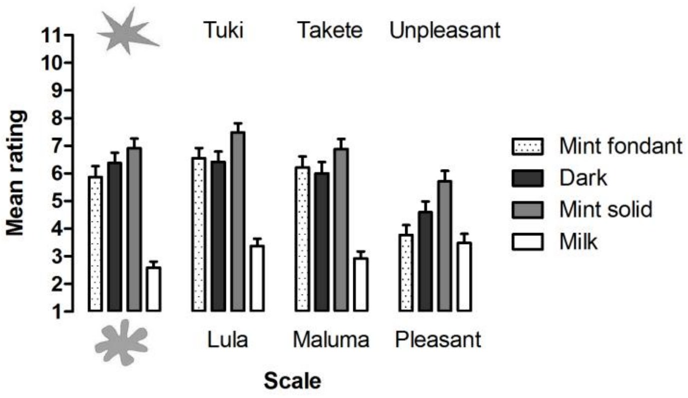


Figure B.6. Participants' mean ratings for each chocolate and scale type in Experiment B.5. Smaller values correspond to the round shape, rounded-words (Lula and Maluma), and Pleasant rating, while larger values correspond to the angular shape, sharp words (Tuki and Takete), and Unpleasant rating. Error bars represent the standard errors of the means.

An analysis of the simple effect of chocolates for the organic-angular scale revealed that milk chocolate (*M*=2.60) was rated as significantly more Round-shaped than mint fondant (*M*=5.87), dark (*M*=6.38), and solid mint (*M*=6.91), *p*<.001 for all comparisons. Respondents also rated the solid mint as significantly more Angular-shaped than mint fondant chocolate, *p*=.046. An analysis of the simple effect of chocolates for the Tuki-Lula scale revealed that milk chocolate (*M*=3.38) was rated as significantly more Lula than mint fondant (*M*=6.56), dark (*M*=6.42), and solid mint (*M*=7.49), p<.001 for all comparisons. Solid mint was rated as significantly more Tuki than dark chocolate, *p*=.016. The difference between respondents' ratings of the mint fondant and solid mint was marginally significant, *p*=.082. An analysis of the simple effect of chocolates for the Maluma-Takete scale revealed that milk chocolate (*M*=2.93) was rated as significantly more Maluma than mint fondant (*M*=6.22), dark (*M*=6.00), and solid mint (*M*=6.89), *p*<.001 for all of the comparisons. The difference between respondents' ratings of the dark and solid mint chocolates was marginally significant, *p*=.064. Finally, a simple effects analysis of chocolates for the Unpleasant-Pleasant scale revealed that milk chocolate (*M*=3.49) was rated as significantly more Pleasant than either dark (*M*=4.60) or solid mint (*M*=5.71) chocolates, *p*=.05 and *p*<.001, respectively. Both mint fondant and dark chocolate were rated as significantly more Pleasant than solid mint, *p*<.001 and *p*=.018, respectively. The difference between respondents' ratings of the pleasantness of the dark and mint fondant chocolates was marginally significant, *p*=.10. The difference between respondents' ratings of the mint fondant and milk chocolates was not significant, *p*=.54.

One-sample t-tests were conducted individually on each chocolate and scale. Specifically, the respondents rated the milk chocolate as being significantly Round-shaped, *t*(44)=-15.71, Maluma, *t*(44)=-12.52, Lula, *t*(44)=-10.21, and Pleasant, *t*(44)=-7.80, *p*<.001 for all comparisons. Respondents also rated the solid mint chocolate as being significantly

Angular-shaped, $t(44)=2.58$, $p=.013$, Tuki, $t(44)=4.58$, $p<.001$, and Takete, $t(44)=2.49$, $p=.017$. Finally, mint fondant and dark chocolates were both rated as being significantly Pleasant, $p<.001$ and $p=.001$, respectively.

The results of Experiment B.5, where the respondents were simply instructed to imagine tasting rather than actually tasting the chocolate samples, are consistent with those of Experiment B.4 in demonstrating, once again, that milk chocolate was clearly associated with the rounder-sounding speech sounds (Lula and Maluma) and organic shape. Here, and in Experiment B.4, the milk chocolate and mint fondant were rated as being Pleasant. The same could not be said for the solid mint chocolate. The respondents in Experiment B.5, however, also rated dark chocolate as being Pleasant, while the participants in Experiment B.4 did not rate the chocolate as being either Pleasant or Unpleasant. The results of Experiment B.5 also demonstrated that solid mint chocolate was consistently rated as being significantly more angular in shape than mint fondant and more ‘Tuki’ than dark chocolate. Similarly, respondents tended to rate the solid mint as being more ‘Tuki’ than the mint fondant, and more ‘Takete’ than the dark chocolate.

B.6. General Discussion

The results of the five experiments reported in Appendix B highlight the existence of reliable crossmodal correspondences between the taste/texture/flavour of chocolates (and, more specifically, their cocoa content) and various qualities of nonsense words and abstract shapes (see also Spence, 2011; Spence & Gallace, 2011). Experiments B.1 through B.3 demonstrated that low cocoa content milk chocolates (with a cocoa content of 30%) and milk chocolate truffles (exact information about the cocoa content of this product is not available) were associated with round shapes and words, whereas higher concentration (70-90% cocoa) dark chocolates were associated with angular shapes and sharp words. Similarly, Experiment B.4 demonstrated that participants matched low (34%) cocoa chocolates and 50% cocoa mint fondant chocolates with round shapes and words, whereas 50% cocoa solid mint and 70% cocoa dark chocolates were matched with angular shapes and sharp words instead.

The findings reported here are consistent with those reported by Thamke, Dürrschmid, and Rohm (2009) who had participants evaluate various chocolates varying in cocoa content. Thamke et al. demonstrated that people rate chocolates having a higher cocoa content as tasting more bitter than chocolates with a lower cocoa content. Additionally, when asked to describe the “mouthfeel” of various chocolates, the participants in Thamke et al.’s study reported that higher cocoa content chocolates (75%) tended to be described as dry, sticky, and mealy, whereas somewhat lower content chocolates (60%) tended to be described as melting or creamy instead. Interestingly, while the slightly lower (50%) cocoa content mint fondant chocolate in Experiments B.4 and B.5 seemed to have been rated similarly to the low (34%) cocoa content milk chocolate (i.e., corresponding with the round shape and words), the solid mint chocolate was rated similarly to the high (70%) cocoa content dark chocolate (i.e., corresponding with the angular shape and sharp words). In fact, despite its lower cocoa content as compared to the dark chocolate, the solid mint chocolate was rated as being more angular and sharp than the dark chocolate. Importantly, the key distinction between the solid mint and mint fondant chocolates is their texture/oral-somatosensory experience. Using the same adjectives as those in Thamke et al.'s study, one might describe the solid mint chocolate as '*dry, sticky, and mealy*', and mint fondant as '*melting or creamy*'. The results of Experiments B.1 through B.5 therefore suggest that the crossmodal correspondences observed here between angularity/sharpness and chocolate may reflect not only the taste (gustatory) attribute of bitterness, but also the texture/oral-somatosensory experience of the chocolate.

While participants in the experiments reported here had no problem differentiating between the milk and dark chocolates (as shown by the significant differences in participants' responses on the various ratings scales), many volunteered the information after completing the experiment that they liked the 70% dark chocolate more than the 90% chocolate sample. It is often said that quality cocoa should have a delicately balanced bitterness (Stark, Bareuther, & Hofmann, 2006), which teeters between pleasurable and unpalatable (at cocoa concentrations of around 85%; see Thamke et al., 2009). This is perhaps why, in Experiment B.4 and B.5, participants' ratings of pleasantness for the dark and solid mint chocolates seemed to be quite varied (cf. Crisinel & Spence, submitted).

Chocolates containing cocoa can be complex, comprising varying levels of bitter, sweet, salty, and even sour tastes. As such, it remains unclear as to what taste/flavour attribute(s) of the chocolates were responsible for the crossmodal correspondences assigned to the samples used (and tasted in a relatively naturalistic manner) in the present experiments. What is clear, though, is that the cocoa content, the source of the bitter taste of dark chocolate, varied, for example, between the three Lindt varieties sampled in Experiment B.1. Aside from their cocoa content, the three chocolates sampled in Experiment B.1 also happened to vary in their sugar content, with 9g sugar in the 30% cocoa milk chocolate, 5g in the 70%, and 2g in the 90% dark chocolate. Guinard and Mazzucchelli (1999) have demonstrated previously that varying the sugar content modulates participants' perception of a chocolate's bitterness, such that a lower sugar content gives rise to higher ratings of bitterness as compared to higher sugar content chocolates (having the same cocoa content). Thus, it is possible that the sweetness of the chocolates tasted in the present study also modulated their perceived bitterness, and perhaps in turn, their crossmodal matches (between taste/flavour and shape/sound). Whether participants' crossmodal correspondences in terms of taste/flavour (with the angularity of shapes and words) were altogether based on the perceived bitterness, sweetness, or some combination of these two tastes, which just happened to be present in (in reciprocal proportions) the chocolate samples tested here represents an intriguing question for future research.

Over the years, many have seen the commercial potential of capitalizing on the findings of research in the area of sound symbolism (e.g., Belli, 2001; Gallace et al., 2011; Klink, 2000, 2001; Spence, in press; Yorkston & Menon, 2004). Indeed, it is well-known that changing various visual aspects of a product's name can influence a consumer's perception of the product itself (Schloss, 1981). The results of the comparison of chocolates in the experiments reported in this Appendix are particularly interesting in this regard, because they show significant differences between the shapes/sounds associated with similarly-placed, commercially-available chocolate products. These results may therefore have implications for the naming of new products in the marketplace. While the name of Cadbury's Koko milk chocolate truffles, for example, would seem to be inspired by one of the ingredients cocoa/cacao (i.e., by semantic association; see Klink, 2001), given the results outlined here, it could be argued that the sound of the brand name may, in this case, be inappropriate (at least from the perspective of sound symbolism research). On the basis of the results reported here, the sound of this particular brand name may set up the wrong expectations in the mind of the consumer. Nestlé's Rolo, on the other hand, is an example of a brand name that does seem to convey the right expectations (chocolates with a chewy caramel centre) from a sound symbolism perspective and happens to be successful in both the UK and US marketplaces.

Cocoa is an inherently bitter ingredient and so, from a sound symbolism perspective, might match the sharp inflection of the sound of the word 'cacao', the substance from which cocoa is derived. It could be argued from this perspective that a company might want to use the sound of a product's brand name precisely in order to accentuate a product's specific sensory properties. Using some variation of the word cocoa might work well for a 70% cocoa product, however, the Koko milk chocolate truffle product from Cadbury isn't particularly high in cocoa content, and hence is unlikely to be experienced as bitter (or sharp). Thus, one might expect this product to fail in the marketplace (all other things being equal), given the negative long-term consequences that disconfirmed expectations are known to have on flavour/product perception (e.g., see Deliza & MacFie, 1997; Spence, in press; Yeomans, Chambers, Blumenthal, & Blake, 2008), the flavour expectation in this case being set-up, in part, by the sound of the product’s name. Moreover, when the separation between consumers’ expectations and the actual product becomes too large then you are likely not only to have disconfirmed expectations, but more detrimentally, a rebound effect (see Schifferstein, 2001).

REFERENCES

Belli, S. (2001). Qual e' takete? Qual e' maluma? [Which one is takete? Which one is maluma?] (2nd Ed.). *La psycolonguistica applicata alla cominicazione pubblicitaria*. Milan: Franco Angeli.

Crisinel, A.-S., and Spence, C. (submitted). The impact of pleasantness ratings on crossmodal associations between food samples and musical notes. *Food Quality and Preference*.

Deliza, R., & MacFie, H. J. H. (1997). The generation of sensory expectation by external cues and its effect on sensory perception and hedonic ratings: A review. *Journal of Sensory Studies*, **2**, 103-128.

Denscombe, M. (2009). Item non-response rates: A comparison of online and paper questionnaires. *International Journal of Social Research Methodology*, **12**, 281-291.

Fónagy, I. (1963). *Die Metaphern in der Phonetik* [The metaphors in phonetics]. The Hague.

Fónagy, I. (2001). A research instrument. In I. Fónagy (Ed.), *Languages within language: An evolutive approach* (pp. 337-357). Amsterdam: John Benjamins.

Gallace, A., Boschin, E., & Spence, C. (2011). On the taste of ‘Bouba’ and ‘Kiki’: An exploration of word-food associations in neurologically normal participants. *Cognitive Neuroscience,* **2**, 34-46.

Guinard, J.-X., & Mazzucchelli, R. (1999). Effects of sugar and fat on the sensory properties of milk chocolate: Descriptive analysis and instrumental measurements. *Journal of the Science of Food and Agriculture*, **79**, 1331-1339.

Hinton, L., Nichols, J., & Ohala, J. J. (Eds.). (1994). *Sound symbolism*. Cambridge: Cambridge University Press.

Klink, R. R. (2000). Creating brand names with meaning: The use of sound symbolism. *Marketing Letters*, **11 (1)**, 5-20.

Klink, R. R. (2001). Creating meaningful new brand names: A study of semantics and sound symbolism. *Journal of Marketing: Theory and Practice*, **9 (Spring)**, 27-34.

Köhler, W. (1929). *Gestalt psychology*. New York: Liveright.

Köhler, W. (1947). *Gestalt psychology: An introduction to new concepts in modern psychology*. New York: Liveright Publication.

Ladefoged, P. (1993). *A course in phonetics (3rd edition)*. New York: Harcourt Brace Jovanovich.

Lefever, S., Dal, M., & Matthiasdottir, A. (2007). Online data collection in academic research: Advantages and limitations. *British Journal of Educational Technology*, **38**, 574-582.

Maurer, D., Pathman, T., & Mondloch, C. J. (2006). The shape of boubas: Sound-shape correspondences in toddlers and adults. *Developmental Science*, **9**, 316-322.

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

Riva, G., Teruzzi, T., & Anolli, L. (2003). The use of the internet in psychological research: Comparison of online and offline questionnaires. CyberPsychology & Behavior, **6**, 73-80.

Sapir, E. (1929). A study in phonetic symbolism. *Journal of Experimental Psychology*, **12**, 225-239.

Schifferstein, H. N. J. (2001). Effects of product beliefs on product perception and liking. In L. Frewer, E. Risvik, & H. Schifferstein (Eds.), *Food, people and society: A European perspective of consumers' food choices* (pp. 73-96). Berlin: Springer Verlag.

Schloss, I. (1981). Chicken and pickles. *Journal of Advertising Research*, **21 (December)**, 47-49.

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

Spence, C. (in press). Managing sensory expectations concerning products and brands: Capitalizing on the potential of sound and shape symbolism. *Journal of Consumer Psychology*.

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

Stark, T., Bareuther, S., & Hofmann, T. (2006). Molecular definition of the taste of roasted cocoa nibs (Theobroma cacao) by means of quantitative studies and sensory experiments. *Journal of Agriculture and Food Chemistry*, **54**, 5530-5539.

Thamke, I., Dürrschmid, K., & Rohm, H. (2009). Sensory description of dark chocolates by consumers. *Food Science and Technology*, **42**, 534-539.

Yeomans, M., Chambers, L., Blumenthal, H., & Blake, A. (2008). The role of expectancy in sensory and hedonic evaluation: The case of smoked salmon ice-cream. *Food Quality and Preference*, **19**, 565-573.

Yorkston, E. A., & Menon, G. (2004). A sound idea: Phonetic effects of brand names on consumer judgements. *Journal of Consumer Research*, **31**, 43-51.