Cross-Modal Information Transfer and the Effect of Concurrent Task-load

Author: Alex Bahrami Balani

Affiliation:
Edge Hill University
Department of Psychology
St Helens Rd.
Ormskirk
L39 4QP
UK

The work has been conducted in the above university.

Corresponding author: Alex Balani

balanial@edgehill.ac.uk

Disclosure of interest:

There exists no conflict of interest in the conduct and reporting of this research (e.g. financial interests in a test or procedure used here or any funding by any organization).

Abstract

Our everyday lives offer plenty of situations where complex processing of information takes place, in which information needs to transfer across modalities to achieve a behavioral goal. The study examined the differential effects on object detection by a visual, verbal or an auditory cue held in Working Memory (WM), and the role of concurrent cognitive task-load on the final detection of that cue. Three experiments, all using same stimuli set but in different modalities, subjects held in memory a representation of a novel cue for a speeded detection in a search display at the end of each trial. The cue stimulus could be an image (visual), the name (verbal) or the sound (auditory) of a common animal or object. A mental arithmetic task was interleaved between the cue presentation and the cue detection. The results showed that information held in WM, either in verbal or auditory form, can efficiently transfer across modalities to complete a visual detection task for a representation of the initial WM-cue. The speed of detection was not affected by the cross-modal transfer of cue information but there was some detrimental effect on detection that could distinctively be attributed to the cognitive task-load. Together, these findings may provide some evidence for the role of Episodic Buffer component of WM (Baddeley, 2000) in integrating multimodal information originated from different sources, hence supporting the notion of the supramodal nature of WM. The results have been discussed in light of Baddeley's (1974, 2000) and Cowan's (1988, 1999) theoretical WM frameworks.

Keywords: Cross-Modal, Object Detection, Concurrent task, Task-load, Episodic Buffer

Human attention facilitates the selection of perceptually salient or behaviorally relevant stimulus representations from among multiple competing representations (Desimone & Duncan, 1995), received through different sensory modalities. A common paradigm used to study such phenomenon is visual search in which a representation of a target stimulus is actively maintained in WM to guide the target detection in a search display (Duncan & Humphreys, 1989; de Fockert, Rees, Frith & Lavie, 2001; Awh, Vogel & Oh, 2006). Evidence from studies using behavioral (Soto, Humphreys & Rotshtein, 2007), ERP (Luria & Vogel, 2011) and fMRI (de Fockert et al., 2001) support the possibility of involvement of WM in such visual selection task. A single neuron study by Chelazzi, Miller, Duncan and Desimone (1998) provided evidence showing how memory guided visual search uses top-down feedback from structures involved in WM to resolve a competition for object detection in primates. However, some other studies have failed to report such WM-involvement (e.g. Kane, Poole, Tuholski & Engle, 2006).

Early seminal research by Atkinson and Shiffrin (1968) on information storage and processing suggested a modal (Baddeley & Hitch, 1974) system receiving and binding sensory information from a variety of sources. A multi-component model of information processing and storage was proposed by Baddeley and Hitch (1974) suggesting separate stores on the basis of the type of information (e.g. phonological, visuo-spatial). Compared to Atkinson and Shiffrin, the current model emphasized the role of executive control and attention. An alternative multidimensional model was suggested by Cowan (1988), called embedded processes model (Cowan, 1999), specifying the focus of attention and the role of Long Term Memory (LTM) in activation of WM (Cowan, 1988, 1999, 2005; Baddeley, 2009). Baddeley & Hitch's model includes a mechanism called central executive to enable control of information. While Baddeley (1986) suggests separate subsystems for phonological and visual stimuli, Cowan (2005) argues for a general, single, central capacity-limited system for information processing and storage, through which incoming information only activates relevant matching features in long term memories. Cowan argues that new

information overwrites or interferes with old information (Adams, Nguyen & Cowan, 2018). In Cowan's (1988) model, focus of attention allows a coherent and organized interpretation of the very limited amount of information that managed to enter the system. Cowan (2001) incorporated the capacity limitation of WM in his model. Baddeley (2000) added a general component called episodic buffer to the model which acted as the link between the stores and associated information from the stores and LTM. Like Cowan's model this new component emphasized focus of attention and integration of information with LTM.

WM is necessary for all visual detection task, where a pre-defined target object is kept online and to be detected at a later stage in a search display that may or may not be presented among other (distractor) objects (e.g. Olivers, Meijer & Theeuwes, 2006; Soto, Heinke, Humphreys & Blanco, 2005). The detection can be less efficient when the target and distrctor objects share one or more features (Maxfield, Staldes & Zelinsky (2014); Duncan and Humphreys, 1989; Cowan, 1988). In any visual search, a template of the cue or target object is activated, and the search is based on the match between the cue template and the actual target. When the cue is an exact picture of the target, a visual representation of the cue is activated to guide detection and the match is direct. However, when the cue is a word or a sound representing the cue object, the cue is defined by its "meaning" (Potter, 1975), in which a representation of the cue object is created as a conceptual self-generated image (Potter, 1993, 2012). This representation acts as a construct, kept in WM, to facilitate the selection of the cue target among competing distractors in the search display (Desimone & Duncan, 1995) and relies on an indirect match between the mental image and the actual target cue. Potter (1993) views this concept as a fleeting form of a very short-term memory (CSTM) that facilitates the recognition of meaningful stimulus (e.g. word or object) which rapidly activates conceptual information and retrieves relevant information from LTM. Potter describes CSTM as a mental buffer and processor in which current perceptual stimuli and their associated concepts from LTM are represented briefly, allowing meaningful patterns or structures to be identified (Potter, 1993, 2012). Expending upon this, an alternative explanation for how information

from different sources are processes has been offered by Ricker, Nieuwenstein, Bayliss and Barrouillet (2018; see also Wyble, Potter, Bowman & Nieuwenstein , 2011).) who suggest that WM consolidation involves a binding of perceptual and semantic features of the item to-be-remembered in WM is bound to a "token representation". This process will need the entry of activated LTM representation into the focus of attention (Ricker & Cowan, 2014) and the binding of perceptual and semantic features of the item (Wyble et al., 2011).

In our everyday lives, we need to process, select and transfer information originating from one sensory modality to another modality, in order to achieve our behavioral goals. Driving a vehicle is such a daily activity where information is smoothly transferred across modalities. In a driving episode, information presented via different modalities such as visual (scenery, GPS maps, warning lights, symbols), verbal (text and numbers), auditory (beeps and sirens, oral guidance from GPS device), or tactile (feeling of uneven surfaces through vibrations). Our attentional selection helps us to prioritize, harness, canalize and process relevant information across modalities in order to achieve that behavioral goal. The common denominators for all these situations are that (i) they contain two or more concurrent tasks which require a degree of consolidation of information transferred between different domains; (ii) WM involvement in conducting the tasks is crucial and (iii) attention to a stimulus in one modality may either interfere with guidance of attention to that stimulus presented in another modality (Rapp & Hendel, 2003). As early as 1962, Gibson reported evidence of cross-modal identification of object between visual and tactile domains. More recently, Hadjikhani and Roland (1988) provided fMRI evidence for the existence of cross-modal transfer of tactile and visual information. The existence of cross-modal (priming) effect has been reported between visual and haptic modalities (Easton, Green & Srinivas, 1997) and between visual and auditory modalities (Schneider, Engle & Debner, 2008). In two recent studies (Balani, Soto & Humphreys, 2010; Kawashima & Matsumoto, 2017) participants held in memory a visual or verbal target object while completing a visual search

task. The cue could be an image of, or the word representing the target stimulus. Memory for the target cue was tested at the end of each trial to ensure that the cue representation was kept active during the trial. Both studies observed that visual and verbal representation of a memory cue could guide attentional selection in a top-down manner. Other similar studies have examined the effect of an additional concurrent cognitive task on the performance of a primary task and have established a detrimental effect of the secondary task-load on detection of the target stimulus in a search display (e.g. Soto & Humphreys, 2008; Vergauwe, Camos & Barrouillet, 2014).

Applying an ecologically valid design, Briem and Hedman (1995) investigated the effect of a secondary concurrent, but irrelevant, task (carrying out a phone conversation) on the simulated driving performance. The secondary task (conversation) had two different levels of difficulty - easy and hard; the difference between the two being that the hard task contained a WM element. Briem and Hedman found that conducting a difficult conversation negatively affected the driving performance, while an easy conversation had no adverse effect. The authors argued that the difficult conversation task (involving WM) was attention demanding and required an active involvement of Supervisory Attentional System (SAS) according to Shallice & Burgess' (1993) model of attentional control. In the same way, Baddeley, Logie, Bressi. Della Sala and Spinnler (1986) considered the SAS as a 'potential framework' for the Central Executive component of WM, which is generally viewed as a broad attentional control function. In their study, Baddeley and colleagues (1986) compared the interfering effect of a concurrent task on three groups of participants; Alzheimer patients, a group of age-matched healthy adults and a group of young healthy adults. The latter two healthy groups are of special interest for the current study. The participants had to complete two concurrent task; (a visual and an auditory) that were dependent of Central Executive component of WM. The main task was a visual pursuit-tracking task. The second concurrent task had three different levels (i) an articulatory suppression task (verbal), (ii) speeded reaction to a tone (auditory) and (iii) Digit Span; a task that involved contribution from both

Central Executive of WM and Phonological Loop. It was observed that, in both young and elderly groups, the addition of a secondary task impaired performance in the primary task. They reported interference between the verbal task (articulatory suppression) and visual task (tracking); and between auditory task (speeded response to a tone) and visual task (tracking). The detrimental effect of the concurrent task on visual tracking was largest when the task involved WM (Digit Span).

In cases where the secondary task uses same modality as the primary task, interference is high. Cocchini, Logie, Della Sala, MacPherson and Baddeley, (2002) had participants perform concurrent combinations of different pairs of tasks; a verbal memory task, a visual memory task, a perceptuo-motor tracking task and combinations of the two memory tasks while performing an articulatory suppression. A concurrent verbal and visual memory task did not result in mutual interference between the tasks with retention of a visual memory. However, when the tasks tapped the same modality (verbal in their example), there was clear disruption from articulatory suppression. Cocchini et al.'s findings showed that visual pre-cues were as effective as auditory pre-cues in a simple visual reaction time task, but that neither was effective for a simple auditory reaction time (RT) task. Ward (1994) reported that visual pre-cues enhanced visual as well as auditory localization, but auditory pre-cues only affected RTs to localize auditory cues and failed to show any effect on visual localization. Ward argued that the exogenous attention system is modality-specific (see also Ward, McDonald, & Lin, 2000). This may suggest that there is some kind of generalization from different forms of stimuli (e.g. visual, verbal, or auditory).

Balani, Soto and Humphreys (2010) examined the effect of top-down processes from WM in modulating the deployment of attention across modalities and demonstrated that object information could transfer between visual and verbal representations of stimuli. In a related work, Soto & Humphreys (2007) had participants hold in memory words corresponding to geometrical shapes and found very similar effects on visual search for a

tilted line embedded in one of four geometrical shapes (one shape could correspond to the word being held). The study demonstrated that both verbal as well as visual WM can guide object search. The findings were consistent with Baddeley's latest views of WM, which highlights that information can be held in a relatively abstract manner, irrespective of the modality of the stimulus (Baddeley, 2000, Bahrami Balani, Soto & Humphreys, 2011; Kawashima & Matsumoto, 2017). This abstract information, subsequently guides search to matching stimuli, whereby the Episodic Buffer can act as a multidimensional and supramodal hub for binding information originating from Visuo-Spatial Sketchpad and Phonological Loop, as well as information retrieved from long-term memory (LTM), to be used in WM for "binding of new information to its context in LTM" (Baddeley, 2012; Baddeley, Allen & Vargha-Khadem, 2010). The question remains as to how a secondary concurrent task affects selection when the cue is presented in a different modality than the visual detection display. The current study intends to examine this question as well as the effect of a secondary cognitive task involving WM (a mental arithmetic task) on the primary task (detection of a pre-determined novel target). As reported above, Briem and Hedman's difficult concurrent task involved a conversation that contained a WM element and Baddeley et al.'s third tasks (Digit Span) also involved an element of WM (mental manipulation of information). The current study has, in the same fashion, a concurrent mental arithmetic task that uses WM resources (DeStefano & LeFevre, 2010) and that has shown to impair performance on the primary task (Harms, 1991; Baldwin & Schieber, 1995).

In three experiments, a pool of complex animate and inanimate stimuli is used to investigate the effect of cue modality on speeded detection by holding in WM a visual, verbal or an auditory representation of the novel cue for the duration of the trial. Furthermore, the consequences of the concurrent task-load on the efficiency of the detection is assessed when the cue stimulus (held in WM) and search array are presented through either the same (Experiment 1) or a different modality (Experiments 2&3). Subjects were asked to keep a representation of a novel cue item for a speeded cross-modal detection in a visual search

display at the end of each trial. In Experiment 1, the initial cue was presented visually (an image); in Experiment 2, it was presented verbally (a name); and in Experiment 3, it was presented aurally (a sound). In addition, the design of the trials aimed to bind two aspects of the relation between concurrent task-load and detection through completion of a mental arithmetic task. This secondary task was interleaved between the cue presentation and cue detection. For the mental arithmetic task the subject had to verify the accuracy of a simple addition or subtraction equation involving two numbers (easy task - low task-load) or three numbers (difficult task - high task-load).

To date, no studies could be found that compared the effects of visual, verbal and auditory cues on object detection using the same stimuli set using a within-subject design. Another novel aspect of this study is having a concurrent task that does not interfere with the primary task. For example, in previous studies by Balani et al. (2010) and Kawashima and Matsumoto (2017), there was a concurrent visual search task interleaved between the cue presentation and the probe display which shared the same pool of stimuli as the cue stimuli (i.e. geometrical shapes, objects, animals). This would raise the possibility that using same stimuli pool for both tasks might have a confounding effect on the tasks. For example, if the initial cue stimulus re-appeared in the concurrent task, it would update the memory for the memorized cue and therefore facilitate the final recall task.

Expanding upon previous research that has examined this effects of verbal-visual stimuli and vice versa (Soto et al, 2007; Balani et al, 2010; Kawashima & Matsumoto, 2017), the objective for the current study is to examine the nature and role of WM in processing information originating from visual, verbal and auditory modalities and to tease apart the effect of modality transfer from the effect of task-load. Hence, the study has the following two aims; (i) to examine the nature of the cross-modal information transfer between auditory, verbal and visual modalities and (ii) explore the effect of a concurrent cognitive task, that taps WM resources, on cue detection. It is expected that the modality through which the cue

is presented would not affect the detection efficiency (speed and accuracy) and it is also expected that cue detection will be less efficient under high task-load than under low task-load, irrespective of its modality.

General Method

Three experiments assessed consequences in speeded detection of presenting a cue stimulus in one modality (visual, verbal or auditory) and detection of its representation in a visual display. This effect was examined when a concurrent cognitive task-load was low or high. In each trial, a novel target cue (henceforth called "cue") was presented for a brief moment at the start of the trial, to be detected in a visual display at the end of the trial. A secondary concurrent task of mental arithmetic was interleaved between the cue presentation and the detection display. The task involved judging the accuracy of a simple equation with addition or subtraction of two digits (low task-load) or three digits (high task-load). DeStefano and LeFevre (2004) have argued that a mental arithmetic task employs all three components of WM, in Baddeley and Hitch's (1974) original WM model, and that mental arithmetic requires central executive resources, even for solving single-digit equations.

The cue item was randomly drawn from a pool of animate (common animals) and inanimate (common man-made) objects; all of which were able to produce a distinctive sound. Experiment 1 used a unimodal paradigm in which all stimuli were presented visually. Experiments 2 and 3 had both a cross-modal paradigm. In Experiment 2, the cue was presented verbally; as a word representing the name of the animal or object, while the search display showed images of the stimuli. In Experiment 3, the cue was presented aurally; as the sound of the animal or object, while the search display was presented visually. In addition, all trials included a secondary mental arithmetic task and the participants were instructed to respond quickly and accurately. Figure 1 illustrates the design of the experiments 1-3.

Figure 1 about here

Participants. Forty-nine university students [41 females; Mean (SD) age=18.93 (1.51), and 8 males; Mean (SD) age=21.63(4.13)] completed all 3 experiments for course credits. All participants had normal or corrected to normal vision and normal hearing. To eliminate "lucky guesses" in the arithmetic and detection task responses, the "kill the twin" correction procedure (Eriksen, 1988) was employed. This resulted in the elimination of 1.3% of all trials. Moreover, two participants' data were excluded due to having chance-level performance, resulting in 47 participants' data in the final analyses [39 females; mean (SD) age=18.95 (1.53) and 8 males; 21.63(4.13)].

Stimuli. The visual stimuli consisted of 24 grayscale digital images of common animals (cat, cow, chicken, crow, dog, duck, frog, horse, pig, robin, rooster, sheep) and everyday objects (phone, car, clock, drum, plane, guitar, horn, bell, ambulance, piano, train, trumpet), each image measured about 4-6 cm in size and shown on a white background. Most of these stimuli were same images used in Balani et al. (2010) and the remaining images were sourced at image-net (.org). All stimuli were converted to grayscale (to reduce color saliency effect) and isolated on a white background using Adobe PS software. The verbal stimuli were printed words representing the same animal or objects. The auditory stimuli were one second epochs with the sound of these animals and objects. The original sounds were downloaded from freesound (.org) website, however, all sounds were digitally edited (e.g. any on-and-off clicking noise removed, sound adjusted for loudness and pitch) using Audacity® software. The sound files were saved in mp3 format (32-bit stereo with 44.1 kHz sampling frequency) and were delivered via stereo over-ear headphones.

Design and procedures. Ethical approval was obtained from the university for the study prior to data collection and all participants provided written consent prior to participation. All experiments were designed, conducted and analyzed on a PC using E-Prime® (Psychology Software Tools 1996-2012), displayed on a 20" LED monitor with refresh rate of 60 Hz on a viewing distance of ~60 cm. All three experiments followed the

same protocol and were conducted in a dimly lit and quiet room. All participants completed all three experiments.

Before each experiment, participants familiarized themselves with the stimuli used in that experiment and completed a practice block. Each experiment had two tasks; (i) to keep a cue in memory for a speeded detection at the end of the trial and (ii) to verify the accuracy of a mental arithmetic task during the interval between the presentation of the cue and its detection. Depending on the difficulty level of the arithmetic task, there were two types of trials with an equal number of each type; "low task-load" and "high task-load".

Low task-load trails. Immediately after the cue presentation, a simple arithmetic equation (addition or subtraction) involving two digits (e.g. 2+ 4=6, or 7-5=3) was shown at fixation. The product of the equation was always between 1 and 9. The task was to verify its accuracy by key press within 3 seconds. Half the equations were correct, and half were incorrect, and their order was random.

High task-load trials. The task and its procedures were the same as in the low task-load except that the arithmetic equation here had three digits (e.g. 3+2+4=8 or 6-1-2=3).

Detection task. All trials ended with a visual detection display showing four images; two animals and two objects, placed around an imaginary square grid; one image in each quadrant with a 4° distance from fixation. These images were located equidistant from the central fixation and from each other. The two exemplars of each category were located diagonally in the opposite quadrant. The task was to decide whether an image representing the cue is present in the search display or not. In half of the trials one of these images corresponded the initial cue. Each experiment had 240 trials. The order in which different trial types were presented was random. Likewise, the order in which the three experiments were conducted by each participant was also randomized. All three experiments were completed on the same day. The approximate time for completion of each experiment was 25 minutes. Participants were given the opportunity to take as long break as they needed between the experiments, however, no participant took longer break than 5-6 minutes. All

responses were collected by number key press. Moreover, all outlying data (±2.5 SD) from the mean were excluded from the datasets, which eliminated ~1% of trials.

Experiment 1: Unimodal visual cue detection

Experiment 1 had a unimodal paradigm where both the WM cue and the stimuli in the search display were grayscale images of objects or animals. In each trial, a cue was kept in WM for a subsequent detection/recognition in a detection display at the end of that trial. During the interval between cue presentation and its detection, there was a mental arithmetic task.

Method

The method and procedures were the same as described in General method section. Each trial started with the presentation of a visual cue and ended with a visual search display. The cue image was drawn randomly from two different types of stimuli; common animal species (e.g. dog, horse) or common everyday objects (e.g. phone, car). There were 12 exemplars of each category. Both the cue type and type of arithmetic task (low or high task-load) were randomized.

Results

RT analysis. A two-way repeated measures ANOVA was conducted on correct detection RTs. The factors were cue-presence (present-absent) and task-load (low-high). There were significant effects of cue-presence; $F(1,46)=18.38$, $p=.001$, $\eta^2_{p=}.286$, (faster detection on cue-present than cue-absent trials; 480.9 ms vs. 538.7 ms); and task-load; $F(1,46)=34.77$, $p=.001$, $\eta^2_p=.430$ (faster detection with low task-load than with high task-load; 483.98 ms vs. 535.73 ms). There was also a significant interaction between the two variables; $F(1,46)=12.46$, $p=.001$, $\eta^2_p=.213$. A simple main effect analysis investigated the nature of the interaction which revealed that when the cue was present, there was no significant difference in detection RT as a factor of task-load ($p>.05$), but when the cue was

absent, higher task-load delayed detection (p=.001). Further, it showed that in both task-load conditions, there were significant differences due to cue-presence (p<.05 for both), showing speedier detection when the cue was present in the final display than when it was absent. Figure 2A illustrates the RT data for Experiment 1.

Figure 2 about here

Error analysis. A two-way repeated measures ANOVA was conducted on the error rates in detection. The factors were cue-presence and task-load. There was a significant main effect of task-load; $F(1,46)=39.47$, p=.001, η^2_p =.462, (lower error rate with low task-load than with high task-load; 11.48% vs.16.03%), but no significant effect of cue-presence and no reliable interaction (p>.05 for both). Figure 2B shows the error data.

Discussion

Experiment 1 examined detection performance in unimodal trials where all stimuli were images. The cue-present trials resulted in speedier detection, which is a common phenomena within such visual search and is reported to depend on the notion that in a target-present visual search, the search terminates as soon as the cue is detected, but if the cue is absent, the search has to inspect all displayed items to complete the task (Wolfe, 1998). Moreover, there was no effect of task-load on RTs in present trials, but in cue-absent trials, higher task-load delayed the detection. Overall the RTs and error rates were lower with low task-load than with high task-load. Experiment 1 has uncovered evidence that two factors can have detrimental effects on the speed of detection for a cue held in WM; high task-load and absence of cue item in the search display. Detection accuracy was lower with higher task-load, but unaffected by the absence of the cue. Experiment 2 investigates whether such factors affect detection in a visual-verbal cross-modal paradigm.

Experiment 2: Cross-modal cue detection (verbal-visual shift)

Experiment 2 used a cross-modal paradigm to investigate the effect of intra-trial modality shift on cue detection when the initial cue was presented verbally (a word), while the search display showed four images, one could be the cue.

Method

The experiment mimicked the method and procedures of Experiment 1 in all aspects but the modality through which the initial cue was presented. Here, the name of the object/animal was printed in black capital letters shown in the middle of screen on a white background in Arial bold font, 28 pt.

Results

RT analysis. A two-way repeated measures ANOVA was conducted on correct detection RTs. The factors were cue-presence and task-load. There were significant main effects of cue-presence; $F(1,46)=52.95$, $p<.001$, $\eta^2_p=.535$, (faster detection when cue was present than absent; 490.96 vs. 585.91 ms); and of task-load; $F(1,46)=11.72$, $p=.001$, $\eta^2_p=.203$, (faster detection with low task-load than high task-load; 517.41 vs. 559.47 ms). There was a reliable interaction between these factors; $F(1,46)=5.2$, $p=.027$. $\eta^2_p=.102$. A simple main effect analysis investigated the nature of this interaction showing that when the cue was present in the display, there was no effect of task-load on RTs ($p>.05$); however, when the cue was absent, higher task-load delayed detection ($p=.001$). Moreover, cue-present trials were faster than cue-absent trials both with both low- and high task-load ($p<.001$ for both). Figure 2C shows the RT data.

Error analysis. A two-way repeated measures ANOVA was conducted on the error rates in detection. The factors were cue-presence and task-load. There were significant main effects of cue-presence; $F(1,46)=6.84$, $p=.012$, $\eta^2_p=.129$, (lower error rate in cue-absent trials than cue-present trials; 12.44% vs.15.36%); and of task-load; $F(1,46)=22.14$, $p=.001$,

η^2_p =.325, (fewer errors with low than with high task-load; 12.03% vs.15.77%). There was no interaction between the two factors (p>.05). Figure 2D shows the error data.

Discussion

Experiment 2 explored whether a verbal-visual modality shift affects detection performance as a factor of task-load. Overall, cue-absent trails and high task-load trials generated slower and less accurate cue detection. The interaction between cue-presence and task-type evidenced that in cue-absent trials, higher task-load is more taxing on the speed of detection, as high task-load uses more of the limited WM-capacity. The main finding here is that a verbal cue can effectively inform detection of a matching representation on the final display. In a seminal work by Potter (1975), no differences in accuracy of target detection were found between when the participants were shown the exact target stimulus (image of a boat), and when they were shown the name of the target (the word boat). The findings in Experiment 2, also corroborates the findings by Balani et al. (2010); Soto and Humphrey's (2007) and Kawashima and Matsumoto's (2017), who also showed that cue information could transfer between verbal and visual representations of stimuli to guide detection. Can the same transfer be observed with auditory modality? Experiment 3 examines this.

Experiment 3: Cross-modal cue detection (auditory-visual shift)

Experiment 3 investigates whether auditory cue information can efficiently transfer to guide cue detection. a cross-modal transfer of auditory stimuli would have any effect on a visual cue detection.

Method

The method in Experiment 3 was identical to that in Experiment 1 and 2 in all aspects except that the cue-stimuli here consisted of one second epochs of pre-recorded sounds of the same animals or objects used in Experiment 1 and 2. All sound clips were made similar in qualities such as pitch, noise level and duration. Over ear stereo headphones were used to

play the sound clips. The participants were asked to adjust the sound level to their desired volume at the start of the experiment.

Results

RT analysis. A two-way repeated measures ANOVA was conducted on the correct detection RTs. The factors were cue-presence and task-load. There were significant main effects of cue-presence; $F(1,46)=49.06$, $p=.001$, $\eta^2_p=.516$, (faster detection when cue was present than absent; 478.87 ms vs. 571.60 ms); and of task-load; $F(1,46)=10.42$, $p=.002$, $\eta^2_p=.185$, (faster detection with low task-load than with high task-load; 504.41 ms vs. 546.07 ms).There was no reliable interaction between the two factors ($p>.05$). Figure 2E illustrates the RT data for Experiment 3.

Error analysis. A two-way repeated measures ANOVA was conducted on the error rates in detection for trials. The factors were cue-presence and task-load. There were significant main effects of cue-presence; $F(1,46)=49.45$, $p=.001$, $\eta^2_p=.518$, (lower error rate in cue-absent trials than in cue-present trials; 5.36% vs.10.86%); and of task-load; $F(1,46)=24.22$, $p=.001$, $\eta^2_p=.345$, (lower error rate with low task-load than with high task-load; 6.06% vs.10.22%). There was a reliable interaction between the two factors $F(1,46)=9.87$, $p=.003$, $\eta^2_p=.177$. A simple main effect analysis investigated the nature of this interaction showing significant differences in detection accuracy as a factor of task-load both when the cue was present ($p=.005$) and when it was absent, ($p=.001$). Similar to Experiment 2, here the overall error rate was slightly higher on cue-present trials than in cue-absent trials. Figure 2F shows the error data.

Discussion

In Experiment 3, the cue was presented aurally, and the detection in one average more accurate compared with Experiment 1 and 2, where the cue was presented visually (as image or printed text). The overall RT data patterns in this experiment remained similar to those in

Experiment 1 and 2; slower detections when (i) the cue stimulus was absent in the search display; and (ii) when the task-load was high. Overall, the speed of detection based on auditory cue information was similar to visual and verbal cues. However, the error data showed an unexpected low error rate when the cue was absent in display. this pattern was slightly different from Experiments 1 and 2. Here, the detection accuracy was higher irrespective of the presence or absence of the cue object in the display, but only when the task-load was low. These findings were somewhat unexpected and not easy to explain. This might be ascribed to some over-reliance on one modality, (visual modality in this case), in both Experiments 1 and 2; where the cue object and the search stimuli were processed visually, and this probably has caused some interference, while in Experiment 3 the cue is a sound but the search display was visual and hence less within modality interference (e.g. Cocchini et al., 2002).

Cross-Modal comparison: Changes in efficiency due to modality shift

Contrasting the experiments. Experiments 1-3 were similar in all aspects bar one; the modality through which the initial cue was presented. In Experiment 1, the cue was presented visually, in Experiment 2, it was presented verbally and in Experiment 3, it was presented aurally. But, in all experiments, all trials ended with a visual display for the detection of the cue stimulus or its representation. Separate analysis of each experiment pointed toward more or less similar patterns in the processing of the cue-stimuli irrespective of modality; both the task-load and absence of the cue (or its representation) in the final display deteriorate detection speed and, in many cases, its accuracy. Does this imply that performance in Experiments 2 and 3 (with a modality shift), would be as efficient as Experiment 1 (with no modality shift)? A comparative statistical analysis of the data from all experiments may answer this question.

RT-analysis. RTs for cue detection from trials with accurate responses were submitted to a separate three-way repeated measure ANOVAs. The factors were cue modality (visual,

verbal and auditory), cue-presence and task-load. It showed significant main effects of task-load; $F(1,46)=33.79$, $p=.001$, $\eta^2_p=.424$, (faster detection with low task-load than with high task-load; 501.92 ms vs. 547.09 ms); and of cue-presence; $F(1,46)=65.04$, $p=.001$, $\eta^2_p=.587$, (faster detection when cue present than when absent; 483.59 ms vs. 565.42 ms). There was no significant main effect of cue modality ($p>.05$) but there were reliable two-way interactions between cue-modality and cue-presence; $F(2,92)=3.9$, $p=.023$, $\eta^2_p=.079$; and between cue-presence and task-load; $F(1,46)=7.46$, $p=<.009$, $\eta^2_p=.139$; and a reliable three-way interaction between cue-modality, cue-presence and task-load. $F(2,92)=4.73$, $p=014$, $\eta^2_p=.093$.

This three-way interaction was further examined by conducting a simple main effect analysis, which showed that, on cue-present trials, there were no significant differences due to cue-modality on either level of the task-load (all $p>.05$). On cue-present trials, there was no difference in detection speed as a factor of the modality of the cue irrespective of task-load (all $p>.05$). On cue-absent trials, with high task-load, there were no RT difference attributable to the cue modality (all $p>.05$), but with low task-load, visual cues generated faster detection than verbal ($p=.04$)) and auditory cues ($p=.03$), but there was no difference between trials with verbal and auditory cues ($p>.05$). Moreover, examining each cue-modality in isolation, the analysis showed that on cue-present trials with visual and verbal (but not auditory) cues, the task-load had no effect on the detection speed ($p>.05$), but when the cue was absent, high task-load slowed down the speed of detection ($p=.001$). With auditory cues, however, there was no difference in detection speed as a factor of task-load irrespective of whether the cue was present or absent (all $p<.01$). The comparative RT data are illustrated in Figure 3A.

Figure 3 about here

Error analysis: Detection error rates were submitted to a three-way repeated measure ANOVA with the factors being cue modality, cue-presence and task-load. It showed significant main effects: of cue modality; $F(2,92)=21.08$, $p=.001$, $\eta^2_p=.314$, (error rates for visual modality=13.75%, for verbal modality=13.9% and for auditory modality=8.11%); of cue-presence; $F(1,46)=8.63$, $p=.005$, $\eta^2_p=.158$, (fewer errors on cue-present than on cue absent trials; 10.63% vs.13.24%) and of task-load; $F(1,46)=84.61$, $p=.001$, $\eta^2_p=.648$ (fewer errors with low task-load than high task-load; 9.84% vs. 14.01%). There was a reliable two-way interaction between modality and cue-presence; $F(2,92)=11.21$, $p=.001$, $\eta^2_p=.196$; and a three-way interaction between modality, cue-presence and task-load; $F(2,92)=4.99$, $p=.009$, $\eta^2_p=.098$. A simple main effect analyses investigated the nature of the latter interaction by examining each modality in isolation. It showed that with visual cues, there were no significant differences in detection error between cue-present and cue-absent trials, irrespective of task-load (all $p>.05$). With verbal modality, when the task-load was low, cue-present trials generated more errors than in cue-absent trials ($p=.01$), but when the task-load was high there was no effect of cue-presence on errors ($p>.05$). With auditory cues, cue-present trials had more errors than cue-absent trials with low as well as high task-load (all $p=.001$). Examining cue-present trial in isolation, with low task-load, there were differences in error rate between visual and auditory modalities ($p=.004$) and between verbal and auditory modalities ($p=.001$), but not between visual and verbal modalities ($p>.05$). But with high task-load, there was no effect of modality on error rates (all $p>.05$). Examining cue-absent trials in isolation, with low task-load, there was a significant error difference between visual and auditory modalities, and between verbal and auditory modalities, (all $p=.001$) but no difference between visual and verbal modalities ($p>.05$). With high task-load, in a similar fashion, there were significant differences in error rate between visual and auditory modalities, and between verbal and auditory modalities, ($p=.001$ for both) but not between visual and verbal modalities ($p>.05$). The error data are illustrated in Figure 3B.

Cost/gain in RT and error rate*:* In Experiment 1, there was no intra-trial modality shift; all stimuli were presented visually, while in Experiment 2 and 3, there was a modality shift within each trial. In Experiment 2 the cue was presented verbally (as printed words) while the search display was presented visually, hence a verbal to visual modality shift. In Experiment 3, the cue was presented aurally, while the search display was, again, presented visually, hence an auditory to visual modality shift. How did these modality shifts affect detection performance? To examine the overall changes in detection performance as a function of a modality shift in Experiment 2 and 3, in each experiment, RTs and error rates were subtracted from RTs and error rates for each matching condition in Experiment 1. These values could exhibit any gains or costs in RT/errors respectively. These values were submitted to two separate 3-way repeated measures ANOVAs with the factors being type of modality shift (i.e. difference in performance relative data with visual cue, when cue was verbal or when it was auditory, respectively), cue-presence and task-load; (for example each visual RT/error value minus verbal RT/error value, and visual RT/error value minus auditory RT/error value).

RT gain/cost: The ANOVA showed a significant main effect of cue-presence; $F(1,46)=6.82$, $P=.012$, $\eta^2_p=.129$ (an overall RT cost of 82 ms when cue was absent than present). There was no difference in cost/gain between verbal and auditory modalities ($p>.05$). There were a two-way interaction between cue-presence and task-load; $F(1,46)=4.76$, $P=.03$, $\eta^2_p=.094$, and a three way interaction between modality, cue-presence and task-load; $F(1,46)=4.71$, $P=.03$, $\eta^2_p=.93$. This three-way interaction was further examined by conducting a simple main effect analysis, which showed that there is a net detection RT cost of 30 ms when the cue was a word than a sound, but only on cue absent trials and under high task-load condition ($p<.05$). There were no other differences due to modality ($p>.05$).

Error gain/cost*:* The ANOVA on error rate differences showed a significant main effect of modality; $F(1,46)=29.51$, $P<.001$, $\eta^2_p=.391$(cues presented verbally - as printed

word - or aurally caused 5.1% fewer errors); as well as a significant main effect of cue-presence; $F(1,46)=14.52$, $P<.001$, $\eta^2_p=.244$ (on average 2.6% fewer errors on cue-present trials compared with cue absent trials). The ANOVA showed a significant two-way interaction between modality and cue-presence; $F(1,46)=5.6$, $P=.02$, $\eta^2_p=.109$, and a three way interaction between all three factors; $F(1,46)=10.73$, $P=.002$, $\eta^2_p=.189$. A simple main effect analysis showed that, when the cue was present, there were higher accuracy gain with auditory cues than with visual cues, but only under low task-load conditions (both $p<.001$). When the cue was absent, auditory modality there were accuracy gains under both low and high task-load (both $p<.001$).

Discussion

The comparative analysis of Experiment 1, 2 & 3 examined similarities and differences in information processing that could be attributed to cue-modality. The major finding was that here were no overall changes in detection speed in any conditions that could be attributed to the shift in cue modality. The analyses did find a reliable interaction between modality task-load or cue-presence. However, additional analysis revealed that these changes in detection speed were unrelated to the cue-modality, but due to the level of task-load and/or cue-presence. Several previous studies have reported an advantage for a direct visual match between cue and target (Wolfe, 2004). However, Potter (1975) who applied the RSVP paradigm using stimuli representing pictures of natural scenes (e.g. two men on a boat), could not find any such advantage for visual stimuli (images) over verbal associates (general name) of the target object. Potter reported as accurate and almost as rapid selection for verbal stimuli as visual stimuli. An ERP study by Nako, Smith & Eimer (2015) has reported a search advantage for visual cues over words cues during search for visual and word objects. Castelhano, Pallatsek and Cave (2008) ascribe this difference in performance to the typicality of the cue. When the cue is an exact picture, the search was more efficient (quicker first fixation) than when it was a basic-level category name (see also Wolfe, Horowitz, Kenner,

Hyle & Vasan, 2004). The comparative analysis of the speed of detection with different cue-modality could not corroborate Nako et al.'s findings. However, auditory cues generated more accurate detections, but only when the task-load was low. This accuracy gain in trials with auditory cues may be because in trials with visual or verbal cues, both the cue stimuli and the stimuli in the search display rely exclusively on the visual modality for processing of the stimuli information (as images or text). This may create some overload in visual processing, compared with trials with only auditory cues where there is no reliance on visual processing, hence fewer errors. It should be noted that contrary to this finding, Cohen, Horowitz and Wolfe (2009) found auditory recognition memory to be less accurate than visual recognition memory. However, the disagreement between studies might be due to the differences in type of stimuli used.

To summarize the findings of the comparative analysis, it seems evident that the differences in detection speed between experiments are due to the detrimental effect of task-load and/or cue-absence, rather than due to the modality through which the cue stimulus is processed. This implies that stimulus information can equally efficiently transfer between visual and verbal modalities as well as between auditory and visual modalities, without cue modality affecting the detection. It also shows that auditory cues tend to generate fewer detection errors than visual and verbal cues, possibly due to less overload interference in visual stimuli processing.

Exploring the overall changes in detection speed and accuracy showed that in experiment 2 and 3, where the cue was shown verbally or aurally, there was an overall delay in detection, compared with Experiment 1, especially when the cue was absent (4 ms for present trials vs. 40 ms for cue absent trials.). However, the study also found a slight gain in accuracy, especially when the cue was present. The accuracy advantage on cue present trial was ~0.4% vs. 5.1% for cue absent trials).

The costs in speed occurred on detection on absent trials in Experiments 2 and 3 where the initial cue was either verbal (a word) or auditory (a sound) and where the search display

contained only visual stimuli, compared to Experiment 1 where both the initial cue and the stimuli in search display were both presented visually. There are two possible explanations for the above phenomena. The speed disadvantage for "visual only" experiment may be attributed to the effect of stimulus similarity (Duncan & Humphries, 1989), while the accuracy gain on cue-absent trials in Experiment 2 and 3, might be due to a speed-accuracy trade off (Wood & Jennings, 1976; Fitts, 1966).

The cost of conducting two tasks concurrently is much larger when the two tasks tap the same modality, as opposed to when two or more different modalities are used (Cocchini et al., 2002; Fougnie & Marois, 2006; Luck & Vogel, 1997). This phenomenon might explain the observed lower detection error rates with auditory cues. In Experiment 1 and 2, the cue uses visual modality and the arithmetic equation rely on the visual (verbal-visual) modality while in Experiment 3, the cue is processed through auditory modality while the arithmetic task uses verbal-visual modality..

General Discussion

In three experiments different versions of the same stimuli set were used as cues to establish that the speed of WM-driven cue detection is unaffected by: (i) the modality through which the WM cue is presented and (ii) whether or not there is an intra-trial modality shift. The study also reported a detrimental effect of a concurrent cognitive task that tapped WM resources. Expanding on previous work (Balani et al., 2010, Kawashima & Matsumoto, 2017), this study discovered that information held in WM, either in verbal form (word) or auditory form (sound), can efficiently be transferred between modalities to complete a visual detection task for a representation of the initial WM-cue. The first experiment had no intra-trial modality shift, as both the cue and search stimuli were images. In Experiment 2, a transfer of information from verbally presented cue to a visual detection of the same cue (or its representation) took place. Likewise, in Experiment 3, there was an intra-trial shift from an auditory cue to visual search. One of the major findings of the study was the lack of

overall decline in the speed of detection in experiments that could be attributed to the effect of cue modality. There were no major differences between experiments in error rates, with one exception; the auditory cues generated slightly more accurate detection in experiment 3 compared with experiments 1 (visual cues) and 2 (with verbal cues). This accuracy advantage was explained by larger performance costs when the two concurrent tasks relied on the same modality, as opposed to when two or more different modalities were involved (Cocchini et al., 2002; Fougnie & Marois, 2006; Luck & Vogel, 1997). In Experiment 1 and 2, both visual and verbal information (in image or text form) relied exclusively on the visual modality to process and transfer cue information, which could generate some processing overload on the visual modality. With auditory cues, however, only the sound was processed and there was no visual interference.

The findings of the current study can be viewed from two distinct but related theoretical perspectives; Baddeley's WM model (1974, 2000) and Cowan's embedded processes model (1988, 1999, 2005; Cowan, Saults & Bulme, 2014). From the former point of view, lack of overall RT differences in cross-modal detection reported in the current study, may lend some support to the proposed supramodal nature of WM (Baddeley et al, 2009) and the notion that Episodic Buffer component of WM integrates all incoming information from different modalities (Baddeley, 2000; Baddeley et al.'s (2009). Episodic Buffer (Baddelley, 2000) is suggested to hold integrated episodes or chunks in a multidimensional code, and as multidimensional representations (Baddeley, 2012) and is assumed that the buffer would depend heavily on the Central Executive component. Another possible source of support is provided by the observed detrimental effect on cue-detection caused by the concurrent cognitive load that taps the WM-resources. Cognitive load can have detrimental effect on detection owing to one or more of the following reasons: (i) due to preventing an effective rehearsal of memory item, leading to decay of information; (ii) due to the notion of limited capacity in number of item than can be stored in WM (Cowan, 2001) or (iii) due to limitation in the focus of attention (Cowan, 1988, 1999). The reported detriment in detection

performance could be ascribed the effect of the concurrent mental arithmetic task, which taxes the limited resources of the central executive function of WM (Baddeley, 2000, 2009) resulting in slower detection when the cognitive task-load is high (with relatively harder arithmetic equation) than when it was low (with easier arithmetic equation).

An alternative point of view is Cowan's (1988, 1999; Cowan et al., 2014) model of WM storage and processing that involves a limited-capacity attentional focus that operates across areas of activated LTM. According to this framework, information from different sources and modalities is temporarily stored within a limited capacity focus of attention. The involvement of LTM which plays a crucial role in explaining some of the current findings. According to Cowan's (1988) model, focus of attention allows a coherent and organized interpretation of the very limited amount of information that managed to enter the system. Cowan (2001) also emphasizes the link between LTM and WM, and the link between central executive and episodic buffer (Baddeley, 2012). Involvement of LTM appears to ease the problem with the capacity limitation of WM and hence improve processing performance. Moreover, LTM plays a crucial role in modulating the effect of the concurrent arithmetic task.

The result indicating the effect of task-load corroborates previous studies by Harms (1991), Baldwin and Schieber (1995) as well as Briem and Hedman (1995); all suggesting that tasks involving WM impair performance on a concurrent primary task. Our findings can therefore, in two distinct ways, offer further support to the role of Episodic Buffer in the multimodal integration of information, as suggested by Baddeley et al. (2009).

Early research has demonstrated that guiding attention to an intended cue uses a combination of bottom-up and top-down attentional guidance (e.g. Treisman & Gelade, 1980; Wolfe & Horowitz, 2004, 2017) and studies have evidenced that cross-modal transfer of object information form one modality (visual) to another (verbal) can occur (Bahrami Balani, et al., 2011). The results from Experiment 2 of the current study demonstrated that verbal object representation can efficiently guide detection of a visual representation of that cue object. This was in line with findings by Balani et al. (2010, 2011) and a more recent study

by Kawashima and Matsumoto (2017), who also reported that visual and verbal representation of a WM-cue can guide attention. Furthermore, Experiment 3 provided new evidence for the existence of cross-modal effects from an auditory WM-cue similar to those from verbal modality. Taken as a whole, this could support to the notion that the effect of guidance from WM cue is coarser and more general, resulting in information being processed in a more abstract, top-down manner, and that this occurs both for verbal and auditory cue modalities. The nature of this cross-modal transfer of information reported here can be accounted for by Baddeley's (2000) revised WM model which suggested that, when searching for a target in a scene (e.g. a search display), representations of an object is kept in "episodic chunks" stored in a multi-modal Episodic Buffer that binds features from different sources and that the central executive function mediates the Episodic Buffer access to information from the Phonological Loop and the Visuo-Spatial Sketchpad (Baddeley, 2000; Baddeley, Hitch & Allen, 2009). This multidimensional integrating role of Episodic Buffer fits well with the results of the current study and provides fresh evidence in support of Baddeley et al.'s (2009) suggested notion of supramodal and domain-general nature of WM. This notion of the existence of such a modality-general attentional resource, was supported by Mesulam (1998) who suggested that associative sensory areas contain ‘‘road maps’’ for binding distributed information in different modalities. Further evidence was also provided by Amedi, von Kriegstein, van Atteveldt, Beauchamp and Naumer (2005) mapping the neural pathways for this phenomenon. Amedi et al. reviewed previous fMRI studies of human cross-modal object recognition which indicated that visual, tactile, and auditory information about objects could activate cortical association areas that were once believed to be modality-specific. Amedi et al. proposed instead a general mechanism for cross-modal object recognition that is dependent on of "associative nodes" for the identification stimuli (e.g. letters and common objects). According to this concept, the information is integrated to form more complex and flexible mappings between information from different modalities (Amedi et al., 2005).

The absence of overall changes in speed of detection in experiments 2 and 3 that could directly be attributed to a modality shift, is in disagreement with some previous research that has reported a latency cost involved in cross-modal detection task. The type of cross-modal cost that was reported in those studies was attributed to its effect on the ability to divide attention (Godefroy, Lhullier & Rousseaux ,1996) or the cost of shift of attention across modalities (LaBerge, 1973). In support of this attentional cost notion, other studies have reported similar cross-modal deficits in attentional shift or deficit in dividing attention between modalities in pathological cases. For example, Ciesielski, Knight, Prince, Harris and Handmaker, (1995) observed cross-modal attentional costs with autistics individuals and Hanewinkel & Ferstl, (1996) reported similar costs with schizophrenic patients. This divergence in the current study's findings from the abovementioned studies may be due to the difference in experimental design. In addition, the current study embedded an additional mental arithmetic task within the trial, which can *per se* have some delaying effect on all trials and therefore might mask any RT cost due to modality change.

Further finding of the current study indicated that the presence of the WM cue (or its representation) in the final search display could facilitate cue detection irrespective of the modality of the cue. Chun and Wolfe (1996) have reported similar facilitatory effect on target-present trials, which was explained by the idea that in a serial search task, as soon as the target is detected the search is abandoned, while in a target absent trials, all available stimuli have to be inspected before a decision could be made about the absence of the target.

One further aspect of the study that should be revisited is that the concurrent mental arithmetic task used the verbal domain and therefore, one might claim that there is an additional modality shift embedded in Experiments 1 and 3, compared to Experiment 2, and this might have affected the performance. The number of modality shifts varied across experiments. Experiment 1 required two modality shifts (visual => verbal => visual); Experiment 2 required only one shift (verbal => visual); and Experiment 3 required two shifts (auditory => verbal => visual). Nevertheless, there are two pieces of evidence that challenge

this claim. Firstly, the mental arithmetic task was part of all experiments, hence any such interference has remained constant across all experiments. Secondly, if there was any interfering effect of an additional modality shift, Experiment 2 should have benefited from having one fewer modality shift. This, however, did not happen. In fact, there was no overall differences in detection efficiency between the experiments, which invalidates the claim.

Two aspects of the current study make it different from previous work. Most prior studies have mainly explored the transfer of cross-modal information between verbal and visual modalities (e.g. Soto & Humphreys, 2007; Balani et al., 2010; Kawashima & Matsumoto, 2017) or haptic and visual modalities (Hadjikhani & Roland, 1988) and in lesser extent, of the auditory modality (Driver & Spence, 1998 & 2004; Maehara & Saito, 2007). But, there has been a lack of cross-modal research involving all three modalities (visual, verbal and auditory) in the same experiment using the same design, procedures, participants and same set of stimuli representing all three modalities. Keeping the stimuli same across all experiments (e.g. dog presented with its image, its printed name or its sound) helped to reduce any potential confounding effect of conceptual or semantic differences. In all experiments, the participants viewed and familiarized themselves with the stimuli set at the beginning. This, therefore, diminished any likely effect of individual differences in semantic knowledge about each stimulus (e.g. the person sees the image of the specific object – piano - and hears the single sound epoch representing that object). By doing so, the conceptual and semantic representation of the stimuli will remain the same for all participants and hence, the representations are less affected by their individual differences in their level of knowledge and familiarity with the stimuli as well as the availability of that information in LTM. Furthermore, use of complex real-world stimuli offered higher level of ecological validity to the study (Newell, 2004) and by using grayscale images, the effect of color saliency was diminished.

In conclusion, this study expanded previous knowledge on the effects of cross-modal WM-cue processing on cue-detection. The study suggests that cross-modal transfer of

information does take place, but any task-load on WM can have detrimental effect on the processing of cross-modal information in WM. It reports some novel effects of cognitive task-load on cross modal processing of cue and their subsequent detection in a search display. Verbal and auditory cues were processed relatively as efficiently as visual cues in guiding attention. Also, high cognitive task-load could slow the speed of detection, particularly when the cue was absent in search display. It also reports some speed-accuracy trade off on cue-absent trials. The detrimental effect of (high) WM task-load on processing information, irrespective of its modality, may be indicative of the involvement of WM structure (and particularly the episodic buffer) in integrating information originating from different modalities. Future designs should examine whether similar effects can be observed with other modalities than visual, verbal and auditory.

Though the current study can shed some light on the concept of cross-modal processing of information, these findings could, equally well, have been conceptualized within both Baddeley's (2000) WM model and Cowan's embedded processes framework (1988, 1999, 2005). The current challenge for both these formulations is that it seems hard to make solid predictions as to (i) what one would expect in a situation where there is a mixture of both executive and perceptional aspects of attention, together with (ii) where there is a switching between two modalities, and (iii) the extent to which it is necessary for the participant to hold instructions in the episodic buffer or central executive in order to perform the task. It is possible, for example, that the participant might simply activate a representation in LTM, to subsequently be triggered via the cue, rather than holding in WM. It remains to devise a detailed model of how all this might happen, but it will depend on the precise setup and the nature of the material and design, which makes the task yet harder.

Acknowledgement

The author would like to thank two anonymous reviewers for their helpful comments on previous version of this manuscript, and Alan Baddeley for valuable advice on a draft of this

manuscript, and Gary Jones for helping with the data collection. The author declares no conflict of interest.

References

Adams, E., Nguyen, A.T. & Cowan, N. (2018). Theories of Working Memory: Differences in Definition, Degree of Modularity, Role of Attention, and Purpose. *Language, Speech, and Hearing Services in Schools. 49 (*3), 340-355. https://doi.org/10.1044/2018_LSHSS-17-0114

Amedi, A., von Kriegstein, K., van Atteveldt, N.M., Beauchamp, MS. & Naumer, M.J. (2005). Functional imaging of human crossmodal identification and object recognition. *Experimental Brain Research, 166* (3-4), 559-571

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In Spence, K. W., & Spence, J. T. *The psychology of learning and motivation* (Volume 2). New York: Academic Press. pp. 89–195.

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

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

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4,* (11): 417-423.

Baddeley, A. (2012). Working Memory: Theories, Models, and Controversies. *Annu. Rev. Psychol. 63*:1–29

Baddeley, A.D., Allen, R.J. & Vargha-Khadem, F. (2010). Is the hippocampus necessary for visual and verbal binding in working memory? *Neuropsychologia, 48*:1089–95

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. A. Bower (Ed.), *Recent advances in learning and* motivation (Vol. 8, pp. 47–90). New York: Academic Press.

Baddeley, A., Hitch, G. J., & Allen, R. J. (2009). Working memory and binding in sentence recall. *Journal of* Memory and Language, 61(3), 438-456. DOI: 10.1016/j.jml.2009.05.004

Baddeley, A., Logie, R., Bressi, S., Della Sala, S. & Spinnler, H. (1986). Dementia and working memory, The *Quarterly Journal of Experimental Psychology Section A, 38*:4, 603 - 618. DOI: 10.1080/14640748608401616

Bahrami Balani, A. B., Soto, D., & Humphreys, G. W. (2011). Separating top-down and bottom-up cueing of attention from response inhibition in utilization behavior. *Neurocase, 18*(2), 98-111. doi.org/10.1080/13554794.2011.568496

Balani, A. B., Soto, D., & Humphreys, G. W. (2010). Working memory and target-related distractor effects on visual search. *Memory & Cognition*, *38(*8), 1058–1076. doi:10.3758/Mc.38.8.

Baldwin, C. L., & Schieber, F. (1995). Dual-task assessment of age differences in mental workload with implications for driving. *Proceedings of the Human Factors and Ergonomics Society 39th Annual Meeting*. Santa Monica, CA: Human Factors and Ergonomics Society. 167–171

Briem, V., & Hedman, L. R. (1995). Behavioural effects of mobile telephone use during simulated driving. *Ergonomics, 38*, 2536–2562.

Castelhano, M.S., Pollatsek, A., & Cave, K. (2008). Typicality Aids Search for an unspecified target, but only in identification, and not in attentional guidance. *Psychonomic Bulletin & Review*, 15(4), 795-801

Chelazzi, L., Duncan, J., Miller, E.K., & Desimone, R. (1998). A neural basis for visual search in inferior temporal cortex. *Nature, 363*:345–347.

Chun, M.M., & Wolfe, J.M. (1996). Just Say No: How are visual searches terminated when there is no target present? *Cognitive Psychology, 30* (1), 39-78. https://doi.org/10.1006/cogp.1996.0002

Ciesielski K.T., Knight, J.E., Prince, R.J., Harris, R.J., & Handmaker, S.D. (1995). Event-related potentials in cross-modal divided attention in autism. *Neuropsychologia, 33*:225-246.

Cocchini. G, Logie, R.H., Della Sala, S., MacPherson, S.E., & Baddeley, A.D. (2002). Concurrent performance of two memory tasks: evidence for domain-specific working memory systems. *Memory & Cognition. 30*(7): 1086-95. doi:10.3758/BF03194326

Cohen, M.A., Horowitz, T.S., & Wolfe, J.M. (2009). Auditory recognition memory is inferior to visual recognition memory. *Proc Natl Acad Sci U S A.;106*(14):6008-10. doi: 10.1073/pnas.0811884106.

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information processing system. *Psychological Bulletin, 104*, 163-191.

Cowan, N. (1999). An embedded-processes model of working memory. In *Models of Working Memory*, ed. A. Miyake, P Shah, pp. 62–101. Cambridge, UK: Cambridge Univ. Press

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

Cowan, N. (2005). *Working Memory Capacity*. Hove, UK: Psychol. Press

Cowan, N., Saults, J.S. & Blume, C.L. (2014). Central and peripheral components of working memory storage. *Journal experimental psychology: General, 143(5)*, 1806-1836, doi:1037/a0036814.

de Fockert J. W., Rees, G., Frith, C. D., & Lavie, N. (2001). *The role of working memory in visual selective attention. Science*, 291 (5509), 1803-1806

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci. 18*, 193–222

De Stefano, D., & LeFevre, J. A. (2004). The role of working memory in mental arithmetic. *Journal of Cognitive Psychology, 16*(3), 353-386.

Driver, J,. & Spence, C. (1998). *Crossmodal attention. Current Opinion in Neurobiology, 8,* 245–253

Driver, J., & Spence, C. (2004). Crossmodal spatial attention: Evidence from human performance. In C. Spence & J. Driver (Eds.), *Crossmodal space and crossmodal attention* (pp. 179-220). Oxford, UK: Oxford University Press.

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review, 96*(3), 433-458.

Easton, R.D., Greene, A.J., & Srinivas, K. (1997). Transfer between vision and haptics: Memory for 2-D patterns and 3-D objects. *Psychonomic Bulletin and Review, 4,* 403–410.

Eriksen, C.W. (1988). A source of error in attempts to distinguish coactivation from separate activation in the perception of redundant targets. *Perception & Psychophysics, 44*(2), 191-193.

Fitts, P. M. (1966). Cognitive aspects of information processing: III. Set for speed versus accuracy. *Journal of Experimental Psychology, 71*(6), 849-857.

Fougnie, D., & Marois, R. (2006). Distinct capacity limits for attention and working memory: Evidence from attentive tracking and visual working memory paradigms. *Psychological Science, 17,* 526–534. doi:10.1111/j.1467-9280.2006.01739.x

Gibson, J.J. (1962). Observations on active touch. *Psychol Rev. 69*:477–491. doi: 10.1037/h0046962.

Godefroy, O., Lhullier, C., & Rousseaux, M. (1996). Non-spatial attention disorders in patients with frontal or posterior brain damage. *Brain, 119,* 191-202.

Hadjikhani, N., & Roland, P.E. (1988). Cross-modal transfer of information between the tactile and the visual representations in the human brain: a positron emission tomographic study. *J. Neurosci.18*:1072–1084.

Hanewinkel, R, & Ferstl., R. (1996). Effects of modality shift and motor response shift on simple reaction time in schizophrenia patients. *J Abnorm Psychol, 105*:459-463.

Harms, L. (1991). Variation in drivers' cognitive load. Effects of driving through village areas and rural junctions. *Ergonomics, 34*(2), 15 1 - 160.

Kane, M. J., Poole, B. J., Tuholski, S. W., & Engle, R. W. (2006). Working memory capacity and the top-down control of visual search: Exploring the boundaries of "executive attention." *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 749–777. doi:10.1037/0278-7393.32.4.749

Kawashima, T. & Matsumoto, E. (2017). Cognitive Control of Attentional Guidance by Visual and Verbal Working Memory Representations. *Japanese Psychological Research. 59*:1, 49–57.

LaBerge, D. (1973) Identification of the time to switch attention: A test of a serial and a parallel model of attention. In S. Kornblum (Ed.), *Attention & Performance IV,* New York: Academic Press, 1975. pp 71–85,

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390,* 279–281. doi:10.1038/36846

Luria, .R, & Vogel, E.K. (2011). Visual search demands dictate reliance on working memory storage. *J Neurosci. 31*(16):6199–6207. http://dx.doi.org/10.1523/JNEUROSCI.6453-10.2011

Maehara, Y., & Saito, S. (2007). The relationship between processing and storage in working memory span: Not two sides of the same coin. *Journal of Memory and Language, 56*(2), 212-228. doi:10.1016/j.jml.2006.07.009

Maxfield, J.T., Stadler, W. & Zelinsky, G.J. (2014). The Effects of Target Typicality on Guidance and Verification in Categorical Search. *Journal of Vision, 13*(9):524–524.

Mesulam M.M. (1998). From sensation to cognition. *Brain 121*:1013-1052.

Nako, R., Smith, T.J. & Eimer, M. (2015). Activation of new attentional templates for real-world objects in visual search. *Journal of Cognitive Neuroscience, 27 (5),* 1–11. doi:10.1162/jocn_a_00747

Newell, F.N. (2004). Cross-modal object recognition. In G.A. Calvert, C. Spence, & B.E. Stein (Eds.), *The handbook of multisensory processes.* Cambridge, MA: MIT Press.

Olivers, C. N. L., Meijer, F., & Theeuwes, J. (2006). Feature-Based Memory-Driven Attentional Capture: Visual Working Memory Content Affects Visual Attention. *Journal of experimental psychology: human perception and performance*, *32*(5), 1243-1265.

Potter, M.C. (1975). Meaning in Visual Search. *Science,187*(4180): 965-6

Potter, M.C. (1993). Very short-term conceptual memory. Memory & Cognition. 21, 156–16110.3758/BF03202727

Potter, M. C. (2012) Conceptual short term memory in perception and thought. *Frontiers in Psychology 3*, 113. doi: 10.3389/fpsyg.2012.00113

Rapp, B., & Hendel, K.S. (2003). Principles of cross-modal competition: Evidence from deficits of attention. Psychonomic Bulletin & Review, 10 (1), 210-219

Ricker, T.J. & Cowan, N. (2014). Differences between presentation methods in working memory procedures: a matter of working memory consolidation. *J. of Exp. Psychol. Learning Memory &Cognition. 40*: 417–428.

Ricker, T.J., Nieuwenstein, M.R., Bayliss, D.M. & Barrouillet, P. (2018). Working memory consolidation: insights from studies on attention and working memory. *Ann. N.Y. Acad. Sci. 1424*, 8–18, doi: 10.1111/nyas.13633

Schneider, T.R., Engel, A.K., & Debner, S. (2008). Multisensory Identification of Natural Objects in a Two-Way Crossmodal Priming Paradigm. *Experimental Psychology, 55*(2) 121-131. DOI 10.1027/1618-3169.55.2.121

Shallice T., & Burgess P. W. (1993). Supervisory control of action and thought selection, in *Attention Selection Awareness and Control - A Tribute to Donald Broadbent*, Eds Baddeley A., Weiskrantz L., editors. (Oxford: Clarendon Press), 171–187.

Soto, D., Heinke, D., Humphreys, G.W. & Blanco, M.J. (2005). Early, involuntary top-down guidance of attention from working memory. *Journal of Experimental Psychology: Human Perception and Performance, 31*, 248-261

Soto, D., & Humphreys, G.W. (2007). Automatic guidance of visual attention from verbal working memory. *Journal of Experimental Psychology Human Perception and Performance 33* (3), 730-737.

Soto, D., & Humphreys, G.W. (2008). Stressing the mind: The effect of cognitive load and articulatory suppression on attentional guidance from working memory. *Perception & Psychophysics, 70* (5), 924-934, doi: 10.3758/PP.70.5.924

Soto, D., Humphreys, G. W., & Rotshtein, P. (2007). Dissociating the neural mechanisms of memory-based guidance of visual selection. *Proceedings of the National Academy of Sciences USA, 104* 17186–1719

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

Vergauwe, E., Camos, V., & Barrouillet, P. (2014). The impact of storage on processing: How is information maintained in working memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40,* 1072–1095. http://doi.org/10.1037/a0035779.

Ward, L. M. (1994). Supramodal and modality-specific mechanisms for stimulus-driven shifts of auditory and visual attention. *Canadian Journal of Experimental Psychology, 48,* 242–259

Ward, L.M., McDonald, J.J., & Lin, D. (2000). On asymmetries in cross-modal spatial attention orienting. *Attention, Perception, & Psychophysics 62* (6), 1258-1264

Wolfe, J. M. (1998). Visual search. In H. Pashler (Ed.), *Attention* (pp. 13–74). Hove, East Sussex, UK: Psychology Press Ltd.

Wolfe, J.M., & Horowitz, T.S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience, 5* 1-7

Wolfe, J.M. & Horowitz, T.S. (2017). Five factors that guide attention in visual search. *Nature: Human Behavior, 1,* 0058.

Wolfe, J.M., Horowitz, T.S., Kenner, N., Hyle M., Vasan, N. (2004). How fast can you change your mind? The speed of top-down guidance in visual search. *Vision Research, 44* 1411-1426.

Wood, C.C. & Jennings, J.R. (1976). Speed-accuracy tradeoff functions in choice reaction time: Experimental designs and computational procedures. *Perception & Psychophysics 19 (*92). https://doi.org/10.3758/BF03199392

Wyble, B., Potter, M. C., Bowman, H., & Nieuwenstein, M. (2011). Attentional episodes in visual perception. Journal of Experimental Psychology. General, 140(3), 488-505. doi: 10.1037/a0023612

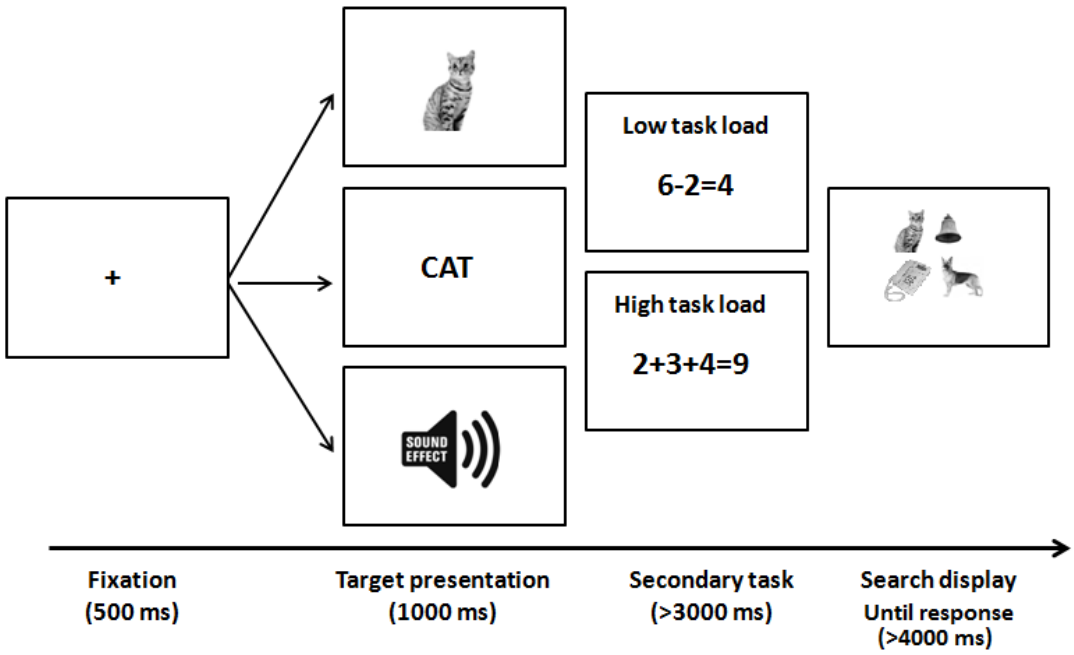


Figure 1: The Schematic depiction of a typical trial in experiments 1-3: Experiment 1 - uses a visual cue item (target for later detection) (A): the cue is a picture of an animal or object. Experiment 2 - uses a verbal cue (B): the cue is a word representing the name of the animal or object. Experiment 3 - uses an auditory cue (C): the cue is the sound of the animal or object. In low task-load condition, following the presentation of cue, a two-digit mental arithmetic equation (addition or subtraction) is shown and in high task-load condition, a 3-digit arithmetic equation is shown. The participant has to verify the accuracy of the equation within 3 seconds. The product of each equation is always less than 9. The arithmetic equation is followed by the search display, where the participant has to verify the presence/absence of the initial cue (or its visual representation) in the search display as quickly and accurately as possible within 4 seconds. The illustration above displays an example of a typical cue-present trial. Both arithmetic equations are correct, and the cue item is present in the search display. The images shown here are for illustration purpose only and the scaling of the illustrations does not reflect the actual relative sizes of the experiment stimuli.

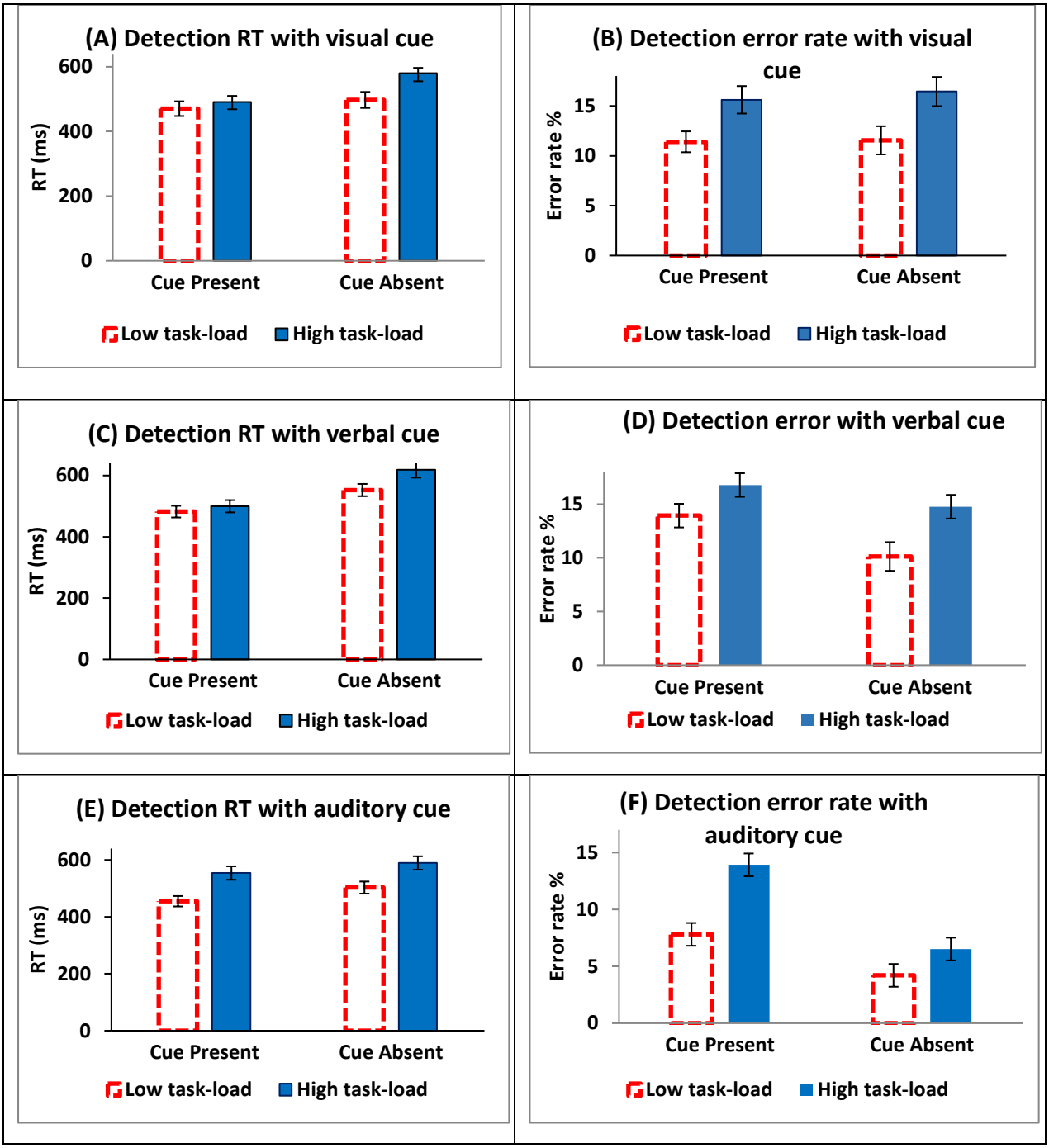
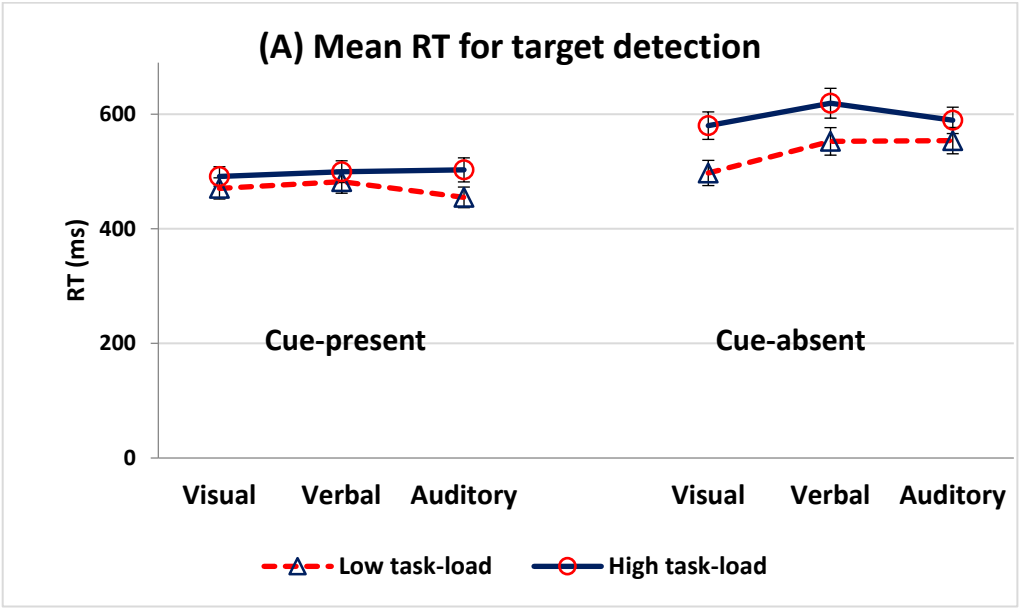


Figure 2: Mean detection RT and error rates as a function of task-load and cue-presence in Experiments 1-3, shown separately for cue-present and cue-absent trials. Graphs A and B show RT and error data, respectively, for Experiment 1 with visual cues. Graphs C and D show RT and error data, for Experiment 2 with verbal cues. Graphs E and F show RT and error data, for Experiment 3 with auditory cues.



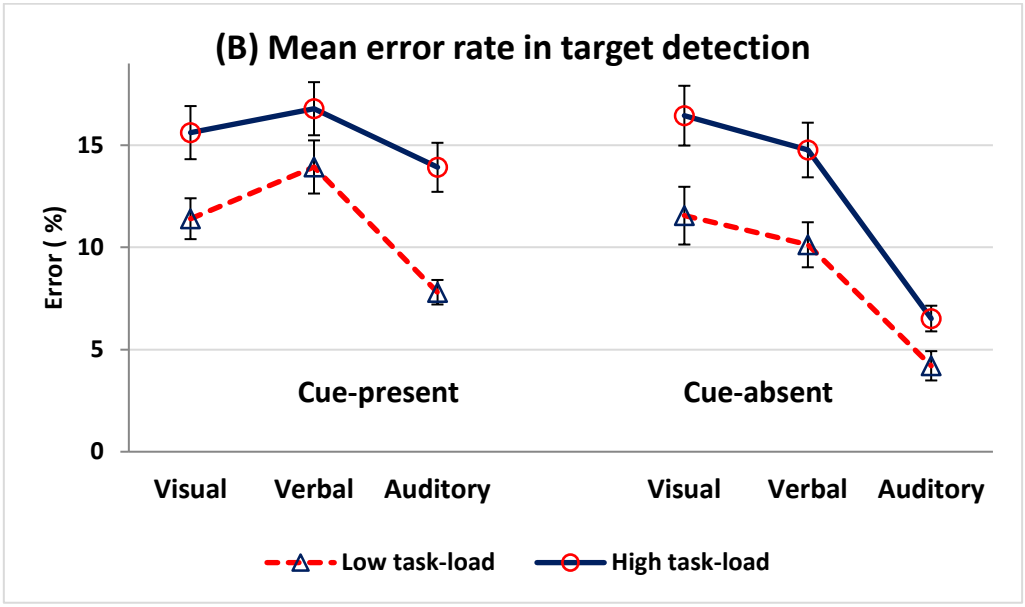


Figure 3: Cross-modal comparison between RTs (Graph A) and error data (Graph B) for the detection of the initial cue in a search display following a visual, a verbal and an auditory cue presentation, both when the task-load is low (red dotted lines) and when it is high (blue continuous lines). Data are presented separately for cue-present (on the left) and cue-absent trials (on the right).