BRILL

Multisensory Research 27 (2014) 91–110

brill.com/msr

Divided Multimodal Attention
Sensory Trace and Context Coding Strategies in Spatially Congruent Auditory and Visual Presentation

Tómas Kristjánsson *, **Tómas Páll Thorvaldsson and Árni Kristjánsson**

Laboratory for Visual Perception and Visuomotor Control, Faculty of Psychology,
School of Health Sciences, University of Iceland, Oddi, 101 Reykjavík, Iceland

Received 7 November 2013; accepted 10 June 2014

Abstract

Previous research involving both unimodal and multimodal studies suggests that single-response change detection is a capacity-free process while a discriminatory up or down identification is capacity-limited. The trace/context model assumes that this reflects different memory strategies rather than inherent differences between identification and detection. To perform such tasks, one of two strategies is used, a sensory trace or a context coding strategy, and if one is blocked, people will automatically use the other. A drawback to most preceding studies is that stimuli are presented at separate locations, creating the possibility of a spatial confound, which invites alternative interpretations of the results. We describe a series of experiments, investigating divided multimodal attention, without the spatial confound. The results challenge the trace/context model. Our critical experiment involved a gap before a change in volume and brightness, which according to the trace/context model blocks the sensory trace strategy, simultaneously with a roaming pedestal, which should block the context coding strategy. The results clearly show that people can use strategies other than sensory trace and context coding in the tasks and conditions of these experiments, necessitating changes to the trace/context model.

Keywords

Multimodal, visual attention, auditory attention, spatial confound

1. Multimodal Attention

A major focus of multimodal research has been the study of multimodal attention (Koelewijn *et al.*, 2010). A key question is whether there is a supramodal

---

* To whom correspondence should be addressed. E-mail: tok1@hi.is

DOI:10.1163/22134808-00002448

attentional process or whether there are independent attentional processes for each modality. This question often takes the form of whether there is a cost of dividing attention between different modalities (Alais *et al.*, 2006). Some researchers have concluded that each modality has an independent pool of attentional resources (Alais *et al.*, 2006; Ferlazzo *et al.*, 2002; Larsen *et al.*, 2003; Shiffrin and Grantham, 1974; Talsma *et al.*, 2006), while others have found significant cross-modal effects of dividing attention between modalities (Driver and Spence, 1994; Koelewijn *et al.*, 2009; Spence and Driver, 1997a; Spence *et al.*, 2001).

Advances in brain imaging technology have made it possible to look at those attentional processes in real time. This has sparked a renewed interest in multimodal attentional research. A lot of evidence indicates that the superior calculus (SC) plays a key role in multimodal integration (Holmes and Spence, 2005; Stein and Meredith, 1993). However this improved technology has also made it clear that feed-forward models, proposed by early researchers of multimodal integration in the brain, are overly simplistic. Instead, complex interactional models, where information from higher brain areas is fed back, influencing early processing areas, have been developed (Driver and Spence, 2000; Vroomen and De Gelder, 2000). For example, Stein *et al.* (2002) showed that when input from the cerebral cortex to SC is reduced or cut off, the SC will not integrate multisensory stimuli but process them separately (see Talsma and Woldorff, 2005; Talsma *et al.*, 2006 for exciting developments from ERP and SSVEP measurements).

2. Methodological Concerns

Spence and Driver (1997b) wrote a comprehensive critique of the methodology prevalent in multimodal attention research. They summarised five common flaws; one of those, involving the spatial cuing effect, often caused problems in this field. Spatial cuing confounds occur when the signals in the two modalities appear in two spatially separate locations making it impossible to know whether attention is divided between the two modalities or the two locations where the signals originate. An example of this can be found in Hafter *et al.* (1998). There, the visual stimuli appeared on a computer screen while the auditory stimuli were presented *via* headphones. This creates a spatial confound, a potentially serious flaw, as people invariably shift their attention to an expected target location (Klein *et al.*, 1992). An additional problem is that attending to separate locations in different modalities has been shown to be less efficient than attending to a single location (Driver and Spence, 1994; Spence and Driver, 1996).

3. Models of Multimodal Divided Attention

Dual-task paradigms are convenient for studying divided attention since they force participants to divide attention between the two tasks (Bonnel and Hafter, 1998; Kristjánsson *et al.*, 2004). In many studies, detection and identification are tested. Detection requires participants to respond to any change in a signal, regardless of the direction or the strength of that change. Identification requires participants to indicate the direction of the change, for example whether a light or a tone increases or decreases in strength. It has long been argued that detection is capacity-free since most early research showed that detection performance in dual-task experiments did not differ from performance on a single task (Alwitt, 1981). Identification has been thought to be capacity-limited since experiments have repeatedly shown that performance drops when attention is divided between the two modalities, compared to single-task performance (Spence and Driver, 1997a). From here on, we will refer to these assumptions as the standard model.

Hafter *et al.* (1998) challenged the standard model by showing that in a divided multimodal attention task, costs can occur in detection and identification can be performed without the cost normally associated with it. In their design a 900 ms pedestal of sound and light was presented; in the middle of the pedestal the sound or the light could briefly increase or decrease in strength. Their initial results were as expected, detection being capacity-free and identification capacity-limited. However when a short gap was introduced before and after the signal, performance both for detection and identification showed costs of divided attention. Moreover Hafter *et al.* (1998) showed that by using a roaming pedestal, neither detection nor identification showed costs of divided attention. These findings violated the predictions of the standard model, which predicts identification costs.

Hafter *et al.* (1998) suggested that two memory processes or strategies are responsible for whether costs are observed, and that there is no fundamental difference between detection and identification. They proposed that participants use a sensory trace strategy in detection and a context coding strategy in identification. The sensory trace strategy involves using transient information from a continuous signal. A gap before and after the signal eliminates information from transients and this forces observers to use a strategy of categorising the different levels of the signal as either up, down or no signal (context coding). Furthermore, they claimed that a roaming pedestal, where the strength of the pedestal varies from trial to trial, prevents the use of the context coding strategy and participants use sensory traces instead. We will refer to these assumptions as the trace/context model.

When the standard model and the trace/context model are compared, several different predictions arise. In detection tasks with a gap before and after the signal the standard model predicts no cost of divided attention, since detection is assumed to be capacity-free, while the trace/context model predicts a cost, since a gap should force participants into using the context coding strategy. The opposite prediction would be made in identification tasks with a roaming pedestal. The trace/context model assumes that there are only two possible signal processing strategies. If there would be more strategies possible in their model then the assumption would not hold that by blocking the sensory trace strategy participants automatically change to the context coding strategy, since participants could then utilize a third strategy. If there are only those two strategies, and the sensory trace strategy can be blocked by using a gap and the context coding strategy can be blocked by using a roaming pedestal, it follows that performance should deteriorate if the possible use of both strategies is simultaneously blocked.

Here we repeat the experiments of Hafter *et al.* (1998) without the spatial confound, in addition to investigating conditions where both sensory trace and context coding strategies are blocked. Although Hafter *et al.* showed convincingly that the standard model did not apply in all situations, their assumption of only two memory strategies was not as strongly supported, so we investigated the possibility that even when both strategies are blocked, by using both gaps and roaming pedestals, performance will not fall significantly under the ideal cost curve. This would argue against the trace/context model and suggest that other strategies can be used. We repeated the gap and the roaming pedestal conditions, both to serve as a baseline for our main question and to test the effects of gaps and roaming pedestals on divided multimodal attention without a spatial confound.

4. Method

4.1. Subjects

Nineteen participants participated in the experiments. Four subjects participated in all three experiments and 15 participants in one experiment each, making the number of participants in each experiment nine. The mean age was 25.6 years (range: 18–30). Twelve participants were male and seven female. Two of the participants in each experiment were the first two authors; the other participants were recruited through the University of Iceland. No performance differences were found between the authors and the naïve participants in any of the experiments. All participants reported normal hearing and normal, or corrected to normal, vision.

4.2. Apparatus and Materials

4.2.1. General

The experiments were conducted in an IAC audiology room with a background luminance of 22.1 cd/m^2. The participants were seated in the middle of the room, 120 cm away from a grey wall. The participants held a keyboard, which they used for responding. A wooden frame, 118 cm in length and 50 cm in height, was situated on the wall that participants faced. On each corner of the frame, a speaker was placed. Glued to the front of each speaker was a light unit consisting of four, super-bright green LED diodes, enclosed in black cylinders, with a CREE XML, 35 mm optic diffused lens covering the top, facing the participants. A single green LED diode was placed in the middle of the frame, serving as a fixation light. An overview of the experimental set-up can be seen in Fig. 1.

4.2.2. Experiment 1

The sound pedestal strength was 59 dB (SPL), the light pedestal strength was 170 cd/m^2. During detection the average sound signal was 5.6 dB (SPL) and the average light signal was 138 cd/m^2, up or down from the pedestal. The

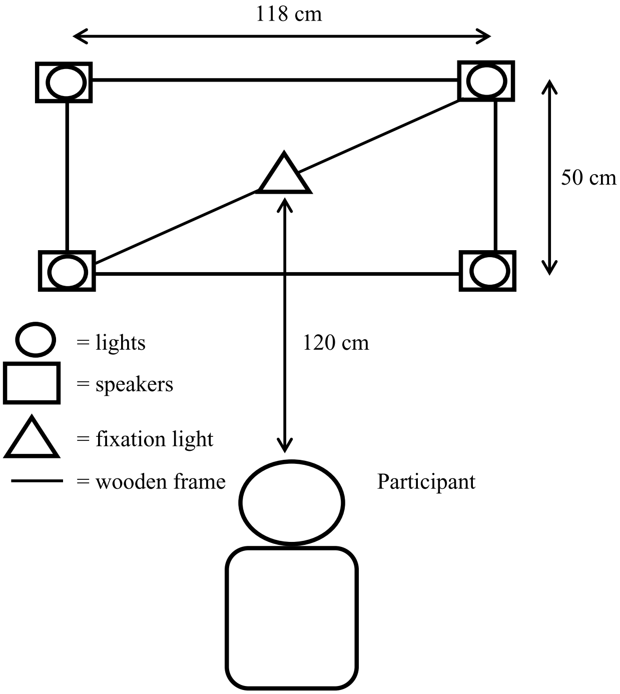

Figure 1. Schematic overview of the experimental set-up.

average sound signal was 6.0 dB (SPL) and the light signal 139 cd/m^2 during identification.

4.2.3. Experiment 2

The weakest sound signal was 42.9 dB (SPL) and the strongest 63.1 dB (SPL). The weakest light signal was 65 cd/m^2 and the strongest was 260 cd/m^2. The average signal during the detection task was 1.7 dB (SPL) for sound and 76 cd/m^2 for light. For identification, the average signal was 3.4 dB (SPL) for sound and 109 cd/m^2 for light.

4.2.4. Experiment 3

The weakest and strongest signals were the same as in Experiment 2. The average sound signal for detection was 5.8 dB (SPL) and 83 cd/m^2 for the light signal. For identification the average sound signal was 5.9 dB (SPL) and 88 cd/m^2 for the light signal.

4.3. Design

There were three within-subject factors, target modality (sound and light), response mode (detection and identification) and instructions about how much attention to devote to each modality [100% sound (s), 100% light (l), 80% (s)/20% (l), 50% (s)/50% (l) and 80% (l)/20% (s)]. The percentages stand for how much attention participants were instructed to pay to each modality. Each participant had several practice blocks, each consisting of ten trials. The signal strength was adjusted after each one until performance was consistently between 70–80% correct. These practice trials were conducted for sound and light separately and used to determine the signal strength for the experimental trials. This was done both for detection and for identification. The experimental trials consisted of 40 trials of 100% light, 40 trials of 100% sound, 100 trials of 80% (s)/20% (l), 100 trials of 80% (l)/20% (s) and 60 trials of 50% (s)/50% (l). The difference in the number of trials between conditions was to ensure sufficient numbers of sound and light trials for data analysis. The same number of trials was used for detection and identification, with half of the participants performing the identification trials first and the other half the detection trials first. During the detection trials there was a 25% chance of the signal going up, 25% chance of the signal going down and 50% chance of there being no signal. During the identification trials there was a 50% chance of the signal going up and 50% chance of the signal going down.

4.4. Procedure

4.4.1. General

Before each session, participants were told the likelihood of the target modality and were instructed to pay attention accordingly. For example, in the 100% light sessions they were told to focus exclusively on the light and that all the signals would be visual. In the 80% light/20% sound dual task session, they were told that 80% of the signals would be visual and 20% auditory. Participants participated in the 100% sessions first, in which they got to know the procedures and the nature of the signals. Five hundred ms before the start of each trial the fixation light was illuminated. Then, the sound and light appeared simultaneously. The total trial time was 900 ms. The participants had 1000 ms to respond, from the time the signal ended. After the second pedestal, all lights and sounds were turned off until the next trial started. During detection, participants responded by pressing the space bar on a keyboard if they thought a change had occurred in either the visual or the auditory modality, but if not, they refrained from responding. During identification, participants responded by pressing the up arrow on the keyboard if they thought the signal had gone up and the down arrow if they thought the signal had gone down. Participation took 1 h and 48 min on average. The signal types are shown in Fig. 2.

4.4.2. Experiment 1

A 380 ms pedestal was followed by a 50 ms gap before the 40 ms signal (in the signal trials) in either the sound or the light modality, then followed by another 50 ms gap and a 380 ms pedestal. The inter-trial interval, that is the time between the offset of the stimuli until the lighting of the fixation light, was 570 ms.

4.4.3. Experiment 2

There was no gap between the pedestal and the signal in Experiment 2 and therefore the pedestal and the signal formed a continuous 900 ms stimulus. When the signal appeared there was a 5 ms onset ramp leading to, and a 5 ms offset ramp leading from the signal. The signal itself was 50 ms, and the pre- and post-signal pedestals were 420 ms. When there was no signal the pedestal was played continuously for 900 ms. The pedestal varied randomly in strength between each trial, independently for sound and light. The inter-trial interval was 575 ms.

4.4.4. Experiment 3

In Experiment 3 the pre- and post-signal pedestals were 375 ms with a 50 ms gap before and after a 50 ms signal. All other aspects of the procedure were the same as in Experiment 2.

5. Experiment 1

In this study we investigate the cost of divided multimodal attention with gaps before and after the signal. This study utilises a methodology adapted from Spence and Driver (1997b) where sound and light are presented from the same location. The trace/context model predicts a cost of divided attention both in

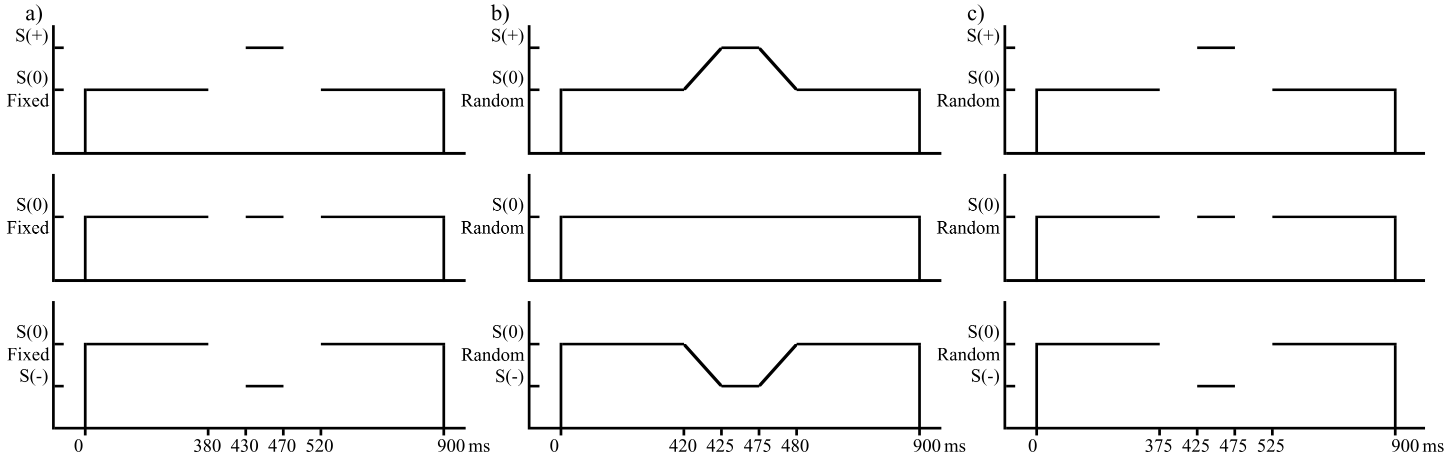

Figure 2. Diagram of the different signals in Experiments 1, 2 and 3. The top panel of each picture (S(+)) shows the signal when strength increased, the middle panel, (S(0)) shows the no-signal trials and the bottom panel (S(−)) shows the signal when strength decreased. Panel a) shows the stimuli for Experiment 1, where gaps preceded and followed the signals but pedestal levels were fixed. Panel b) shows the stimuli for Experiment 2, where pedestal levels were random and panel c) shows the stimuli for Experiment 3 where gaps and random pedestals were used.

detection and identification while the standard model predicts a cost in identification but no cost in detection.

5.1. Results

We calculated d' for each participant in each condition and those d' values were then averaged and plotted in an AOC graph. The d' values for detection were calculated with a standard yes–no formula, $d' = (z(\text{Hit}) - z(\text{FA}))$ and $C = -(z(\text{Hit}) + z(\text{FA}))/2$ while identification data was treated as 2AFC data and therefore calculated with $d'_{\text{FC}} = (z(\text{Hit}) - z(\text{FA}))/\sqrt{2}$ and $C = -(z(\text{Hit}) + z(\text{FA}))/\sqrt{2}$ (Kingdom and Prins, 2010; McNicol, 1972). Zero values for false alarms or misses were adjusted in accordance with suggestions from Macmillan and Creelman (2004). The AOC graph describes joint performance as a function of attentional instructions (Hafter *et al.*, 1998). Figure 3 shows the AOC graph for Experiment 1. The dotted lines describe performance of an ideal participant. The d' values should cluster around the corner where the straight lines meet, if divided attention is capacity free. This has been called the independence point. The curved line is an ideal line, which d' values should fall close to if divided attention is capacity limited.

As can be seen in panel a) in Fig. 3, the squares fall on or below the curved ideal limited capacity line. This is in accordance with both the trace/context model and the standard model which both predict a cost of divided attention in identification with gap conditions. However the circles that represent the detection condition fall between the ideal cost curve and the independence point. This makes interpretation difficult as the trace/context model predicts a cost while the standard model predicts no cost in this condition. We therefore plot performance separately for light and sound in panel b).

The results for the detection condition become much clearer in panel b) of Fig. 3. For detection of sound the 50% point is higher than the 80% point and the 20% point is lowest of all. This creates a small downward trend in d' values, representing a cost of divided attention, confirmed by ANOVA analysis (see below). However there is no such downward trend in the graph for the detection of light, indicating no cost of divided attention. The bias was measured as the criterion, C. For detection, $C = 0.16$ for sound and $C = 0.25$ for light. For identification, $C = -0.37$ for sound and $C = -0.12$ for light. This indicates a small bias towards not responding during detection and that this bias was stronger for sound than light. There was a substantial bias towards responding ‘up’ for sound with a smaller bias, in the same direction for light. A significance test (ANOVA) confirmed that for detection, there was a cost of divided attention for sound $[F(3, 24) = 3.00,\ p = 0.05]$ but not for light $[F(3, 24) = 1.14,\ p = 0.35]$. For identification a significance test confirmed a cost, both for sound $[F(3, 24) = 7.52,\ p \leqslant 0.01]$, and for light $[F(3, 24) = 4.98,\ p \leqslant 0.01]$.

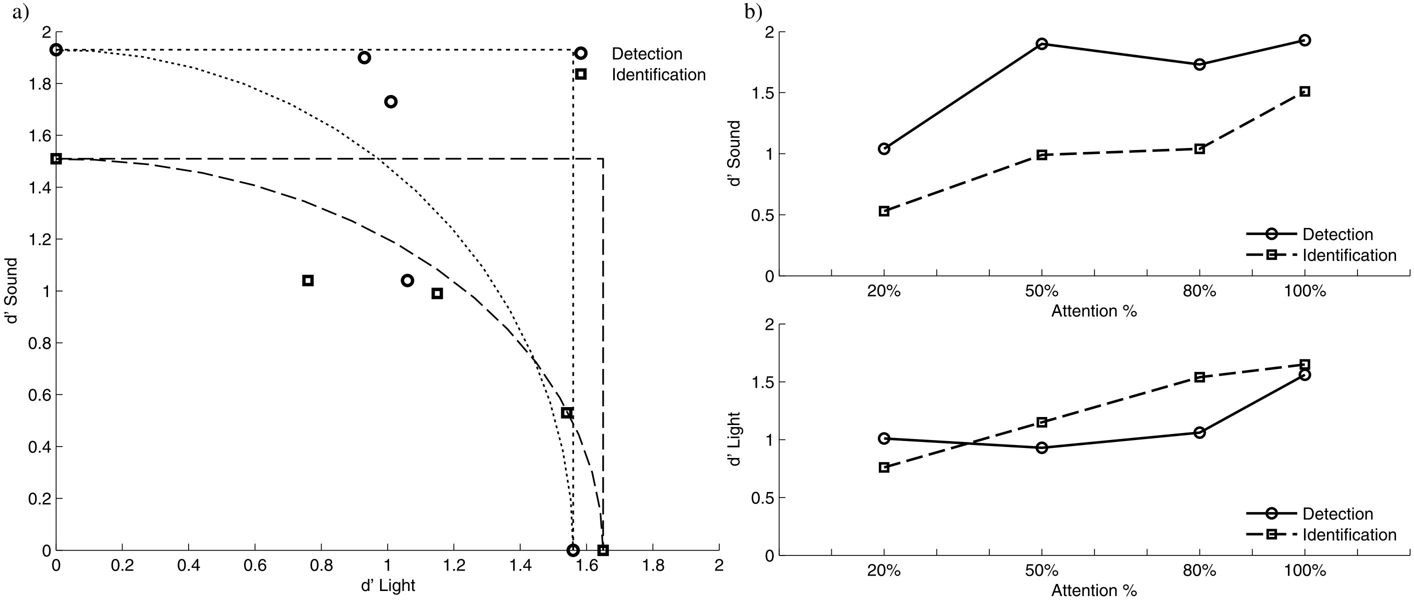

Figure 3. The result of Experiment 1. Panel a) shows the AOC graph for Experiment 1. The curved lines represent ideal performance if there is a cost of divided attention while the point where the straight lines meet represents the independence point, where d' values should cluster around if there is no cost of divided attention (cf. Hafter *et al.*, 1998). The short-dotted lines connecting the circles on the x- and y-axes represent the ideal lines for detection while the long-dotted lines connecting the squares on the axis are the ideal lines for identification. The squares and circles represent different attentional instructions, one marker for each instruction. The highest markers (squares and circles) of each type on the y-axis represent performance in 100% sound condition (single task) then in descending order on the y-axis the markers represent the dual-task conditions 80% sound/20% light, 50% sound/50% light, 20% sound/80% light. The markers on the x-axis represent performance in the 100% light condition (single task). Panel b) shows separated graphs for light and sound. The attentional percentages on the x-axis represent how much attention participants were instructed to pay to the modality and the likelihood of the signal occurring in that modality.

5.2. Discussion

Experiment 1 served the purpose of investigating the costs of divided attention in the paradigm used by Hafter *et al.* (1998), but this time without the spatial confound that might have affected their results (cf. Spence and Driver, 1997b). As expected, identification showed a cost of divided attention, in accordance with both the trace/context model and the standard model. This is a good indicator that despite the methodological changes from previous experiments, the same effect was measured. The results from the detection condition were more ambiguous. Performance for the sound signals showed a cost, which fits nicely with the trace/context model. Performance for the light signals, however, showed no cost of divided attention, which would fit better with the standard model. This could be the result of the light signal being easier or harder than the sound signal and that the results show a floor or ceiling effect. This is unlikely however, as d' values were close to one, indicating a performance of around 75% correct.

6. Experiment 2

In Experiment 2, the roaming pedestal condition from Hafter *et al.* (1998) was recreated, but importantly, again without the spatial confound (Spence and Driver, 1997b). As explained above, no cost of divided attention would be predicted by the trace/context model in either detection or identification, while according to the standard model we should expect no cost for detection but a cost for identification.

6.1. Results

As in Experiment 1 an AOC graph was plotted from the average d' values. Panel a) in Fig. 4 shows that for detection there is a very small cost of divided attention although the point representing 80% light/20% sound falls on the ideal cost curve. The three detection points do not follow the cost curve but do not quite reach the level of independence, indicating a minor cost of divided attention. The results for identification are harder to read from the AOC graph as two points fall above the ideal cost curve while one point falls below. Panel b) in Fig. 4 separates performance for light and sound to see how the cost or no cost appears in each modality. Panel b) shows that in the detection condition there is little or no cost of divided attention between 100%, 80% and 50% attention in either modality. However there is a significant drop in performance when 20% of attention is paid to either modality.

For identification there is a clear cost of divided attention for light but no such cost for sound. For detection $C = 0.28$ for sound and $C = 0.35$ for light, while for identification $C = -0.18$ for sound and $C = -0.26$ for light.

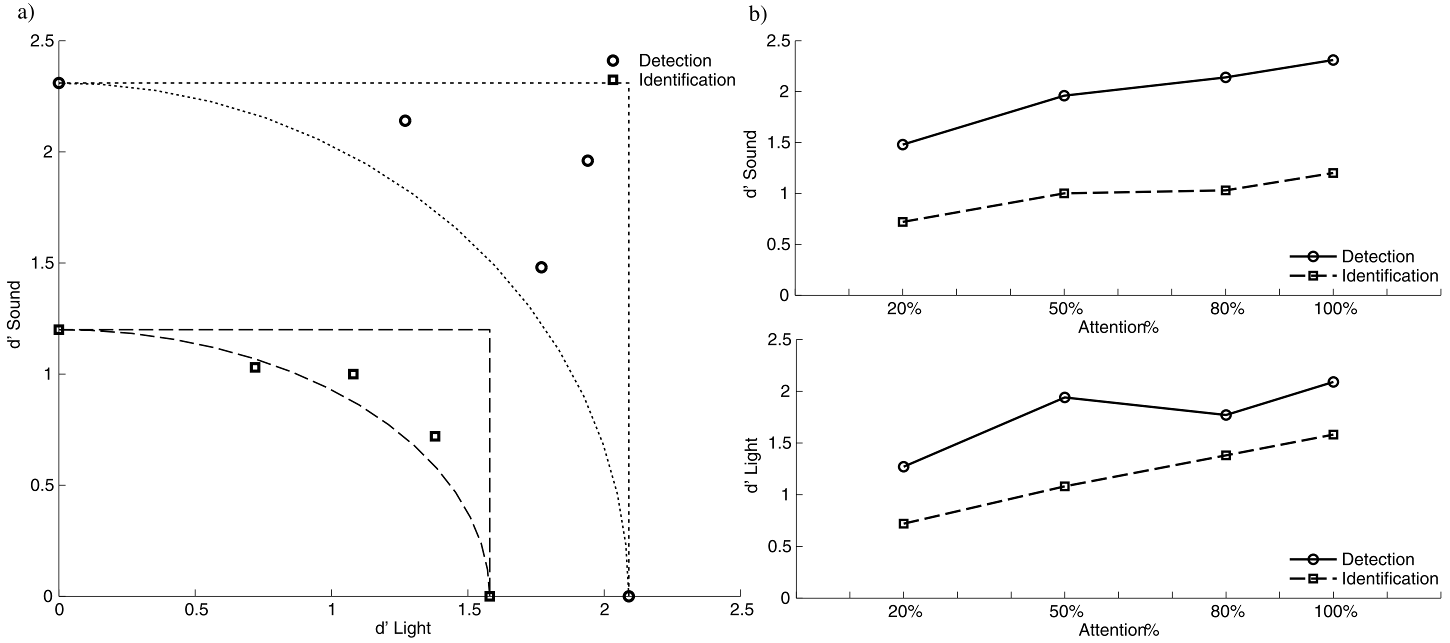

Figure 4. The results of Experiment 2. The AOC graph in panel a) uses the same ideal lines and markers as in panel a) in Fig. 3. Panel b) separates performance for light and sound. Axes and lines in panel b) are the same as in panel b) in Fig. 3.

This indicates no-response bias in detection and an ‘up’ bias in identification, with a stronger bias for light signals in both conditions. A significance test showed that in detection, sound did not show a cost of divided attention, $[F(3,24) = 1.87,\ p = 0.16]$, however there was a significant effect found for light $[F(3,24) = 3.58,\ p = 0.02]$. For identification $[F(3,24) = 1.12,\ p = 0.36]$ for sound and $[F(3,24) = 7.33,\ p \leqslant 0.01]$ for light, confirming that there was a cost for light but not for sound.

6.2. Discussion

The trace/context model and the standard model are in agreement that there should be no cost of divided attention in the detection condition. The significance tests confirm this for sound, but there was an effect for light. Figure 4b shows that there are small differences between the 100%, 80% and 50% points for light in the detection condition. However there is a large drop when it comes to the 20% point. There are at least three possible explanations for a cost of divided attention at 20% attention. First, the signals used in this experiment are weaker than those used by Hafter *et al.* (1998), and it is possible that for weaker signals a cost of divided attention will appear at lower levels of attentional deployment. Another possibility is that the difference between our results and earlier ones is methodological. In our experiment there was no spatial confound. This spatial confound may have masked a drop in performance at a lower level of attentional deployment in Hafter *et al.* (1998). Third, as there are fewer signals of the 20% modality, a frequency effect might skew the results (Wolfe *et al.*, 2005). For identification one modality shows a cost of divided attention while the other does not. There was no cost of divided attention for light in Experiment 1 and we speculated that this might be the result of the light signal being easier or harder than the sound signal. In Experiment 2 it was the sound signal that showed no cost of divided attention but the light signal showed a clear cost. This indicates that it is not the level of difficulty that affected the results but that there may be a difference in how different conditions affect performance in each modality.

7. Experiment 3

In Experiment 3 we tested the main question of the project. According to the trace/context model, a gap before and after the signal removes any transient information and blocks the sensory trace strategy. The trace/context model assumes that this causes participants to use a context coding strategy. The model further argues that a roaming pedestal blocks the use of the context coding strategy and causes participants to use the sensory trace strategy. There is no room within the trace/context model for more than those two strategies, so the purpose of this experiment was to investigate what happens if both strategies are blocked through the use of a gap and a roaming pedestal. Will performance deteriorate or will participants be able to perform equally well as in Experiments 1 and 2, which would indicate that the trace/context model needs revision?

7.1. Results

The AOC graph for Experiment 3 is plotted in panel a) in Fig. 5. The identification squares all fall close to the ideal cost curve but the detection circles fall midway between the ideal cost curve and the independence point, very similarly to Experiment 2. No points fall significantly below the curve, indicating that the task demands were not too high. As in Experiments 1 and 2, panel b) shows light and sound performance separately.

From Fig. 5 it is clear that the drop in performance is not any larger than in Experiments 1 or 2. There is an indication of a cost of divided attention for both sound and light in the identification condition. However there is a smaller cost, if any, in both modalities in detection. This indicates that Experiment 3 did not produce a larger cost of divided attention than Experiments 1 or 2.

For detection $C = 0.06$ for sound and $C = 0.20$ for light. For identification $C = -0.08$ for sound and $C = -0.24$ for light. This indicates minor biases in the same direction as in Experiment 2. A significance test showed that for detection neither sound nor light reached critical values [$F(3, 24) = 1.88$, $p = 0.15$] for sound and [$F(3, 24) = 0.40$, $p = 0.76$] for light. For identification, both modalities showed significant costs of divided attention [$F(3, 24) = 5.09$, $p \leqslant 0.01$] for sound and [$F(3, 24) = 1.51$, $p = 0.05$] for light.

7.2. Discussion

These results clearly show that performance *does not* deteriorate when sensory trace and context coding strategies are both blocked. This contradicts the predictions derived from the trace/context model. Furthermore, the results in both conditions and both modalities fit with the standard model but not with the trace/context model. Participants were able to perform both tasks, indicating that other strategies are available for performing the tasks and that the trace/context model needs modification. The fact that in the first two experiments the pattern was different for each modality raises interesting questions. Does the gap affect sound more than light and does a roaming pedestal affect light more than sound? Also, does the combination of gap and roaming pedestal cancel out effects from the other? These results would suggest so, but further research is clearly needed. What this other strategy (or strategies) is that the participants used in this experiment is not clear. The fact remains that observers can use one or more strategies that Hafter *et al.* (1998) did not account for, and for which there is no room in their model. This needs to be taken into account for a comprehensive model of divided multimodal attention.

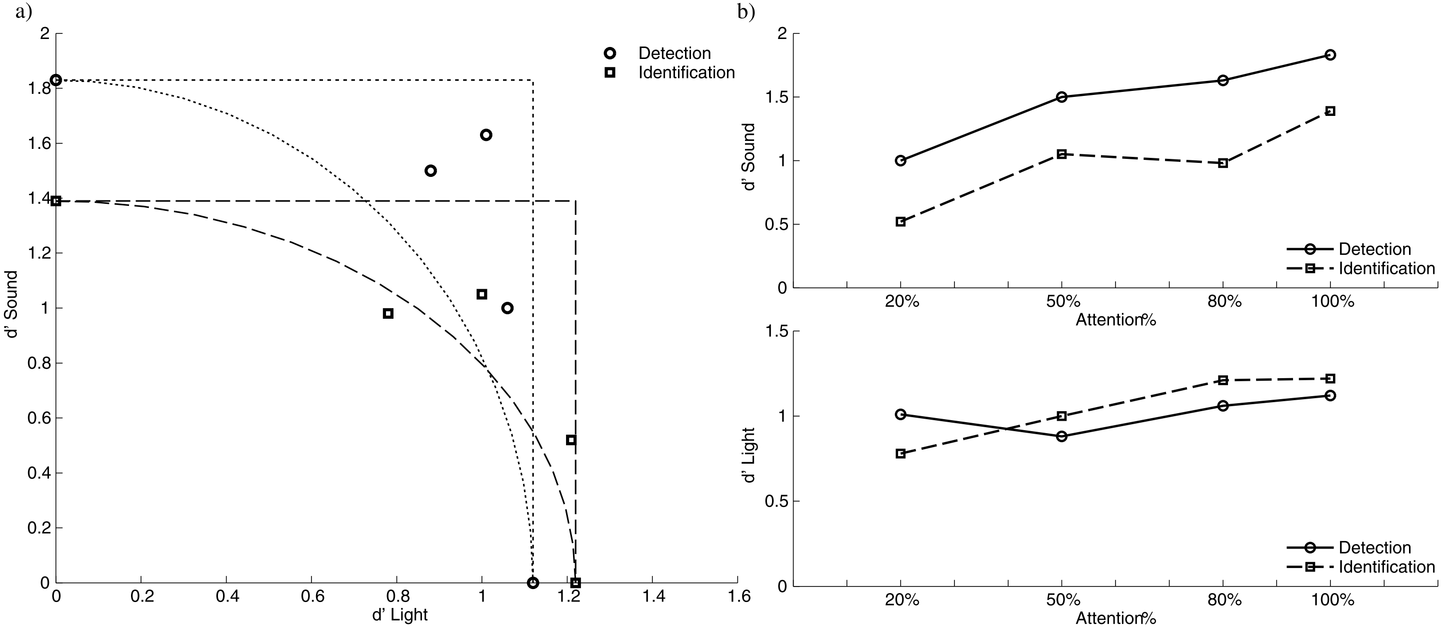

Figure 5. Results from Experiment 3. AOC graph in panel a) and a separated performance on sound and light in panel b).

8. General Discussion

A considerable amount of research has accumulated over the last 60 years in the field of divided multimodal attention. Already in 1969, ideas about a supramodal attentional system had been proposed (Gibson, 1969). Investigators are divided into two camps, the supramodal camp, assuming a cost of dividing attention (e.g., Beer and Roder, 2005; Massaro and Warner, 1977) and the modality independence camp, claiming no cost of divided attention (e.g. Alais *et al.*, 2006; Duncan *et al.*, 1997). Methodological differences are a potential reason for this. It is quite possible, that both supramodal and modality independent systems exist, and that different experimental methods tap into those systems differently, but the aforementioned spatial confound may have contaminated previous results. Klein *et al.* (1992) found that, in cuing situations, people invariably shift their attention to a spatial location, even when the cue carried no spatial information. Also, Spence and Driver (1997b) found that presenting light and sound at different spatial locations exaggerated the effect of dividing attention between two modalities, compared to presenting the light and sound from the same spatial location. This makes repeating many older studies very important in order to find which effects are the result of divided multimodal attention and which are merely the effect of a spatial confound. This was one of the main aims of our experiments, in addition to exploring the assumptions of the model of Hafter *et al.* (1998).

Our results from Experiments 1 and 2 indicate that there is a difference between detection and identification, although not the simple difference that the standard model predicts, but rather an asymmetrical difference between sound and light. Also, that a gap before and after the signal, or a roaming pedestal, changes how detection and identification affects the cost of divided attention. However, our results suggest that the difference between detection and identification and the effect of a gap or a roaming pedestal are not as clear or as simple as previous models have suggested and neither model fits completely with our results. The standard model predicts nine out of the 12 results (modalities × answer-mode × experiment) while the trace/context model predicts seven. More revealingly, where the predictions of the two models are different, our results fit the standard model in four cases out of six. Where both models agree our results fit those predictions in five cases out of six, the only exception being that light detection in Experiment 2 showed a cost of divided attention contrary to both models. As discussed above, this was as a result of a significant drop in performance in the 20% attention condition and possible reasons have been discussed. To our knowledge, most research on the standard and trace/context models has not separated performance on the light and sound stimuli; our results suggest that there is an asymmetrical relationship between conditions and modalities that raises interesting questions; did the participants use different strategies for the light and sound signals? Or does the gap or roaming pedestal affect each modality differently? Results from the experiments of Ward (1994) and Spence and Driver (1997a) suggest that the interaction of the two modalities may not be symmetrical. The current results do not speak directly to this question, nor did we study interactions specifically, but the results suggest that in future studies these interactions should be taken into account.

Although a bias was present in all three experiments between sound and light signals, this bias cannot be attributed to a criterion shift, as the bias did not change between different attentional instructions. Also, there was little or no difference in bias between the experiments, showing that the gap and/or pedestal did not cause a criterion shift.

In Experiment 3, performance for detection and identification was in line with the predictions of the standard model, detection not showing a cost of divided attention in either modality and identification showing a cost in both modalities. Most importantly, performance, both for detection and identification, did not fall below the ideal cost curve. This result convincingly shows that participants could perform the task and therefore use other strategies than sensory trace or context coding suggested by Hafter *et al.* (1998).

Our results suggest that Hafter *et al.* were right in that the difference between detection and identification is not universal and can be altered by manipulating the task demands. However, our results do not support their assumption that there are only two strategies available and that participants automatically switch between them if one is blocked. Furthermore, our results suggest that without a spatial confound the interaction between attentional instructions and performance might not be as simple as both the standard model and Hafter *et al.* suggest, especially at lower levels of attention. Our results suggest, at least with a weak signal, that there will be a cost of divided attention on lower levels of attentional deployment, even in conditions that otherwise show no cost of divided attention.

As there were several differences between our methods and those used by Hafter *et al.* (1998) our results should be interpreted with some caution, but nonetheless, they show that the simple two-strategy model, proposed by Hafter *et al.* (1998) does not generalize to our tasks. Participants only had to make one response in each trial in these experiments [while responding to both modalities in Hafter *et al.* (1998)]; this change might affect the results and needs further exploration. However, if the trace/context model is to explain how divided multimodal attention works, slight variations should not change the results. A simpler explanation is that more than two strategies can be used.

Further research is needed, especially since much of the older research suffers from the spatial confound. Among the remaining questions are what strategies participants use, how many there are, and their nature. Also, how does signal strength affect the patterns of cost in divided multimodal attention? Another issue that needs to be addressed is possible effects of signal frequency on performance, since the signals in the 20% condition were less frequent. As Wolfe *et al.* (2005) showed, error rates increase for less frequent signals. Therefore there is a possibility that the drop in performance is not due to a cost of divided attention but rather an effect due to signal frequency. The less frequent signals also decrease the likelihood of repeated signals in the 20% modality, which decreases chances of priming effects that may speed up processing in the 80% modality where repeated signals in the same modality are more likely. Priming can speed reaction times (Kristjánsson *et al.*, 2008) and improve accuracy (Ásgeirsson *et al.*, 2014; see e.g. Lamy and Kristjánsson, 2013, for review) and it is an interesting avenue for further research to explore the possible priming effects in multi-modal tasks.

Acknowledgements

A.K. was supported by grants from the Research fund of the University of Iceland, and the Icelandic research fund. T.K. was supported by an innovation grant from the Icelandic research fund. We thank Mr Gudmundur Bjorn Birkisson M.S. Software Engineering, for invaluable support and the reviewers for constructive comments and help.

References

Alais, D., Morrone, C. and Burr, D. (2006). Separate attentional resources for vision and audition, *Proc. R. Soc. B Biol. Sci.* **273**(1592), 1339–1345.

Alwitt, L. (1981). Two neural mechanism related to modes of selective attention, *J. Exp. Psychol. Hum. Percept. Perform.* **7**, 324–332.

Ásgeirsson, Á. G., Kristjánsson, Á. and Bundesen, C. (2014). Independent priming of location and color in identification of briefly presented letters, *Atten. Percept. Psychophys.* **76**, 40–48.

Beer, A. L. and Roder, B. (2005). Attending to visual or auditory motion affects perception within and across modalities: an event-related potential study, *Eur. J. Neurosci.* **21**, 1116–1130.

Bonnel, A. M. and Hafter, E. R. (1998). Divided attention between simultaneous auditory and visual signals, *Percept. Psychophys.* **60**, 179–190.

Driver, J. and Spence, C. (1994). Spatial synergies between auditory and visual attention, in: *Attention and Performance: Conscious and Nonconscious Information Processing*, C. Umiltá and M. Moscovitch (Eds), pp. 311–331. MIT Press, Cambridge, MA, USA.

Driver, J. and Spence, C. (2000). Multisensory perception: beyond modularity and convergence, *Curr. Biol.* **10**, R731–R735.

Duncan, J., Martens, S. and Ward, R. (1997). Restricted attentional capacity within but not between sensory modalities, *Nature* **387**, 808–810.

Ferlazzo, F., Couyoumdjian, A., Padovani, T. and Belardinelli, M. O. (2002). Head-centred meridian effect on auditory spatial attention orienting, *Q. J. Exp. Psychol. A* **55**, 937–963.

Gibson, E. J. (1969). *Principles of Perceptual Learning and Development.* Appleton-Century-Crofts, New York, NY, USA.

Hafter, E. R., Bonnel, A.-M., Gallun, E. and Cohen, F. (1998). A role for memory in divided attention between two independent stimuli, in: *Psychophysical and Physiological Advances in Hearing*, A. R. Palmer, A. Rees, A. Q. Summerfield and R. Meddis (Eds), pp. 228–238. Whurr Publishing, London, UK.

Holmes, N. P. and Spence, C. (2005). Multisensory integration: space, time and superadditivity, *Curr. Biol.* **15**, 762–764.

Kingdom, F. A. A. and Prins, N. (2010). *Psychophysics: A Practical Introduction*, pp. 178–187. Academic Press, London, UK.

Klein, R. M., Kingstone, A. and Pontefract, A. (1992). Orienting of visual attention, in: *Eye Movements and Visual Cognition: Scene Perception and Reading*, K. Rayner (Ed.), pp. 46–65. Springer-Verlag, New York, NY, USA.

Koelewijn, T., Bronkhorst, A. and Theeuwes, J. (2009). Competition between auditory and visual spatial cues during visual task performance, *Exp. Brain Res.* **195**, 593–602.

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

Kristjánsson, Á., Ingvarsdóttir, Á. and Teitsdóttir, U. D. (2008). Object- and feature-based priming in visual search, *Psychon. Bull. Rev.* **15**, 378–384.

Kristjánsson, Á., Vandenbroucke, M. and Driver, J. (2004). When pros become cons for anti versus prosaccades, *Exp. Brain Res.* **155**, 231–244.

Lamy, D. and Kristjánsson, Á. (2013). Is goal-directed attentional guidance just intertrial priming? A review, *J. Vis.* **13**, 14.

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

Macmillan, N. A. and Creelman, C. D. (2004). *Detection Theory: A Users Guide*, p. 21. Lawrence Erlbaum Associates, Mahwah, NJ, USA.

Massaro, D. W. and Warner, D. S. (1977). Dividing attention between auditory and visual perception, *Percept. Psychophys.* **21**, 569–574.

McNicol, D. (1972). *A Primer of Signal Detection Theory*, pp. 53-56. George Allen & Unwin, London, UK.

Shiffrin, R. M. and Grantham, D. W. (1974). Can attention be allocated to sensory modalities? *Percept. Psychophys.* **15**, 460–474.

Spence, C. and Driver, J. (1996). Audiovisual links in endogenous covert spatial attention, *J. Exp. Psychol. Hum. Percept. Perform.* **22**, 1005–1030.

Spence, C. and Driver, J. (1997a). Audiovisual links in exogenous covert spatial orienting, *Percept. Psychophys.* **59**, 1–22.

Spence, C. and Driver, J. (1997b). On measuring selective attention to an expected sensory modality, *Percept. Psychophys.* **59**, 389–403.

Spence, C., Nicholls, M. E. and Driver, J. (2001). The cost of expecting events in the wrong modality, *Percept. Psychophys.* **63**, 330–336.

Stein, B. E. and Meredith, M. A. (1993). *The Merging of the Senses.* MIT Press, Cambridge, MA, USA.

Stein, B. E., Wallace, M. W., Stanford, T. R. and Jiang, W. (2002). Cortex governs multisensory interactions in the midbrain, *Neuroscientist* **8**, 306–314.

Talsma, D., Doty, T. J., Strowd, R. and Woldorff, M. G. (2006). Attentional capacity for processing concurrent stimuli is larger across sensory modalities than within a modality, *Psychophysiology* **43**, 541–549.

Talsma, D. and Woldorff, M. G. (2005). Selective attention and multisensory integration: multiple phases of effects on the evoked brain activity, *J. Cogn. Neurosci.* **17**, 1098–1114.

Vroomen, J. and De Gelder, B. (2000). Sound enhances visual perception: cross-modal effects of auditory organization on vision, *J. Exp. Psychol. Hum. Percept. Perform.* **26**, 1583–1590.

Ward, L. M. (1994). Supramodal and modality-specific mechanisms for stimulus-driven shifts of auditory and visual attention, *Can. J. Exp. Psychol.* **48**, 242–259.

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