



## **Divided multimodal attention**

Sensory trace and context coding strategies in  
spatially congruent auditory and visual presentation

Tómas Kristjánsson & Tómas Páll Þorvaldsson

**Lokaverkefni til BS-gráðu**  
**Sálfræðideild**  
**Heilbrigðisvísindasvið**



**HÁSKÓLI ÍSLANDS**

**Divided multimodal attention**  
***Sensory trace and context coding strategies in***  
***spatially congruent auditory and visual presentation***

Tómas Kristjánsson & Tómas Páll Þorvaldsson

Lokaverkefni til BS-gráðu í sálfræði

Leiðbeinandi: Árni Kristjánsson

Sálfræðideild

Heilbrigðisvísindasvið Háskóla Íslands

Júní 2013

Ritgerð þessi er lokaverkefni til BS gráðu í sálfræði og er óheimilt að afrita ritgerðina á nokkurn hátt nema með leyfi rétthafa.

© Tómas Kristjánsson & Tómas Páll Þorvaldsson, 2013

Prentun: Háskólaprent  
Reykjavík, Ísland 2013



Previous unimodal and multimodal research suggests that detection is a capacity-free process while identification is a capacity-limited process. The trace/context model assumes that different memory strategies are responsible for this difference, and not inherent differences between identification and detection. The model assumes that people use two strategies, a sensory trace or a context coding strategy and that if one is blocked, people will automatically use the other. Most previous experiments have presented lights and sounds in separate locations, creating the possibility of a spatial confound, which introduces an alternative interpretation of the results. This paper outlines a series of experiments, investigating divided multimodal attention, without the spatial confound, and challenges the assumption of the trace/context model, that only two strategies are available to participants. Our critical experiment involved a gap, 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; necessitating changes to the trace/context model.

Sensory and perceptual processes have been the subject of research since the beginning of scientific psychology (Freedheim & Weiner, 2003; Leahey, 2012). Throughout the twentieth century, research focused on each sensory modality in isolation but as our understanding grew, the importance of the interaction between sensory modalities became apparent. Although multimodal research started in the sixties (Brown & Hopkins, 1967; Taylor, Lindsey & Forbes, 1967; Tulving & Lindsey, 1967), unimodal research still dominated the field of sensation and perception. Multimodal research grew slowly through the second part of the twentieth century but in the last twenty years this growth has been exponential (Koelewijn, Bronkhorst & Theeuwes, 2010).

Multimodal perception plays a crucial role in our everyday life, as we try to sort out and give meaning to the millions of stimuli that bombard our senses every day. In the laboratory visual or auditory stimuli can be presented in isolation, which gives researchers great control but in the real world, things are not as clear and simple. We perceive most, if not all, events with multiple senses simultaneously. This requires integration between the senses involved and this integration, although normally beneficial, can in some circumstances create a perception that is not part of the sensory

input from any of the senses involved. A clear example of this is the McGurk effect. McGurk and MacDonald (1976) had observers watch a video of lips forming one sound, while listening to a different sound, the observers perceived a sound that neither the lips nor the audio produced. This showed that the visual information a person gets from seeing someone speak, changes the way that the sound is perceived (Calvert, Spence & Stein, 2004).

### **Multimodal attention**

A major focus of multimodal research has been the study of multimodal attention (Koelewijn et al., 2010). Attention plays a key role in selecting which stimuli are attended to and which are ignored, thereby greatly influencing our understanding of the world. Two main questions have been pervasive in the literature about multimodal attention. The first question pertains to the automaticity of multimodal attention. This debate is focused on exogenous (stimulus driven, bottom-up) factors of attention, how automatic they are and how they interact with endogenous (voluntarily driven, top-down) factors of attention. This is the multimodal version of the age old top-down, bottom-up debate that goes on in most areas of cognitive sciences. The second question is whether there is a supramodal attentional process or whether there are independent attentional processes for each modality. This question often takes the form of asking whether there is a cost of dividing attention between different modalities (Alais, Morrone & Burr, 2006).

The automaticity of multimodal attention has been investigated extensively (e.g. Ward, 1994; Santangelo & Spence, 2007; Cosman & Vecera, 2009) but no clear consensus has yet been reached. On the one hand many researchers have shown that irrelevant stimuli in one modality can significantly impact search, discrimination and/or recognition of target stimuli in a different modality (Mazza, Turatto, Rossi & Umiltá, 2007; Theeuwes, 2004; Spence & Driver, 1997*a*). On the other hand, several researchers have shown that this is not a completely automatic process that can be modified or extinguished by different task demands (Koelewijn, Bronkhorst & Theeuwes, 2009*a*; Boot, Brockmole & Simons, 2005; Santangelo, Belardinelli & Spence, 2007). Some researchers have tried to reconcile those opposing views by suggesting interacting models of endogenous and exogenous mechanisms (Muller & Rabbitt, 1989; Prime, McDonald, Green & Ward, 2008; Ward, 1994). It seems reasonable to assume that, although exogenous attentional orientation is not truly automatic, there are two

interacting systems at play. The nature of this interaction is, however, still not fully understood.

A different line of enquiry focuses on the nature of attentional mechanisms. A dividing question has been, whether there is a supramodal attentional pool, or if each sensory modality has an independent pool of attentional resources? To answer this, researchers have investigated the effect of dividing attention between modalities, measuring effects upon performance in each modality. Some researchers have concluded that each modality has an independent pool of attentional resources (Alais et al., 2006; Shiffrin & Grantham, 1974; Larsen, McIlhagga, Baert & Bundesen, 2003; Ferlazzo, Couyoumdjian, Padovani, & Belardinelli, 2002) while others have found significant crossmodal effects of dividing attention between modalities (Driver & Spence, 1994; Spence & Driver, 1997*a*; Spence, Nicholls & Driver 2001; Koelwijn, Bronkhorst & Theeuwes, 2009*b*).

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 researches. A lot of evidence indicates that the superior colliculus (SC) plays a key role in multimodal integration (Holmes & Spence, 2005; Stein & 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 & Spence, 2000; Vroomen & De Gelder, 2000). For example, Stein, Wallace, Stanford and Jiang (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. Other brain areas that have attracted interest from researchers include the lingual gyrus (Macaluso, Frith & Driver, 2000), the thalamus (Frith & Friston, 1996) and the superior temporal sulcus (Driver & Spence, 2000). Further research will advance our understanding of how these various brain areas interact in integrating multimodal stimuli, how attention can be deployed and how the brain divides attention between different sensory systems.

### **Models of multimodal divided attention**

Various versions of dual-task paradigm have been used to study divided attention (Bonnell & Hafter, 1998). The paradigm requires participants to perform two tasks

simultaneously. This forces them to divide their attention between the two tasks and is therefore a convenient method of investigating divided attention. The tasks can either be unimodal (e.g. Bonnel, Stein & Bertucci, 1992) or multimodal, where the two tasks are presented in two different modalities (e.g. Bonnel & Hafter, 1998).

Two conditions, detection and identification, are commonly tested within the dual-task paradigm. Detection requires participants to respond to any change in a signal, regardless of the direction or the strength of that change. Identification, on the other hand, 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 a capacity-free process while identification is capacity-limited (Bonnel & Hafter, 1998). Detection, being a capacity-free process, refers to the fact that most early research showed that detection performance in dual-task experiments did not differ from performance on a single task (Alwitt, 1981). In the context of divided multimodal attention, this means that performance on detection tasks does not drop as less attention is paid to a modality. Identification is said to be capacity-limited because experiments have repeatedly shown that performance drops when attention is divided between the two modalities, compared with single task performance (Spence & Driver, 1997a). In sum, detection has been thought of as capacity-free process that does not show a cost of divided attention while identification has been thought of as capacity-limited process showing costs of divided attention. From here on out, we will refer to these assumptions, and the predictions derived from them, as the standard model.

These assumptions were challenged by Hafter, Bonnel, Gallun and Cohen (1998). They showed that in a divided multimodal attention task a cost in detection can be observed, but no cost in identification. In their design a 900 ms pedestal of sound and light was played, in the middle of the pedestal the sound or the light could increase or decrease in strength for a very short time. The detection task was to detect if change occurred and the identification task was to discriminate whether the signal strength increased or decreased. The results were as expected, detection being capacity-free and identification being capacity-limited. However when a short gap was introduced before and after the signal, performance both in detection and identification showed costs of divided attention. This violates the basic assumption that detection is capacity-free. Moreover Hafter et al. (1998) showed that by presenting a roaming pedestal, without the gaps introduced in the previous experiment, neither detection nor identification showed

any costs of divided attention. This violated the predictions of the standard model, where an identification cost is predicted.

Haft et al. (1998) showed in a simple but convincing manner that the standard model was incorrect; however their own interpretations and assumptions regarding the processes underlying this deviation from the older assumptions were not as convincing as their experimental results. They assume that two memory processes or strategies are responsible for whether costs are observed in earlier experiments, but not a fundamental difference between detection and identification. They argue that participants use a sensory trace strategy in detection and a context coding strategy in identification. The sensory trace strategy uses transient information from a continuous signal. By using a gap before and after the signal, participants do not receive information from transients and Haft et al. assume that this forces them to use a context coding strategy. Context coding, they argue, is a strategy of categorising the different levels of the signal as either up, down or no signal. They claim that a roaming pedestal, where the strength of the pedestal varies from trial to trial, and therefore the level of the signal that should be categorised as up or down, prevents the use of the context coding strategy and participants use sensory traces instead. From here on out 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 while the trace/context model predicts a cost. In the same vein, in identification tasks with a roaming pedestal the standard model predicts a cost of divided attention while the trace/context model does not. The trace/context model seems to assume that there are only two possible signal processing strategies. If there would be more strategies in their model then the assumption, that by blocking the sensory trace strategy participants automatically change to the context coding strategy, would not hold since participants could then utilize a third strategy. The same goes for the context coding strategy. Another possibility is that there are only two strategies available to observers but that those strategies are not the sensory trace and context coding strategies. 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 both possibilities are simultaneously blocked.

## **Methodological concerns**

Another cause for concern is the methodology adopted by most, if not all, researchers in the field of multimodal attention. Spence and Driver (1997*b*) wrote a comprehensive critique of the methodology prevalent in this field. They summarised five common flaws in methodology which, according to them, apply to a majority of the research in the field. The five flaws are the apparent benefit of knowing target modality, possible response priming artifacts, possible criterion shifts, expectancy effects and spatial cuing effects (Spence & Driver, 1997*b*). The first two flaws mainly concern cuing paradigms and are therefore not very relevant here. The possible criterion shift can be examined by employing signal detection theory, which Hafter et al. use, although they did not report the bias in different modalities and experimental conditions. Expectancy effects are one of the most common methodological flaws in multimodal divided attention experiments. However this flaw has only been reported to affect response times and not accuracy, which is the measure used by Hafter et al.. The last flaw however is a major concern regarding Hafters et al. methodology and most other research in this field. Spatial cuing effects occur when the signal in the two modalities appear in two spatially separated locations. When this is done, it is impossible to know if attention is divided between the two modalities or between the two locations where the signals originate. This creates a spatial confound. In Hafter et al. (1998) the visual stimuli appeared on a computer screen while the auditory stimuli were presented via headphones. This means that there may be an alternative interpretation of their results. This is a serious flaw, as people invariably shift their attention to an expected target location (Klein, Kingstone & Pontefract, 1992). In addition, attending to separate locations in different modalities has been shown to be less efficient than attending to a single location (Driver & Spence, 1994; Spence & Driver, 1996). Brain research shows that a simultaneous sound and light from separate locations produce a suppressive neural response in the superior colliculus while a simultaneous sound and light from the same location produce a multiplicative neural response (Driver & Spence, 2000). This spatial confound can be found in almost all multimodal experiments, both those supporting the standard model and the trace/context model. Therefore it is imperative to repeat the experiments behind those models without the spatial confound to truly understand the mechanisms involved.

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. Our main hypothesis is 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 repeat the gap and the roaming pedestal conditions, both to serve as a baseline for our main hypothesis and to see the effects of gaps and roaming pedestals on divided multimodal attention without a spatial confound.

## Experiment 1

In this study we investigate the cost of divided multimodal attention with gaps before and after the signal, without the spatial confound that has been present in most previous studies. This study utilises an adapted methodology from Spence and Driver (1997*b*) where sound and light are presented from the same location. The trace/context model would predict a cost of divided attention both in detection and identification while the standard model would predict a cost in identification but no cost in detection.

### Method

**Subjects.** Six subjects participated in this study, two were the authors, and the other four were naive participants recruited through the University of Iceland. Five

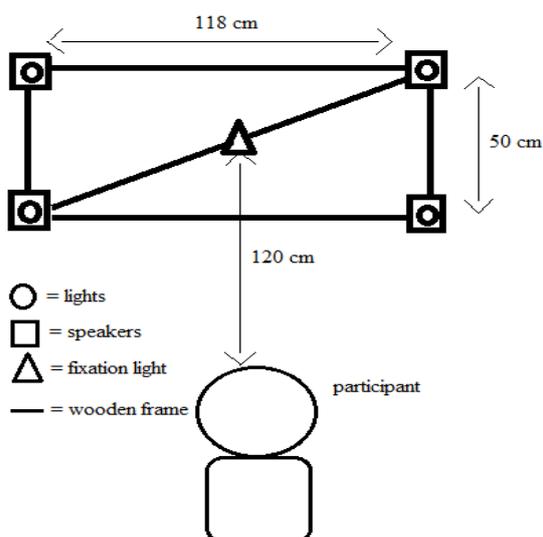


Figure 1. Schematic overview of the experimental set-up.

participants were male and one was female. The mean age was 25.3 years, with a range of 18 – 28 years. All reported normal hearing and normal, or corrected to normal, vision. The authors participated in all three experiments. No difference was found in performance between them and naive participants.

**Apparatus and materials.** All the experiments were conducted in an IAC audiology room with a background

luminance of  $22.1 \text{ cd/m}^2$ . The participants were seated in the middle of the room, 120 cm away from a grey wall. The experimenters sat behind the participants, controlling the experiment on a computer whose screen faced away from the participants but was the source of the luminance in the room. The participants held a QWERTY 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. The chair that participants sat in was adjusted for each one, so that their eyes faced the middle of the wooden frame. On each corner of the frame, a Cambridge Soundworks speaker was placed, connected to a 5.1 sound system and two separate soundcards in the computer. Glued to the front of each speaker was a light unit that consisted of four, super-bright green LED diodes, enclosed in black cylinders, with a CREE XML, 35mm optic diffused lens covering the top, facing the participants. A single super-bright 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 figure 1. The lights were connected through an Arduino micro-computer, to the main computer and the lights had an independent 7,5V power source. The main computer ran the experiments through Matlab R2011B. Sound strength was measured with a Brüel & Kjær, sound level meter, type 2250L and light strength was measured with a Cambridge Research system, ColorCal II colorimeter. The sound pedestal strength was 59 dB (SPL), the light pedestal strength was  $170 \text{ cd/m}^2$ . The average sound signal was 5.7 dB (SPL) and the average light signal was  $138 \text{ cd/m}^2$ , up or down from the pedestal, in detection. The average sound signal was 6.0 dB (SPL) and the light signal  $146 \text{ cd/m}^2$  in identification.

**Design.** There were three within-subjects factors, target modality (sound and light), answer mode (detection and identification) and attentional instructions (100% sound (s), 100% light (l), 80%(s)/20%(l), 50%(s)/50%(l) and 80%(l)/20%(s)). 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 in detection and identification, with half of the participants taking the identification trials first and half of the participants taking the detection trials first. In 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 (no change). In the identification trials there was a 50% chance of the signal going up and 50% chance of the signal going down, there were no trials without a change.

**Procedure.** Before each session, participants were told the likelihood of the target modality and 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 same vein, participants were instructed to focus exclusively on the sound in 100% sound sessions. In the dual task sessions, participants were told how likely it was for the signal to appear in each modality. Thus, in 80% light/20% sound they were told that 80% of the signals would be visual and 20% auditory, therefore it would be best to focus mainly on the light while keeping their ears open for the occasional sound signal. Since participants took the 100% sessions first they got to know the procedures and the nature of the signals, making it easier for them to understand and follow the instructions in the dual task sessions. 500 ms before the start of each trial the fixation light was illuminated and stayed on until the end of the trial. Then, the sound and light appeared simultaneously, first, a 380 ms pedestal followed by a 50 ms gap before the 40 ms signal (in the signal trials) in either the sound or the light modality, followed by another 50 ms gap and a 380 ms pedestal. The total time of the trial from the onset of the sound and light until its offset was 900 ms. The different signals (“up”, “down” and no-signal) are shown in figure 2.

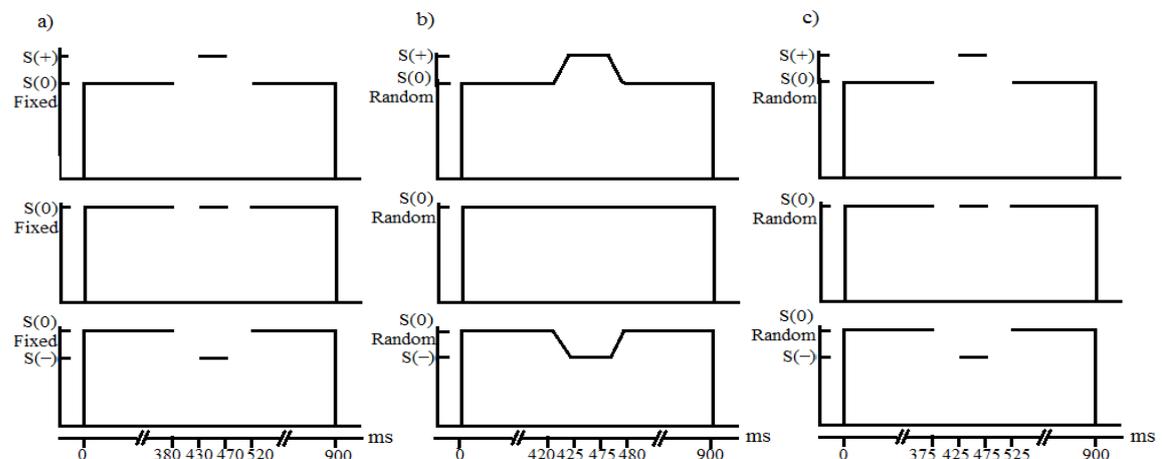


Figure 2. Diagram of the different signals in experiments 1, 2 and 3. The top panel of each picture (S(+)) shows the signal when strength increased, the middle panel, (S(0)) shows the no-signal trials and the bottom panel (S(-)) shows the signal when strength decreased. Part a) of the picture shows the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

The participants had 1000 ms to respond, starting when the signal ended. This created a 570 ms interval where all lights and sounds were turned off, then the fixation light illuminated again and the next trial started. In the detection trials participants responded by pressing the space bar on a keyboard if they thought a change had occurred, but refrained from responding if they thought that no change had occurred. In the identification trials participants responded by pressing the up arrow on a keyboard if they thought the signal had gone up and the down arrow if they thought the signal had gone down. After each session, participants were encouraged to take a break for as long as they needed to clear their head and refocus before the next session started. The time the whole experiment took for each participant varied, but the average time was 1 hour and 48 minutes. After the experiments all participants reported that they had understood and been able to follow the instructions.

## Results

We calculated  $d'$  for every 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 we used  $d'_{FC} = (z(\text{Hit}) - z(\text{FA}))/\sqrt{2}$  and  $C = -(z(\text{Hit}) + z(\text{FA}))/\sqrt{2}$  (McNicol, 1972; Kingdom & Prins, 2010). Zero values in false alarm or misses in the 2X2 tables 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 are ideal lines that 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 on (or close to it) if divided attention is capacity limited.

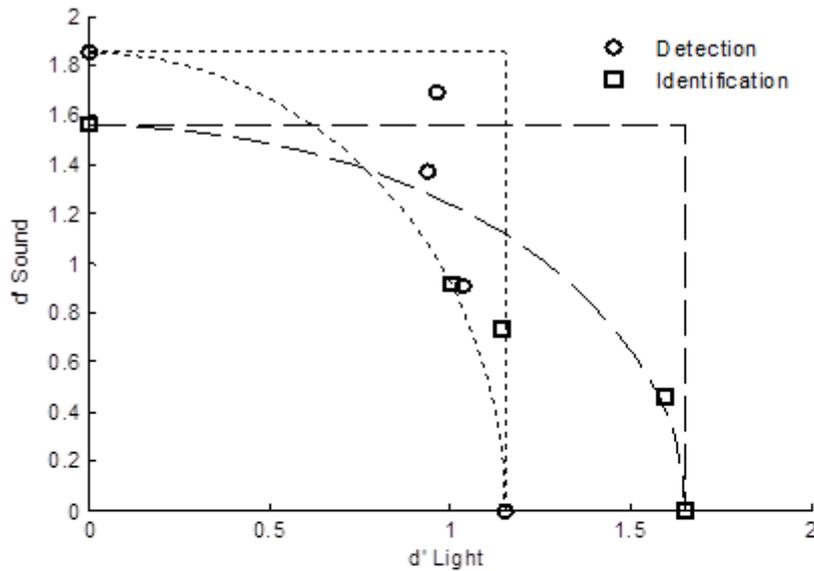


Figure 3. AOC graph for experiment 1. The curved lines represent ideal performance if there is a cost of divided attention while the point where the straight lines meet represent the independence point, where  $d'$  values should cluster around if there is no cost of divided attention. The short-dotted lines connecting the circles on the x and y-axis represent the ideal lines for detection while the long-dotted lines connecting the squares on the axis are the ideal lines for identification. The circle and square on the x-axis represent performance on the 100% light, single task condition. The circle and square on the y-axis represent performance on the 100% sound, single task condition.

As can be seen in figure 3, the squares fall on or below the curved ideal capacity limited 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. Therefore, in figure 4 we plot the performance separately for light and sound.

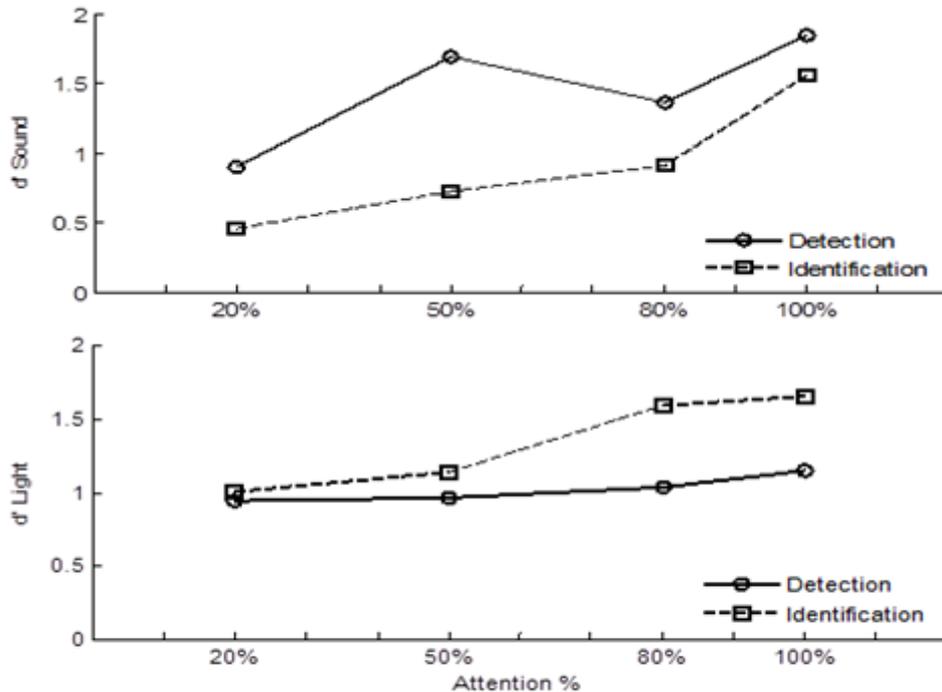


Figure 4. Performance in experiment 1, with a separated graph for light and sound. The attentional percentages on the x-axis represent how much attention participants were instructed to pay to the modality and the likelihood of the signal occurring in that modality.

As we saw from the AOC graph there is a clear cost of divided attention in the identification condition. The ambiguous results for the detection condition become much clearer in figure 4. There is a clear downward trend in the line representing the detection of sound although the 50% point is higher than the 80% point. Both points are lower than the 100% point and the 20% point is lowest of all, creating a small but clear downward trend, representing a cost of divided attention. However there is no such downward trend in the graph for the detection of light, indicating no cost of dividing attention. The bias was measured as the criterion,  $C$ . For detection,  $C = 0.181$  for sound and  $C = 0.267$  for light. For identification,  $C = -0.527$  for sound and  $C = 0.049$  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 while no such bias was present for light. A significance test (ANOVA) confirmed that for detection, there was a cost of divided attention for sound [ $F(3,20) = 3.10, p = .05$ ] but not for light [ $F(3,20) = .41, p = .75$ ]. For identification a significance test confirmed the cost for sound [ $F(3,20) = 7.81, p \leq .01$ ], light however did not reach significance [ $F(3,20) = 1.87, p = .17$ ].

## **Discussion**

Experiment 1 served the purpose of investigating the cost of divided attention with the Hafter et al. paradigm without the spatial confound that might have affected their results (cf. Spence & Driver, 1997*b*). As expected, identification showed a cost of divided attention, which is in accordance with both the trace/context model and the standard model. Although the significance test for light did not confirm this, the p-value was low and figure 3 and 4 both indicate a cost. This is a good indicator that despite the methodological changes we made from previous experiments, we were still measuring the same effect. The results from the detection condition were more ambiguous. Performance for the sound signals showed signs of 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 floor or ceiling effects. This is a possibility, but an unlikely one, as  $d'$  values were close to one, indicating a performance around 75% correct. Also, identification showed a clear downward trend in the light signal suggesting that the light signals were not impossible to perceive.

## **Experiment 2**

In experiment 2, the roaming pedestal condition from Hafter et al. (1998) was recreated, again, without the spatial confound. In this experiment, the trace/context model would predict no cost of divided attention in neither detection nor identification. The standard model would again predict that detection would be capacity free while identification would be capacity limited.

## **Method**

**Subjects.** Six subjects participated in this experiment, four males and two females. Two were the authors, two had participated in experiment 1 and two were new, recruited from the University of Iceland. Only the authors knew the purpose of the experiment, the other four were naive participants. Their age ranged from 22-28 with an average age of 27.0 years. All reported normal hearing and normal, or corrected to normal, vision.

**Apparatus and materials.** Apparatus and materials were the same as in

experiment 1 with the exception of maximum and minimum signal strengths. The weakest sound signal was 42.9 dB (SPL) and the strongest 63.1 dB (SPL). The weakest light signal was 65 cd/m<sup>2</sup> and the strongest was 260 cd/m<sup>2</sup>. The average signal strength in detection was 1.7 dB (SPL) for sound and 76 cd/m<sup>2</sup> for light. For identification the average signal strength was 3.3 dB (SPL) for sound and 111 cd/m<sup>2</sup> for light.

**Procedure.** 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, see also in figure 2. The pedestal varied randomly in strength, independently for sound and light. The participants had 1000 ms to respond, starting from the end of the signal. This created a 575 ms interval between trials, where all lights and sounds were off. Otherwise the procedure was the same as in experiment 1.

## Results

As in experiment 1 an AOC graph was plotted from the average  $d'$  values. Figure 5 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. Taken together 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. Therefore in figure 6 we separate performance for light and sound to see how the cost or no cost appears in each modality.

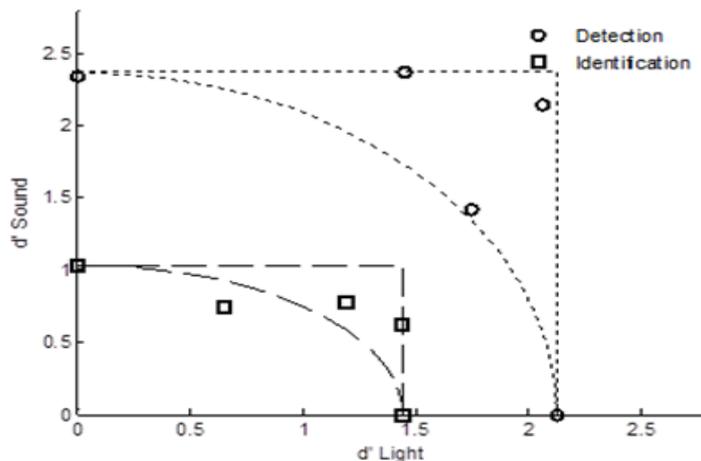


Figure 5. AOC graph for experiment 2. Ideal lines and markers are the same as in figure 3.

Figure 6 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.

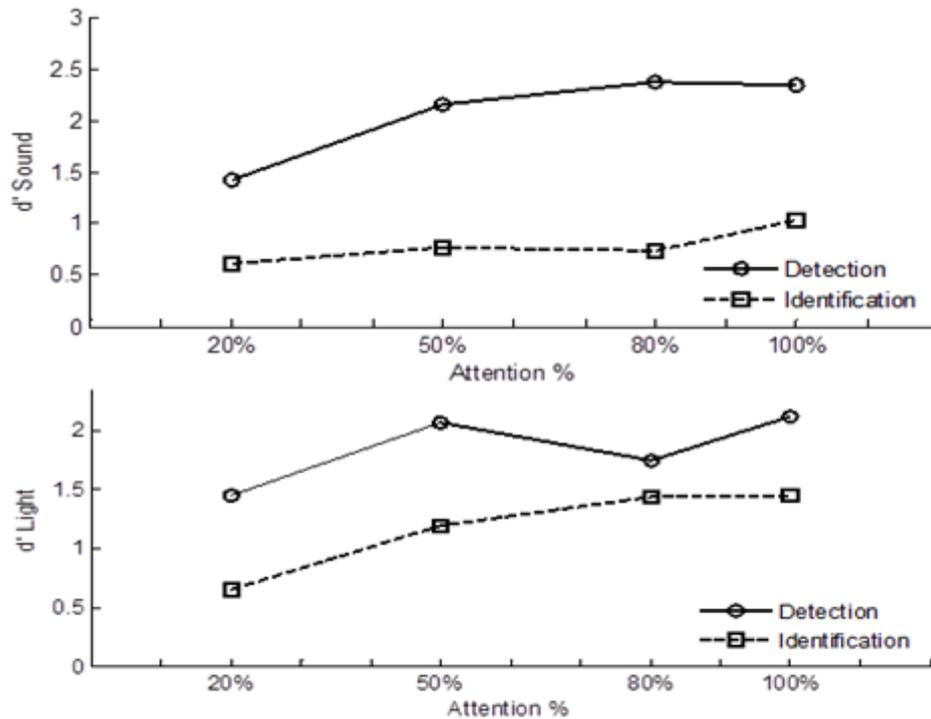


Figure 6. Performance in experiment 2, with a separated graph for light and sound. Axis and lines are the same as in figure 4.

In identification there is a clear cost of divided attention for light but no such cost for sound. In detection  $C = 0.275$  for sound and  $C = 0.418$  for light. In identification  $C = -0.202$  for sound and  $C = -0.263$  for light. This indicates a no-response bias in detection and an 'up' bias in identification, with a stronger bias for light signals in both conditions. A significance test confirmed that overall, detection did not show a cost of divided attention,  $[F(3,20) = 1.96, p = .15]$  for sound and  $[F(3,20) = 1.41, p = .27]$  for light. For identification  $[F(3,20) = .64, p = .60]$  for sound and  $[F(3,20) = 5.11, p \leq .01]$  for light, confirming that there was a cost for light but not for sound.

## 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. Our significance tests confirm this although figure 6 indicates a cost at 20% attention in both modalities. There are at least two possible explanations for a cost 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 and earlier results is methodological. In our experiment there was no spatial confound which was present in most, if not all, earlier experiments. This spatial confound may have masked a drop in performance at a lower level of attentional deployment. In 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 is a difference in how different conditions affect performance in each modality.

### **Experiment 3**

In experiment 3 we tested the main hypothesis of this series of experiments. 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. As discussed before there is no room in the trace/context model for more than those two strategies. Therefore the purpose of this experiment was to investigate what happens if both strategies are blocked by using a gap and a roaming pedestal. Would performance deteriorate or would participants be able to perform comparatively with experiments 1 and 2 which would indicate that there are more than those two strategies and the trace/context model is incorrect.

#### **Method**

**Subjects.** Six subjects participated in this experiment, four males and two females. Two were the experimenters, two had participated in experiments 1 and 2 and two were new participants recruited from the University of Iceland. The four participants that participated in all three experiments had the order of experiments randomised. Their age ranged from 18-28 with an average age of 24.6 years. All reported normal hearing and normal, or corrected to normal, vision.

**Apparatus and materials.** Apparatus and materials were the same as in experiments 2 except that the average signal in detection was 5.8 dB (SPL) for the sound signal and 83 cd/m<sup>2</sup> for the light signal. For identification the average signal was 6.1 dB (SPL) for the sound signal and 86 cd/m<sup>2</sup> for the light signal.

**Procedure.** In this experiment the stimulus was a 900 ms sound and light presented simultaneously as in experiments 1 and 2. In experiment 3 the pre and post signal pedestals were 375 ms with a 50 ms gap before and after a 50 ms signal in the middle of the 900 ms stimulus, see also in figure 2. As in experiment 2 the pedestal strength varied randomly. The interval between trials was the same as in experiment 2. All other aspects of the procedure were the same as in experiment 1 and 2.

## Results

The AOC graph for experiment 3 is plotted in figure 7. The points all fall closely to the ideal cost curve both in detection and identification. No points fall significantly below the curve, indicating that the task demands were not too high. Even though results from the AOC graph were clear we plotted light and sound separately for comparison in figure 8.

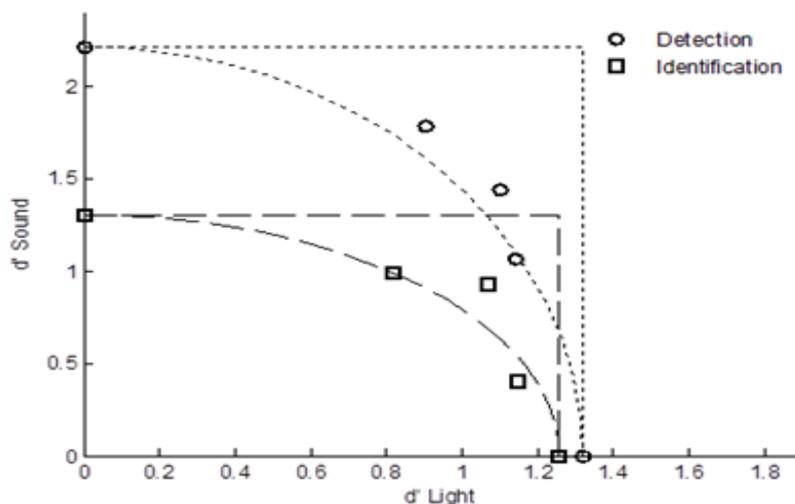


Figure 7. AOC graph for experiment 3. Ideal lines and markers are the same as in figure 3.

From figure 8 it is clear that a very comparable cost of divided attention occurs both for detection and identification. The  $d'$  values of sound in detection are all higher than the comparable identification points, however the drop in performance remains constant between detection and identification so there is a cost of divided attention in

both conditions in both modalities. This drop does not exceed the predicted drop of the ideal cost curve. There is less drop in  $d'$  values for light than for sound, but there is still a consistent downward trend, both in detection and identification for light.

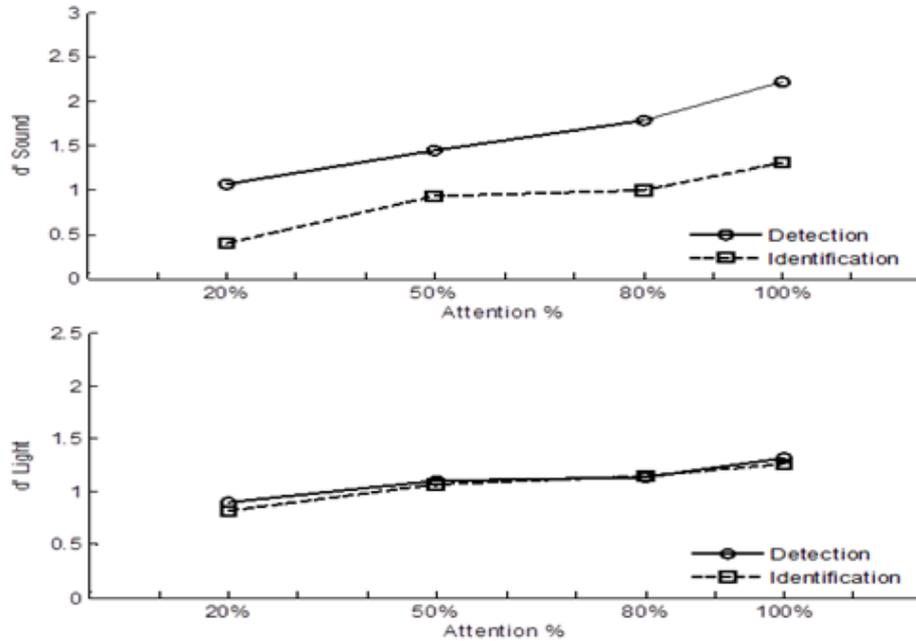


Figure 8. Performance in experiment 3, with a separated graph for light and sound. Axis and lines the same as in figure 4.

For detection  $C = 0.084$  for sound and  $C = 0.180$  for light. For identification  $C = -0.109$  for sound and  $C = -0.247$  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 [ $F(3,20) = 2.08, p = .14$ ] for sound and [ $F(3,20) = .97, p = .43$ ] for light. For identification [ $F(3,20) = 7.05, p \leq .01$ ] for sound and [ $F(3,20) = 1.51, p = .24$ ] for light, indicating a cost of divided attention for sound but not for light.

## 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. Participants were able to perform both tasks, indicating that there are other strategies available and that the trace/context model needs modification. The fact that in all three experiments there was a different pattern 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 each other out? These results would suggest so, but further

research is clearly needed. What this other strategy or strategies are that the participants used in this experiment is not clear. The fact remains that one or more strategies, that Hafter et al. (1998) did not account for, which there is no room for in their model, are available and need to be taken into account for a comprehensive model of divided multimodal attention.

### **General discussion**

There is a considerable amount of research that 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). This has been debated since, with no clear answer on the horizon. Scholars in this field continue to be divided into two camps, the supramodal camp, claiming a cost of divided attention (e.g. Massaro & Warner, 1977; Beer & Roder, 2005) and the modality independence camp, claiming no cost of divided attention (e.g. Alais, Morrone & Burr, 2006; Duncan, Martens & Ward, 1997). One reason, why these camps are as divided now, as they were 40 years ago is that they continue to conduct methodologically different experiments. This in itself is not a fundamental flaw as it is quite possible, and even likely, that both supramodal and modality independent systems exist, and that different experimental methods tap into those systems differently. A more serious problem is that there has been a prevalent fundamental methodological flaw in the experimental design of both camps. This flaw that experimenters have failed to take into account is the possible effect of a spatial confound. In most earlier research, light and sound were presented from different locations (Spence & Driver, 1997*b*). This makes it possible that any effect found is not the result of dividing attention between modalities, but the result of dividing attention between two spatial locations. Klein, Kingstone and Pontefract (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 (1997*b*) found that presenting light and sound at different spatial locations exaggerated the effect, compared to presenting the light and sound from the same spatial location. This casts doubt on any experimental finding that is based on the method of presenting light and sound separately. This has been the most common method, especially since computers became dominant in experimental settings, where visual stimuli are presented on a computer

screen while auditory stimuli are presented through headphones or speakers. This makes repeating many of those 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 implied by 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, is not as clear or as simple as previous models have suggested. One difference between our results and previous ones is that performance on sound and light signals did not follow the same pattern of cost or no cost within the same conditions. Both in experiment 1 and 2, the sound signal performance followed the predictions of the trace/context model. In experiment 1, where a gap was introduced, sound showed a cost of divided attention, even in detection. In experiment 2, with a roaming pedestal, performance for sound signals showed no cost of divided attention, even in identification. On the other hand, the light signal performance, in experiments 1 and 2, followed the predictions of the standard model, although it did not reach statistical significance in experiment 1 for identification. In the detection condition in experiment 1, performance in light signals showed no cost of divided attention even though there was a gap before and after the signal. In the identification condition in experiment 2, performance in light signals showed signs of a cost of divided attention even though the pedestals were roaming. This 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 (1997*a*) suggest that the interaction of the two modalities may not be symmetrical.

Although a bias was present in all three experiments and slight difference between the bias in 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 in detection and identification were very similar. This might seem at odds with the results of experiments 1 and 2, however if the gap and the roaming pedestal affect each modality differently there is a possibility that those effects cancelled each other out, explaining the similarity. The fact remains that performance, both in detection and identification did not fall below the ideal cost curve. With significance tests even indicating that there might be no cost for light. 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 main hypothesis is thus supported.

The standard model assumed that detection and identification were fundamentally different; detection was assumed to be a capacity free process occurring early on in the sensory process. Identification was assumed to require higher mental processing, occurring centrally and thus be a capacity limited process (Gallun, Hafter & Bonnel, 1998). Hafter et al. (1998) showed that the assumptions of the standard model were not universal and that the patterns of detection and identification could be reversed by introducing a gap or a roaming pedestal. They argued that the difference lies in different strategies used in the tasks but not in a difference between detection and identification. 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.

Further research is needed, especially since most of the older research suffers from the spatial confound. Amongst the questions that remain to be answered are, what strategies participants use, how many and their nature? How signal strength affects the patterns of cost in divided multimodal attention? And, how the nature of the task affects the results? In our experiments the signal was a change in strength, but it needs to be

investigated whether our results only apply to strength discrimination or if they apply to other tasks as well. Another issue that needs to be addressed is the possible effect of signal frequency on performance. Since the frequency of the signal followed the attentional instructions the signals in the 20% condition were less frequent. As Wolfe, Horowitz and Kenner (2005) showed, error rates increase when the signal is less frequent. 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 the frequency of the signals. Also, a systematic investigation of the effects of spatial confound on the results of divided multimodal attention research needs to be conducted in a similar manner to the experiments of Spence and Driver (1997*b*), where they investigated the effects of spatial confound in cuing paradigms. But one thing is certain, for this field to move forward and start building a better understanding of the true nature of the underlying mechanisms in multimodal divided attention, the methodology, especially in regards to spatial confounds, needs to be improved.

**Author note:** We would like to thank Árni Kristjánsson for invaluable support, through the experimental and the writing process. His comments, criticisms and help made this paper possible. Also, we would like to thank Guðmundur Björn Birkisson for countless hours of technical help, both in the preparation, building the experimental set-up and programming for the experiments.

## References

- Alais, D., Morrone, C. & Burr, D. (2006). Separate attentional resources for vision and audition. *Proceedings of the Royal Society B-Biological Sciences*, 273(1592), 1339–1345.
- Alwitt, L. (1981). Two neural mechanism related to modes of selective attention. *Journal of Experimental Psychology: Human Perception and Performance*, 7(2), 324-332.
- Beer, A. L. & Roder, B. (2005). Attending to visual or auditory motion affects perception within and across modalities: An event-related potential study. *The European Journal of Neuroscience*, 21(4), 1116–1130.
- Bonnel, A. M. & Hafter, E. R. (1998). Divided attention between simultaneous auditory and visual signals. *Perception and Psychophysics*, 60(2), 179–190.
- Bonnel, A-M, Stein, J. F. & Bertucci, P. (1992). Does attention modulate the perception of luminance changes? *The Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology*, 44(4), 601-626.
- Boot, W. R., Brockmole, J. R. & Simons, D. J. (2005). Attention capture is modulated in dual-task situations. *Psychonomic Bulletin & Review*, 12(4), 662-668.
- Brown, A. E. & Hopkins, H. K. (1967). Interactions of the auditory and visual sensory modalities. *Journal of the Acoustical Society of America*, 41(1), 1–7.
- Cosman, J. D. & Vecera, S. P. (2009). Perceptual load modulates attentional capture by abrupt onsets. *Psychonomic Bulletin & Review*, 16(2), 404–410.
- Calvert, G. A., Spence, C., & Stein, B. E. (Eds.). (2004). *The Handbook of Multisensory Processes* (pp. 3–25). Cambridge: MIT Press.
- Driver, J. & Spence, C. (1994). Spatial synergies between auditory and visual attention. In C. Umiltá & M. Moscovitch (eds.), *Attention and performance: conscious and nonconscious information processing* (pp. 311–331). Cambridge: MIT Press.
- Driver, J. & Spence, C. (1998). Cross-modal links in spatial attention. *Philosophical Transactions of the Royal Society B-Biological Sciences*, 353(1373), 1319–1331.

- Driver, J. & Spence, C. (2000). Multisensory perception: Beyond modularity and convergence. *Current Biology*, *10*(20), R731–R735.
- Duncan, J., Martens, S. & Ward, R. (1997). Restricted attentional capacity within but not between sensory modalities. *Nature*, *387*, 808–810.
- Ferlazzo, F., Couyoumdjian, A., Padovani, T. & Belardinelli, M. O. (2002). Head-centred meridian effect on auditory spatial attention orienting. *The Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology*, *55*(3), 937–963.
- Freedheim, D. K. (vol. ed.) & Weiner, I. B. (series ed.). (2003). *Handbook of Psychology: Volume 1, History of Psychology*. New Jersey: John Wiley & Sons.
- Frith, C. D. & Friston, K. J. (1996). The role of the thalamus in “top down” modulation of attention to sound. *Neuroimage*, *4*, 210-215.
- Gallun, E., Hafter, E. R. & Bonnel, A-M. (1998). The role of memory in the dual-task: Evidence from a frequency/amplitude judgement. *Journal of the Acoustical Society of America*, *103*(5), 3018-3019.
- Gibson, E. J. (1969). Principles of perceptual learning and development. New York: Appleton-Century-Crofts. In Freides, D. (1974). Human information processing and sensory modality: Cross-modal functions, information complexity, memory, and deficit. *Psychological Bulletin*, *81*(5), 284-310.
- Hafter, E. R., Bonnel, A-M., Gallun, E. & Cohen, F. (1998). A role for memory in divided attention between two independent stimuli. In A. R. Palmer, A. Rees, A. Q. Summerfield, & R. Meddis (eds.), *Psychophysical and Physiological Advances in Hearing* (pp. 228-238). London: Whurr Publishing.
- Holmes, N. P. & Spence, C. (2005). Multisensory Integration: Space, Time and Superadditivity. *Current Biology* *15*(18), 762-764.
- Kingdom, F. A. A. & Prins, N. (2010). *Psychophysics: A Practical Introduction* (pp. 178-187). London: Academic Press.

- Klein, R. M., Kingstone, A. & Pontefract, A. (1992). Orienting of visual attention. In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading* (pp. 46-65). New York: Springer-Verlag.
- Koelewijn, T., Bronkhorst, A. & Theeuwes, J. (2009a). Auditory and visual capture during focused visual attention. *Journal of Experimental Psychology -Human Perception and Performance*, 35(5), 1303–1315.
- Koelewijn, T., Bronkhorst, A. & Theeuwes, J. (2009b). Competition between auditory and visual spatial cues during visual task performance. *Experimental Brain Research*, 195 (4), 593–602.
- Koelewijn, T., Bronkhorst, A. & Theeuwes, J. (2010). Attention and the multiple stages of multisensory integration: A review of audiovisual studies. *Acta Psychologica*, 134(3), 372–384.
- Larsen, A., McIlhagga, W., Baert, J. & Bundesen, C. (2003). Seeing or hearing? Perceptual independence, modality confusions, and crossmodal congruity effects with focused and divided attention. *Perception & Psychophysics*, 65(4), 568–574.
- Leahey, T. H. (2012). *A History of Psychology: From Antiquity to Modernity: International Edition* (7<sup>th</sup> edition). New Jersey: Pearson.
- Macaluso, E., Frith, C. D. & Driver, J. (2000). Modulation of human visual cortex by crossmodal spatial attention. *Science*, 289, 1206–1208.
- Macmillan, N. A. & Creelman, C. D. (2004). *Detection theory: A users guide* (pp. 21). New Jersey: Lawrence Erlbaum Associates.
- Massaro, D. W. & Warner, D. S. (1977). Dividing attention between auditory and visual perception. *Perception & Psychophysics*, 21(6), 569–574.
- Mazza, V., Turatto, M., Rossi, M. & Umiltá, C. (2007). How automatic are audiovisual links in exogenous spatial attention? *Neuropsychologia*, 45(3), 514–522.
- McGurk, H. & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264, 746–748.

- McNicol, D. (1972). *A Primer of Signal Detection Theory* (pp. 53-56). New Jersey: Routledge.
- Müller, H. J. & Rabbitt, P. M. A. (1989). Reflexive and voluntary orienting of visual-attention - time course of activation and resistance to interruption. *Journal of Experimental Psychology -Human Perception and Performance*, 15(2), 315–330.
- Prime, D. J., McDonald, J. J., Green, J. & Ward, L. M. (2008). When cross-modal attention fails. *Canadian Journal of Experimental Psychology*, 62(3), 192-197.
- Santangelo, V. & Spence, C. (2007). Multisensory cues capture spatial attention regardless of perceptual load. *Journal of Experimental Psychology -Human Perception and Performance*, 33(6), 1311–1321.
- Santangelo, V., Belardinelli, M. O. & Spence, C. (2007). The suppression of reflexive visual and auditory orienting when attention is otherwise engaged. *Journal of Experimental Psychology -Human Perception and Performance*, 33(1), 137-148.
- Shiffrin, R. M. & Grantham, D. W. (1974). Can attention be allocated to sensory modalities? *Perception & Psychophysics*, 15(3), 460–474.
- Spence, C. & Driver, J. (1996). Audiovisual links in endogenous covert spatial attention. *Journal of Experimental Psychology -Human Perception and Performance*, 22(4), 1005–1030.
- Spence, C. & Driver, J. (1997a). Audiovisual links in exogenous covert spatial orienting. *Perception & Psychophysics*, 59(1), 1–22.
- Spence, C. & Driver, J. (1997b). On measuring selective attention to an expected sensory modality. *Perception & Psychophysics*, 59(3), 389–403.
- Spence, C., Nicholls, M. E. & Driver, J. (2001). The cost of expecting events in the wrong modality. *Perception & Psychophysics*, 63(2), 330–336.
- Stein, B. E., & Meredith, M. A. (1993). *The merging of the senses*. Cambridge: MIT.
- Stein, B. E., Wallace, M. W., Stanford, T. R. & Jiang, W. (2002). Cortex governs multisensory interactions in the midbrain. *Neuroscientist*, 8(4), 306–314.

- Taylor, M. M., Lindsay, P. H. & Forbes, S. M. (1967). Quantification of shared capacity processing in auditory and visual discrimination. *Acta Psychologica*, 27, 223–229.
- Theeuwes, J. (2004). Top-down search strategies cannot override attentional capture. *Psychonomic Bulletin & Review*, 11(1), 65-70.
- Tulving, E. & Lindsay, P. H. (1967). Identification of simultaneously presented simple visual and auditory stimuli. *Acta Psychologica*, 27, 101–109.
- Vroomen, J. & de Gelder, B. (2000). Sound enhances visual perception: Cross-modal effects of auditory organization on vision. *Journal of Experimental Psychology -Human Perception and Performance*, 26(5), 1583–1590.
- Ward, L. M. (1994). Supramodal and modality-specific mechanisms for stimulus-driven shifts of auditory and visual attention. *Canadian Journal of Experimental Psychology*, 48(2), 242–259.
- Wolfe, J. M., Horowitz, T. S. & Kenner, N. M. (2005). Rare items often missed in visual searches. *Nature*, 435, 439-440.