Temporal Multisensory Processing and its Effects on Attention

van der Burg, E.

2009

Summary

The research described in this thesis addresses perception across the senses. Specifically it investigates how auditory processing affects attention in the visual domain. The thesis is divided in two different sections. Section I is concerned with audiovisual integration in multiple object environments and shows that sound affects the competition between visual events (Chapters 2-7) when the two events are synchronized. Section II is concerned with the attentional blink between the auditory and visual modality and shows the conditions in which auditory processing takes away attention from visual processing (Chapters 8-9).

Chapter 1 introduces multisensory integration, and discusses integration when information from different sensory modalities is spatially aligned. We discuss in detail multisensory integration when information from different senses is presented in synchrony (temporal alignment). Furthermore, the introduction of this thesis highlights the role of attention in multisensory processing and discusses possible neural mechanism underlying multisensory integration.

In the first section, Chapter 2 addresses whether searching for an object within a cluttered, continuously changing environment can benefit from non-spatial auditory signals. Participants were asked to search for a horizontal or vertical line segment amongst other line segments of various orientations, all continuously changing color. We observed that searching for the target in this cluttered environment can be a very time consuming process. However, the addition of a simple auditory pip drastically decreased search times when the target color change was accompanied by an auditory signal. The sound ("pip") appeared to cause the visual target to pop out – hence we termed this phenomenon "pip and pop". Search benefits were observed even though the sound contained no information regarding the location or identity of the visual object. The experiments also showed that the effect is not due to general alerting (as it does not occur with visual cues), nor due to top-down cueing of the visual change (as it still occurs when the pip is synchronized with distractors on the majority of trials). Instead, the spatially non-specific sound appears to affect the competition between multiple visual events by increasing the salience of one of these events.

Chapter 3 investigates the exact nature of the signal necessary to obtain auditory driven visual search. Search for a visual target improved when a target

luminance change was uniquely synchronized with an auditory signal, but only when the audio-visual events were square-wave modulated, but not when they were sine-wave modulated. Equivalent improvements in search performance in the square-wave condition were observed using a difference modulation whose temporal profile is the difference between square and sine wave modulations. This indicates that the waveform effect is not due to overall differences in amplitude or the presence of discrete vs. continuous temporal information. Instead, additional results indicate that the visual search benefits of audiovisual temporal alignment are contingent upon the abrupt change of the audiovisual signal, with visual transients being particularly important. These results imply that temporal alignment alone is not sufficient for audio-visual integration to occur, and suggest that transient sensory information is necessary to avoid false matches.

In Chapters 2 and 3, we report evidence that abrupt audiovisual synchrony guides attention. Chapter 4 provides evidence that it does this in an automatic fashion. Participants were asked to make a temporal order judgment (TOJ) about which of two dots (left or right) appeared first. Lateral to each of the dots, nine irrelevant distractors continuously changed color. Prior to the presentation of the first dot, an auditory signal was synchronized with the color change of one of these distractors, either on the same side or opposite side of the first dot. Even though both the tone and the distractors were completely irrelevant to the task, TOJs were affected by the location of the synchronized distractor. TOJs were not affected when the tone was absent, or synchronized with distractors on both sides. The experiments also rule out an alternative explanation that the observed effects were due to a response bias, as a synchronized distractor had an identical effect when participants were asked to make a simultaneity judgment (SJ) instead of a temporal order judgment. We conclude that audiovisual synchrony guides attention in an automatic fashion.

In Chapter 5, we investigated the neural mechanisms underlying attentional guidance by audiovisual stimuli in multiple object environments. Event-related potentials (ERPs) were measured during visual search through dynamic displays consisting of line elements that randomly changed orientation. Consistent with Chapters 2 and 3, search improved when a target event was synchronized with a spatially uninformative auditory signal, compared to when the auditory signal was absent, or when the auditory signal was synchronized with an orientation change of a distractor. Important, the ERP data revealed clear evidence that the observed search benefits were due to an early multisensory interaction over the left parieto-occipital electrodes (~50 ms post-stimulus onset). This multisensory integration appeared to drive a lateralized effect over the occipital cortex (~ 80 ms) that we attribute to an increase in visual salience. This saliency boost is followed by an enhanced N2pc and SPCN effects reflecting boosts in attentive and visual short term memory processes. The latency of the early parieto-occipital and occipital effects is consistent with the notion that auditory signals can affect visual events at an early perceptual processing stage. Moreover, in case that a single distractor orientation change was accompanied by an auditory signal, we found a similar multisensory interaction, a similar early lateralized saliency boost effect as well as a similar N2pc as in the case a target orientation change was accompanied by an auditory signal. These effects corroborate the findings reported in Chapters 2 and 4, in which we found behavioral evidence that the pip and pop effect is due to automatic capture. We conclude that due to early multisensory interaction, the auditory signal boosts the salience of a synchronized visual event, which increases the chances of selection in a multiple object environment, and that this occurs automatically.

In Chapter 6, we investigated whether this pip and pop effect is restricted to *audio*visual synchrony or can extend to other modalities. The experiment in Chapter 6 was identical to the experiment in Chapter 2 (Experiment 1), except that the auditory signal was replaced by a tactile signal. Consistent with Chapters 2 and 3, search times and search slopes were substantially reduced when the target color change was accompanied by a signal from another modality. However, in this case search benefits were due to tactile-visual synchrony instead of audio-visual synchrony. The search benefits illustrate that tactile-visual synchrony guides attention in multiple object environments by increasing the saliency of the visual event. Therefore, we conclude that the pip and pop effect is not limited to the audiovisual domain but is a true multisensory phenomenon.

The previous Chapters (2-6) illustrate that visual search is markedly improved when a target color, or orientation change is synchronized with a spatially non-informative auditory, or tactile signal. Other studies investigating visual attention have indicated that automatic attentional capture is susceptible to the extent to which attention is distributed across the display (i.e. the size of the attentional window). In Chapter 7, we investigated whether this is also the case for the pip and pop effect by manipulating the extent to which participants divide their attention across the visual field. We demonstrated that participants were overall faster in detecting a synchronized audiovisual event when they divided their attention across the visual field relative to a condition in which they focused their attention. We argue that audiovisual integration, although largely automatic (see Chapters 2, 4-5), partly depends on top-down settings such as the size of the attentional window (see also Chapter 2).

In the second empirical section, we investigated the influences of audition on visual attention using a different paradigm, rapid serial visual presentation (RSVP). In RSVP observers are asked to detect targets in rapid streams of distractors, all presented at the same location. When there are two targets, the second is often missed, a phenomenon referred to as the "attentional blink". Whereas the attentional blink is a robust phenomenon within sensory modalities, the evidence for cross-modal attentional blinks is rather mixed. In Chapter 8 we tested whether the earlier-found absence of an auditory-visual attentional blink is due to the efficient use of echoic memory, allowing for the postponement of auditory processing until the visual target has been detected. Observers were asked to detect target beeps, followed by a visual target letter. Contrary to the hypothesis, forcing participants to immediately process the auditory target, by either presenting interfering sounds during retrieval or by making the first target directly relevant for a speeded response to the second target, did not result in a return of a cross-modal attentional blink. The findings argue against echoic memory as an explanation for efficient cross-modal processing. Instead, we hypothesized that a cross-modal attentional blink may be observed when the different modalities use common representations, such as semantic representations. In support of this, a deficit for visual letter recognition returned when the auditory task required a distinction between spoken digits and letters.

In Chapter 8, we showed that an attentional blink between the auditory and visual modality occurred in the case we used semantic stimuli, indicating that the cross-modal attentional blink is due to semantic interference. In Chapter 9, we investigated whether audiovisual semantic interference also occurs under conditions in which participants were asked to ignore simultaneously presented letters. More specifically, we investigated the role of attention in audiovisual semantic interference, by using an attentional blink paradigm in which the auditory and visual event were sometimes relevant to the task, and sometimes irrelevant to the task. Participant's task was to make an unspeeded response to the identity of a visual target letter. This target letter (T2) was preceded at various SOAs by a synchronized audiovisual letter-pair, which was either congruent (e.g. hearing an "F" and viewing an "F") or incongruent (e.g. hearing an "F" and viewing a "Z"). An attentional blink was found in all three experiments indicating that the audiovisual letter pairs were processed. T2 performance was dependent on the congruency of the audiovisual letter-pair when participants were asked to report one of the audiovisual letters, but no such congruency effect was observed when participants were asked to ignore the audiovisual letters. These results indicate that attention to at least one modality is necessary to establish audiovisual semantic interference.