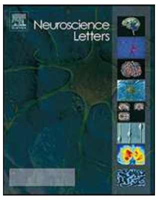
Neuroscience Letters 450 (2009) 60–64

Contents lists available at ScienceDirect

Neuroscience Letters

journal homepage: www.elsevier.com/locate/neulet



Poke and pop: Tactile–visual synchrony increases visual saliency

Erik Van der Burg[a,*], Christian N.L. Olivers[a], Adelbert W. Bronkhorst[a,b], Jan Theeuwes[a]

[a] Vrije Universiteit Amsterdam, The Netherlands
[b] TNO Human Factors, Soesterberg, The Netherlands

ARTICLE INFO

Article history:
Received 27 June 2008
Received in revised form 8 October 2008
Accepted 4 November 2008

Keywords:
Multisensory integration
Attention
Touch
Vision
Pip and pop effect

ABSTRACT

The majority of studies investigating interactions between vision and touch have typically explored single events, presenting one object at a time. The present study investigates how tactile–visual interactions affect competition between multiple visual objects in more dynamic cluttered environments. Participants searched for a horizontal or vertical line segment among distractor line segments of various orientations, all continuously changing color. Search times and search slopes were substantially reduced when the target color change was accompanied by a tactile signal. These benefits were observed even though the tactile signal was uninformative about the location, orientation, or color of the visual target. We conclude that tactile–visual synchrony guides attention in multiple object environments by increasing the saliency of the visual event.

In our everyday life, we receive a bulk of information via different senses. These different sensory inputs often interact when presented in close temporal or spatial proximity [see 23, 36, 38, for reviews]. For instance, in a noisy environment, we can understand a speaker better when we observe his or her lip movements [27]. Most studies have focused on interactions between audition and vision, but interactions between vision and touch have also been reported [e.g. 6, 16, 22, 24, 34]. For instance, Spence et al. [24] have shown that responses to visual targets were faster when such targets were preceded by a tactile signal from the same location (compared to signals coming from a different location), indicating that spatially informative tactile cues affect spatial visual selection.

Studies investigating interactions between vision and touch have typically used stimuli consisting of single tactile and/or visual events, and thus say little about multisensory interactions in more cluttered, multiple-object environments. An exception is a study by Lindeman et al. [8]. They found that visual search through multiple objects improved when a tactile signal informed participants about the location of the visual target (compared to a tactile signal absent condition). The finding from Lindeman et al. is consistent with other studies showing that auditory [e.g. 17], or visual [e.g. 19, 29, 31] top-down knowledge about the location of a visual target affects visual selection. Here we show that tactile signals do not need to support such top-down knowledge about the target's location in order to guide selection in a multiple-object environment, as long as the tactile signal is temporally synchronized with the target event.

Recently, within the auditory and visual domains, Van der Burg et al. [33] found that auditory signals affected the selection of targets in visual search, even when the auditory event carried no information on the target's location or identity. Participants searched for a horizontal or vertical line segment among up to 48 distractor lines of various orientations, all continuously changing color. They found that search times as a function of set size were substantially reduced when the target color change was accompanied by an auditory signal (compared when no such signal was present). Van der Burg et al. labeled this search benefit the "pip and pop" phenomenon, and suggested that audiovisual synchrony guides attention by increasing the saliency of the visual event [see also 32].

The current study is a follow-up to this auditory–visual "pip and pop" investigation. We investigated whether tactile–visual synchrony similarly modulates spatial visual selection in multiple object displays. We used the paradigm of Van der Burg et al. [33], except that we replaced the auditory signal by a tactile signal. We show that this tactile signal drastically decreases search times as well as search slopes for a synchronized visual object that is normally very difficult to find.

Nine participants (5 female, mean age 19.9 years; range 16–24 years) participated for €7 an hour. One participant was excluded from further analysis because of an overall error rate of 25%.

The experiment was run in a dimly lit, air-conditioned cabin. Participants were seated at approximately 80 cm from the monitor. The tactile stimulus was a vibration with a 50 ms duration. The

---

[*] Corresponding author at: Cognitive Psychology, Vrije Universiteit Amsterdam, Boechorststraat 1, 1081BT Amsterdam, The Netherlands. Tel.: +31 20 598 6744.
E-mail address: e.van.der.burg@psy.vu.nl (E. Van der Burg).

0304-3940/$ – see front matter © 2008 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.neulet.2008.11.002

tactor was a mobile phone vibrator attached to the back of the left hand. Participants wore closed headphones (Sennheiser HD202) to prevent them from hearing the vibrator.[1] The visual search displays consisted of 24, 36, or 48 red (13.9 cd m^{-2}) or green (46.4 cd m^{-2}) line segments (length 0.57° visual angle) on a black (<0.05 cd m^{-2}) background. Color was randomly determined for each item. All lines were randomly placed in an invisible 10×10 grid ($9.58° \times 9.58°$, 0–0.34° jitter) centered on a white (76.7 cd m^{-2}) fixation dot, with the constraint that the target was never presented at the four central positions, to avoid immediate target detection. The orientation of each line deviated randomly by either plus or minus 22.5° from horizontal or vertical, except for the target which was horizontal or vertical. The displays changed continuously in randomly generated cycles of 9 intervals each. The length of each interval varied randomly between 50, 100 or 150 ms with the constraint that all intervals occurred equally often within each cycle and that the target change was always preceded by a 150 ms interval and followed by a 100 ms interval. At the start of each interval, a randomly determined number of search items changed color (from red to green or vice versa), with the following constraints: when set size was 24, the number of items that changed was either 1, 2, or 3. When set size was 36 either 1, 3, or 5 items changed, and when it was 48 either 1, 4, or 7 items changed. Furthermore, the target always changed alone, and could only change once per cycle, so that the average frequency was 1.11 Hz. The target could not change during the first 500 ms of the very first cycle of each trial. On average, the first target color change was 750 ms after the display onset, which was independent of the set size. For each trial, 10 different cycles were generated, which were then repeated after the 10th cycle if the participant had not yet responded.

The set size was either 24, 36, or 48. The important manipulation involved the presentation of a tactile signal (present or absent), of which the onset was always synchronized with the target color change. Dependent variables were the reaction time (RT) and accuracy. Note that the RT reflects the time between the search display onset and the response to the target, because the target was present when the search display appeared. Each trial began with a fixation dot presented for 1000 ms at the center of the screen. The search display was presented until participants responded. Participants were asked to fixate on the fixation dot. Participants were instructed to press the n- or m-key with their right hand on the standard keyboard as fast and accurately as possible when the target orientation was horizontal or vertical, respectively. Target orientation was balanced and randomly mixed within blocks of 48 trials each. Participants received four tactile signal absent blocks, and four tactile signal present blocks, presented in counterbalanced, alternating order, preceded by two practice blocks. Participants received feedback about their overall mean accuracy and overall mean RT after each block.

The results are presented in Fig. 1. RT data from practice blocks, erroneous trials, and trials in which participants responded slower or faster than 2.5 times the S.D. (4.0%) were excluded. All data were subjected to a repeated-measures Univariate Analysis of Variance (ANOVA) with set size (24, 36 vs. 48) and tactile signal presence (present vs. absent) as within-subject variables. The reported values for p are those after a Huynh–Feldt correction for sphericity violations, with alpha set at .05. The overall mean error rate was 3.2%. There were no reliable error effects (all $ps > .12$).

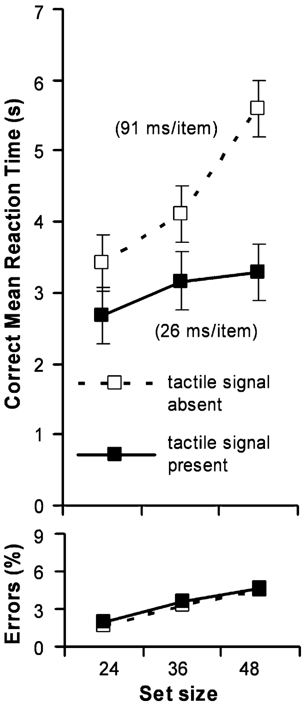


Fig. 1. Correct mean reaction time (RT) and mean error percentages, as a function of set size, and the presence of the tactile signal. Note that the reaction time reflects the time from the search display onset. The first target color change (and tone onset) was between 500 and 900 ms later. The error bars represent the .95 confidence intervals for within-subject designs, following Loftus and Masson [9]. The confidence intervals are those for the interaction between tactile signal presence and set size.

On average, RTs were faster when the tactile signal was present (3043 ms) than when the tactile signal was absent (4376 ms), $F(1, 7) = 22.7$, $p < .005$. Furthermore, search was more efficient in the tactile signal present condition than in the tactile signal absent condition, as confirmed by a significant two-way interaction between set size and tactile signal presence, $F(2, 14) = 11.7$, $p < .005$. In the tactile signal absent condition, the average search slope measured 91 ms/item, as RTs increased significantly with increasing set size, $F(2, 14) = 33.4$, $p < .005$. In the tactile signal present condition, the average search slope measured 26 ms/item. However, this set size effect on RTs was not reliable, $F(2, 14) = 2.7$, $p = .117$.

The substantial search slope and the overall somewhat long search times in the tactile signal present condition suggests that the target did not pop out on each trial. Similar to the Van der Burg et al. [33] study, observers presumably waited for the first synchronized event before they started to search. On average, the first synchronized occurred 750 ms after the display onset. The overall search times in each condition may thus be regarded as 750 ms shorter than is plotted in Fig. 1. Fig. 2 represents the RT distributions for the tactile signal present and absent condition, pooled across set size, and locked to the first target color change (which was also the time of the first tactile signal in the tactile signal present condition; bin size was 200 ms). As is clear from Fig. 2, the tactile signal present condition shows a marked peak at around 900 ms, compared to the tactile signal absent condition [see also

---

[1] To exclude the possibility that participants could *hear* the tactor, a control experiment (eight participants; 5 female, mean age 22.3 years; range 18–34 years) was conducted in which the tactor was attached to the back of the first author's left hand rather than the participant's hand. While wearing the headphones, the participant saw a fixation dot (1000 ms), followed by two 1200 ms intervals visually indicated by the digits "1" and "2" in the center of the screen. On each trial, a tactile signal (duration 50 ms) was presented during either of the two intervals. Participants were instructed to try and listen in which interval the tactor was switched on, and to make an unspeeded response by pressing the 1- or 2-key on the numeric keypad. Overall accuracy was 52.5%. Perceptual sensitivity (d') measured .105, and the bias (c) was .008. None of these measures differed significantly from chance performance (all $ts < 1$, all $ps > .45$). We conclude that participants were unable to hear the tactile signal.

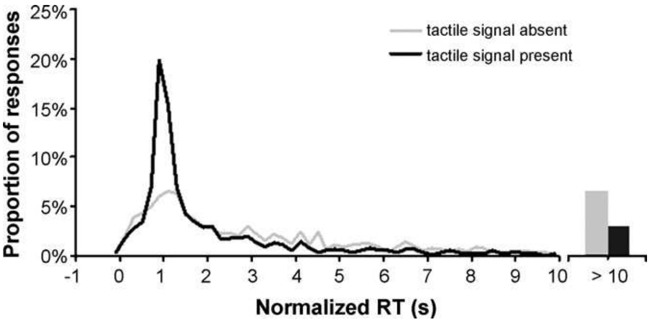


Fig. 2. Reaction time (RT) distributions. Here the proportion of responses is plotted as a function of the normalized RT (bin width is 200 ms). The normalized RT is the time to respond to the visual target from the first target color change.

33]. This peak was on average the time the second tactile signal could occur. On most trials in this condition, this second tactile signal occurred too late to affect the response, but it is possible that occasionally, because of eye blinks or other factors, observers waited for the second tactile signal. Therefore, it looks as if on the majority of trials, the target popped out after the first tactile signal. In the tactile signal absent condition, the RT distribution spanned a broader range, including substantially more responses of 10 s and longer.

The present findings are clear. Despite the fact that when the target changed color it was the only item changing at that time (i.e. the target change represented a unique event), finding the target required strong attentional effort when there was no tactile signal. Apparently, the temporally adjacent dynamics camouflaged the unique target event [see, e.g. 33, 35]. In the tactile present condition, the synchronized tactile stimulus caused a dramatic improvement in detecting the visual event. Thus, a tactile signal improves spatial visual search in dynamic and cluttered visual search displays, despite the fact that the tactile signal contained no information about the location, orientation, or color of the visual target.

One explanation for the search benefits is that the tactile signal acted as a warning signal on when to expect a visual color change. Even though this is certainly a possibility, we think it is unlikely that alerting or arousal can explain the present findings. First, most theories place warning effects at a postperceptual response level, later in the information-processing stream than the current selection effects are expected to reside [5,10,20]. Note that the tactile stimulus contained no information about which response should be prepared. Second, if the tactile stimulus affected only non-specific response preparation, then we would have expected effects only on overall RTs, not in terms of search slopes. Third, we observed benefits in the order of seconds (see the higher set sizes), while warning signals have been shown to improve RTs by a fraction of a second. Fourth, in an identical experimental set-up, Van der Burg et al. [33] have demonstrated that visual warning signals (with the same temporal information as the tactile signal in the present study) are not sufficient to aid search. Fifth, in the present study, effects were observed when that tactile signal was synchronized with the visual event. Because a state of alertness, or arousal needs time to develop [see, e.g. 18], one would not have expected to find a benefit when the signals are synchronized, is inconsistent with an alerting or arousal hypothesis. In an additional control experiment,[2] we tested whether participants perceived the tactile and visual target events as concurrent. Note that participants perceived the events as fully synchronized. All in all then, what we propose is that tactile–visual synchrony guides attention in multiple object displays. The present finding extends the pip and pop effect [33], by showing that other modalities than audition affect visual selection of synchronized visual objects.

The present study is not the first to show effects of tactile stimulation on spatial visual processing. Whereas earlier studies [e.g. 7, 24] demonstrate performance benefits when touch and vision were spatially correlated, the present study revealed evidence that this spatial accordance is not always necessarily to affect spatial visual processing. Note however that, in the present study, it is not quite true that the tactile and visual events were never spatially correlated. After all, the tactor was attached to the left hand, which may have resulted in benefits only when the target happened to appear in the left visual field. To test for any such hemisphere effects, we compared performance in the tactile present condition for targets presented on the left of fixation to that for targets presented on the right, but found no reliable difference, $t(7) = .175$, $p > .85$.

The present study is also not the first to show effects of non-spatial tactile signals on visual processing. Keetels and Vroomen [6] asked participants to make a temporal order judgment (TOJ) about which of two visual stimuli appeared first. TOJs were improved when the first visual stimulus was preceded by a tactile signal while the second visual stimulus was followed by a tactile signal, as compared to a condition in which tactile signals were absent. Moreover, consistent with the present findings, Keetels and Vroomen have shown that this “tactile–visual temporal ventriloquism” was also present when the tactile signals were synchronized with the visual events, and importantly, that spatial discordance between the two modalities did not affect the phenomenon. While Keetels and Vroomen [6] have shown that tactile signals affect visual processing when using single visual events at a time, the present study provides evidence that tactile–visual synchrony modulates the competition between visual objects in multiple object environments.

Whereas we believe that the present findings are due to a pop-out of the visual target, the non-zero search slope and the somewhat long search times in the tactile signal present condition suggest that some mental effort is nevertheless required for the search. Note that all subjects showed a decreased search slope in the tactile present condition compared to the tactile absent condition. However, some observers were more efficient than others. Van der Burg et al. [33, Experiment 4] have shown that, in the audiovisual domain, the pip and pop phenomenon is susceptible to eye-movements or eye-blinks, explaining some individual differences. For instance, on some trials, some observers missed the first target color change due to an eye-movement, explaining the overall somewhat long search times, because the second target event occurred on average 900 ms later. Another feasible explanation is that the size of an attentional window modulates capture by tactile–visual synchrony. Recently, within the visual domain, Belopolsky et al. [2] have shown that the size of

---

[2] The perceptual synchrony between the tactile and visual events was investigated by a control experiment. Participants (eight participants; 1 female, mean age 31.6 years; range 25–35 years) were asked to make a simultaneity judgment (SJ), about whether or not the tactile signal and a single target color change were simultaneously presented. Such a simultaneity judgment task has been shown to be independent of response biases [32]. In this control experiment, the search display was static (set size was always 36), except that the target (the horizontal or vertical line) changed color once. Important, a single tactile signal (50 ms duration) preceded or followed this target color change by a randomly determined SOA (150, 50, 25, 8 or 0 ms). The point of subjective simultaneity turned out to be on average 0.6 ms, indicating that the visual event had to lead the tactile event by 0.6 ms for simultaneity to be reached. We conclude that participants perceived the tactile and visual target color change as simultaneous.

the attentional window affects capture by color singletons [see also 30]. In the Belopolsky et al. study, search performance for a color singleton improved when participants adopted a diffuse attentional window compared to the condition that participants adopted a small, focused attentional window. With regard to the present study, participants might have missed the visual target event if they adopted a small attentional window (e.g. due to attention towards an irrelevant distractor color change), explaining the absence of capture at that specific moment. Moreover, neurophysiological [28] as well as behavioral [1] studies corroborate the notion that attention is required to obtain multisensory integration. All in all then, it is clear that the present bottom-up signals are not always sufficiently strong to result in automatic attentional capture. Until further research is conducted, we propose here that tactile–visual synchrony guides attention, but only when participants are in a distributed state of attention.

Since both auditory and tactile signals evoke what appears to be an identical effect, one might expect this to be mediated by a neural mechanism, which is influenced by multiple sensory inputs. For instance, Meredith and Stein [13] demonstrated multisensory integration in single superior colliculus (SC) neurons [see 26, for a recent review]. The idea is that each multisensory neuron has several receptive fields for each specific modality (e.g. vision, audition, and touch), which are in spatial register with each other. As a result, two stimulus modalities will be defined as originating from the same source location, and perceived as a unified percept. Importantly, if two modalities originate from two different locations (as in the present study), then one stimulus falls within and the other outside the neuron's receptive field, explaining the absence of multisensory enhancement. Therefore, we consider this explanation as an unlikely underlying mechanism for the perceived auditory and tactile effects on vision.

Alternatively, within the auditory and visual domains, substantial evidence has emerged for the idea that the auditory signal enhances the perception of visual stimuli automatically [15,25,32,33,37], and that this enhancement could be due to auditory activation in the early visual cortex [3,4,14,21]. Within the tactile and visual domains, Violentyev et al. [34] observed that participants reported seeing two flashes on the majority of trials when a single flash was accompanied by two tactile signals. This "touch-induced visual illusion" occurred even though the tactile signals were completely irrelevant to the task, suggesting a strong automatic component for the reported phenomenon. Neurological studies [11,12] corroborate the notion that tactile signals affect visual perception, by demonstrating early tactile activation in the visual cortex. Consistent with Violentyev et al., we propose that the tactile signal boosts the saliency of a concurrently presented visual event, resulting in a salient emergent feature that pops out from the cluttered visual environment, and guides attention to the relevant location. All in all then, we believe that the visual cortex receives auditory as well as tactile information, explaining the pip and pop effect as well as the poke and pop effect, respectively.

Acknowledgements

This research was supported by a Dutch Technology Foundation STW grant (07079), a division of NWO and the Technology Program of the Ministry of Economic Affairs (to Jan Theeuwes and Adelbert W. Bronkhorst), and a NWO-VIDI grant 452-06-007 (to Christian N.L. Olivers). Correspondence concerning this article should be addressed to Erik van der Burg, Department of Cognitive Psychology, Vrije Universiteit Amsterdam, The Netherlands. E-mail: e.van.der.burg@psy.vu.nl.

References

[1] A. Alsius, J. Navarra, R. Campbell, S. Soto-Faraco, Audiovisual integration of speech falters under attention demands, Current Biology 15 (2005) 839–843.
[2] A.V. Belopolsky, L. Zwaan, J. Theeuwes, A.F. Kramer, The size of an attentional window modulates attentional capture by color singletons, Psychonomic Bulletin & Review 14 (2007) 934–938.
[3] A. Falchier, S. Clavagnier, P. Barone, H. Kennedy, Anatomical evidence of multimodal integration in primate striate cortex, Journal of Neuroscience 22 (2002) 5749–5759.
[4] M.H. Giard, F. Peronnet, Auditory–visual integration during multimodal object recognition in humans: a behavioral and electrophysical study, Journal of Cognitive Neuroscience 11 (1999) 473–490.
[5] S.A. Hackley, F. Valle-Inclán, Which stages of processing are speeded by a warning signal? Biological Psychology 64 (2003) 27–45.
[6] M. Keetels, J. Vroomen, Tactile–visual temporal ventriloquism: no effect of spatial disparity, Perception & Psychophysics 70 (2008) 765–771.
[7] S. Kennett, C. Spence, J. Driver, Visuo-tactile links in covert exogenous spatial attention remap across changes in unseen hand posture, Perception & Psychophysics 64 (2002) 1083–1094.
[8] R.W. Lindeman, Y. Yanagida, J.L. Sibert, R. Lavine, Effective vibrotactile cueing in a visual search task, in: R.M.e. al. (ed.), Human Computer Interaction, IOS Press, 2003, pp. 89–96.
[9] G.R. Loftus, M.E.J. Masson, Using confidence intervals in within-subject designs, Psychonomic Bulletin & Review 1 (1994) 476–490.
[10] S.A. Los, M.L.J. Schut, The effective time course of preparation, Cognitive Psychology 57 (2008) 20–55.
[11] E. Macaluso, C.D. Frith, J. Driver, Crossmodal spatial influences of touch on extrastriate visual areas take current gaze direction into account, Neuron 34 (2002) 647–658.
[12] E. Macaluso, C.D. Frith, J. Driver, Modulation of human visual cortex by crossmodal spatial attention, Science 289 (2000) 1206–1208.
[13] M.A. Meredith, B.E. Stein, Interactions among converging sensory inputs in the superior colliculus, Science 221 (1983) 389–391.
[14] S. Molholm, W. Ritter, M.M. Murray, D.C. Javitt, C.E. Schroeder, J.J. Foxe, Multisensory auditory–visual interactions during early sensory processing in humans: a high-density electrical mapping study, Brain Research Cognitive Brain Research 14 (2002) 115–128.
[15] C.N.L. Olivers, E. Van der Burg, Bleeping you out of the blink: sound saves vision from oblivion, Brain Research 1242 (2008) 191–199.
[16] F. Pavani, C. Spence, J. Driver, Visual capture of touch: out of body experiences with rubber gloves, Psychological Science 11 (2000) 353–359.
[17] D.R. Perrott, K. Saberi, K. Brown, T.Z. Strybel, Auditory psychomotor coordination and visual search performance, Perception & Psychophysics 48 (1990) 214–226.
[18] E.A. Phelps, S. Ling, M. Carrasco, Emotion facilitates perception and potentiates the perceptual benefits of attention, Psychological Science 17 (2006) 292–299.
[19] M.I. Posner, Orienting of attention, Quarterly Journal of Experimental Psychology 32 (1980) 3–25.
[20] M.I. Posner, S.J. Boies, Components of attention, Psychological Review 78 (1971) 391–408.
[21] C.E. Schroeder, J.J. Foxe, Multisensory contributions to low-level, 'unisensory' processing, Current Opinion in Neurobiology 15 (2005) 454–458.
[22] D.I. Shore, N. Simic, Integration of visual and tactile stimuli: top-down influences require time, Experimental Brain Research 166 (2005) 509–517.
[23] C. Spence, Audiovisual multisensory integration, Acoustical Science and Technology 28 (2007) 61–70.
[24] C. Spence, M.E.R. Nicholls, N. Gillespie, J. Driver, Cross-modal links in exogenous covert spatial orienting between touch, audition, and vision, Perception & Psychophysics 60 (1998) 544–557.
[25] B.E. Stein, N. London, L.K. Wilkinson, D.D. Price, Enhancement of perceived visual intensity by auditory stimuli: a psychological analysis, Journal of Cognitive Neuroscience 8 (1996) 497–506.
[26] B.E. Stein, T.R. Stanford, Multisensory integration: current issues from the perspective of the single neuron, Nature Reviews Neuroscience 9 (2008) 255–266.
[27] W.H. Sumby, I. Pollack, Visual contribution to speech intelligibility in noise, The Journal of the Acoustical Society of America 26 (1954) 212–215.
[28] D. Talsma, T.J. Doty, M.G. Woldorff, Selective attention and audiovisual integration: is attending to both modalities a prerequisite for early integration? Cerebral Cortex 17 (2007) 691–701.
[29] J. Theeuwes, Cross-dimensional perceptual selectivity, Perception & Psychophysics 50 (1991) 184–193.
[30] J. Theeuwes, Perceptual selectivity for color and form, Perception & Psychophysics 51 (1992) 599–606.
[31] J. Theeuwes, E. Van der Burg, The role of spatial and non-spatial information in visual search, Journal of Experimental Psychology: Human Perception and Performance 33 (2007) 1335–1351.
[32] E. Van der Burg, C.N.L. Olivers, A.W. Bronkhorst, J. Theeuwes, Audiovisual events capture attention: evidence from temporal order judgments, Journal of Vision 8 (2008) 1–10.
[33] E. Van der Burg, C.N.L. Olivers, A.W. Bronkhorst, J. Theeuwes, Pip and pop: nonspatial auditory signals improve spatial visual search, Journal of Experimental Psychology: Human Perception and Performance 34 (2008) 1053–1065.
[34] A. Violentyev, S. Shimojo, L. Shams, Touch-induced visual illusion, NeuroReport 16 (2005) 1107–1110.

[35] A. Von Mühlenen, M.I. Rempel, J.T. Enns, Unique temporal changes is the key to attentional capture, Psychological Science 16 (2005) 979–986.
[36] J. Vroomen, B. De Gelder, Perceptual effects of cross-modal stimulation: ventriloquism and the freezing phenomenon, in: G. Calvert, C. Spence, B.E. Stein (Eds.), Handbook of Multisensory Processes, MIT Press, 2004, pp. 141–150.
[37] J. Vroomen, B. De Gelder, Sound enhances visual perception: cross-modal effects of auditory organization on vision, Journal of Experimental Psychology: Human Perception and Performance 26 (2000) 1583–1590.
[38] R.B. Welch, D.H. Warren, Immediate perceptual response to intersensory discrepancy, Psychological Bulletin 88 (1980) 638–667.