Flash illusions induced by visual, auditory, and audiovisual stimuli

Deborah Apthorp

Research School of Psychology, Australian National University, Canberra, Australian Capital Territory, Australia
School of Psychology, University of Wollongong, New South Wales, Australia

David Alais

School of Psychology, University of Sydney, New South Wales, Australia

Lars T. Boenke

School of Psychology, University of Sydney, New South Wales, Australia
Leibniz Institute for Neurobiology (LIN), Magdeburg, Germany

When two objects are flashed at one location in close temporal proximity in the visual periphery, an intriguing illusion occurs whereby a single flash presented concurrently at another location appears to flash twice (the visual double-flash illusion: Chatterjee et al., 2011, Wilson & Singer, 1981). Here, for the first time, we investigate the time course of the effect, and directly compare it to the time course of the auditory (sound-induced flash illusion) effect, for both fission (single test flash, double inducer) and fusion (double test flash, single inducer) conditions, across stimulus onset asynchronies (SOAs) of 30 to 250 ms. In addition, using a novel audiovisual stimulus, we directly compare the cue strength of the two modalities, and whether they are additive in effect. The results show that the time course of fission and fusion is different for visual inducers, but not for auditory inducers. In audiovisual conditions, in situations of uncertainty, observers tended to follow the more reliable (auditory) cue. There was little evidence for a superadditive effect of auditory and visual cues; rather, observers tended to follow one or the other modality. The results suggest that the visually induced flash illusion and the auditory-induced effect may both stem from perceptual uncertainty, with the difference in time courses attributable to the lower temporal resolution of vision compared to audition.

Introduction

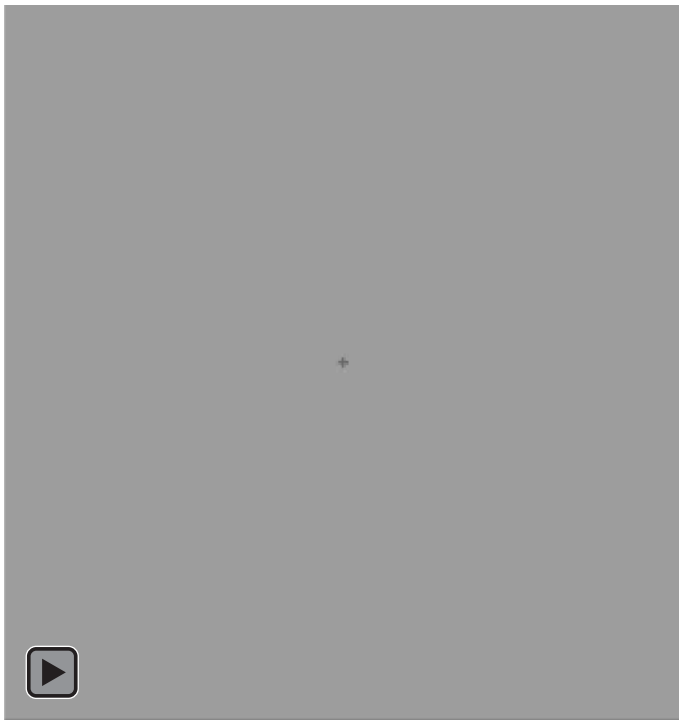
The visually induced double flash illusion (VIFI) is less well known than the sound-induced flash illusion (SIFI; Shams, Kamitani, & Shimojo, 2000; Wilson, 1987), but it offers a compelling example of visual temporal interactions over considerable spatial distances (Chatterjee, Wu, & Sheth, 2011; Leonards & Singer, 1997; Wilson, 1987; Wilson & Singer, 1981). The effect occurs when two flashes in quick succession are presented in one area of the visual field (e.g., the upper visual field), simultaneously with a single flash in another area (e.g., the lower visual field). The single flash is often perceived as double (see Quicktime Movie Demo 1).

Previous studies have shown that the VIFI is robust for separations of up to 20° of visual angle (Wilson & Singer, 1981), occurs both in central and peripheral vision (Wilson, 1987), can be induced by oriented stimuli such as Gabors, persists across different contrast polarities of inducer and test, and survives dichoptic presentation (Chatterjee et al., 2011).

Although the VIFI and the SIFI share a similar visual phenomenology, it is not clear whether the two effects share common mechanisms. Previous attempts to explain the auditory-induced effect have mainly invoked multisensory processes (Kawabe, 2009; Miller & D'Esposito, 2005; Mishra, Martinez, & Hillyard, 2008; Shams et al., 2000; Shams, Kamitani, & Shimojo,

Citation: Apthorp, D., Alais, D., & Boenke, L. T. (2013). Saccade adaptation goes for the goal. *Journal of Vision, 13*(5):3, 1–15, http://www.journalofvision.org/content/13/5/3, doi:10.1167/13.5.3.

doi: 10.1167/13.5.3 Received August 29, 2012; published April 1, 2013 ISSN 1534-7362



Movie 1. An example of the visually induced double flash illusion (VIFI). When fixating in the center of the image, the observer may perceive a double flash both above and below fixation, when in fact there is only a single flash below fixation. In this example, the lower flash is synchronous with the first upper flash.

2002), such as the modality-appropriateness hypothesis, the discontinuity hypothesis (that the effect of the discontinuous modality is stronger than that of the continuous one; Shams et al., 2002), the directed-attention hypothesis, and the information reliability hypothesis (Andersen, Tiippana, & Sams, 2004). Recent accounts have tended to focus on Bayesian integration theories, (e.g., Andersen, Tiippana, & Sams, 2005; Shams, Ma, & Beierholm, 2005). Chatterjee et. al. (2011) suggest that the emergence of a purely visual-induced version may require a reworking of the theories behind the SIFI; however, it remains a possibility that the two seemingly similar effects may be at least partly unrelated.

Evidence for a common mechanism between VIFI and SIFI is somewhat equivocal. The two effects seem to share a similar time scale, occurring over a period of around 100 ms separation between the onsets of the two inducing stimuli, although the timing of the visual effects in particular has never been extensively explored. Wilson (1987) compared both auditory and visual versions of the illusions (to our knowledge, the only study so far to do so), and considered that they were modulated by the same mechanism, suggesting superior temporal sulcus (STS) as a possible site, since it is thought to be responsible for temporal perception and has large receptive fields (>20°). Interestingly, though, the authors framed the effect in terms of a masking phenomenon rather than an illusion, and the study is not widely discussed in studies of the SIFI. In seeking a physiological site for the effect, Wilson cites Phillips, Zeki, and Barlow (1984) who suggested that distinct, mutually inhibitory types of neuron in STS that could be responsive to "flickering" or "non-flickering" visual stimuli. However, this study offered a model framework rather than physiological data. The more recent work on perception of the SIFI centers on decisions about the number of flashes presented, rather than whether a stimulus is flickering; but, STS is still a strong candidate for both illusions, as it is also considered a multisensory area (Meredith, Nemitz, & Stein, 1987; Shahin, Bishop, & Miller, 2009; Watkins, Shams, Tanaka, Haynes, & Rees, 2006).

Other evidence suggests the VIFI and the SIFI differ in important ways. Chatterjee et al. (2011), for example, cite the fact that they did not find fusion for VIFI (two flashes seen as one), although this is a common finding in the SIFI (Andersen et al., 2004, 2005; Kawabe, 2009; Mishra et al., 2008; Shams et al., 2005) as evidence for different mechanisms. It is relevant here to note that some studies using apparently very similar stimuli do not report fusion (Innes-Brown & Crewther, 2009; Shams et al., 2000; Watkins et al., 2006). The reason for this inconsistency is not clear. The stimulus onset asynchrony (SOA) between the inducing beeps is typically about 50–60 ms in most recent studies; timing between visual double flashes is usually about the same, and to our knowledge the SOA between stimuli has never been systematically manipulated for VIFI. For the SIFI, Shams et al. reported that the effect declined from 70 ms separation onwards (Shams et al., 2002) and persisted when auditory and visual stimuli are separated by up to about 100 ms, consistent with the integration time of multisensory neurons (although superior colliculus neurons can also show integration periods up to 300 ms; Meredith et al., 1987). A similar tolerance of around 50–100 ms is reported for the visually induced effect (Wilson & Singer, 1981). In addition, the SIFI showed similar effects when the inducing sound was presented before or after the test flash. However, to our knowledge the only study to have compared the two effects with similar stimuli was that of Wilson (1987, a study seldom cited in discussions of the SIFI), although he did not examine the time courses of the auditory and visual effects. Indeed, the time course of the visual illusion with regard to the SOA of the inducing flashes has never been explored, and the timing of the two effects has never been compared in a single study.

In this study, we measured the time course of the VIFI and the SIFI by varying the SOA of the inducing flashes. First, we examined whether the visually induced effect persisted when the first inducing flash occurred before the target flash. Shams et al. (2002) found a symmetrical effect for the SIFI (see their figure 7) when the inducing beeps occurred either before or after the test flash, but this has not been investigated with the VIFI. Second, we directly compared the time course for auditory and visual inducers presented separately or together. Thus for the first time we used an identical paradigm to study the time course of both auditory- and visual-induced flash illusions, the relative strength of each cue, and whether they might be sub- or superadditive. The results help to elucidate the extent to which common mechanisms underlie both effects, and how they interact. We used an objective approach to measure the effect. Fusion trials (a single inducing flash synchronous with either the first or second of two test flashes) and fission trials (a single test flash synchronous with the first or second inducer flash) were interleaved with "catch" trials in which single or double flashes will be presented at both locations, and participants were informed that there were equal numbers of double- and single-flash trials, to attempt to minimize any biased response strategy. We also adjusted for any existing bias (to respond "one" or "two") by adjusting participants' responses according to their level of accuracy on a separate 'single-flash/double-flash' discrimination task with no inducing flashes. To foreshadow our results, for the VIFI, we found a symmetrical effect for inducing flashes presented before or after the test flash; the VIFI was present for both fission and fusion, but declined more rapidly for fusion across longer SOAs. Auditory-induced effects were stronger, did not differ between fission and fusion, and also declined across SOAs to around 110 ms. Audiovisual inducers produced a very similar effect to auditory-only inducers, showing the effect was not additive across modalities, and when the two were in conflict, observers tended to follow the auditory rather than the visual inducers. We examine the results in terms of a temporal uncertainty hypothesis.

Methods

Observers

The ten observers (six males) were two of the authors (LTB and DMA), two experienced but naïve observers, and six naïve undergraduate student participants who were compensated for their time. All had normal or corrected-to-normal vision. Ten participated in the V-induced conditions, eight in the AV-induced and eight in the A-induced conditions; the two fewer participants were from the undergraduate group. All gave informed written consent. The study was approved by the University of Sydney Ethics Board and conformed to the Declaration of Helsinki.

Apparatus and stimuli

Stimuli were programmed in Matlab version 7.9.0.529 (R2009b) using Psychtoolbox (Brainard, 1997; Pelli, 1997), and displayed on a Sony Trinitron Multiscan G500 CRT monitor with a refresh rate of 100 Hz. The monitor was linearized in software to correct for display gamma, and a Bits++ digital-to-analogue converter (Cambridge Research Systems, Rochester, Kent, UK) was used to enable finer control of contrast levels. Visual stimuli were light-colored Gaussian blobs with a standard deviation of 0.74°, presented for a single frame on a gray background, 5.67° above and below a dark fixation cross. (Note that although a frame is generally reported as 10 ms for a 100 Hz refresh rate, the actual duration of the flash is estimated to be much shorter, in the region of 2–3 ms, due to the phosphor decay rate of the CRT monitor; Elze, 2010). Thus we report SOAs rather than inter-stimulus intervals (ISIs), as the onset timing of each stimulus is known more precisely than the offset. The position of the flashes on the screen was jittered slightly on each trial to avoid luminance adaptation effects, and the luminance of each flash was also varied randomly over a range of Michelson contrast ranging from 0.3 to 0.7, to avoid participants guessing the number of flashes based on luminance alone. Minimum luminance was 0.26 cd/m^2 and maximum luminance was 67.3 cd/m^2, with mean luminance at 33.7 cd/m^2. Auditory stimuli were brief tones of 1568 Hz, presented at $\sim$ 75 dB (measured from head position) for 10 ms with raised cosine on/off ramps, and were presented from a single speaker placed on top of the monitor. Participants sat in a darkened room 57 cm from the display, supported by a Headspot headrest (University of Houston Center for Optometry; http://www.opt.uh.edu/uhcotech/Headspot/).

Procedure

Participants were instructed to fixate on a fixation cross in the center of the screen and report whether one or two flashes were seen in the lower visual field on each trial. They were informed that there would be an equal number of trials with one or two flashes in the lower visual field, and that the distractor flashes in the upper visual field or the auditory beeps were irrelevant to the

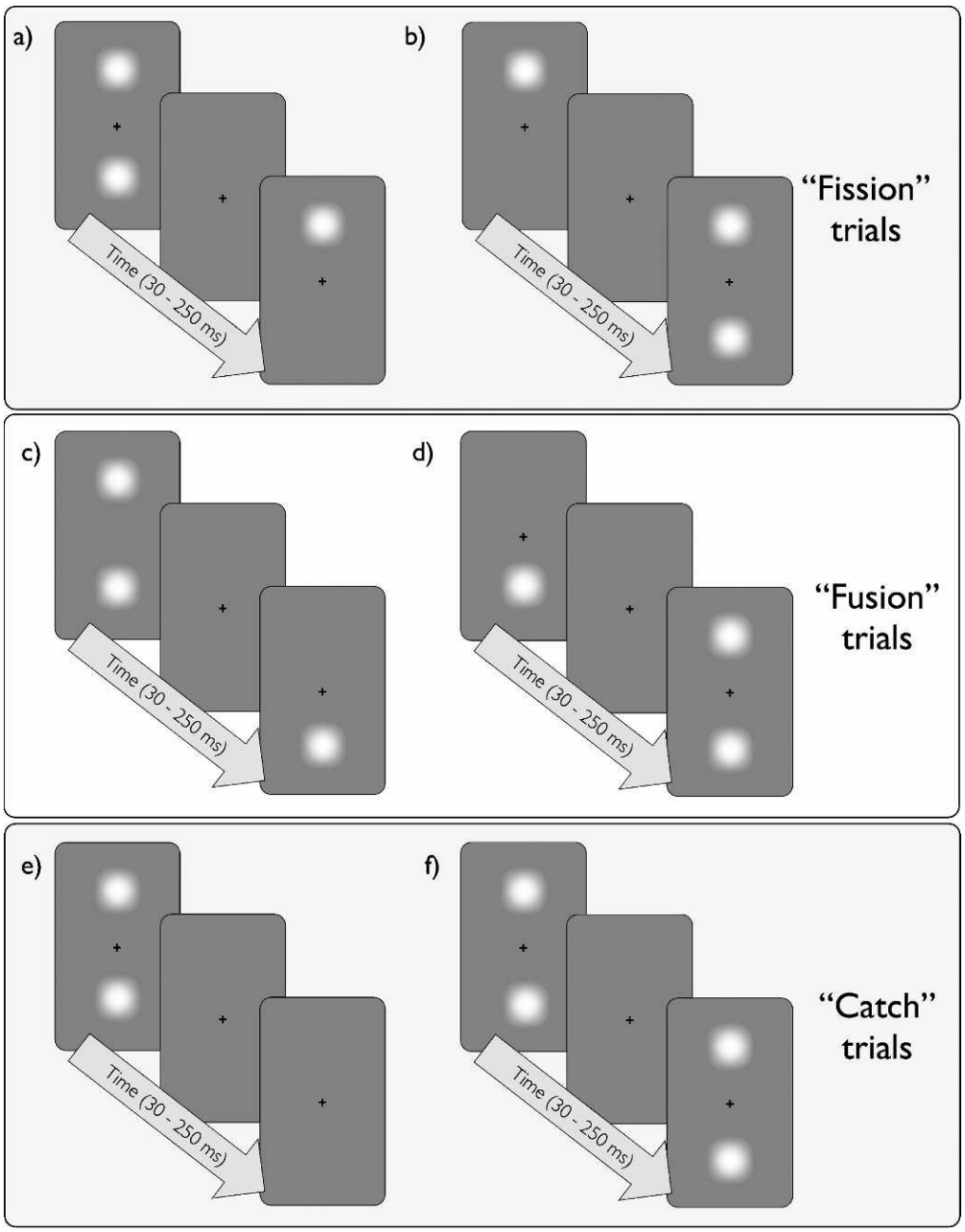


Figure 1. Conditions in the visually induced flash illusion (VIFI) experiment. (a) Fission with test synchronous with first inducing flash; (b) Fission with test synchronous with second inducing flash; (c) Fusion with inducer synchronous with first test flash; (d) Fusion with inducer synchronous with second test flash; (e) "Catch 1" trials; (f) "Catch 2" trials. Note that the lower flash/es were always the test stimuli, while the upper flash/es were always the inducing stimuli.

report and should be ignored. Thus, they would direct selective visual attention to the lower visual field where the target flashes would be presented. For the V-induced condition, there were six trial types (illustrated in Figure 1). There were two "fission" conditions, in which a single test flash was synchronous with either the first or the second inducer flash, and two "fusion" conditions, in which a single inducing flash was synchronous with either the first or the second of two test flashes. In addition there were two sets of "catch" trials in which the inducing and test flashes were both presented either once or twice, to control for expectation effects, lapses, and response bias. In addition, there were separate blocks of control trials in which only the test flashes were presented, without inducers. SOAs between the double flashes (both test and inducer) varied between 30 ms and 250 ms, in steps of 20 ms. Each experimental block contained randomly inter-

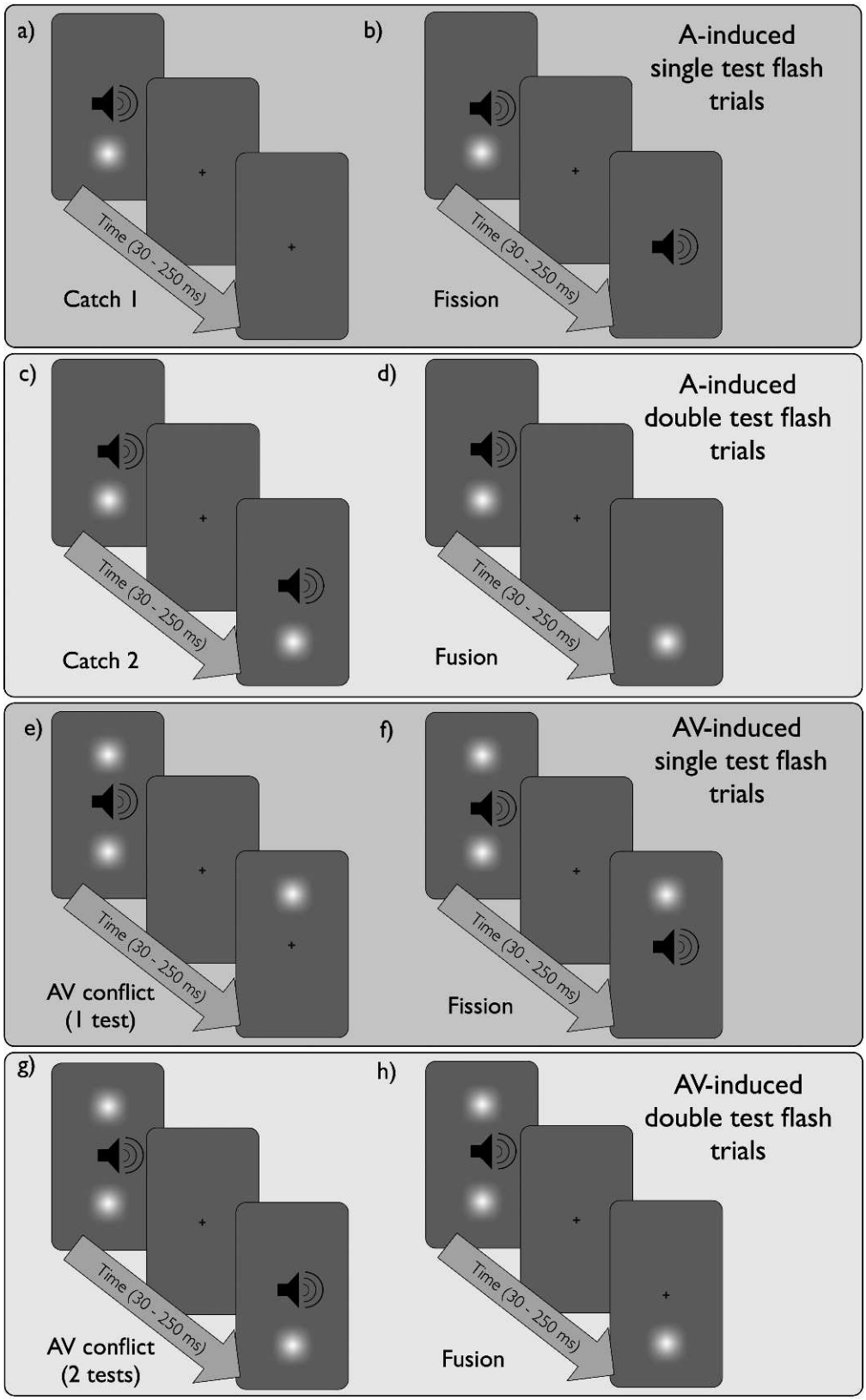


Figure 2. Conditions in the A-induced (a–d) and AV-induced (e–h) experiments. (a) "Catch 1" trials, analogous with 1e; (b) "fission" trials, analogous with 1a; (c) "Catch 2" trials, analogous with 1f; (d) "fusion" trials, analogous with 1c; (e) AV conflict: double visual inducer, single test flash and single sound, synchronous with the test. (f) AV-induced fission: Single test flash and two sounds, synchronous with visual inducers (upper). (g) AV conflict: single visual inducer, two sounds, synchronous with the test flashes (lower).

→

leaved trials of 10 of each trial type at each of the 12 SOAs. Each subject ran at least four blocks, giving at least 40 trials for each data point.

In the A-induced trials, the syntax of the experimental design was identical to the V-induced, with the difference that the (irrelevant) inducers were now auditory instead of visual. In these trials, inducing beeps were always synchronous with the first test flash, since preliminary data analysis (see also Shams et al., 2002) indicated that there was no difference between first- and second-flash conditions. Thus there were four trial types in the SIFI conditions: fission, fusion, and "catch" trials where there was either a single inducing beep synchronous with a single test flash or a double inducing beep synchronous with a double test flash (see Figure 2a through 2d).

In the AV-induced conditions, test flashes, as above described for the A-induced condition, were always synchronous with the first inducing flash/beep. In the fusion conditions, the single inducing flash was synchronous with the first test flash. In these conditions, the beep/s was/were synchronous with the *inducing* flashes in the upper field (Figure 2f and 2h); in the "AV conflict" conditions, the beep/s were synchronous with the *test* flash or flashes in the lower visual field (Figure 2e and 2g), while the inducing flashes remained as above.

Results

First, the results were adjusted for individual ability to veridically report the number of flashes without inducing flashes or sounds at each temporal separation; in other words, performance was measured relative to individual baselines collected in the control conditions. Then the adjusted proportion of veridical trials was converted into a measure of the proportion of trials on which an effect (either fission of fusion) was evident (that is, the size of the induced effect compared to nonveridical reports without inducer). The results for the visual-only inducer conditions are plotted in Figure 3. Conditions in which the test flash was synchronous with either the first or the second inducing flash were not significantly different. This is in line with the auditory-only inducer condition (see above and Shams et al., 2002). For the "fission" conditions, a two-way, repeated-measures ANOVA, with interval (first or second) and SOA (with 12 levels) as factors, revealed no significant difference between test flashes which were synchronous with first and second inducer flashes, $F(1,9) = 0.03$, $p = 0.878$, nor was there any interaction between SOA and interval, $F(11,99) = 2.11$, $p = 0.095$, using the Greenhouse-Geisser correction for departure from sphericity. These conditions were thus pooled across first and second intervals, and a one-way, repeated-measures ANOVA revealed a significant effect of SOA, $F(1,9) = 7.17$, $p = 0.007$, using Greenhouse-Geisser corrections. Comparing the mean at each SOA to 0 (100% correct, adjusting for performance in the control conditions), and controlling for multiple comparisons, there was a significant effect at SOAs up to 110 ms. This is, interestingly, quite comparable with the data for the SIFI provided in Shams et al. (2002).

Turning to the visual-only fusion conditions, it is clear from the results (see Figure 3a and 3c) that the effect of fusion occurs most strongly at the very shortest inter-stimulus interval of 20 ms, dropping very quickly to baseline at around 50–70 ms. Again, there was no significant difference between conditions where the single inducer was synchronous with the first or second test flash [no main effect of order, $F(1, 9) = .77$, $p = 0.402$, and no interaction between order and SOA, $F(11, 99) = 1.05$, $p = 0.408$], so these values were pooled. The overall one-way ANOVA on the adjusted values for fusion was significant, $F(1,9) = 10.44$, $p = 0.001$, corrected using the Greenhouse-Geisser method. Comparing the mean at each SOA to 0, again, and correcting for multiple comparisons, there was a significant effect only at SOAs of 30, $t(9) = 5.03$, $p = 0.012$, adjusted; 70, $t(9) = 5.04$, $p = 0.012$; and 110 ms, $t(9) = 4.09$, $p = 0.042$, but the differences from 0 at the longer two SOAs (70 and 110) were very small (0.03 and 0.025, respectively).

In a further analysis, we compared the results for fission and fusion conditions in a two-way, repeated-measures ANOVA, with condition (fission vs. fusion) and SOA (12 different SOAs) as factors. There was a main effect of SOA, $F(1,9) = 22.68$, $p = 0.001$, but no main effect of condition, $F(1,9) = 4.39$, $p = 0.07$; however, there was a significant interaction between condition and SOA, $F(11,99) = 3.11$, $p = 0.048$, corrected. In essence, these results show that the two effects showed a different time course, with the peak of the fusion effect occurring at very short SOAs of 30 ms, whereas the peak of the fission effect occurs at 50 ms and declines more slowly till around 110 ms. This would explain why fusion was not shown in the Chatterjee et al. (2011) study, which used SOAs of 67 ms.

However, it could be noteworthy that the difference in the time course between fission and fusion might be

←

(h) AV-induced fusion: single sound, synchronous with visual inducing flash (upper). Note: The location of speakers is only for reference, as the real location was always above the monitor (see text).

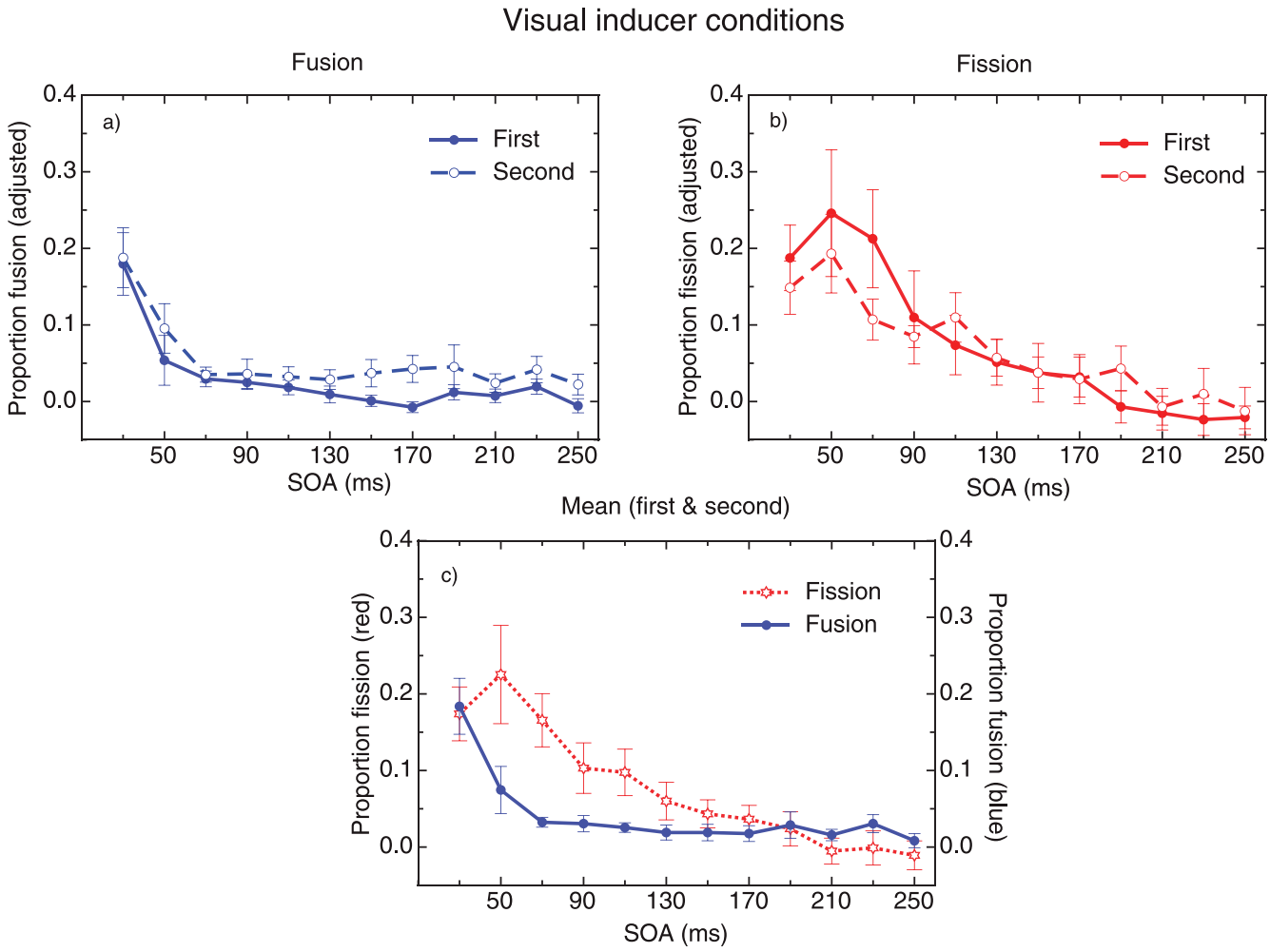


Figure 3. Results for 10 participants in the V-induced conditions, adjusted for performance using the control conditions (see beginning of Results section). (a) Fusion conditions, comparing the inducer synchronous with first or second test flash; (b) Fission conditions, comparing the test synchronous with first or second inducer flash; (c) Mean of first and second conditions for both fission and fusion. Error bars show ±1 standard error.

introduced by the baseline correction. One important difference in the baseline for the fusion condition (double-flashes) is that with smaller SOAs only one stimulus might be perceived regardless of the presence of the single inducing flash. Thus we examined the *uncorrected* data in an identical analysis to the analysis above. Interestingly, although the main effect of SOA is still significant, $F(1,9) = 20.26$, $p < 0.0001$, there is no main effect of condition, $F(1,9) = 1.16$, $p = 0.309$, and importantly, no interaction between condition and SOA, $F(11,99) = 1.97$, $p = 0.174$, Greenhouse-Geisser corrected. This suggests that the difference in time course between fission and fusion for the VIFI might be due to an inequality in the baseline conditions (poor resolution for double flashes at the shortest SOAs) rather than a fundamental difference between the two conditions.

In the A-induced (SIFI) conditions (Figure 4a), where inducers were auditory tones only, the fission and fusion conditions were not significantly different in their time course—there was no main effect of condition (fission vs. fusion), $F(1, 7) = 0.02$, $p = 0.899$, and no interaction between condition and SOA, $F(11,77) = 0.64$, $p = 0.788$. There was a main effect of SOA, $F(11,77) = 15.95$, $p < 0.001$, which showed significant linear ($p = 0.005$) and quadratic ($p = 0.001$) trends, indicating that the effect reduced over time and then flattened out, as expected. For fission, comparing the effects to 0 again, effects for up to 90 ms SOA were significant, whereas for fusion the effects were only significant up to 70 ms.

In the AV-induced conditions, where auditory tones were synchronous with the visual inducing flashes (AV-induced fission and fusion; Figure 2f and 2h; Figure 4b) the results were very similar—there was no main effect of condition, $F(1, 7) = 1.52$, $p = 0.257$, and no interaction between condition and SOA, $F(11,77) = 0.96$, $p = 0.491$, though there was a main effect of SOA,

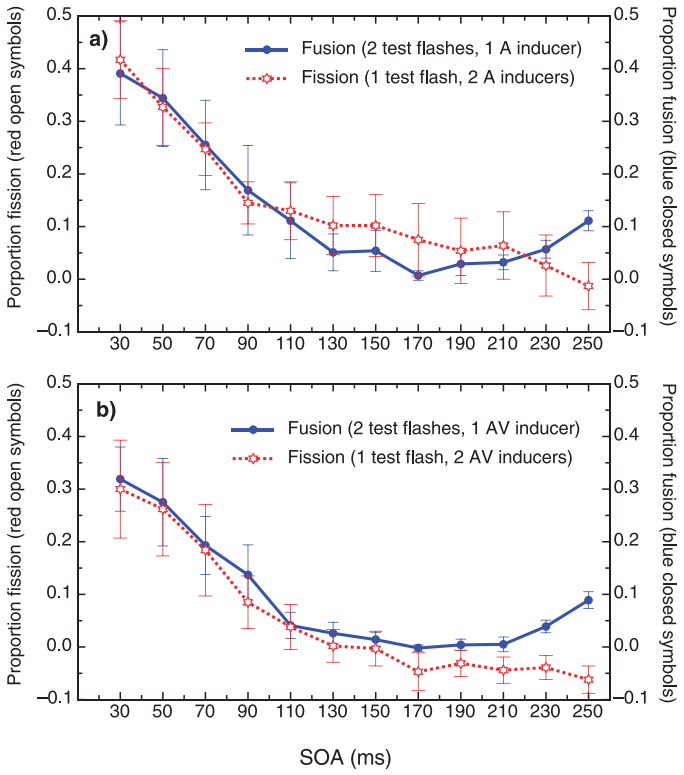


Figure 4. Auditory-only (SIFI) (a) and audiovisual (b) conditions for fusion (blue solid lines) and fission (red dashed lines) conditions, averaged over eight observers, and adjusted for control performance as above. Error bars show ±1 standard error.

$F(11, 77) = 12.63$, $p = 0.002$, Greenhouse-Geisser corrected. This main effect showed significant linear ($p = 0.006$) and quadratic ($p = 0.001$) trends, indicating that the effect again reduced over time and flattened out, as above. Comparing the effects against 0 as above, only the two shortest were significant for fission (up to 50 ms) and the three shortest for fusion (up to 70 ms). There were also significant effects for fusion at 230 ms and 250 ms, probably due to observers forming a response decision before the second test flash occurred. The similarity between A-induced and AV-induced effects suggests that once auditory inducers are present, additional visual inducers have little or no extra effect. These conditions are compared more directly below.

As outlined in the Introduction, we were also interested to see how the different inducer modalities compared to each other (visual inducers, auditory inducers, and audiovisual inducers). Thus, two-way, repeated-measures ANOVAs were used to compare these three conditions across all SOAs, for both fission and fusion conditions. These analyses were carried out only on the eight participants who completed all three conditions. For the fission conditions (Figure 5a), there was a main effect of modality, $F(2,14) = 3.89$, $p = 0.045$, and of SOA, $F(11,77) = 10.93$, $p = 0.005$ (corrected

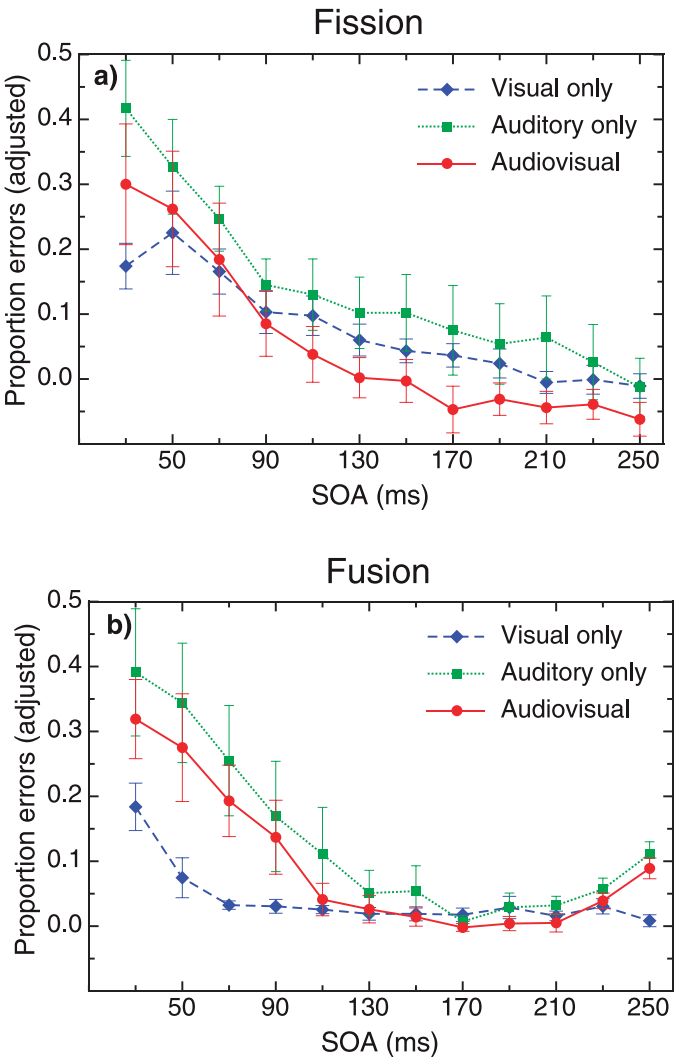


Figure 5. Results for the visual-only (VIFI), auditory-only inducer (SIFI), and audiovisual-induced conditions, adjusted for bias using the control conditions. Error bars show ±1 standard error. (a) Fission trials, with a single test flash and double inducer; (b) Fusion conditions, with a double test flash and single inducer (visual, auditory, or combined AV).

using Greenhouse-Geisser), and a significant interaction between modality and SOA, $F(22, 154) = 2.16$, $p = 0.004$. Essentially, the time course of the effect was different over the different modalities. A supplementary analysis of the audiovisual and auditory conditions alone showed a significant main effect of modality, $F(1,7) = 11.1$, $p = 0.013$, and of SOA, $F(11,77) = 10.89$, $p = 0.005$, corrected, but no interaction between modality and SOA, $p = 0.646$, corrected. Somewhat surprisingly, the purely A (auditory double) inducers produced more nonveridical reports than combined AV inducers across the entire time course, but the time course of the effects did not differ.

Turning to the fusion conditions (Figure 5b), there was no main effect of modality, $F(2,14) = 2.69$, $p = 0.11$,

but there was a significant main effect of SOA, $F(11,77) = 8.05$, $p = 0.002$, corrected, and a significant interaction between modality and SOA, $F(22,154) = 3.81$, $p = 0.018$, corrected. The interaction showed significant linear ($p = 0.042$) and quadratic ($p = 0.005$) trends. Again, a supplementary analysis of the auditory and audiovisual conditions showed a main effect of SOA, $F(11,77) = 9.54$, $p = 0.001$, corrected, but no main effect of modality ($p = 0.128$) and no interaction between modality and SOA ($p = 0.704$).

Also of interest is a comparison across modalities of all the conditions in which the inducer was congruent with the test flash (in the visual conditions, these were the "catch trials"; in the A-induced conditions they were labeled as "congruent" trials, and in the AV-induced as "AV conflict," since the auditory inducer conflicted with the visual inducer. Note that "catch" trials were not included in previous data analyses). This enables us to examine whether an additional flash or sound leads to more veridical performance compared to baseline performance. In other words, were participants more likely to correctly distinguish between single and double flashes with the extra information provided by inducing sounds or flashes, which are congruent with the test flash/es? First, looking at the A- and V-induced trials in which there was a single test flash (these were referred to as "fission" trials in conditions where there were double inducers), the results are averaged across all trials, as SOAs were not relevant where there was only a single test flash and a single co-occurring inducer. Results are plotted in Figure 6a. Adding a single inducing flash to the control (referred to as "catch 1" trials), interestingly, did not increase veridical reports over baseline ($p = 0.943$), i.e., observers were not more likely to veridically report a single flash in the lower visual field if it was accompanied by a single inducing flash. Adding sound produced slightly better performance than baseline, but this difference was not significant ($p = 0.196$).

More informative in addressing whether participants will follow auditory or visual information when both are present but in conflict, the AV conflict condition consisted of a single test flash, a single irrelevant beep (congruent with the test flash; see Figure 2a) and a double irrelevant flash. In this case, SOA is relevant. Results are plotted in Figure 6a. A one-way, repeated-measures ANOVA on these values showed no significant main effect of SOA, $F(11,77) = 2.109$, $p = 0.11$, corrected. Interestingly, though, all of the means are below 0, meaning participants were more likely to report (veridically) a single test flash in this condition than in the single-flash control condition, regardless of SOA.

In the case of double test flashes, Figure 6b plots the effect (compared to baseline) in the conditions where there were two inducing flashes or two

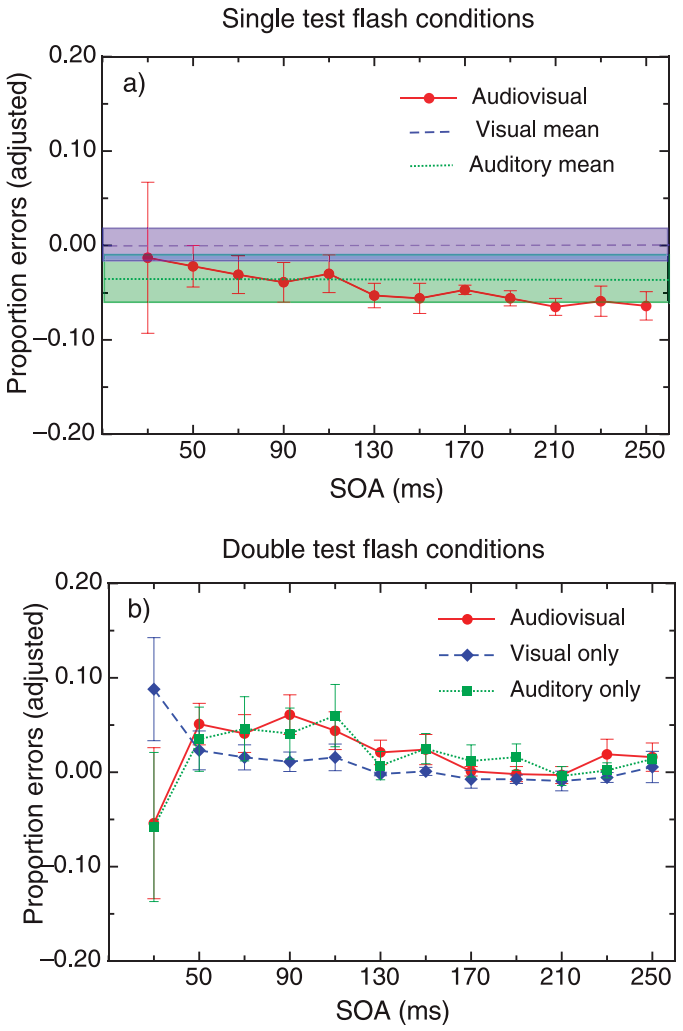


Figure 6. Comparison of all different modality conditions in which the inducers were congruent with the test flashes, for single (a) and double (b) test flash conditions. (a) Visual and auditory conditions ("catch 1;" see Figures 1e and 2a) consisted of a single test flash occurring simultaneously with single inducing flash or beep (shown by blue and green lines—shaded area shows ±1 standard error; SOA is irrelevant in these conditions, and thus means are plotted across all SOAs for illustrative purposes). AV conflict conditions (see Figure 2e), shown by green dotted lines, consisted of a single test flash, a single beep synchronous with the test flash, but two inducing flashes . (b) As above, visual- and auditory-only conditions consisted of two test flashes and two inducing flashes or beeps (Figures 1f and 2c). In the AV conflict conditions, there was a single inducting flash in the upper visual field, as in the VIFI "fusion" conditions (see Figure 1c), but two beeps synchronous with the test flashes (see Figure 2g); it is clear that the participants follow the auditory information in this case, and the "fusion" effect is abolished. Note that all conditions are corrected for baseline levels of veridical report.

inducing sounds congruent with the double test flash ("check 2 conditions"; Figures 1f and 2c); or in the AV-induced ("AV conflict") conditions, two inducing sounds and a single inducing flash; see Figure 2g. Only at the shortest SOA (30 ms) is there a difference between the conditions; here, adding two inducing sounds with the single inducing flash produces slightly better performance than baseline, while adding additional congruent flashes seems, surprisingly, to produce slightly worse performance than baseline. A two-way, repeated-measures ANOVA on the congruent data showed no significant main effect of modality or SOA ($p > 0.5$), but a significant interaction between modality and SOA, $F(2,22) = 3.266$, $p = 0.017$, corrected for nonsphericity. Post hoc comparisons at the shortest SOA showed A-induced and AV-induced conditions did not differ significantly, $t(7) = 0.168$, $p = 0.871$; there was a trend towards a significant difference between AV and V-only conditions, $t(7) = 2.197$, $p = 0.064$, but only the difference between A-induced and V-induced was significant, $t(7) = 2.6$, $p = 0.035$.

Taken all together, the results seem to show that combining both visual and auditory inducers does not produce an additive effect (Figure 5), but in fact produces an effect with a strength somewhere between the auditory and visual effects, suggesting that in this condition observers might have been following either the visual or the auditory cues, rather than combining them. For instance, do individuals whose visual temporal resolution is less precise have more of a tendency to follow the auditory stimuli in audiovisual trials? This is not possible to distinguish on the level of the grand average, and so the question was addressed with a correlation analysis. If this were the case, then there should be a negative correlation between proportion of nonveridical responses at 30 ms SOA in the visual double flash control condition, indexing the level of individual temporal resolution of double-flashes, and proportion of "two flashes" responses in the conflict trials in the AV conditions (where there was a single test flash in the lower visual field accompanied by a single beep, but two inducing flashes in the upper visual field; Figure 2e). Conversely, there might also be a similar correlation between nonveridical responses on the double flash control conditions and "two flashes" responses in the "fusion" conflict conditions (where there were two test flashes, two beeps, but only a single inducing flash in the upper visual field). In this case, the correlation would be positive, since "two flashes" would represent a veridical answer. The most useful method here is to examine correlations between the double flash control conditions and the *differences* between V-induced conditions and AV-conflict conditions for the same visual situation (see Figures 1a and 2e for fission, Figures 1c and 2g for fusion); in other words, observers might have been relatively more influenced by the visual cue if their visual temporal resolution were more precise or less influenced by the irrelevant auditory information (i.e., fewer nonveridical responses in the control condition), but this was not the case; $r(7) = -0.532$, $p = 0.174$ for fission; $r(7) = 0.546$, $p = 0.161$ for fusion. Thus it does not seem that, at the shortest SOA, observers with higher temporal resolution are more influenced by the visual cue in situations of cue conflict. Correlations at longer SOAs were considered problematic because of restriction of range effects (most values were very close to 0 across observers).

The difference between different inducer modalities in the "Fission" conditions (Figure 5a) also becomes most apparent at the shortest SOA (30 ms), where visual inducers seem to show a somewhat weaker effect whereas auditory and audiovisual inducers show similar effects. Thus we tested whether individual participants' responses at the shortest SOA (30 ms) for the AV-induced condition were more closely correlated with their responses on the A-induced or V-induced conditions. There was a significant correlation between AV-induced and A-induced conditions, $r(7) = 0.925$, $p = 0.001$, but no significant correlation between AV-induced and V-induced conditions, $r(7) = 0.433$, $p = 0.283$. These correlations were significantly different, $p = 0.014$ (two-tailed). The correlation between AV- and A-induced conditions was still significant at 50 ms, $r(7) = .760$, $p = 0.029$, but had disappeared by 70 ms, $r(7) = -0.615$, $p = 0.105$.

Control experiment: Disks compared to Gaussian blobs

In almost all previous experiments, both with the SIFI and VIFI, the test stimuli have been hard-edged disks rather than Gaussian blobs. Although our stimuli would have had similar temporal profiles to those used earlier, it could be the case that the broader range of spatial frequencies in hard-edged disks could render them more salient to the observer, possibly increasing the inducing effect of visual flashes. Thus we repeated the experiment with six observers (three new), using the same equipment and procedure as in the earlier trials, but interleaving trials with hard-edged disks and trials with Gaussian blob stimuli randomly, and only sampling SOAs up to 130 ms where we expected the strongest effects, based on previous results. Moreover, we included a shorter SOA since we were interested whether, due to the limits of temporal resolution, this would increase or decrease the illusion.

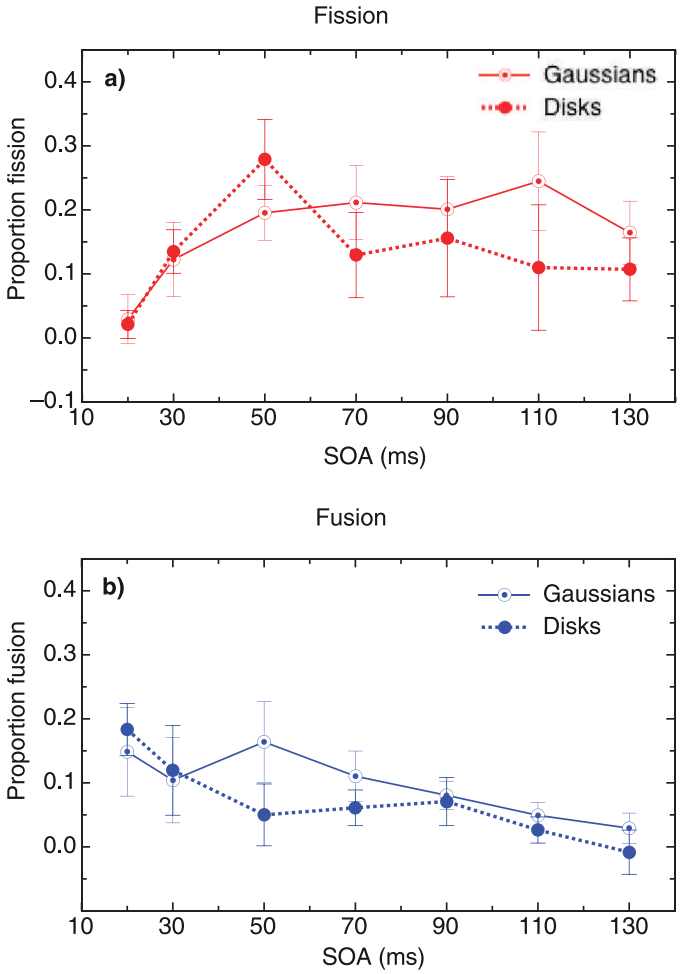


Figure 7. Results from the control experiment comparing the effect of hard-edged disks with Gaussian blob stimuli in the VIFI for (a) fission and (b) fusion. Solid lines show Gaussians and dotted lines show disk stimuli. The stimuli did not differ significantly, and, interestingly, the results from the original experiment were replicated, with fission and fusion showing different time courses. We also included a shorter (20 ms) SOA, which shows relatively stronger fusion and weaker fission.

Results

There was no significant difference between disk and Gaussian stimuli for either fission or fusion conditions (see Figure 7); main effect of stimulus was $F(1,5) = 0.44$, $p = 0.538$ for fission, $F(1,5) = 2.45$, $p = 0.178$ for fusion. Both showed a significant effect of SOA, $F(6, 30) = 2.73$, $p = 0.031$ for fission, and $F(6, 30) = 2.98$, $p = 0.021$ for fusion. In neither condition was there a significant interaction between SOA and stimulus type. Again, the time course for fission and fusion was significantly different, and showed the same pattern as in the main experiment—there was a significant interaction between condition (fission vs. fusion) and SOA, $F(6,30) = 3.54$, $p = 0.024$, corrected. Interestingly, at the shortest SOAs of 20 ms, not previously tested, fission decreased further and fusion increased (compared to 30 ms SOA).

Any difference between the stimulus types could possibly have been obscured by a difference in temporal resolution at baseline (without inducing flashes). As this would also be highly informative about the relative salience of the stimuli (hard-edged disks, if more salient, should be more easy to distinguish at baseline), we also performed a repeated-measures ANOVA on the baseline data for double flashes, comparing Gaussians and disks. There was no significant effect of stimulus, $F(1,5) = 1.29$, $p = 0.308$, and no interaction between stimulus and SOA, $F(6,30) = 1.63$, $p = 0.252$, corrected. Participants were also not significantly more or less likely to veridically report a single flash in the disk compared to Gaussian control conditions, $t(5) = 0.43$, $p = 0.688$.

Discussion

We set out to explore the time course of both the auditory-induced and visually induced flash and fusion illusions, and the influence of the modality of the inducers. In addition, for the first time, we directly compared auditory and visual effects for both fission and fusion, and the relative contribution of both cues to the combined effect. In general, auditory inducers produced stronger effects than visual inducers, and in a bimodal-inducer situation, when the inducers conflicted temporally, participants' responses tended to follow the auditory cue. Both fission and fusion illusions declined with temporal separation, up to around 100 ms for most conditions, though over a shorter time for purely visual fusion conditions.

Visual inducers

Firstly, in the visual-only condition, similar to the results found by Shams et al. (2002), there was no difference in the visual fission effect when the test flash was synchronous with the second, rather than the first inducing flash, suggesting that the inducers are effective within a certain time window rather than depending on a certain phase relationship with the test flash (pilot testing suggested that placing the test flash between the two inducers in time produced similar results). Similarly, for the fusion effect, there was no difference between conditions where the inducer was synchronous with the first or the second test flash. Again, this is in line with the notion that inducers in general need not be phase-locked with the test flashes. One explanation advanced by Chatterjee et al. (2011) is that there might

be recurrent processing from higher visual areas which modulate the effect. If only online low-level visual processes were involved, it would be expected that the second inducing flash would affect a test flash seen earlier, but not one seen at the same time as the second inducer. If this were the case, it seems unlikely that recurrent processing would be a sufficient explanation of the effect.

The second interesting result from the visual-only inducers is that we found a fusion effect at the shortest SOAs, declining to baseline at around 60 ms. This would explain why a fusion effect was not seen in Chatterjee et al.'s (2011) experiments, as their stimuli were only presented with a SOA of 67 ms. The fact that the time course of this effect is significantly different from that of the fission effect indicates that the results cannot be merely due to response bias, since if observers had resorted to a biased response strategy (e.g., always responding "one," or always "two"), this should not have differed markedly with SOA for the two different conditions, which were presented in equal numbers of trials, randomly interleaved. The issue of response bias is discussed further below.

Auditory inducers

Since the temporal resolution of audition is much greater than vision, it is not surprising that auditory inducers produced a much stronger effect than visual inducers at the shortest SOAs. It seems likely that, given uncertainty about the number of events at these short separations, auditory would be weighted more heavily (Alais & Burr, 2004; Ernst & Bülthoff, 2004). It is interesting that the induced fusion effect was as strong as fission in both the auditory and the AV conditions, in contrast to the visual-only effect and to several previous studies of the SIFI (Innes-Brown & Crewther, 2009; Shams et al., 2000; Watkins et al., 2006). The reasons for this are not clear; Innes-Brown et al. (2009) suggest that their lack of fusion effects might be due to either timing characteristics (their auditory stimuli were presented slightly before the flashes) or monitor technology (stimuli were presented on a CRT as opposed to LCD monitors used in some of the studies that found fusion, which may have led to more easily distinguished flashes). Since, like Innes-Brown et al., we used a CRT (at an even higher refresh rate—100 Hz), the timing offset explanation seems more likely. It could also be that our auditory stimuli had higher salience compared to our visual stimuli, since we used Gaussian blobs on a gray background rather than hard-edged white discs on a black background as in many previous experiments. The latter would open an interesting line of thought since it would open an interpretation that the relative salience of irrelevant inducer dimension and relevant target dimension influence the strength of the illusion (the illusion is greater for inducers that are more salient relative to the targets). Chatterjee et al. (2011) used both hard-edged disks and Gabor stimuli, and found similar results, although the effect for the Gabor stimuli was somewhat smaller. However, this does not speak to possible timing differences between detection rates for the different stimuli, which could occur with higher salience of disk stimuli; in our control experiment, we directly compared the two types of stimuli and did not find differences.

Audiovisual inducers

It is interesting that introducing visual inducers concurrent with the auditory inducers (Figure 2f and 2h; Figure 5, green dotted lines) did not produce an additive or superadditive effect, but rather seemed to result in effects somewhere between auditory-only and visual-only inducers, at least at the shorter SOAs. This seems to indicate that observers may have been following a single modality (either auditory or visual) in making a perceptual decision, rather than combining the modalities. This is discussed further below.

Also informative are the conditions in which visual and auditory stimuli were in conflict (Figure 2e and 2g; Figure 6, green dotted lines). In these cases, the visual stimuli were identical to those in the visual-inducer conditions (Figure 1a and 1c), but sounds were added that were congruent with the test flashes, either single or double. It is clear that in the conflict conditions, the observers tend to follow the auditory cue: That is, they are more likely to veridically report a single or double test flash if it is accompanied by a congruent (single or double) sound. In both single and double test flash conditions, the results very closely follow the auditory-only inducer (congruent) conditions (blue lines) across all SOAs, where there was a single or double test flash synchronous with single or double beeps (that is, "catch" trials). This suggests that the visual inducers could be weaker in salience than the auditory inducers, a conclusion which is also supported by the smaller number of nonveridical reports for the visual-only inducer conditions (Figure 5). As mentioned above, the spatial characteristics of the visual stimuli may have been a factor here, although Chatterjee et al. (2011) found the effect was robust across a number of spatial manipulations, including Gabor stimuli not dissimilar to our Gaussians; however, it is worth noting that their effects for Gabors were weaker than those for hard-edged stimuli (p. 7, Figure 5).

It is worth noting that adding synchronous visual flashes (the "catch 1" and "catch 2" conditions; Figure 6, red lines) also produces veridical reports that are not

different from baseline condition. The only exception is that veridical reports are slightly reduced for the shortest SOA in the visual double test flash condition (Figure 6b). This is probably due to the poor temporal resolution of vision making discrimination of double inducing flashes difficult at the shortest SOA, thus reducing their effect. This can also be seen in the reduced visual effect at 30 ms SOA in Figure 5a.

Fission compared to fusion

As outlined in the Introduction, one of our main interests was the nature of the fission and fusion effect. Are they related by common mechanisms or not? The results of the ANOVAs comparing fission and fusion showed that the time course of these effects was significantly different for visual inducers (i.e., there was a significant interaction between condition and SOA) but not for auditory inducers (SIFI) or combined (AV) inducers. Fusion is also weaker than fission for visual but not for auditory or AV inducer conditions. This pattern of results is consistent with the suggestion by Chatterjee et al. (2011) that the visually induced and auditory induced illusions are caused by at least partly different underlying mechanisms. However, the difference may also be caused by the differences in temporal acuity for visual and auditory inducers, in combination with differences in the baseline conditions; an analysis of the uncorrected fission and fusion results for the VIFI showed the time courses were no longer different. Thus a more parsimonious explanation might be that the two effects stem from a common mechanism.

It could be asked whether the occurrence of fission and fusion effects are largely due to high perceptual uncertainty, such as occurs at small SOAs (temporal uncertainty window), especially for visual targets. This might be also one explanation as to why the flash illusion has been observed for visual targets but not for auditory targets (although see Andersen et al., 2004). In this case, observers might develop a higher probability to follow the number of events in the irrelevant dimension, especially when the saliency is higher in the irrelevant dimension, to one of the two response categories (one flash or two flashes) because although the inducers were not attended, they may have processed by the system to a certain degree. If, for the visual inducer case, the location could not well enough established, this would lead to a random response. Thus, it might be likely that within the temporal uncertainty window, participants bind the wrong information (temporal misbinding compared to feature misbinding). Since the global brain state is not stationary over trials, this may result to a drop in performance across all trials. Part of the difference between fission and fusion might be explained by the relationship of total activation within this uncertainty window, at least within a single modality. For example, they might follow an internal (motor) preference to one of the two response keys. A systematic bias to report one of the responses (e.g., two flashes) would result in more "fission" reports than expected and less fusion reports (or vice versa for bias to respond to one flash). When plotted across SOA, this account predicts that fission and fusion would develop from larger to smaller SOAs in the opposite direction (one increasing, the other decreasing). The data show this is clearly not the case (Figure 5) and therefore a systematic response bias to one specific button press during an increase of uncertainty (based on smaller SOAs) is not a tenable account. Instead it may be that the concurrent information available through the irrelevant sensory channel becomes harder to ignore when the salience is higher compared to that of the target.

Another consequence of perceptual uncertainty at short SOAs could be that responses become more variable or even random. Random responding at the smallest SOA of 30 ms would predict the proportion of veridical responses to be at chance level, and responses under high uncertainty would tend to give a similar result (Boenke, Deliano, & Ohl, 2009). This pattern of results can be seen in the visual condition of Figure 5a and is also compatible with the integration model proposed by Shams et al. (2005). In this model, the number of inducers is integrated with the number of targets in a Bayesian fashion with the number of reported targets tending to follow the number of inducer events (see also Chatterjee et al., 2011 for a discussion on this point and the failure to find this for the visual inducer condition, p.13).

Visual compared to auditory and AV inducers

Comparing the fission and fusion effects between the three modalities shows they follow different time courses, as evidenced by the significant interactions in the ANOVAs comparing modalities. This interaction is attributable to vision differing from the other two conditions, since a separate ANOVA showed that auditory and AV time courses did not differ significantly for either fission or fusion. Looking at the data (Figure 5a), it is clear that the fission effect peaks later for vision than for the auditory and AV conditions, at around 50 ms SOA, which again is consistent with a poorer temporal resolution for vision than for audition; that is to say, the visual inducer is at the smallest SOA not individuated and has therefore no impact on the reported number of targets. This way of thinking would be consistent with the notion that inducers need to be present at "later" stages of

processing as two identifiable tokens, which, in turn, implies that if the explanation of perceptual uncertainty holds, this uncertainty is relatively late in the processing stream. This also suggests that conscious perception of two flashes may be necessary for the visually induced fission effect to occur.

Since the effect of the audiovisual inducers does not seem to be additive, it seems likely that individual observers might have had a tendency to follow either vision or audition. A given observer, for instance, might have a tendency to use the information available in the auditory modality whereas another might use the visual modality. These kinds of differences between observers have been observed in audiovisual temporal tasks such as synchrony or temporal order judgments (Boenke et al., 2009; Stone et al., 2001). It is also possible that a given observer might alternate between relying on vision or audition across trials, particularly if stimulus conditions are not reliable or do not favor the participant's preferred modality.

Conclusions

We investigated, for the first time, the differences and similarities between sound-induced and vision-induced fission and fusion illusions, the relative influence of sound and vision on visual event perception, and the time course of both types of illusion. Visually induced illusions showed a different time course to those induced by sound, and unlike previous researchers (Chatterjee et al., 2011; although see Wilson & Singer, 1981). We found that viewing both single and double ("incongruent') inducing flashes as distractors could affect the perception of the number of target flashes. These effects had different time courses, with "fusion" peaking earlier and declining sooner that "fission" for vision, which might explain why previous research has failed to find fusion for visually induced flash illusions. For the sound-induced effect—the traditional SIFI—we did not find a difference between the time course or strength of the illusions of fission and fusion. In general, sound-induced effects were stronger, but combining auditory and visual inducers did not enhance the effect. Adding conflicting auditory and visual cues resulted in effects that suggested that observers would, in general, follow the auditory temporal cues. In other words, the effect of visual inducers was negated by introducing sounds that were synchronous with the test flashes. Overall, the evidence suggests that the two effects may well stem from a similar mechanism that is involved in reducing perceptual uncertainty; the difference in time course is most likely to be due to differing temporal resolutions, and thus different levels of temporal certainty, for vision and audition.

Keywords: temporal vision, illusory flash, fission, fusion, cross-modal perception

Acknowledgments

The authors would like to thank Dr. Bhavin Sheth and an anonymous reviewer for their very helpful and constructive comments on the manuscript. This research was supported by the Australian Research Council's Discovery Project no. DP 120101474, awarded to David Alais.

Commercial relationships: none.
Corresponding author: Deborah Apthorp.
Email: deborah.apthorp@anu.edu.au.
Address: Research School of Psychology, Australian National University, ACT 0200, Australia.

References

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology, 14*(3), 257–262.

Andersen, T., Tiippana, K., & Sams, M. (2004). Factors influencing audiovisual fission and fusion illusions. *Cognitive Brain Research, 21*(3), 301–308.

Andersen, T., Tiippana, K., & Sams, M. (2005). Maximum likelihood integration of rapid flashes and beeps. *Neuroscience Letters, 380*(1–2), 155–160.

Boenke, L. T., Deliano, M., & Ohl, F. W. (2009). Stimulus duration influences perceived simultaneity in audiovisual temporal-order judgment. *Experimental Brain Research, 198,* 233–244.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*(4), 433–436.

Chatterjee, G., Wu, D.-A., & Sheth, B. R. (2011). Phantom flashes caused by interactions across visual space. *Journal of Vision, 11*(2):14, 1–17, http://www.journalofvision.org/content/11/2/14, doi:10.1167/11.2.14. [PubMed] [Article]

Elze, T. (2010). Misspecifications of stimulus presentation durations in experimental psychology: A systematic review of the psychophysics literature. *PLoS ONE, 5*(9), e12792.

Ernst, M. O., & Bülthoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences, 8*(4), 162–169.

Innes-Brown, H., & Crewther, D. (2009). The impact of spatial incongruence on an auditory-visual illusion. *PLoS ONE, 4*(7), e6450.

Kawabe, T. (2009). Audiovisual temporal capture underlies flash fusion. *Experimental Brain Research, 198*(2–3), 195–208.

Leonards, U., & Singer, W. (1997). Selective temporal interactions between processing streams with differential sensitivity for colour and luminance contrast. *Vision Research, 37*(9), 1129–1140.

Meredith, M. A., Nemitz, J. W., & Stein, B. E. (1987). Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. *The Journal of Neuroscience, 7*(10), 3215–3229.

Miller, L. M., & D'Esposito, M. (2005). Perceptual fusion and stimulus coincidence in the cross-modal integration of speech. *Journal of Neuroscience, 25*(25), 5884–5893.

Mishra, J., Martinez, A., & Hillyard, S. A. (2008). Cortical processes underlying sound-induced flash fusion. *Brain Research, 1242,* 102–115.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*(4), 437–442.

Phillips, C. G., Zeki, S., & Barlow, H. B. (1984). Localization of function in the cerebral cortex. Past, present and future. *Brain, 107*(Pt 1), 327–361.

Shahin, A. J., Bishop, C. W., & Miller, L. M. (2009). Neural mechanisms for illusory filling-in of degraded speech. *NeuroImage, 44*(3), 1133–1143.

Shams, L., Kamitani, Y., & Shimojo, S. (2000). Illusions: What you see is what you hear. *Nature, 408*(6814), 788.

Shams, L., Kamitani, Y., & Shimojo, S. (2002). Visual illusion induced by sound. *Cognitive Brain Research, 14*(1), 147–152.

Shams, L., Ma, W. J., & Beierholm, U. (2005). Sound-induced flash illusion as an optimal percept. *NeuroReport, 16*(17), 1923–1927.

Stone, J. V., Hunkin, N. M., Porrill, J., Wood, R., Keeler, V., Beanland, M., et al. (2001). When is now? Perception of simultaneity. *Proceedings of the Royal Society B: Biological Sciences, 268*(1462), 31–38.

Watkins, S., Shams, L., Tanaka, S., Haynes, J.-D., & Rees, G. (2006). Sound alters activity in human V1 in association with illusory visual perception. *NeuroImage, 31*(3), 1247–1256.

Wilson, J. T. (1987). Interaction of simultaneous visual events. *Perception, 16*(3), 375–383.

Wilson, J. T., & Singer, W. (1981). Simultaneous visual events show a long-range spatial interaction. *Perception and Psychophysics, 30*(2), 107–113.

Note: Data availability

Raw data for individual subjects, along with both uncorrected and corrected means, and a figure showing the uncorrected means for the main experiment, are available free from Figshare. The citation details are:

Flash illusions induced by visual, auditory, and audiovisual stimuli.

Deborah Apthorp. figshare.

http://dx.doi.org/10.6084/m9.figshare.155719