Flicker flutter: Is an illusory event as good as the real thing?

Tracey D. Berger
Psychology and Neural Science, New York University, New York, NY, USA

Marialuisa Martelli
Psychology and Neural Science, New York University, New York, NY, USA

Denis G. Pelli
Psychology and Neural Science, New York University, New York, NY, USA

Verghese and Stone (1995) showed that reducing the perceived number of objects by grouping also reduces objective performance. Shams, Kamitani, and Shimojo (2000) showed that a single flash accompanied by multiple beeps appears to flash more than once. We show that objective orientation-discrimination performance depends solely on the perceived number of flashes, independent of the actual number of beeps and flashes. Thus the unit of perceptual analysis seems to be a perceived event, independent of how it is induced.

Keywords: cross modal interaction, vision, audition, illusion, event perception, object recognition, auditory driving

Introduction

While visual science has learned something about how observers detect features, we still have almost no inkling of how vision combines them to make multi-feature judgments. There have been tantalizing hints that our visual system is predisposed to see the world as made of discrete objects, and that the contribution of a particular image component to our judgment can depend strongly on whether it is perceived as an object in its own right or as part of a larger object (e.g., see McDermott, Weiss, & Adelson, 2001). Verghese and Stone (1995, 1996, 1997) used grouping to change the number of perceived objects and found that objective speed discrimination performance was determined by the perceived number of objects. We wondered whether the current emphasis on objects, which are spatially discrete, might not be overlooking a similar role for events, which are temporally discrete.

"The world presents nothing but continuity and flux, yet we seem to perceive activity as consisting of discrete events that have some orderly relations (Zacks & Tversky, 2001)." Miller and Johnson-Laird (1976) suggest that events are dynamic objects. Quine (1985) proposes that we treat events and objects alike as bounded regions of space-time. In their review, Zacks and Tversky note that a Quine event is an important unit of analysis in philosophy, linguistics, and cognitive psychology. Verghese and Stone changed the perceived number of objects. If one could change the perceived number of events, would objective performance be determined by the perceived or actual number of events?

Shams, Kamitani, and Shimojo (2000) discovered the illusory flash effect, which is a compulsory integration of visual and auditory information. A spot is flashed once, accompanied by one to three beeps. When the flashed spot is accompanied by more than one beep, it appears to flash twice. The extra perceived flash is illusory. This is much like auditory driving, in which the apparent frequency of a flickering visual stimulus can be driven up or down by an accompanying auditory stimulus presented at a different rate (Gebhard & Mowbray, 1959; Shipley, 1964; Welch, Duttonhurt, & Warren, 1986).

Vision usually dominates over hearing (Posner, Nissen, & Klein, 1976), but in the illusory flash effect, hearing alters the visual experience. Vision dominates many spatial judgments, presumably because other senses provide less spatial information (Kitagawa & Ichihara, 2002). For any task for which multiple cues are available, the optimal combination weighs the various cues in accord with their signal-to-noise ratio for the given task (Ernst & Banks, 2002; Clarke & Yuille, 1990). When combining across senses, this is called the modality appropriateness hypothesis (Welch & Warren, 1980; also see Massaro, 1985; Meredith & Stein, 1983; Stein & Meredith, 1993; Frassinetti, Pavani, & Lødavas, 2002). It has been suggested that the auditory stimuli are powerful in the illusory-flash and auditory-driving effects because flicker is temporal. Since the auditory system is the most precise sense for temporal judgments, it is the most influential (Welch, Duttonhurt, & Warren, 1986).

Even though hearing has excellent temporal resolution, it is surprising that the observer gives hearing so much weight in an explicitly visual task. Observers are unable to ignore the irrelevant sound cue. They perceive a
visual flash rate that depends on a combination of the auditory and visual cues. Optimal cue combination is more successful in accounting for the familiar examples of sound affecting vision when the sound is relevant and disambiguates an ambiguous visual stimulus, helping the observer select among several valid interpretations. For example, in the Metzger (1934) illusion, a dot moves smoothly from left to right across the display. At the same time, another dot moves smoothly from right to left. The dots’ trajectories are ambiguous, consistent with either passing through or bouncing off one another. If a click is played at the moment that they meet, they are seen as bouncing off one another (Sekuler, Sekuler, & Lau, 1997).

The observer's inability to separate sight from sound may reflect a physiological convergence. Molholm, Ritter, Murray, Javitt, Schroeder, and Foxe (2002) report auditory effects in early responses in visual cortex, and suggest that this may reduce the behavioral reaction time to auditory and visual stimuli presented simultaneously.

In the illusory flash effect, sound dominates and causes the observer to perceive an extra event that isn’t there. Verghese and Stone (1995, 1996, 1997) showed that when the perceived number of objects is reduced by grouping, performance is determined by the perceived number. If events are like objects, might performance depend on perceived number of events, however induced?

We developed two tasks based on a modified version of the illusory-flash/auditory-driving paradigm. Both tasks use a grating patch instead of the original white spot. A subjective matching task replicates the classic auditory-driving experiments, measuring how many apparent flashes the illusion produces. However, the auditory-driving literature does not say whether objective performance is affected. Our objective orientation-discrimination task measures how performance is affected by number of illusory and actual flashes. The results tell us how the illusion affects objective performance and whether the effect is mediated by the number of perceived events.

### Methods

Observers were three undergraduate students (MSX, MAF, and TDB) and one postdoctoral researcher (MLM) at New York University. Their vision was normal or corrected to normal. The observers were initially shown the original Shams et al. (2000) illusion and asked to report how many times the disc flashed. Another observer reported not being able to see the illusion and was not included in the study. Observers TDB and MLM are authors.

The stimuli were created by a Power Macintosh computer using MATLAB with the Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997). The background luminance of the CRT used to display the stimuli was 10 cd/m². The display resolution was 1,024 x 768 pixels at 75 Hz, 28 pixels/cm. The viewing distance was 50 cm.

The visual stimuli were grating patches displayed at full contrast. The patches were 2.2 c/deg sinusoidal gratings vignetted by a Gaussian envelope. The Gaussian envelope was circularly symmetric, 1.14 deg wide (full width at half height). The patches appeared to be 2-3 periods wide. The observer fixated a small dot in the center of the screen. The center of the grating patch was 1.14 deg to the left of fixation.

The grating patch was normally presented in one-frame 13-ms flashes, at the specified flash rate throughout the 1- or 2 s interval. The other frames were blank. The first flash was centered in the first flash period (1/rate) at the beginning of the 1- or 2 s interval. We also measured performance with longer flash durations. The flash rate ranged from 1 to 5.8 Hz, but was usually 4 Hz. In most conditions, the visual stimulus was accompanied by several 30-ms 3-kHz auditory beeps, presented at the specified beep rate during the 1- or 2 s interval. The first beep was centered in the first beep period (1/rate) at the beginning of the 1- or 2 s interval. The beep rate ranged from 0 to 33.3 Hz. The extremes are special: 0 Hz is silence and at 33.3 Hz, the 30-ms beeps abut and merge to produce a continuous pure tone. The beeps and flashes were synchronous only when the beep rate was an odd integer multiple of the flash rate, or the flash rate was an odd integer multiple of the beep rate. The number of flashes or beeps in an interval is equal to the rate multiplied by the interval duration.

#### Experiment 1: Matching Flash Rate

Perceived rate of flashes is defined here as the rate of physical flashes in silence that the observer finds to be the best visual match to the condition under study. Observers were asked to make judgments about the flashing rate of a grating. Observers were told that we were studying an illusion whereby beep rate affects perceived flash rate. Each observer was tested at 20 different beep rates (0, 2, 2.5, 3, 3.5, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, and 33.3 Hz) in random order. Each beep rate was tested in a five-trial run. Results were averaged from three runs for each beep rate.

Each trial consisted of two intervals separated by a blank interlude. The first is the reference interval and the second is the match interval. In the reference interval, a grating flashed at 4 Hz for 1 or 2 seconds accompanied by a series of beeps. After a 500-ms silent blank interlude, the match interval began, displaying another flashing grating for 1 or 2 seconds, with no sound. Then the observer was provided a slider bar that controlled the flash rate of the second grating on future presentations. The task was to match the apparent flash rate of the two gratings. Moving the slider bar up/down caused an increase/decrease in the flash rate of the match grating in the next trial.
During the next trial, the observer reviewed the resulting goodness of the match, and made further adjustments if necessary. The observer had five trials per condition to fine-tune the adjustment. Most of our results are for 1-s intervals; one observer (TDB) ran all the conditions with 1- and 2-s intervals.

**Experiment 2: Discriminating Orientation**

The same stimuli were used as in the matching task, except that we slightly tilted the grating in one of the two intervals. The flash rate was 4 Hz. The beep rate was 0, 2, 2.5, 3, 3.3, 4, 6, 8, 10, 12, 16, or 33.3 Hz. Each run consisted of 40 trials. Each observer completed at least three runs of at least 6 of the 12 beep rates, in random order. Proportion correct was averaged across runs.

Each trial consisted of two intervals. In the first interval, a grating flashed at 4 Hz for 1 s, accompanied by beeps. After a 500-ms silent blank interlude, the second interval presented another flashing grating accompanied by the same number of beeps for 1 s. The grating in one of the two intervals was tilted 1° clockwise from vertical. The order of the two intervals was random. The observer’s task was to indicate which interval had the vertical grating, by clicking the computer mouse once for the first interval or twice for the second. After each trial, a correct response was rewarded by a computer-synthesized voice that said “right.” Observers ran 300 practice trials to learn the task before collecting the data reported here. One observer also ran the task with a flash rate of 1 Hz, with flash durations of 13, 52, and 1,000 ms.

**Table 1. Conditions for Experiment 3**

<table>
<thead>
<tr>
<th>Observer</th>
<th>Perceived flashes (Hz)</th>
<th>Vary beeps</th>
<th>Vary both ( R )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Vary beeps (Hz)</td>
<td>Beep (Hz)</td>
<td>Flash (Hz)</td>
</tr>
<tr>
<td>MSX</td>
<td>3</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>4</td>
<td>20</td>
</tr>
<tr>
<td>MAF</td>
<td>2.5</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>6.3</td>
<td>4</td>
<td>20</td>
</tr>
<tr>
<td>TDB</td>
<td>2</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>5.8</td>
<td>4</td>
<td>22</td>
</tr>
</tbody>
</table>

Each observer was tested in the orientation discrimination task with three different perceived flash rates achieved in two different ways: by varying the beep rate accompanying a 4-Hz flashing grating patch or by varying flash rate accompanied by an equal number of beeps.

**Experiment 3: Discriminating at Matched Rates**

We hypothesize that discrimination performance is determined by the perceived flash rate, independent of how it is achieved (by adding beeps to induce illusory flashes or by adding physical flashes). To directly compare the effects of adding beeps and adding flashes, we tested discrimination performance of three observers as a function of flash rate (with one beep per flash) and as a function of beep rate (with 4-Hz flash rate). We tested each observer at three rates for each manipulation (Table 1). All conditions in each row have the same perceived flash rate. We used the matching task to measure the perceived flash rate.

Each run consisted of eight interleaved trials of each condition. Each observer completed 25 runs (200 trials per condition). Pilot data showed that results without interleaving are similar, but show small shifts in level of performance from day to day.

**Results**

We begin by measuring the effect of beep rate on the perceived flash rate (Figure 1) and the proportion correct (Figure 2). Perceived rate of flashes is defined here as the rate of physical flashes in silence that the observer finds to be the best visual match to the condition under study. We never asked observers to report a number; they merely adjusted a rate to achieve a perceptual match.

If there were no illusion, the observers would ignore the sound and veridically match the 4-Hz flash rate of the reference grating, unaffected by beep rate. Figure 1 shows, as expected, that all four observers did produce a veridical match at beep rates of 0, 4, and 33.3 Hz. The zero beep rate at the left of each figure is silence. At the highest beep rate, 33.3 Hz, the 30-ms beeps abutted and became a 1-s pure tone. The two conditions, silence and pure tone, are similar in leaving the 1-s interval undivided. At 4 Hz, there is one beep per flash, and the match is veridical. With fewer or more beeps, observers perceived the flash rate to be as low as 2.3 Hz or as high as 6.1 Hz. The average across observers (solid black line) has its low of 2.7 Hz at 2-Hz beep rate, and its high of 5.5 Hz at 22-Hz beep rate. Perceived flash rate increases monotonically with beep rate over the range 2 to 22 Hz. Higher beep rates are less effective. Increasing the duration of both intervals from 1 to 2 s (doubling the number of physical flashes and beeps) did not change the perceived rate, so it doubles the number of perceived flashes. This effect of doubling the duration indicates that the illusion affects perception of rate (which didn’t change), independent of total number (which doubled). As noted in "Discussion," these results are consistent with the auditory-driving literature.
Figure 2. Orientation discrimination. Proportion correct for four observers is plotted as a function of the number of beeps accompanying the flashing grating. The vertical dashed line at 4 Hz is one beep per flash. The horizontal dashed line at 0.72 proportion correct represents average proportion correct in silence.

Figure 3. Comparing the effects of adding flashes and beeps. The vertical scale is proportion correct orientation discrimination of a 1° tilt for stimuli with various beep rates (accompanied by a 4-Hz flash rate). The horizontal scale is proportion correct for stimuli with various flash rates (accompanied by an equal beep rate). Each point represents one perceived flash rate induced in two different ways (Table 1). The error bars at the point representing 5.8 Hz perceived flash rate for TDB represent ±1 SE for that measurement. The error bars for this point are representative of the results for all observers.
perceived number of events, regardless of how that number is induced. This proves our claim, but it's interesting to probe a little deeper.

Figure 4 shows how proportion correct (right scale) grows with number of perceived flashes, comparing the effects of adding beeps and adding flashes. The data are the same as in Figure 3. For modeling purposes, below, we also transform proportion correct to \( d' \) and display it as a left scale. The left axis is \( d' \) on a log scale and the bottom axis is \( n \), the perceived number of events (in the 1-s interval), also on a log scale. To our surprise, the log-log slope varies across observers. Two observers have a slope of about 1/2 and the third has a slope of about 1. Even so, for each observer, the results for the variable flash (solid line) and 4-Hz flash conditions (dashed line) are virtually identical. Thus, for each observer, proportion correct (and \( d' \)) depends only on the number of perceived flashes, whether induced by adding beeps or flashes.

Is it really just the number of events that matters? Would increasing the duration of a single flash provide the same benefit? One observer (TDB) also ran the orientation-discrimination task with flashes of various durations (in silence) at a flash rate of 1 Hz. The proportions correct ±SE for flash durations of 13, 52, and 1,000 ms are not significantly different: 0.63 ± 0.04, 0.65 ± 0.03, and 0.64 ± 0.04, respectively. For 13-ms flashes, increasing the flash rate from 1 to 4 Hz did increase the proportion correct from 0.64 ± 0.04 to 0.74 ± 0.05. These results indicate that orientation discrimination depends on number, independent of duration. This is the temporal analog of the spatial finding by Verghese and Stone (1995) that speed discrimination depends on number of gratings, independent of area.

### Discussion

We used two tasks: matching and discrimination. Subjectively, our matching task results replicate the findings of the auditory-driving literature (Gehard & Mowbray, 1959; Shipley, 1964; Welch, Duttonhurt, & Warren, 1986). With many physical flashes, several extra perceived flashes can be induced. When beeps are less frequent than flashes, observers perceive fewer flashes than are shown. Furthermore, we found that perceived rate is independent of duration and total number of events.

Our objective results are all new. Orientation discrimination improves with more beeps than flashes, and worsens with fewer beeps than flashes. When there is one beep per flash, and no illusion, performance increases with number of flashes. There is no effect of flash duration.

Combined, the results from the two tasks show that objective performance is determined by perceived number of flashes, independent of how the perceived number is induced. Our objective finding parallels a physiological result: Shams, Kamitani, and Shimojo (2001) find similar visual evoked potentials for real and illusory flashes. Perceived, not actual, events are the coin of the realm.

Verghese and Stone (1995) find that speed discrimination improves with number of drifting grating patches, but does not improve with grating area. Based on their results, they suggest that increasing the number of objects improves speed discrimination because each perceived object yields an independent estimate of speed. They suppose that each estimate has statistically independent error and that discrimination is based on the average of the estimates (Verghese & Stone, 1996, 1997). This predicts that the just noticeable difference is proportional to \( 1/\sqrt{n} \), where \( n \) is the number of
perceived objects. They found support for this only up to 2 objects. For larger $n$ the measured improvement fell far short of the independence prediction.

We wondered whether the independence model might account for our results. The average of $n$ independent identically distributed random variables (one estimate of orientation per perceived event) has a signal-to-noise ratio proportional to $\sqrt{n}$. If the observer uses that quantity to make his or her decisions, then the observer’s $d'$ will be proportional, too, $d' \propto \sqrt{n}$ (Green & Swets, 1974). That predicts a log-log slope of 0.5. Two of our observers have nearly that slope, whereas the third has a slope of about 1. Thus, the independence model agrees with two of our three observers.

Verghese and Stone used Gestalt grouping to change the perceived number of objects. We used sound to change the perceived number of events. Orientation discrimination improves with the perceived number of events, no matter how they are induced.

### Conclusion

For better or worse, sound alters not only the visual experience but visual performance as well. Our ability to describe the world is determined by number of perceived events. This supports the extension of the independence model from objects to events. The full potency of the induced event is strong evidence that perceived events are a fundamental unit of perception.

### Acknowledgments

Thanks to Barbara Tversky for making us think about events in the first place, to Gregory Murphy for supervising the writing of Berger’s honor’s thesis, to Najib Majaj for being so hard to convince, to Robin Nixon for suggesting Figure 3, to Souheil Inati, Bhavin Sheth, and Ragnar Steingrimsson for their helpful questions about synchrony and duration, to Ioana Apetroaia Fineberg and Josh Fineberg for “tone,” to Michael Landy and John Foxe for pointers to the literatures on cue combination and physiology, and to Diana Balmori and Cesar Pelli for suggesting improvements to the figures and abstract. A special thanks to the two anonymous reviewers for suggesting we interleave conditions and vary flash duration. Some of these results were presented at the Vision Sciences Society meeting in Sarasota, FL, May 2001 (Berger & Pelli, 2001), and in Berger’s 2002 undergraduate psychology honor’s thesis at New York University. This project was supported by National Institutes of Health grant EYO4432 to Denis Pelli. Commercial relationships: none.

### References


