Seeing Sound: Changing Visual Perception Through Cross-Modal Interaction.

Tracey D. Berger

New York University

Department of Psychology

Faculty Sponsor : Denis Pelli
Psychology Department
New York University
6 Washington Place
New York, NY 10003

T.D. Berger (2002) *Seeing sound: Changing visual perception through cross-modal interaction*. New York University, Department of Psychology honors thesis.
http://psych.nyu.edu/pelli

Abstract

A single flash of light accompanied by a series of beeps gives the impression of flickering (Shams, Kamitani, & Shimojo, 2000). Thus, an unambiguous visual stimulus appears different when paired with an auditory stimulus. Similarly, with a flashing pattern, we find that beeps can increase or decrease perceived number of flashes. Furthermore, the perceived number or flashes can either improve or impair objectively measured ability to judge the pattern's orientation. Extra sound-induced flashes facilitate performance as much as an additional patch of pattern. Although much prior research has asserted visual dominance in cross-modal interaction, our results indicate that in certain tasks the auditory sense can alter visual perception and performance.

As we move through the world we are continually taking in information - what we see, feel, hear, etc. - and relating this new information to previous memories. Sometimes the information received by the various sensory modalities can be in conflict. For example, you may see a ventriloquist's dummy's lips moving, but hear the sound coming from the ventriloquist. However, according to your visual system, the sound seems to be coming from the dummy's moving mouth. According to your auditory system, the sound seems to be coming from the ventriloquist's mouth. Your perceptual system is faced with the challenge of incorporating these inconsistent signals into a coherent impression. Partly because the senses have mostly been studied independently of one another, the way that integration of information across modalities occurs in situations such as this one is not well understood.

"The illusory flash effect" is based on the integration of the visual and auditory senses (Shams, Kamitani, & Shimojo, 2000). A disc is flashed once and is accompanied by one to three beeps. When the flashed disc is accompanied by more than one beep, it appears to flash twice. The extra perceived flash is illusory (Shams et al., 2000).

We wondered whether the extra perceived flash is equivalent to an additional visual event. Does the extra "event" contribute to how well we see, as measured by objective performance? In order to address these questions, two tasks were developed for this study based on a modified version of the original illusion. A flashing gabor patch was used in the place of the original flashing white disc. The modified illusion is called *flicker flutter*.

Other cross-modal research. Other illusions demonstrate that auditory stimuli can affect the interpretation of an ambiguous visual situation. One example is an illusion in which two circles appear on opposite sides of a display. The circles move smoothly toward one another, meeting in the middle. At the midpoint, the circles' paths are unclear and they appear to either pass through one another or bounce off each other. (Metzger, 1934). The display is visually ambiguous in that either interpretation is supported by the visual stimuli. If a beep or click is played at the time that the circles appear to pass through one another, the circles are interpreted as bouncing off one another (Sekuler, Sekuler, & Lau, 1997). The sound disambiguates an ambiguous display. By comparison, the visual display in the Shams et al. (2000) illusion is unambiguous. Each presentation contains only one flash. However, when more than one beep is played, the observer sees two flashes. Instead of causing the observer to choose between two different, equally valid interpretations of the display, the sound actually causes the observer to see something that is not there.

Visual dominance. Many studies support the concept of visual dominance in cross-modal interaction. In other words, if a person receives conflicting information from two different sensory modalities, and one of the modalities is vision, then the visual input will dominate. A classic example is the McGurk Effect, discovered by McGurk and MacDonald (1976). Observers seeing a woman mouthing the syllable "va," and hearing an auditory recording of a woman pronouncing "ba," report hearing "va." The inconsistent auditory stimulus is overridden by the visual stimulus which induces a consistent auditory experience. Other trials with different syllables resulted in reported syllables that combined the visual and auditory input.

Other cross-modal research has demonstrated that visual adaptation can produce auditory aftereffects. One such experiment explored how adaptation to a moving visual stimulus could affect the perceived location of a subsequently presented auditory stimulus, and how the reverse situation affected the perceived location for a visual stimulus (Kitagawa & Ichihara, 2002). Kitigawa & Ichihara (2002) point to the fact that visual adaptation was more likely to affect auditory perception than vice versa as proof that the visual system is more precise and influential in terms of spatial perception. One explanation of this result is that the internal representations (spatial) that were best suited to decision making were visual, so the visual system dominated.

One of the reasons that the illusory flash effect is so intriguing is that it contradicts the usual finding of visual dominance in cross-modal interaction (Shams et al., 2000). Despite unambiguous visual information, the auditory stimulus changes the visual experience rather than vice versa. It is due to this unexpected result that further exploration of this illusion seems necessary. Through a deeper understanding of this illusion, perhaps we can develop a better grasp of how cross-modal integration works.

We used a subjective matching task to measure how many apparent flashes could be produced by the flicker flutter illusion. This task was used to identify how different conditions affected subjective perception of flashes. By knowing exactly how many illusory flashes were being perceived in different conditions, it was possible to explore how performance in an objective task was affected by different numbers of temporal objects (illusory or actual flashes) and different numbers of spatial objects (gabor patches). By comparing these effects we could answer the first question of whether extra perceived flashes are equivalent to additional events. We used an objective orientation discrimination task to address this question as well as the second question of whether this change in perceived events actually affects how well you see.

Matching Task

Method

Subjects. Participants were three undergraduate students and one post-doctoral student at New York University. All had normal vision or wore corrective lenses during their participation.

Apparatus. The stimuli were generated and presented on a Power Macintosh computer using MATLAB software. The background luminance of the computer monitor used to display the stimuli was 7.7 cd/m^2.

Stimuli. The visual stimuli were gabor patches displayed at full contrast. The patches were 2.2 c/deg. sinusoidal gratings presented in a gaussian envelope. The gaussian envelope was 1.14 degrees wide (full width at half height). The gaussian envelope made the grating appear as small circular patches, roughly 6 stripes wide, as shown in Figure 1. The center of the gabor was located at 1.14 degrees of visual angle from a fixation point. The viewing distance was 50 cm.

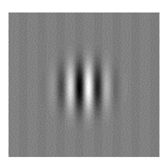

Figure 1: Gabor patch used as visual stimulus.

In most conditions, the visual stimuli were accompanied by auditory stimuli which were high-pitched beeps of 3 kilohertz intensity and a duration of 30 ms. Between conditions the frequency of beeps varied between the following values (in Hz): 2, 2.5, 3, 3.5, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 40, and 50

Procedure. For this task, participants were asked to make judgments about the flashing rate of two gabors. The participant sat 50 cm from the computer screen and fixated on a small dot in the center of the screen. Participants completed the twenty conditions of the task described above in random order. Each condition consisted of five trials. Each trial consisted of two intervals: the reference interval and the match interval. In the reference interval, a gabor flashed at 4 Hz for one second accompanied by beeps which began and ended at the same time as the flashing. After a break of 500 ms, the match interval began, displaying another flashing gabor for one second, with no sound. Then the participant was provided a slider bar that controlled the flash rate of the second gabor for future presentations. The task was to match the apparent flash rate of the two gabors. Moving the slider bar up/down caused an increase/decrease in the rate of flashing of the match gabor. In the next trial the gabor in the match interval flashed at the newly set rate. The observer had five trials per condition to fine-tune the adjustment. One observer also ran all conditions with the duration of both the reference and the match interval doubled.

Results

The results for the matching task are shown in Figure 2. Absence of illusory effect in this task would have resulted in an exact match (4 Hz) between flash rate of the reference and match gabors, with no effect of the beep rate. However, all of the observers slowed the flashing rate of the match gabor when there were fewer beeps than flashes in the reference interval and speeded up the flashing rate of the match gabor when there were more beeps than flashes. The cross- over from slower to faster occurred when the number of flashes and beeps were equal. The observers slowed the flashing rate of the match gabor down to as low as 2.6 Hz with fewer beeps than flashes and sped it up to as much as 6 Hz with more beeps than flashes. Increasing the duration of the intervals from one second to two seconds did not change these results. Beeps more frequent than the flash rate raised the perceived image frequency by up to two flashes per second and beeps less frequent than the flash rate decreased perceived image frequency by up to one flash per second. When there are four flashes in one second, the observer may perceive up to six. When there are eight flashes in two seconds, the observer may perceive up to twelve. As the beep rate approached 50 Hz, the observers began making veridical matches.

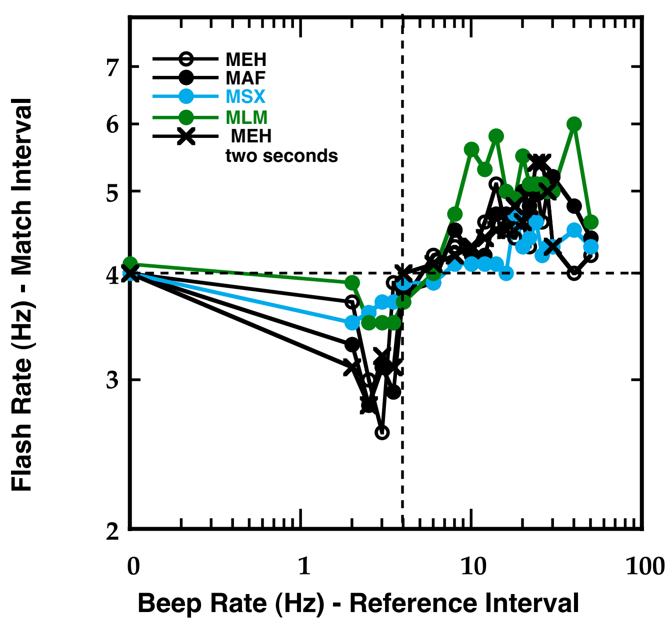

Figure 2: **Matching results for four observers. Includes two seconds duration results for one observer. The observer's final adjustment of flash rate in the match interval is plotted as a function of the beep rate in the reference interval. The dashed horizontal line at 4 Hz image frequency represents a veridical match**

between the reference and match gabors. The vertical dashed line represents equal number of flashes and beeps in the reference interval. With fewer beeps than flashes, subjects matched with fewer flashes; with more beeps than flashes, subjects matched with more flashes.

Discussion

This extends the Shams et al. (2000) result in two ways. They used only one flash and got only one extra perceived flash. We used many flashes and got several extra perceived flashes. Furthermore, when beeps were less frequent than flashes, observers perceived fewer flashes than were shown.

These findings lead to a different conclusion as to the strength of the illusion than that of Shams et al. (2000). Their experiment used a single flash, and the beeps doubled the number of perceived flashes, from one to two. Our matching task results indicate that the increase is limited to one to two extra flashes per second. The fact that increasing the duration of each interval to two seconds (thereby increasing the number of flashes and beeps in each interval) does not change this limit indicates that it is not the number of beeps and flashes, but rather the ratio between the two that is critical

In order to address the question of whether this change in perceived events actually affects performance in a perceptual task, observers were asked to perform an objective orientation discrimination of stimuli exhibiting the flicker flutter illusion.

Orientation Discrimination Task

Method

Subjects. The same observers who participated in the matching task experiment served in this task to allow for comparisons of results between tasks.

Apparatus. The same apparatus was used as in the matching task.

Stimuli. The same stimuli were used as in the matching task. In conditions involving the synchronous presentation of two gabors, each gabor was displaced at 1.14 degrees of eccentricity from a central fixation point along the horizontal meridian. Between conditions the frequency of the beeps were varied between the following values (in Hz): 0, 2, 3.3, 4, 8, 16, and 40.

Procedure. For this task, the participant was asked to determine which of two gabors was tilted. The participant completed each of the 7 conditions described above at least three times in random order and performance was averaged across runs. Each condition consisted of 40 trials. Each trial consisted of two intervals. The participant fixated on a small dot in the center of the screen. In the first interval a gabor flashed at 4 Hz for one second, accompanied by beeps. The beeps began and ended at the same time as the gabor began and stopped flashing. After a break of 500 ms, the second interval began. The participant was presented with another flashing gabor accompanied by the same number of beeps for one second. One of the two gabors was tilted one degree away from vertical. The task of the participant was to indicate whether the tilted gabor was in the first or second interval by clicking the computer mouse once for the first interval or twice for the second. The interval containing the tilted gabor was randomized. After each trial, the participant was told if they had answered correctly. One observer also ran all of the conditions with two gabors in each interval; see description of stimuli. In this case, the two gabors in the same interval flashed at the same time and had the same degree of tilt. Due to the difficulty of the discrimination required, all observers initially completed a number of practice runs in order to learn the task using gabors tilted 1.5 degrees from vertical.

Results

The results for the orientation discrimination task for conditions involving one gabor per interval are summarized in Figure 3. Performance was assessed according to proportion correct of a 40 trial run. Baseline performance, or performance with no sound varied across observers. With more beeps than flashes, all of the observers improved from their respective baselines. In conditions with the highest facilitation, the sound improved performance by up to 8%. This facilitation did not occur at the same point for all observers; for two the most facilitation was at 8 Hz beep frequency, and for the other two the most facilitation was at 16 Hz. Three of the observers showed a loss of facilitation at the next highest beep number from the condition with their best performance.

Performance suffered when there were fewer beeps than flashes. The sound in these conditions inhibited performance by up to 10%. Observers also did worse when the number of beeps was equivalent to the number of flashes.

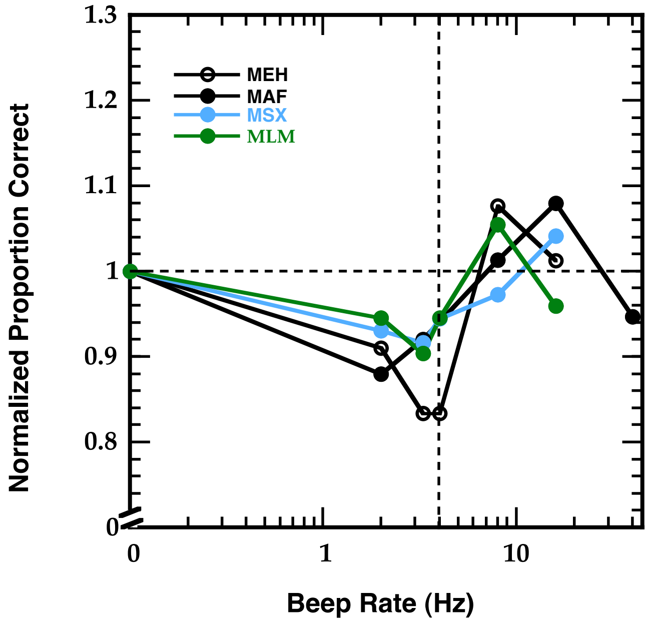

Figure 3: Orientation discrimination results for four observers. Observers' proportion correct is plotted as a function of the number of beeps accompanying the flashing gabors. Proportion correct was normalized by dividing performance in each condition by baseline performance. The horizontal dashed line at 1 indicates baseling performance. Points below this line denote worse performance than with no sound, points above it denote better performance. The vertical dashed line at 4 Hz indicates the same number of beeps and flashes per interval. Points to the left of this line denote conditions with fewer beeps than flashes, points to the right denote conditions with more beeps than flashes.

A comparison between the results of one observer's performance with one gabor per interval versus performance with two gabors is shown in Figure 4. The observer performed worse in conditions with fewer beeps than flashes, and improved in conditions with more beeps than flashes. The curve of performance with two gabors is shifted up. Performance in all conditions with two gabors per interval was better than the corresponding conditions with only one gabor per interval. Performance with 8 beeps, 4 flashes, and one gabor per interval is equivalent to baseline performance with two gabors per interval.

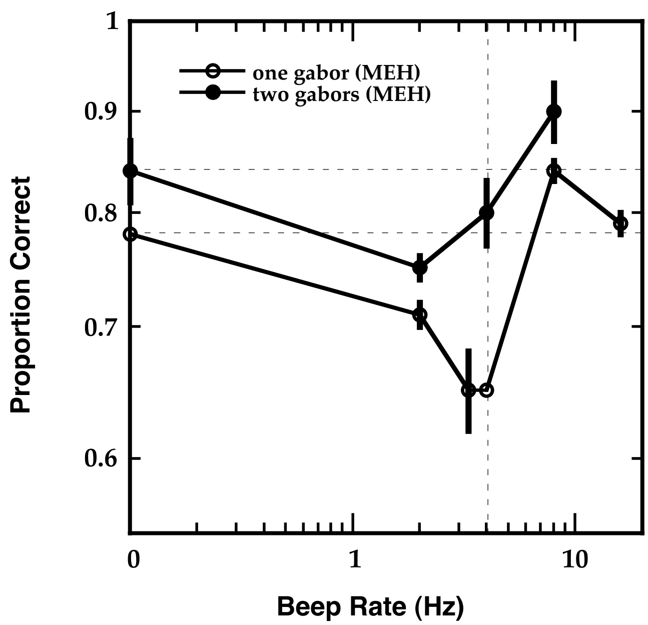

Figure 4: Orientation discrimination results for one observer with both one and two gabors in each interval. The horizontal dashed lines serve as guides to compare performance with no sound with other conditions. The vertical dashed line indicates the same number of flashes and beeps in each interval.

Discussion

The increase in proportion correct for all observers occurred in conditions with more beeps than flashes. Based on the matching task results, observers perceive extra flashes in these conditions. In conditions with fewer beeps than flashes, performance suffered. In these conditions, observers perceive fewer flashes than are actually shown. As the subjective perceived number of flashes increases/ decreases, objective performance improves/ declines. Therefore, sound can either facilitate or worsen objective performance.

The relationship between the amount of facilitation/ hindrance in the orientation discrimination task and the increase/ decrease in perceived number of events evidenced by the matching task is shown in Figure 5. If the facilitation were perfectly proportional to the perceived number increase, all the points would fall on a diagonal line with a slope of 1. The correlation has a slope of .54. Thus, these two measures are clearly related, although the correlation is not perfect.

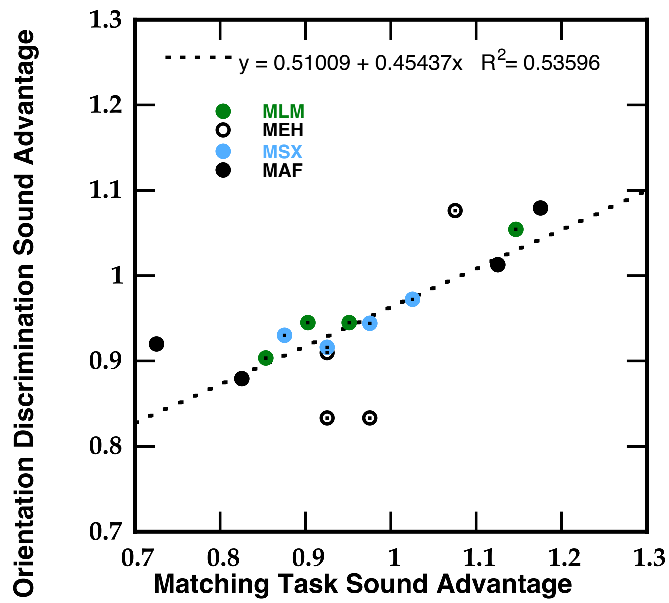

Figure 5: Correlation between number of perceived flashes in matching task and performance in the orientation discrimination task. The orientation discrimination sound advantage was figured by dividing each observer's performance in all conditions by that observer's baseline performance. The matching task sound advantage is figured by the same method using the matching task data. On the horizontal axis, points below 1 indicate performance in matching task conditions with fewer beeps than flashes (less perceived flashes); points above 1 indicate performance in conditions with more beeps than flashes (more perceived flashes). On the vertical axis, points below 1 indicate performance in orientation discrimination conditions with fewer beeps than flashes (worse), and points above 1 indicate performance in conditions containing more beeps than flashes (improved). The positive slant indicates that as the number of perceived flashes increased/ decreased (inferred from matching task results), performance in the objective task improved/ worsened.

As was shown in Figure 4, adding a second gabor patch per interval facilitated performance as much as adding sound to conditions including only one gabor per interval. Hence, extra perceptual flashes facilitated performance as much as additional objects did. Also, changing beep rate affected performance with two gabors per interval in the same way that it affected performance with one gabor per interval. This was shown by the fact that the curve of performance was shifted up for the two gabor conditions. Due to this shift in curve, we can conclude that the effects of extra temporal events and the effects of extra spatial events are independent.

General Discussion

The matching task results replicated the Shams et al. (2000) finding that sound can increase perceived number of flashes for an unambiguous stimulus. Our results further indicate that sound can also decrease the perceived number of flashes. The results from the orientation discrimination task indicate that changing the perceived number of flashes can either help or hinder performance in an objective perceptual task. These findings are consistent with the results of studies done by Preeti Verghese and Leland Stone (1995; 1996; 1997) that showed that the more discrete parts a visual display is perceived as having, the better an observer performs in an objective perceptual task. Verghese and Stone (1995) found that observers' ability to perform speed discrimination improved as they were presented with increasing numbers of discrete moving stimuli (1-6), but did not improve when observers were presented with one moving stimulus with an increasingly large surface area (1-6 times original size). More objects helped in motion discrimination in the Verghese and Stone study much as more perceived flashes helped in orientation discrimination in this study. In a later study, Verghese and Stone (1997) proposed the explanation that increasing the number of objects may help in speed discrimination because each discrete object provides its own speed information with statistically independent errors, and discrimination was based on a summation of all component parts.

Verghese and Stone's work provides a basis for the explanation that additional illusory flashes help in orientation discrimination because all flashes are perceived as discrete events. According to this explanation, each discrete flash can be expected to contribute individual orientation information, and discrimination is based on a summation of the information from each component event. As the number of perceived flashes increases, the observer gets more individual indicators of orientation to sum, and performance improves. Likewise, when sound decreases the number of perceived flashes, the observer has fewer individual indicators of orientation and performance suffers.

Another experiment conducted by Verghese and Stone (1996) explored how merely changing one's perception of a display could change one's ability to make discriminations about it. In this case, observers saw a display of moving gabors patches either as discrete items or, with depth, as pieces of one large, partially occluded patch. The display itself remained unchanged, only the observer's perception of how many components it had changed. Verghese and Stone (1996) found that the more components that a display was perceived as having, the better observers performed in discrimination tasks. Thus, increasing perceptual number of components improved performance just as increasing actual number of components improved performance. This is similar to the way that performance in the objective discrimination task in this study was affected by either changing the actual number of components through changing the number of gabors per interval or by changing the perceived number of components through the manipulation of beep rate. In fact, facilitation from sound (extra illusory flashes) and facilitation from an additional object (2 gabors in each interval) was equivalent (see Figure 4). Verghese and Stone (1996) have already shown that number of perceptual objects can facilitate the visual system just as actual objects can. The results of this study indicate that events, be they perceptual or actual, may be arrayed in different ways yet affect perception in similar ways. As evidenced by the comparison between performance with two gabors versus performance with one gabor with added perceptual flashes, the facilitation induced by perceptual events arrayed temporally (illusory flashes) was similar to the facilitation induced by actual events arrayed spatially (two gabors). Illusory events in time can affect perception as much as actual events or objects in space.

But why are there more perceptual events in the first place? Other research on cross-modal interaction has found visual dominance. However, the bulk of this research has been centered on spatial discriminations. Visual dominance may reflect the fact that vision is better suited to make spatial discriminations than other senses (Kitagawa & Ichihara, 2002). According to Massaro's Fuzzy Logical Model of Perception (FLMP), after incoming sensory information is registered and combined, the resulting aggregate is compared to existing representations in the mind (Massaro, 1985). Thus, since the visual system's representations are a better fit for spatial information, visual stimuli tend to have more of an effect on overall perception than auditory stimuli.

One explanation for why the auditory stimulus in the flicker flutter illusion has the effect it does on overall perception may hinge on the nature of the illusion itself. What affects performance in both tasks is the perceived number of flashes, which is a temporal distinction. Perhaps as the aggregate of flashes and beeps is being compared to internal representations, the auditory representations are more influential due to the fact that the auditory system is better suited to make temporal distinctions than the visual system. Hence, visual dominance may just be visual spatial dominance whereas other senses may dominate in other paradigms and tasks such as audition does for temporal distinctions. Further research may explore this issue and identify those areas in which other senses dominate in cross-modal interaction, thus providing us with a wider understanding of how this integration occurs.

Some aspects of the findings are not accounted for in the above explanation of how sound can affect perception of visual stimuli. In the orientation discrimination task, all of the observers performed worse when four flashes were accompanied by four beeps than when accompanied by no sound. Future examination of this illusion could explore why performance suffered in this condition, even though the perceived number of events was not affected. Perhaps when the number of events remains constant, the distraction of information from two different modalities hinders performance.

Another issue addressed by this study is how an event affects perception. In the discussion of these findings, “event” has been taken to mean an object in space and time. We can increase events by adding an additional spatial object (gabor) or, an additional temporal presentation (flash) of an object. According to the comparison of results in the orientation discrimination task in Figure 4, both of these dimensions can affect visual perception. Furthermore, these mechanisms appear to be independent of one another. Performance can be facilitated by the addition of another spatial event (Verghese & Stone, 1995), and further facilitated by the addition of another temporal event. Temporal and spatial aspects have been separated here, even though in real life, events unfold as a combination of spatial and temporal aspects (Michotte, 1963). In other words, if a ball moving through space is defined as an event, then the difference between where it started and ended up is the spatial aspect, and the time the movement took is the temporal aspect. In this experiment, we have shown that both of these aspects of an event can independently affect perception.

The discovery of “the illusory flash effect” revealed that more information was needed about how exactly cross-modal interactions work, and that visual dominance theories were insufficient (Shams et al., 2000). As more work is done to investigate how we use cross-modal interactions in making decisions about the information we are presented with, we can hope to gain a deeper understanding of how perception works in the real world. Concentration on separate modalities may give us valuable insights, but life is inherently a multimodal experience and therefore, full comprehension of perception must be multimodal as well.

Acknowledgements

This work was conducted through a collaboration between myself, Denis Pelli, and Marialusia Martelli. We are currently preparing a manuscript for publication.

References

Kitagawa, N., & Ichihara, S. (2002). Hearing visual motion in depth. *Nature, 416*, 172-174.

Massaro, D. W. (1985). Attention and perception: An information-integration perspective. *Acta-Psychologica, 60*, 211-243.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*, 746-748.

Metzger, W. (1934). Beobachtungen ueber phaenomenale Identitaet. [Observations on phenomenal identity]. *Psychologische Forschung, 19*, 1-60.

Michotte, A. (1963). *The perception of causality*. New York: Basic Books.

Sekuler, R., Sekuler, A. B., & Lau, R. (1997). Sound alters visual motion perception. *Nature, 385*, 308.

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

Verghese, P., & Stone, L.S. (1995). Combining Speed Information across Space. *Vision Research, 35*, 2811-2823.

Verghese, P., & Stone, L.S. (1996). Perceived visual speed constrained by image segmentation. *Nature 381*, 161-163.

Verghese, P., & Stone, L.S. (1997). Spatial layout affects speed discrimination. *Vision Research, 37*, 397-406.