

Audiovisual bounce-inducing effect: When sound congruence affects grouping in vision

MASSIMO GRASSI AND CLARA CASCO
University of Padua, Padua, Italy

Two disks moving from opposite points in space, overlapping, and stopping at one another's starting point can be seen as either bouncing off one another or streaming through one another. With silent displays, observers report streaming, whereas, if a sound is played when the disks are in the overlap region, observers report bouncing. The change in perception is thought to be modulated by a lack of attention that inhibits the integration of the motion signal when disks overlap and by the sound that increases the congruence of the display, in comparison with a real elastic bounce. Here, we accompanied the disks' motion with either a bounce-congruent sound (a billiard ball) or with bounce-incongruent sounds (a water drop, a firework). When the sound was switched on 200 msec before the disks' overlap, (1) all the audiovisual displays induced more bounce responses than did the silent display, but (2) the bounce-congruent sound induced more bounce responses than did the bounce-incongruent sounds. However, when the sound was switched on at the disks' overlap, only the first result was observed. These results highlight both the role of attention and that of sound congruence.

One of the figures depicted most often in psychology textbooks illustrates two lines crossing each other (e.g., Gray, 2002). This figure is used to explain a perceptual grouping principle, the Gestalt principle of continuity (Wertheimer, 1923): Although the four segments composing the cross can be grouped in several ways, what we see is two lines crossing each other. Grouping by continuity applies not only to static figures, but also to objects in motion, provided that the two objects are "passing simultaneously over the same point" (Koffka, 1935, p. 301). This is, for example, the case with Metzger's motion display (Metzger, 1934), which shows two identical objects (e.g., two disks) that move along the azimuth with uniform rectilinear motion and opposite directions: The disks start their motion, overlap, and stop at one another's starting point (see Figure 1).

This motion display was used by Metzger (1934) to show that the continuity principle applies to objects' motion. Vision here faces an *inverse optics* problem (Marr, 1982), because the objects' 2-D motion pattern is equally representative of two very different events in the real, 3-D world. In both events, the observer is looking at two objects placed at different depths so that the retinal images of both have identical sizes. In one event, the objects start their motion, overlap (i.e., one object occludes the other), then stream past one another (trajectories A–B and C–D, respectively, in Figure 1). In the other possible event, on the contrary, after the occlusion, the objects reverse their motion and return to their original starting position (trajectories A–D and C–B, respectively, in Figure 1). In summary, the disks in Metzger's display could be perceived

as either bouncing off or streaming through each other. But the bistability of the display remains potential when it comes to the observers' responses. Observers, in fact, group by continuity of motion and report perceiving the streaming percept much more often than the bouncing percept (fewer than 0%–20% of bounce responses; Bertenthal, Banton, & Bradbury, 1993; Grassi & Casco, 2009; Kawabe & Miura, 2006; Kawachi & Gyoba, 2006; Remijn & Ito, 2007; A. B. Sekuler & Sekuler, 1999; Watanabe & Shimojo, 1998). In the recent past, however, R. Sekuler, Sekuler, and Lau (1997) showed that grouping by continuity can be extinguished cross-modally: It is sufficient to play a brief sound when the disks overlap to increase the number of bounce responses from 10%–20% to 80%–90% (Grassi & Casco, 2009; Grove & Sakurai, 2009; Kawabe & Miura, 2006; Kawachi & Gyoba, 2006; Remijn, Ito, & Nakajima, 2004; R. Sekuler et al., 1997; Watanabe & Shimojo, 2001a, 2001b; Zhou, Wong, & Sekuler, 2007). Here, we refer to this change in the observer response as the *audiovisual bounce-inducing effect* (ABE).

The literature suggests that the origin of the ABE is double. The first component at the basis of the effect is thought to be attentional. When observers look at the silent display, grouping by continuity of motion occurs thanks to attention. Attention integrates the disks' local motion signals when the disks overlap, thus favoring the perception of streaming (Kawabe & Miura, 2006; Watanabe & Shimojo, 1998, 2005). In audiovisual displays, the sound is presented when the integration process is occurring—that is, when the disks overlap. Therefore, it subtracts part of the attentional resources that are neces-

M. Grassi, massimo.grassi@unipd.it

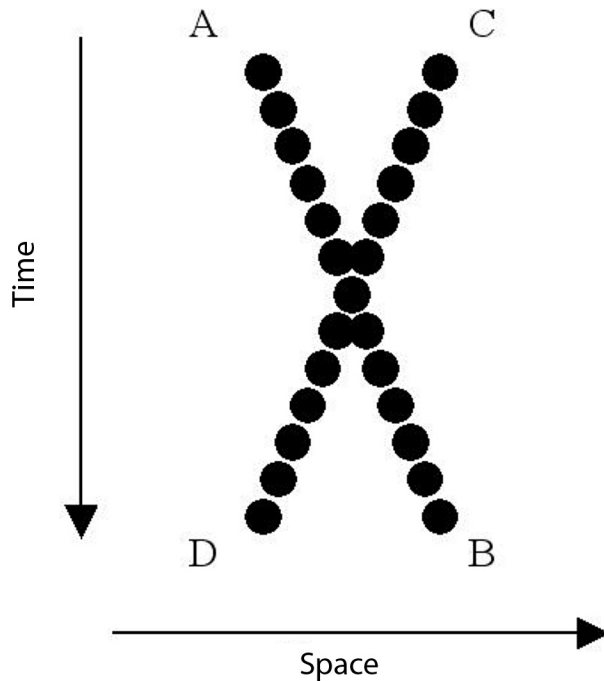


Figure 1. Two-dimensional representation of Metzger's motion display. The objects' motion is grouped as A–B and C–D with silent displays and as A–C and C–D if a sound is played when the disks overlap. In our experiments, the disks' actual motion was horizontal.

sary for the execution of the integration process.¹ Consistent with the attentional hypothesis, bounce responses can also be induced with brief tactile or visual stimulations, as long as they are delivered simultaneously with the disks' overlap (Kawabe & Miura, 2006; Watanabe & Shimojo, 1998, 2005). A perfect temporal coincidence is, however, not necessary for perceiving the ABE. In fact, the bounce percept is predominant (although less compelling) even when sounds are switched on before or after the disks' overlap (e.g., Remijn et al., 2004; R. Sekuler et al., 1997). The width of the temporal window of the bounce percept is about 200 msec (Watanabe & Shimojo, 2005), and the ABE is extinguished if the sound is presented out of this window (e.g., 300 msec before/after; Watanabe & Shimojo, 2001b).

Recently, however, Grassi and Casco (2009) showed that the attentional component cannot account for the whole effect and that (the lack of) attention is necessary (but not sufficient) for a compelling ABE. The authors found that, in order to observe an enhanced ABE, the sound also needs to be congruent, in comparison with that of a real, elastic bounce. They argued that the intensity profile (i.e., the envelope) of real bounce sounds (i.e., impact sounds; Gaver, 1993b; see below) is invariably characterized by an abrupt amplitude attack followed by a gradual amplitude decay. In the experiment, Grassi and Casco accompanied Metzger's display with sounds whose envelope was either bounce congruent (i.e., abrupt amplitude attack followed by a gradual amplitude decay) or bounce incongruent

(a realistic impact sound reversed in time) and found that the former induced the ABE much more than did the latter. In one experiment in particular, this was observed even though the bounce-incongruent sound was 20 dB higher in level than the bounce-congruent sound and the bounce-incongruent sound was judged by all the participants to be much more salient than the bounce-congruent one (Grassi & Casco, 2009, Experiment 2). In the present study, we further investigated the role of this nonattentional component of the ABE.

The soundscape we live in is full of a variety of sounds, and many of these do not result from the contact between solid objects such as in elastic impacts (i.e., bounces). Humans categorize environmental sounds according to the physical event at the origin of the sound, by listening to differences in timbre between the sounds (Gaver, 1993a, 1993b). Gaver's (1993b) perceptual taxonomy of environmental sounds divides sounds into three categories: those resulting from the interaction of vibrating solid objects (e.g., impact sounds), those resulting from the interaction of liquids (e.g., dripping), and aerodynamic sounds (e.g., explosions). Although our capability to identify a given sound source event varies from sound to sound (Ballas, 1993; Gygi, Kidd, & Watson, 2004), listeners do not seem to confuse sounds across these three categories (Gaver, 1993b). Interestingly, the envelope of isolated impact, liquid, and aerodynamic sounds is similar, and it is invariably characterized by an abrupt amplitude attack, followed by a gradual amplitude decay (Gaver, 1993b). In the present study, we accompanied Metzger's motion display with an impact sound (a billiard ball striking a second billiard ball), a liquid sound (a water drop falling into water), or an aerodynamic sound (the explosion of a firework). Although the three sounds have similar envelopes (i.e., abrupt amplitude attack, gradual amplitude decay; see Figure 2), only the billiard sound is congruent with the bouncing event of solid objects and should, therefore, promote (more than the other sounds) the perception of bouncing.

The present study was divided into two parts. The first part consisted of the main experiment, subdivided into Experiments 1A and 1B and followed by a brief sound recognition task and by two sound categorization tasks. In Experiment 1A, participants reported their perception (i.e., streaming or bouncing) by looking at the Metzger display accompanied by the sound of the billiard ball, that of the water drop, that of the firework, or no sound. The participants were provided with no prior information about the sounds used in the experiment. The results of Experiment 1A were qualified by those of Experiment 1B, in which the participants performed the same task but the sounds' envelopes were filled with noise in order to clear the timbre differences between the sounds. The participants performed two sound categorization tasks successively, the first with the sounds used in Experiment 1A, the second with the sounds used in Experiment 1B. The sound categorization tasks were also preceded by a free sound recognition task. In Experiments 1A and 1B, we expected all audiovisual motion displays to induce the ABE,

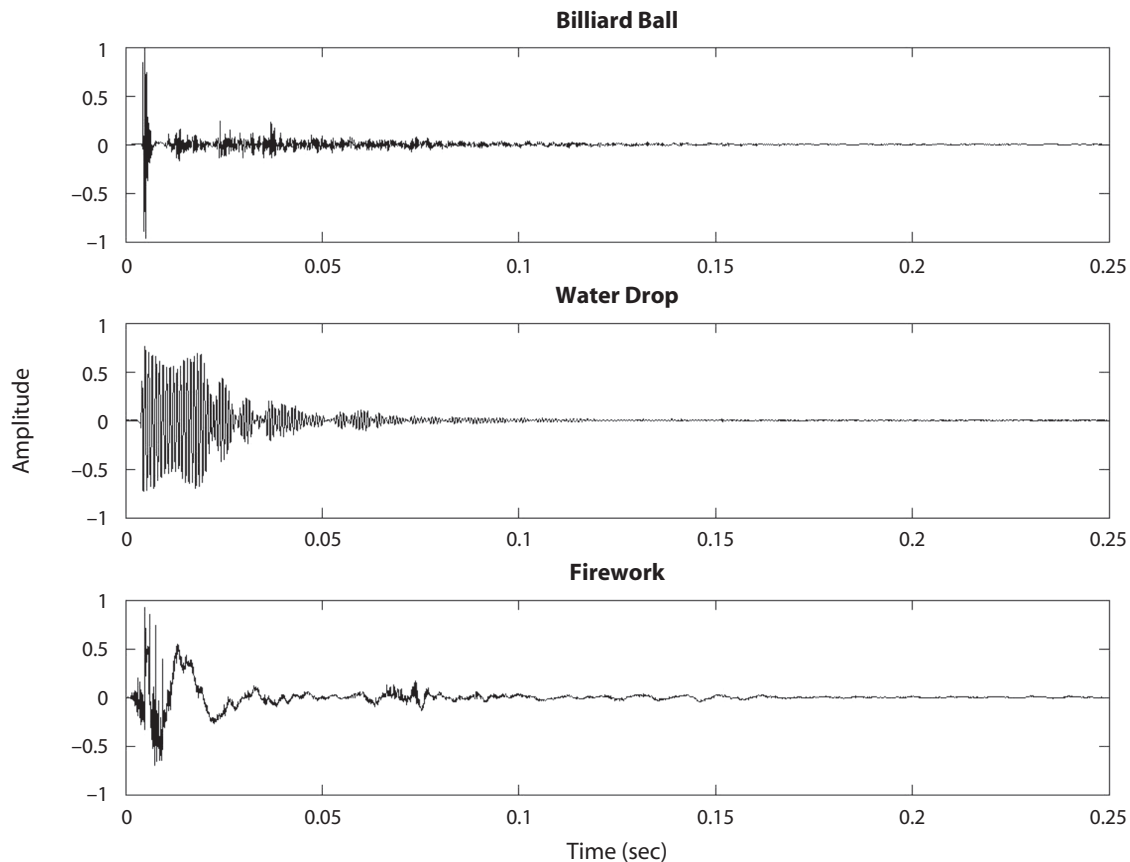


Figure 2. Waveform of the natural sounds used in the experiment, equalized by root-mean square amplitude.

in comparison with the silent version of it. In addition, in Experiment 1A only, we expected the billiard sound to induce (more than the other sounds) the bouncing percept.

The second part of the present study consisted of two additional control experiments, Experiments 2 and 3, that were carried out post hoc, following the reviewers' suggestions, and that were used to further assess the findings and assumptions of the main experiment.

EXPERIMENT 1

Method

Participants. Sixteen participants (3 males) with normal or corrected-to-normal vision and normal hearing participated in the experiment. They were all naive as to the purpose motivating the study.

Apparatus. We wrote our experiments in MATLAB (The MathWorks, Inc., Natick, MA) using the Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997). The software was running on a Pentium IV computer connected to an NEC Multisync FP950 monitor (100-Hz refresh rate). Sounds had a sample rate of 44.1 kHz and 16 bits resolution. The output of the sound card (M-AUDIO Fast Track Pro) was passed to two M-AUDIO Studiophile AV 30 amplified speakers. The speakers were placed at the left/right of the monitor's sides, and the speakers' drivers were aligned with the monitor's horizontal midline so that sounds were perceived as originating from the monitor's center (Bertelson & Aschersleben, 1998). The experiment was conducted in a dark and silent (below 35 dBA at the listener's ear) room. During the experiments, the sounds' peak pressure at the listener's ear was 95 ± 2 dBA.

Visual stimulus. The motion display was the same for all the experiments. Two black disks (0.41° of visual angle) moved within a white (67.0 cd/m^2) square (6.19°) placed in the center of a black (0.09 cd/m^2) background. A black fixation cross was placed 2.69° above the center of the white square and completed the disks' motion context. The background, white square, and fixation cross were continuously present during the experiment. The disks' motion started at the beginning of each trial from the left/right extremities of the square's horizontal axis. The disks moved horizontally with uniform rectilinear motion ($2.07^\circ/\text{sec}$) from two opposite positions in space: They overlapped partially, then completely, and finally continued their motion and stopped at one another's starting point. The disks' motion lasted 1.4 sec. The disks disappeared after the motion.

Sound stimulus: Experiment 1A. In Experiment 1A, the motion display was accompanied by an impact sound (billiard display), a liquid sound (water drop display), an aerodynamic sound (firework display), or no sound. Sounds were taken from a database (SoundIdeas). The impact sound was a billiard ball striking another billiard ball; the liquid sound was one drop of water falling into water; the aerodynamic sound was a firework explosion. Original sound samples had different durations and sound power, but all the envelopes were characterized by an abrupt amplitude attack followed by a gradual amplitude decay (Figure 2). The intensity peak of all the sounds occurred within ~ 10 msec from the sound's onset. Sound samples were equalized in root-mean square power by attenuating (or amplifying) with custom raised cosine modulators the intensity of the sounds' tails.² This operation did not affect the shape of the sounds' envelope, which continued to be characterized by an abrupt amplitude attack and by a gradual amplitude decay. The equalization also returned sounds of similar durations, all within 500 msec. The audiovisual displays were constructed in

such a way that the sound's intensity peak occurred ~200 msec before the disks' overlap.³

Sound stimulus: Experiment 1B. In Experiment 1B, the motion displays were accompanied by a filled-with-noise version of the sounds used in Experiment 1A. The envelopes of billiard, water drop, and firework sounds were extracted as follows. The original sound waves were full-rectified, and the resulting signals were low-pass filtered at 50 Hz and returned the sounds' envelopes. These envelopes were used to modulate the amplitude of a 500-msec-long pink noise. This is a classic procedure for neutralizing a sound's meaning (Fastl & Zwicker, 2006). In Experiment 1B, the three filled-with-noise sounds substituted for the billiard, the water drop, and the firework sounds used in Experiment 1A.

Sound stimulus: Sound categorization tasks. These experiments used the audiovisual displays of either Experiment 1A or 1B, with the exclusion of the silent display.

Procedure. The experiment was divided into five sessions. Experiments 1A and 1B were followed by the free sound recognition task. At the end of the free sound recognition task, the participants performed the two sound categorization tasks. The participants performed Experiments 1A and 1B in counterbalanced order. Also, the two sound categorization tasks were performed in counterbalanced order. In all the experiments, the participants viewed the display binocularly from a distance of 95 cm that was kept constant by means of a chinrest; moreover, they were asked to look at the fixation cross at the beginning of each trial.⁴ In Experiments 1A and 1B, the participants looked at the display and reported whether they perceived the disks as streaming or bouncing by pressing the appropriate buttons on the computer keyboard. In the free sound recognition task, each sound was played once in random order, and the participants were asked to describe as accurately as possible the event at the origin of the sound. In the sound categorization tasks, the participants were asked to categorize the sounds accompanying the audiovisual displays presented in Experiment 1A (or 1B). The participants were asked to categorize the sound while looking at the motion display and were given three categories: impact, liquid, or aerodynamic sounds. In the free recognition task and in the sound categorization tasks, the participants received no feedback on their responses. In Experiments 1A and 1B, there were 20 trials for each of the four audiovisual displays, and the order of the trials was random. In the sound categorization tasks, there were 5 trials for each audiovisual display in Experiment 1A (or 1B).

Results and Discussion

The participants' responses obtained in Experiments 1A and 1B were transformed into percentages of bounce re-

sponses separately for billiard, water drop, and firework displays, as well as for the corresponding filled-with-noise displays. Percentages of bounce responses were subjected to a 2 (Experiment 1A vs. Experiment 1B) \times 4 (displays) two-way ANOVA. The number of bounce responses did not differ in the two experiments [$F(1,15) = 0.007, p > .05$], but it differed across displays [$F(3,45) = 13.98, p < .0001$]. The two-way interaction was also significant [$F(3,45) = 5.28, p < .0001$]. With a set of pairwise comparisons, we tested whether bounce responses were more frequent for the audiovisual than for the silent displays in Experiment 1A, and, within the same experiment, whether bounce responses were more frequent for the billiard than for the water drop and the firework displays.⁵ Billiard, water drop, and firework displays induced a higher number of bounce responses than did the silent display ($ps = .001, .002, \text{ and } .019$, respectively). Moreover, the billiard display induced a higher number of bounce responses than did the water drop and firework displays ($ps = .021 \text{ and } .004$, respectively; see Figure 3, left).

The same pairwise comparisons were performed on the responses collected in Experiment 1B. Here, the number of bounce responses collected with silent displays was larger than that in Experiment 1A (see Figure 3, right), so that only the filled-with-noise water drop display induced a greater ABE than did the silent display ($p = .046$). Most important, the filled-with-noise billiard display did not induce more bounce responses than did either the filled-with-noise water drop ($p = .951$) or the filled-with-noise firework ($p = 1$) display (see Figure 3, right).

Free sound recognition task. Many participants recognized the billiard sound (13 out of 16 participants), whereas the water drop sound was recognized by only 4 participants and the fireworks sound was not recognized as such. The participants, however, confused this sound with either the sound of a rifle shot (8 out of 16) or the sound of a cannon shot (4 out of 16).

Sound categorization tasks. The participants' responses to the sound categorization tasks were coded as percentages of correct categorization for each participant and sound. Three one-sample t tests showed that the partici-

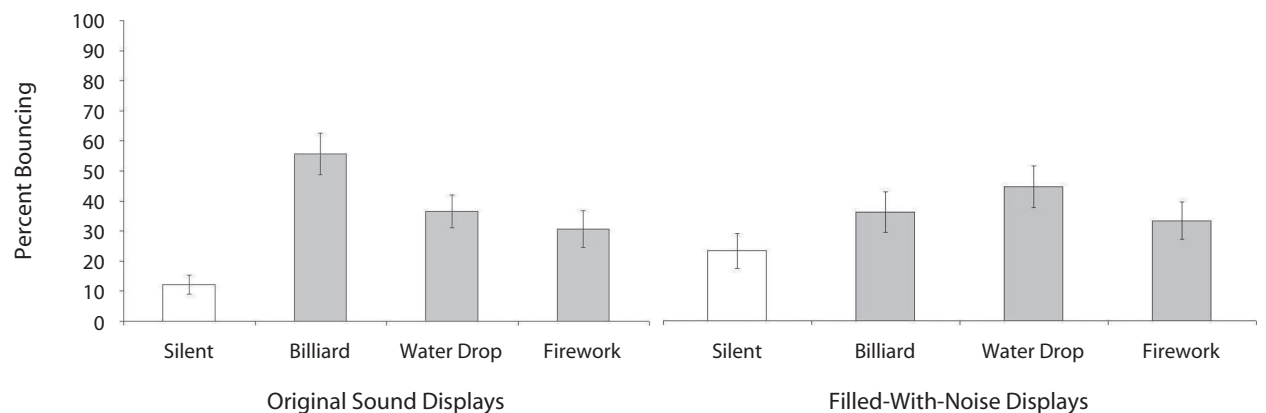


Figure 3. Experiment 1: Mean percentages of bounce responses for silent and audiovisual displays in Experiments 1A (left) and 1B (right). Error bars are ± 1 standard error of the mean.

pants' correct categorizations exceeded chance level (i.e., 33%) for all the sounds when the displays were accompanied by the original sounds (all $t_s > 6.4$, all $p_s < .0001$). With the original sounds, correct categorizations ($M = 92.0\%$, $SD = 3.5\%$) ranged from a minimum of 81.3% (the water drop) up to a maximum of 98% (the billiard ball). We ran a further analysis on the categorization data. The sessions comprised five repetitions of the sounds. Therefore, the participants may have recognized the sounds at their second, third, fourth, and fifth presentations. The responses to the five trials could not, therefore, be independent. For this reason, for each participant, we analyzed only the responses that corresponded to the first presentation of each sound. Correct categorizations exceeded chance level (all $t_s \geq 3$, all $p_s < .05$) for both the water drop sound ($M = 69\%$, $SD = 47\%$) and the firework sound ($M = 94\%$, $SD = 25\%$) and were at ceiling (i.e., $M = 100\%$) for the billiard sound. The data are presented in Figure 4.

The same analyses were performed on the data for the filled-with-noise sounds. Here, correct categorizations were at chance level for the filled-with-noise billiard sound [$t(15) = 1.69$, $p > .05$] and were below chance for the filled-with-noise water drop sound [$t(15) = -5.27$, $p < .0001$]. However, they exceeded chance for the filled-with-noise firework sound [$M = 71.3\%$, $SD = 33.44\%$; $t(15) = 4.57$, $p < .0001$]. These analyses were conducted also on the responses given to the first presentation of the sounds. Correct categorizations were at chance for the filled-with-noise billiard sound ($t = -1.78$, $p > .05$; $M = 19\%$, $SD = 40\%$) and were at floor for the filled-with-noise water drop sound ($M = 0\%$). However, they exceeded chance for the filled-with-noise firework sound ($t = 9.72$, $p > .0001$; $M = 75\%$, $SD = 45\%$). A closer look at the responses to this particular categorization task revealed that 60% of responses given by the participants were "aerodynamic."⁶

In Experiment 1A, the billiard display induced more bounce responses than did the water drop and the firework displays. This difference, however, disappeared in Experiment 1B when the sounds' timbres were cleared. In the free

sound recognition task, many participants recognized the billiard sound, whereas the water drop and the firework sounds were not recognized as much. However, in the sound categorization task, the participants were nevertheless able to categorize correctly the event at the origin of the sounds when the sounds were those used in Experiment 1A, but this ability disappeared when the sounds' timbres were cleared. A possible explanation of the results of Experiment 1A is the following. In Experiment 1A, the sounds' onset preceded the disks' overlap by 200 msec, which was an interval sufficient to allow the participants to recognize the sound (a brain network identified as the *auditory-what* is activated as few as ~70 msec from a sound's onset; Murray, Camen, Gonzalez Andino, Bovet, & Clarke, 2006) and successively evaluate the likelihood of perceiving a bouncing (or streaming) event by evaluating the congruence of the integrated audiovisual percept in comparison with an elastic impact (e.g., Ernst & Banks, 2002).

EXPERIMENT 2

In the discussion of the results of Experiment 1A, we hypothesized that the difference in the ABE observed with the billiard, the water drop, and the firework sounds could be due to a fast sound recognition process that enabled the participant to evaluate the congruence of the integrated audiovisual percept in comparison with an elastic impact before the disks' overlap. Therefore, if our speculation is correct, the difference in the number of bounce responses observed with the three sounds in Experiment 1A should disappear by making the sound's onset coincident with the disk's overlap. To test this prediction, in Experiment 2, the sounds in Experiment 1A were switched on in temporal coincidence with the disks' overlap.

Method

Participants. One group of 12 new participants (with normal or corrected-to-normal vision and normal hearing) participated in each experiment. They were all naive as to the purpose motivating the experiments.

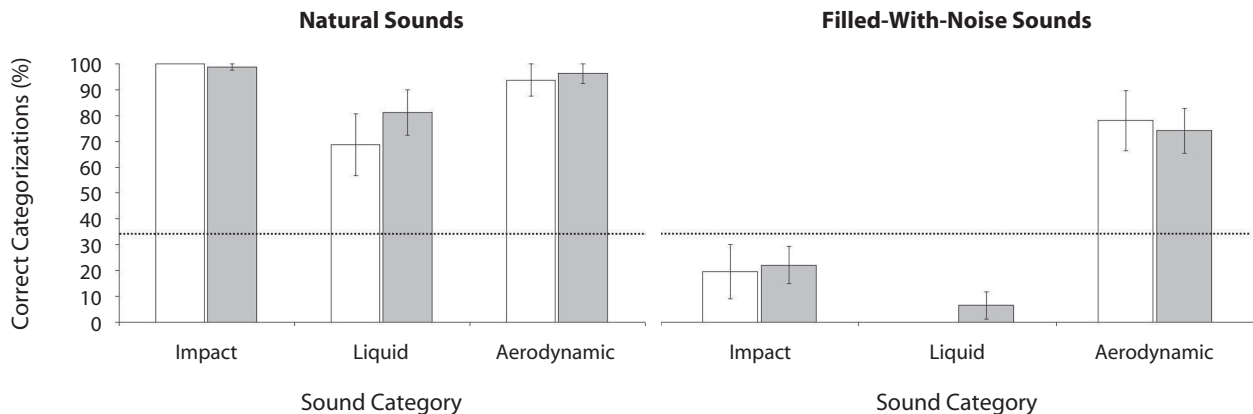


Figure 4. Sound categorization tasks: Correct categorizations for impact, liquid, and aerodynamic sounds with the natural sounds (left) and the filled-with-noise sounds (right). The graphs show the responses to the first presentation of the sound (white bars; see the text for further details) and the responses on all the trials (gray bars). In the graphs, the horizontal dotted lines represent chance level (i.e., 33%), and error bars are ± 1 standard error of the mean.

Stimuli, Apparatus, and Procedure. The apparatus and the motion display were identical to those in Experiment 1A. However, in audiovisual displays, sounds were switched on in coincidence with the disks' overlap. The task was identical to that in Experiment 1A.

Results and Discussion

The results are shown in Figure 5. The participants' responses were transformed into percentages of bounce responses separately for silent, billiard, water drop, and firework displays. Percentages of bounce responses were subjected to a one-way ANOVA with displays (4) as a factor. The number of bounce responses was different across the four displays [$F(3,33) = 52.47, p < .0001$]. A set of pairwise comparisons revealed that billiard, water drop, and firework displays induced a higher number of bounce responses than did the silent display (all $ps < .0001$). However, the billiard display did not induce more bounce responses than either the water drop ($p > .05$) or the firework ($p > .05$) displays.

The results of Experiment 2 showed that the effect of the sounds' congruence/incongruence disappears when they are switched on in coincidence with the disks' overlap.

EXPERIMENT 3

In Experiment 3, we investigated whether the three sounds used in Experiment 1A were equally fast at capturing the participants' attention. The ABE is hypothesized to have an attentional component (Kawabe & Miura, 2006; Watanabe & Shimojo, 1998, 2005), and the simple reaction time is an indication of the capacity of a stimulus to capture attention (Jonides, 1981). In Experiment 3, the participants were asked to react as quickly as possible to the sounds' onsets for the displays used in Experiment 1A. The aim of Experiment 3 was control (at least partially) whether there was any difference in the capacity of the three sounds to capture attention—in particular, the billiard sound.

Method

Participants. One group of 12 new participants (with normal or corrected-to-normal vision and normal hearing) participated in the experiment. They were all naive as to the purpose motivating the experiment.

Stimulus, Apparatus, and Procedure. The participants viewed 16 displays: 1 unimodal, 15 audiovisual. In the audiovisual displays, in contrast to those in Experiment 1A, sounds were switched on in synchrony, ± 100 msec before/after the disks' overlap, or ± 200 msec before/after the disks' overlap. In Experiment 3, we extended the range of times when sounds could be switched on, because we wanted the sounds' onsets to be unpredictable. Each display was presented 10 times to the participants, for a total of 160 trials. The displays were presented in random order, and the participants were asked to press the space bar as soon as they could hear the sound accompanying the display. Silent displays were used as catch trials: The participants did not have to respond to these stimuli. Reaction times were calculated as the interval between the sound's onset and the space bar pressure.

Results and Discussion

The participants' responses to catch trials, reactions smaller than 120 msec (i.e., anticipations), and reactions

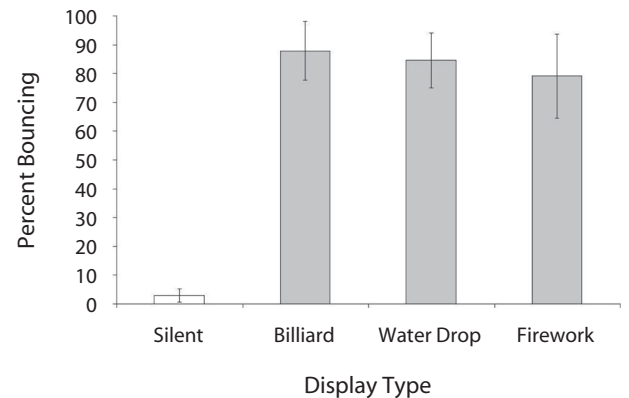


Figure 5. Experiment 2: Mean percentages of bounce responses for silent and audiovisual displays. Error bars are ± 1 standard error of the mean.

greater than 500 msec (i.e., delayed responses) were excluded from the analysis. The excluded data were ~6% of the data collected. Reaction times were subjected to a one-way ANOVA with sounds (3) as a factor. The participants' reactions were independent of the sound type [$F(1,22) = 2.81, p > .05$]. Mean reaction times to the three sounds were $M = 289.6$ msec ($SD = 85.0$ msec) for billiards, $M = 291.8$ msec ($SD = 90.2$ msec) for water drops, and $M = 297.5$ msec ($SD = 89.9$ msec) for fireworks.

The results of Experiment 3 show that the participants were equally fast at responding to the onset of the billiard, the water drop, and the firework sounds. The absence of a difference in the reaction times, however, may not necessarily imply the lack of differential attentional processing of the three sounds. Nevertheless, there is clear evidence that the sounds' timbre is extracted without attention focused on it (Jacobsen, Schröger, & Alter, 2004; Jacobsen, Schröger, & Sussman, 2004; Tervaniemi, Winkler, & Näätänen, 1997). Therefore, the results of the present experiment, together with the evidence of the preattentive processing of timbre, lead us to conclude that it is unlikely that the three sounds had a differential capacity to subtract attentional resources.

GENERAL DISCUSSION

The results of Experiment 3 suggest that, as far as we could test, the natural sounds used in the present study did not differ in the capacity to capture the participants' attention. Nevertheless, in Experiment 1A, the billiard display induced a greater ABE than did the water drop and the firework displays. In Experiment 1B, on the contrary, when the timbre of the sounds was cleared, the results changed: All the displays accompanied by sound returned the same number of bounce responses. In the free sound recognition task, many participants recognized the billiard sound, whereas the water drop and the firework sounds were not recognized as much. However, in the sound categorization task in Experiment 1A, the participants easily categorized all the sound source event types. The sounds, nonetheless, were not categorized when the sounds' timbres were

cleared—that is, when the sounds' envelopes were filled by the pink noise. Moreover, Experiment 2 showed that the effect of the sounds' congruence/incongruence observed in Experiment 1A disappeared when sounds were switched on in coincidence with the disks' overlap.

The results of the present study do not agree with the view that attention alone is at the origin of the ABE. In the literature, this view is supported by the finding that the subtraction of attention is sensory aspecific: Indeed, signals delivered to sensory modalities other than audition (e.g., touch) also induce the bounce percept (Watanabe & Shimojo, 2005). However, the results of our previous study (i.e., Grassi & Casco, 2009) suggested that a second component is involved in the ABE—namely, the congruence of the audiovisual percept in comparison with a real elastic bounce. In that study, the effect of this second component could be observed when we compared the effect of a bounce-congruent sound (abrupt attack at 75 dBA, gradual decay to 55 dBA) with the effect of a bounce-incongruent sound (abrupt attack at 75 dBA, gradual increment to 95 dBA) and found that, even though the sounds had identical abrupt onsets and the latter sound was evaluated by the participants as more salient than the former, the former collected more bounce responses than did the latter. The results of the present study also suggest that attention alone is insufficient for explaining the ABE; moreover, they suggest that the congruence of the audiovisual percept, in comparison with a real elastic bounce, plays a role in the ABE. In the present experiment, all the sounds were characterized by similar envelopes. Therefore, they all reduced the attentional resources necessary for the motion integration in a similar fashion (see the results of Experiment 3). (In this regard, the reader should keep in mind that the billiard sound was also the least intense sound of the lot and that the ABE is known to be directly correlated with the sound's intensity; Dufour, Touzalin, Moessinger, Brochard, & Despres, 2008; Grassi & Casco, 2009; Watanabe & Shimojo, 2001b). However, the billiard sound, in particular, fitted with a bouncing event of solid objects; thus, this display induced more bounce responses than did the water drop and firework displays. The results of Experiment 1B support this conclusion: When the timbre of the sound was cleared, the filled-with-noise billiard sound display did not induce more bounce responses than did the other two filled-with-noise audiovisual displays. In brief, we believe that the results of the present study highlight the role of sound congruence more clearly than do those in our previous study (Grassi & Casco, 2009). In that experiment, the ABE was modulated by modulating the intensive content of sounds and, as we wrote previously, the magnitude of the ABE is known to be related to the sounds' intensity. Here, however, the ABE was modulated by manipulating the nonintensive content of sounds (i.e., the frequency content)—that is, the timbre of the sounds. In the present Experiment 1A, the sounds' onset preceded the disks' overlap by 200 msec, which was an interval sufficient to allow the participants to recognize the billiard sound and successively evaluate the likelihood of perceiving a bouncing (or streaming) event by evaluating the congruence of the integrated audiovi-

sual percept in comparison with an elastic impact (e.g., Ernst & Banks, 2002). A common network of brain areas within the *what* auditory processing stream (e.g., the right posterior superior, the right middle temporal, and the left inferior frontal cortices) is activated as little as ~70 msec after a sound's onset (Murray et al., 2006). Sounds' timbre is preattentively determined from relatively short (150-msec) sound samples (Tervaniemi et al., 1997). As a corollary, in a four-alternative forced choice task, timbre recognition requires only as few as one cycle of a sound wave to be performed over chance level, whereas, for example, other elementary properties of sounds (e.g., pitch) require many cycles to be recognized with the same level of accuracy (Robinson & Patterson, 1995a, 1995b). Moreover, the results of Experiment 2 support our explanation: When the sounds were switched on in coincidence with the disks' overlap (i.e., too late for the sound-congruence/bounce-likelihood comparison), the effect of the sound timbre disappeared. The exact origin of the sound congruency component of ABE is, however, still unknown. The simplest hypothesis could be that the sound biases the participant's response toward the bounce response. All the studies on the ABE (including ours) have relied on self-report responses (i.e., "do you perceive streaming or bouncing?"). All the studies, therefore, have collected responses that *are* open to biases in signal detection terms (Green & Swets, 1966).

Although we have provided evidences for a sound congruency component of the effect, the attentional component, both covert and overt, cannot be denied either. We would like here to speculate about the relationship between the two components. The present and the previous studies seem to suggest how the attentional and the sound congruency components of the ABE interact and suggest that the two components may interact in an additive (i.e., not multiplicative) fashion. This is supported by the results of the experiments in which only one of the two components was acted on: They all showed an increment in the number of bounce responses of the multimodal condition, in comparison with the unimodal (i.e., visual only) condition. For example, the tactile stimulation used by Watanabe and Shimojo (2005) subtracted attentional resources; however, it did not increment the congruence of the event in comparison with a real, elastic bounce. Nonetheless, the authors observed an increment in the number of bounce responses. Along the same line are the results of Grassi and Casco (2009) when they used a bounce-incongruent sound with abrupt amplitude attack. They also observed an increment in the number of bounce responses. Finally, when the two components are both active, it is possible to observe more bounce responses than when, for example, one component is acting alone (see the results of Experiment 1A, as well as those in Grassi & Casco, 2009, Experiment 2). However, the independency of the two components seems difficult to prove because, as far as we can judge, the attentional component, in particular, cannot be completely eliminated.

To conclude, our results add information on how the human perceptual and cognitive system combines signals from multiple senses to form a coherent and unified per-

cept of the outside world. Several previous studies have obtained data on cross-modal binding, but the studies were often concentrated on how vision affects audition. The ABE is interesting because it allows investigation of the modulation of vision by audition. The present results agree with the view that the audiovisual integration occurs at multiple processing stages, ranging from early sensory to semantic and higher conceptual (or decisional) processes. At the behavioral level, multisensory integration information facilitates categorization of objects or novel events in our environment (Beauchamp, Argall, Bodurka, Duyn, & Martin, 2004; Beauchamp, Lee, Argall, & Martin, 2004; Gottfried & Dolan, 2003; Laurienti, Kraft, Maldjian, Burdette, & Wallace, 2004; Molholm, Ritter, Javitt, & Foxe, 2004). The present results allow us to speculate not only that multiple stages of processing may show audiovisual integration, but also that integration is possible between different levels of representation: They show that high-level auditory information may modulate the response of a very low-level mechanism responsible for the binding of elementary motion signals (e.g., Parovel & Casco, 2006).

AUTHOR NOTE

The authors thank Natalia Bak and Cristina Venturini for helping with the data collection and reviewers for improving the manuscript substantially. A demo of the stimuli used in Experiment 1B and Experiment 2 can be seen at the following Web page: www.psy.unipd.it/~grassi/grassicascodemos.html. Correspondence concerning this article should be addressed to M. Grassi, Dipartimento di Psicologia Generale, Università di Padova, Via Venezia 8, 35131 Padua, Italy (e-mail: massimo.grassi@unipd.it).

REFERENCES

- BALLAS, J. A. (1993). Common factors in the identification of an assortment of brief everyday sounds. *Journal of Experimental Psychology: Human Perception & Performance*, *19*, 250-267. doi:10.1037/h0090367
- BEAUCHAMP, M. S., ARGALL, B. D., BODURKA, J., DUYN, J. H., & MARTIN, A. (2004). Unraveling multisensory integration: Patchy organization within human STS multisensory cortex. *Nature Neuroscience*, *7*, 1190-1192. doi:10.1038/nn1333
- BEAUCHAMP, M. S., LEE, K. E., ARGALL, B. D., & MARTIN, A. (2004). Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron*, *41*, 809-823. doi:10.1016/S0896-6273(04)00070-4
- BERTELSON, P., & ASCHERSLEBEN, G. (1998). Automatic visual bias of perceived auditory location. *Psychonomic Bulletin & Review*, *5*, 482-489.
- BERTENTHAL, B. I., BANTON, T., & BRADBURY, A. (1993). Directional bias in the perception of translating patterns. *Perception*, *22*, 193-207. doi:10.1068/p220193
- BRAINARD, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433-436.
- DUFOUR, A., TOUZALIN, P., MOESSINGER, M., BROCHARD, R., & DESPRES, O. (2008). Visual motion disambiguation by a subliminal sound. *Consciousness & Cognition*, *17*, 790-797. doi:10.1016/j.concog.2007.09.001
- ERNST, M. O., & BANKS, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*, 429-433. doi:10.1038/415429a
- FASTL, H., & ZWICKER, E. (2006). *Psychoacoustics: Facts and models* (3rd ed.). New York: Springer.
- GAVER, W. W. (1993a). How do we hear the world? Explanations in ecological acoustics. *Ecological Psychology*, *5*, 285-313. doi:10.1207/s15326969eco0504_2
- GAVER, W. W. (1993b). What in the world do we hear? An ecological approach to auditory event perception. *Ecological Psychology*, *5*, 1-29. doi:10.1207/s15326969eco0501_1
- GOTTFRIED, J. A., & DOLAN, R. J. (2003). The nose smells what the eye sees: Crossmodal visual facilitation of human olfactory perception. *Neuron*, *39*, 375-386. doi:10.1016/S0896-6273(03)00392-1
- GRASSI, M., & CASCO, C. (2009). Audiovisual bounce-inducing effect: Attention alone does not explain why the discs are bouncing. *Journal of Experimental Psychology: Human Perception & Performance*, *35*, 235-243. doi:10.1037/a0013031
- GRAY, P. (2002). *Psychology* (4th ed.). New York: Worth.
- GREEN, D. M., & SWETS, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.
- GROVE, P. M., & SAKURAL, K. (2009). Auditory induced bounce perception persists as the probability of a motion reversal is reduced. *Perception*, *38*, 951-965. doi:10.1068/p5860
- GYGI, B., KIDD, G. R., & WATSON, C. S. (2004). Spectral-temporal factors in the identification of environmental sounds. *Journal of the Acoustical Society of America*, *115*, 1252-1265. doi:10.1121/1.1635840
- JACOBSEN, T., SCHRÖGER, E., & ALTER, K. (2004). Pre-attentive perception of vowel phonemes from variable speech stimuli. *Psychophysiology*, *41*, 654-659. doi:10.1111/1469-8986.2004.00175.x
- JACOBSEN, T., SCHRÖGER, E., & SUSSMAN, E. (2004). Pre-attentive categorization of vowel formant structure in complex tones. *Cognitive Brain Research*, *20*, 473-479. doi:10.1016/j.cogbrainres.2004.03.021
- JONIDES, J. (1981). Voluntary versus automatic control over the mind's eye movement. In J. [B.] Long & A. [D.] Baddeley (Eds.), *Attention and performance IX* (pp. 187-203). Hillsdale, NJ: Erlbaum.
- KAWABE, T., & MIURA, K. (2006). Effects of the orientation of moving objects on the perception of streaming/bouncing motion displays. *Perception & Psychophysics*, *68*, 750-758.
- KAWACHI, Y., & GYOBA, J. (2006). Presentation of a visual nearby moving object alters stream/bounce event perception. *Perception*, *35*, 1289-1294. doi:10.1068/p5594
- KOFFKA, K. (1935). *Gestalt psychology*. New York: Harcourt Brace.
- LAURIENTI, P. J., KRAFT, R. A., MALDIJIAN, J. A., BURDETTE, J. H., & WALLACE, M. T. (2004). Semantic congruence is a critical factor in multisensory behavioral performance. *Experimental Brain Research*, *158*, 405-414. doi:10.1007/s00221-004-1913-2
- MARR, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. New York: Freeman.
- METZGER, W. (1934). Beobachtungen über phänomenale Identität. *Psychologische Forschung*, *19*, 1-60. doi:10.1007/BF02409733
- MOLHOLM, S., RITTER, W., JAVITT, D. C., & FOXE, J. J. (2004). Multisensory visual-auditory object recognition in humans: A high-density electrical mapping study. *Cerebral Cortex*, *14*, 452-465. doi:10.1093/cercor/bhh007
- MURRAY, M. M., CAMEN, C., GONZALEZ ANDINO, S. L., BOVET, P., & CLARKE, S. (2006). Rapid brain discrimination of sounds of objects. *Journal of Neuroscience*, *26*, 1293-1302. doi:10.1523/JNEUROSCI.4511-05.2006
- PAROVEL, G., & CASCO, C. (2006). The psychophysical law of speed estimation in Michotte's causal events. *Vision Research*, *46*, 4134-4142. doi:10.1016/j.visres.2006.08.005
- PELLI, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437-442.
- REMIJN, G. B., & ITO, H. (2007). Perceptual completion in a dynamic scene: An investigation with an ambiguous motion paradigm. *Vision Research*, *47*, 1869-1879. doi:10.1016/j.visres.2007.03.017
- REMIJN, G. B., ITO, H., & NAKAJIAMA, Y. (2004). Audiovisual integration: An investigation of the "streaming-bouncing" phenomenon. *Journal of Physiological Anthropology & Applied Human Science*, *23*, 243-247. doi:10.2114/jpa.23.243
- ROBINSON, K. L., & PATTERSON, R. D. (1995a). The duration required to identify the instrument, the octave, or the pitch-chroma of a musical note. *Music Perception*, *13*, 1-15.
- ROBINSON, K. L., & PATTERSON, R. D. (1995b). The stimulus duration required to identify vowels, their octave, and their pitch chroma. *Journal of the Acoustical Society of America*, *98*, 1858-1865. doi:10.1121/1.414405
- SEKULER, A. B., & SEKULER, R. (1999). Collisions between moving visual targets: What controls alternative ways of seeing an ambiguous display? *Perception*, *28*, 415-432. doi:10.1068/p2909

- SEKULER, R., SEKULER, A. B., & LAU, R. (1997). Sound alters visual motion perception. *Nature*, **385**, 308. doi:10.1038/385308a0
- TERVANIEMI, M., WINKLER, I., & NÄÄTÄNEN, R. (1997). Pre-attentive categorization of sounds by timbre as revealed by event-related potentials. *NeuroReport*, **8**, 2571-2574.
- WATANABE, K., & SHIMOJO, S. (1998). Attentional modulation in perception of visual motion events. *Perception*, **27**, 1041-1054. doi:10.1068/p271041
- WATANABE, K., & SHIMOJO, S. (2001a). Postcoincidence trajectory duration affects motion event perception. *Perception & Psychophysics*, **63**, 16-28.
- WATANABE, K., & SHIMOJO, S. (2001b). When sound affects vision: Effects of auditory grouping on visual motion perception. *Psychological Science*, **12**, 109-116. doi:10.1111/1467-9280.00319
- WATANABE, K., & SHIMOJO, S. (2005). Crossmodal attention in event perception. In L. Itti, G. Rees, & J. K. Tsotsos (Eds.), *Neurobiology of attention* (pp. 538-546). San Diego: Elsevier.
- WERTHEIMER, M. (1923). Untersuchungen zur Lehre von der Gestalt: II. *Psychologische Forschung*, **4**, 301-350. doi:10.1007/BF00410640
- ZHOU, F., WONG, V., & SEKULER, R. (2007). Multi-sensory integration of spatio-temporal segmentation cues: One plus one does not always equal two. *Experimental Brain Research*, **180**, 641-654. doi:10.1007/s00221-007-0897-0

NOTES

1. In the present article, here as well as later in the text, with the word *reduction* we refer to a partial reduction and not to a complete elimination of attentional resources.
2. The exception was the billiard ball sound, which, despite the amplification, was 5 dB less powerful than the other sounds.
3. We tested the timing accuracy of audiovisual displays with an oscilloscope. On average, the sound preceded the disks' overlap by 200 ± 5 msec. Moreover, so that the timing and quality of the audiovisual displays would not be different from those of the silent displays (because of the call of the sound card), all the displays actually played a sound. The sound played during the silent display had null amplitude.
4. We did not control, however, whether fixation was actually kept by the participants.
5. Here and in the successive analyses, the probability shown by repeated statistical tests was adjusted with the Bonferroni correction for the number of tests.
6. An informal listening to the sounds revealed that the timbre of the firework sound and that of the pink noise actually sounded quite similar.

(Manuscript received May 12, 2009;
revision accepted for publication September 16, 2009.)