Audiovisual Bounce-Inducing Effect: Attention Alone Does Not Explain Why the Discs Are Bouncing

Massimo Grassi and Clara Casco
University of Padova

Two discs moving from opposite points in space, overlapping and stopping at the other disc’s starting point, can be seen as either bouncing or streaming through each other. With silent displays, observers report the discs as streaming, whereas if a sound is played when the discs touch each other, observers report the discs as bouncing. The origin of the switch from streaming to bouncing response is not known yet. The sound either shifts perception toward that of an impact-elastic event (i.e., a bounce) or subtracts the attention that is necessary to perceive the discs as streaming. We used either impact-similar (abrupt amplitude attack, gradual decay) or impact-dissimilar sounds (gradual amplitude attack, abrupt decay) and found that the first sounds induce the bouncing response, whereas the latter, although as distracting as the first, render streaming and bouncing responses equally frequent at most. We interpret the audiovisual bouncing effect as resulting from attention subtraction, which raises the number of bounce responses in comparison with silent displays, and from perception, which further increments the number of bounce responses and turns the response into a strong bounce response.

Keywords: crossmodal perception, audiovisual perception

The last two decades of multisensory perception studies have shown that, overall, vision prevails over hearing in spatial tasks, whereas hearing prevails over vision in temporal tasks (see Calvert, Spence, & Stein, 2004, for an overview). There is, however, an exception such as that shown by Sekuler, Sekuler, and Lau (1997), where the presentation of a sound reverses the perceived direction of motion of two visual objects (hereafter referred to as audiovisual bounce-inducing effect [ABE]). The visual display shows two identical discs that move along the azimuth with uniform rectilinear motion and opposite directions: The discs start their motion, overlap, and stop at each other’s starting point. Metzger (1934) proposed this display as an example of bistable motion, because discs can be perceived either as bouncing off or streaming through each other. The double perceptual solution of Metzger’s display can be better understood if the display is translated into a three-dimensional event. The observer is looking at two objects placed at different depths so that their retinal image has identical size. A first possible event is that the objects start their motion, overlap (i.e., one object completely occludes the other), then stream past each other. The alternative possible event differs only in the postoverlap trajectory: After the occlusion, the objects reverse their motion (i.e., they bounce off one another) and return to their original starting position.1 These two events are different; however, they would produce an identical two-dimensional motion pattern on the retina.

Although Metzger’s (1934) display is bistable, the streaming percept largely prevails, and observers perceive the discs as streaming much more often than bouncing (less than 20% bounce responses; Berthenthal, Banton, & Bradbury, 1993; Kawabe & Miura, 2006; Kawachi & Gyoba, 2006; Remijn & Ito, 2007; Sekuler & Sekuler, 1999; Watanabe & Shimojo, 1998). Several authors suggest that, because momentum makes objects move in the same direction as in the past, the visual system may be unwilling to reverse the discs’ direction of motion (Sekuler & Sekuler, 1999; Watanabe & Shimojo, 1998; Watanabe & Shimojo, 2001a, 2001b). However, Sekuler et al. (1997) showed that it is sufficient to present a brief sound simultaneously with the discs’ touch to increase the number of bounce responses from 20% to 80% (Kawabe & Miura, 2006; Kawachi & Gyoba, 2006; Kawachi, Kawabe, & Gyoba, 2007; Remijn, Ito, & Nakajima, 2004; Sekuler et al., 1997; Watanabe & Shimojo, 2001a, 2001b; Zhou, Wong, & Sekuler, 2007).

Since the ABE was observed, two alternative explanations have been advanced to account for the change in the observers’ response when the sound was added to the silent display. Sekuler et al. (1997) suggested that the sound alters the visual motion perception. The authors argued that, in the real world, when two objects collide an impact sound occurs (Gaver, 1993a, 1993b). The temporal coincidence between sound and discs’ touch increases the realism of the display in comparison with a real, elastic impact—therefore, the perception of the discs as bouncing. Alternatively, the sound may simply subtract attentional resources from vision. According to some authors, attention would integrate the discs’ local motion signals, and

---

1 Strictly speaking, the verb “to bounce” is inappropriate for describing this event. Two real objects bouncing (e.g., such as two pool balls) cannot occupy the same position in space.
this integration process would promote the perception of streaming when observers look at the silent display (e.g., Kawabe & Miura, 2006; Watanabe, 2001; Watanabe & Shimojo, 1998). In the audiovisual display, the sound is played when the discs touch one another; therefore, it subtracts attentional resources for the integration process underlying the perception of the streaming postoverlap trajectory. In practice, the sound promotes the perception of bouncing because it inhibits the perception of streaming. Coherent with the attentional hypothesis, bounce responses are also predominant when brief tactile or visual stimulations are delivered simultaneously with the discs’ touch (Kawabe & Miura, 2006; Watanabe, 2001; Watanabe & Shimojo, 1998). Here, we report an experimentum crucis demonstrating that attention alone is insufficient to explain the ABE.

The frequency content of impact sounds varies widely, whereas their intensity profile (i.e., the envelope) is invariably characterized by an abrupt change in sound pressure level followed by a gradual decay (Gaver, 1993b). Many studies have investigated the perception of synthetic sounds whose envelope is similar to that of impact sounds. In the literature, these sounds are called damped sounds (see Figure 1). The perception of damped sounds has often been compared with the perception of ramped sounds (i.e., a damped sound played backward in time). Although ramped and damped sounds have identical durations, identical average sound pressure levels, and identical change in sound pressure levels, the first are perceived as longer (DiGiovanni & Schlauch, 2007; Grassi & Darwin, 2006; Schlauch, Ries, & DiGiovanni, 2001) and louder (Stecker & Hafter, 2000), and as characterized by a greater change in loudness (Neuhoff, 1998, 2001) than the latter. In brief, ramped sounds are perceptually more salient than damped sounds (Grassi & Darwin, 2006).

In Experiments 1 and 2, the discs’ motion display was accompanied by either a ramped or a damped sound. Because the damped sound resembles an impact sound more than the ramped sound does, it should improve the realism of the bouncing event and induce more bounce responses than the ramped sound. However, because the ramped sound is perceptually more salient than the damped sound, it should subtract more attentional resources and induce more bounce responses than the damped sound.

![Figure 1. Schematic representation of the audiovisual displays used in Experiments 1 and 2.](image-url)
Experiment 1

In this experiment observers viewed Metzger’s (1934) display and were asked to report their perception (i.e., streaming or bouncing). Displays could be silent, accompanied by a ramped sound, or accompanied by a damped sound. Rampd and damped sounds differ in how the sound pressure is distributed over time: The damped sound’s peak pressure is located in the onset region, whereas the ramped sound’s peak pressure is located in the offset region. In order to have the sound pressure peak in coincidence with the discs’ touch, sounds were either switched on at discs’ touch or switched off at discs’ touch (see Figure 1).

Method

Participants. Ten participants with normal or corrected-to-normal vision and hearing participated in the experiment. They were all naive as to the purpose of the study.

Apparatus. We wrote our experiments in Matlab (Mathworks; Natick, MA) using the Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997). The software was running on a Pentium IV computer connected to a 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 computer soundcard was passed to two Philips MMS110 amplified speakers. Speakers were placed at the left/right of the monitor’s sides, and 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 listener’s ear) room. During the experiment, the maximum sound pressure at the listener’s ear was 87 dBA.

Stimuli. Participants evaluated one silent, four damped (i.e., display accompanied with a damped sound), and four ramped (i.e., display accompanied with a ramped sound) displays. The visual stimulus was identical for all displays. Two black discs (0.41° visual angle) moved within a white (67.0 cd/m2) square (6.19°) placed in the center of a black (0.09 cd/m2) background. A black fixation cross was placed at 2.69° above the center of the white region and completed the discs’ motion context. Background, square region, and fixation cross were continuously present during the experiment. The discs’ motion started at the beginning of each trial from the left/right extremities of the square’s horizontal axis. Discs moved horizontally with uniform rectilinear motion (2.07°/s) from two opposite positions in space: They overlapped partially, then completely, and finally continued their motion and stopped at the other disc’s starting point. The discs’ motion lasted 1.4 s. The discs disappeared after the motion.

The sounds accompanying the audiovisual displays were either a tone or a noise. The tone was a harmonic complex obtained by summing 10 sinusoids with frequencies corresponding to harmonics 2 to 10 of a fundamental sinusoidal tone of 250-Hz. All sinusoids had identical phases and amplitudes. The noise was obtained by filtering a Gaussian noise with a band-pass filter and attenuating all frequencies out of the 250–2,500 Hz range. Therefore, the frequency content of tone and noise spanned over an identical frequency range. Sounds’ envelopes could be either damped or ramped in sound pressure level. Damped sounds were obtained by modulating the amplitude of a 100-ms long, amplitude-steady sound with an exponential ramp that decreased 40 dB over its duration. The resulting sound was gated on and off with two 5-ms raised cosine ramps. The ramped sounds were damped sounds played backward in time. Damped and ramped sounds could be switched on either 100 ms before the discs’ touch or simultaneously with the discs’ touch2 (see Figure 1).

Procedure. Participants viewed the display binocularly from a distance of 95 cm that was kept constant by means of a chin rest. In the experiment there were 20 trials for each of the nine displays, and the order of trials was random. After each trial, participants were asked to report whether they had perceived the discs as streaming or bouncing by pressing the appropriate button on the computer keyboard.

Results and Discussion

Data gathered from the audiovisual displays were subjected to a 2 (sound timbres) × 2 (onset times) × 2 (envelopes) analysis of variance (ANOVA). The magnitude of the ABE was identical regardless of the frequency content of the sound (F < 1) and the temporal position of the sound onset, F(1, 9) = 1.66, p > .05. Participants showed a much greater ABE with damped than with ramped displays, F(1, 9) = 159.77, p < .0001, ηp2 = .947. Two- and three-way interactions did not return a significant result (all Fs < 1). With two planned comparisons we tested whether bounce responses were more frequent for damped (or ramped) displays than for silent displays. Damped displays gathered a much higher number of bounce responses than silent ones, F(1, 9) = 160.27, p < .0001. ηp2 = .947, whereas ramped displays gathered a number of bounce responses identical to silent ones, F < 1 (see Figure 2).

The results showed that the ABE is independent of the frequency content of the sound and of a perfect timing between discs’ touch and sound’s onset (e.g., Remijn et al., 2004; Sekuler et al., 1997; Watanabe, 2001). However, the ABE is dependent on the sound’s envelope because damped (and not ramped) sounds induced it.

In the current experiment, participants did not perceive the ABE even though a potentially subtracting attention sound (i.e., the ramped sound) was presented simultaneously with the discs’ touch. It may be argued that ramped sounds do not subtract attentional resources as much as damped sounds do. In particular, their low sound pressure onset might capture less attention than the high sound pressure onset of damped sounds. Simple reaction times are a measure of stimulus-driven attentional capture (Jonides, 1981). Recently, Grassi and Darwin (2006) showed that the simple reaction time to a sound’s onset (but also the first spike of the auditory nerve; see Heil, 1997a, 1997b) is proportional to the sound pressure of the sound at onset. In Experiment 2, ramped sounds were amplified to make their sound pressure at onset identical to that of damped sounds.

2 We tested the timing accuracy of audiovisual displays with an oscilloscope. On average the sound preceded the discs’ touch by 2 ms (SD = 4.22). Moreover, in order to avoid the situation where timing and quality of audiovisual displays were different from those of silent displays (because of the call of the soundcard), all displays actually played a sound. The sound played during silent displays had null amplitude.
Experiment 2

In the current experiment participants observed either silent, ramped, or damped displays. In ramped displays, the sound pressure level of ramped sounds was amplified so that their onset level was identical to that of damped sounds in damped displays (see Figure 1). Because of the amplification, the average sound pressure level of ramped sounds was much higher than the average sound pressure level of damped sounds. According to the attentional explanation, ramped and damped displays should return, at least, the same number of bounce responses. In contrast, the perceptual explanation predicts a higher number of bounce responses for damped than for ramped displays.

Method

Participants. Fourteen new participants with normal or corrected-to-normal vision and hearing participated in the experiment. They were all naive to the purpose of the study.

Apparatus, stimuli, and procedure. The visual stimulus, apparatus, procedure, and observer’s task were identical to those in Experiment 1. Participants evaluated five displays: one unimodal, four audiovisual. The timbres used to create the sounds of the audiovisual displays were the complex tone and the bandpass noise used previously. Sounds could be either damped or ramped sounds. Damped sounds were obtained by modulating the amplitude of a 100-ms long, amplitude-steady sound with an exponential ramp that decreased 20 dB over its duration. The resulting sounds were gated on and off with two 5-ms raised cosine ramps. Rampend sounds were time-reversed damped sounds amplified by 20 dB. The sound pressure level of the damped sounds ranged from 75 dBA to 55 dBA (65 dBA average level), whereas the sound pressure level of ramped sounds ranged from 75 dBA to 95 dBA (85 dBA average level). In the current experiment, the sounds were always switched on simultaneously with the discs’ touch.

Results

Data gathered with audiovisual displays were subjected to a 2 (sound timbres) × 2 (envelopes) ANOVA. The magnitude of the ABE was identical regardless of the frequency content of the sound \( F < 1 \), and participants showed a greater ABE for damped than for ramped displays, \( F(1, 13) = 15.32, p = .002, \eta^2_p = .541 \). The two-way interaction did not return a significant result, \( F(1, 13) = 1.93, p < .05 \). Furthermore, two planned comparisons showed that both ramped and damped displays gathered a number of bounce responses higher than silent displays: respectively, \( F(1, 13) = 9.67, p = .008, \eta^2_p = .427 \), and \( F(1, 13) = 56.70, p < .0001, \eta^2_p = .814 \). Rampend displays returned an equal number of bouncing and streaming responses (see Figure 2). Indeed, two one-sample \( t \) tests \( (H_0 = 50\%) \) showed that in ramped displays streaming and bouncing response were equally frequent \( (t < 1 \), statistical power \( p = .020 \)), whereas, in damped displays, the first response occurred more frequently than the latter, \( t(13) = 5.79, p < .0001, d = 1.54 \).

The results of the current experiment confirm those of Experiment 1. Bounce responses were higher with damped than with ramped displays. However, because in the current experiment ramped sounds were amplified, here ramped displays gathered more bounce responses than the ramped displays of Experiment 1.

Discussion of Experiments 1 and 2

In Experiments 1 and 2 participants returned more bounce responses with damped displays than with ramped displays. In Experiment 2, in particular, this result was obtained even though the average sound pressure level of ramped sounds was 20 dB higher than that of damped sounds. The results collected so far suggest that the subtraction of attention caused by the sound (i.e., the ramped) is not sufficient to induce the ABE (Experiment 1) even when the sound is relatively loud (Experiment 2). On the contrary, an impact realistic sound (i.e., the damped) is sufficient to induce a compelling ABE (Experiment 1) even when the sound is relatively quiet (Experiment 2). In synthesis, impact realistic sounds are sufficient to induce a clear bounce response; in contrast, impact unrealistic sounds (although possibly as distracting as impact realistic sounds) do not induce a clear bounce response as much.

Ramped sounds are known to be more salient than damped sounds (DiGiovanni & Schlauch, 2007; Grassi & Darwin, 2006; 3 In all experiments, the probability returned by repeated statistical tests was adjusted with the Bonferroni correction.)
Neuhoff, 1998, 2001; Schlauch, Ries, & DiGiovanni, 2001; Stecker & Hafter, 2000). For this reason, we assumed that they could not subtract more attentional resources than damped sounds. However, in Experiments 1 and 2, sounds were not presented alone but within the peculiar context of a visual motion display, and this might affect the perceptual properties of the sounds. Experiment 2 had a further assumption. The simple reaction time is an indication of the capacity of a stimulus to capture attention (Jonides, 1981), and the simple reaction time to a sound is a function of the sound pressure level at the sound’s onset (Grassi & Darwin, 2006). For this reason, the ramped and damped sounds of Experiment 2 should capture attention equally fast. But, once again, this assumption may not hold when sounds are presented within the context of a visual motion display.

Experiments 3 and 4 used the displays of Experiments 1 and 2. The experiments aimed at investigating the difference in perceptual salience between ramped and damped sounds when they are heard within the context of the discs’ motion display. Moreover, participants performed a simple reaction time task (i.e., they looked at the displays and reacted to the sounds’ onset) to estimate how fast ramped and damped displays captured attention.

Experiment 3

Experiments 3 and 4 were designed to assess the capability of ramped and damped sounds to capture attention (i.e., to subtract attentional resources) when they are heard within the context of the discs’ motion display used in Experiments 1 and 2. In the first part of the experiment, participants looked at the displays and were asked to react as fast as possible to the sound’s onset. In the second part of the experiment, participants looked at pairs of audiovisual displays (one ramped and one damped) and were asked to judge in which display the sound was longer, louder, and characterized by the greater change in loudness.

Method

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

Apparatus, stimuli, and procedure: Reaction time. The visual stimulus, sounds accompanying the audiovisual displays, and apparatus were identical to those used in Experiment 1. Participants viewed 21 displays: 1 unimodal, 20 audiovisual. In the audiovisual displays, in contrast to those in Experiment 1, the sounds were switched on either in synchrony, \(\pm 100\) ms before/after, or \(\pm 200\) ms before/after the discs’ touch. In the current experiment, we extended the range of times when sounds could be switched on because we wanted the sounds’ onset to be unpredictable for the participant. Each display was presented 10 times to the participant, for a total of 210 trials. Displays were presented in random order, and the participant was asked to press the spacebar as soon as she or he could hear the sound accompanying the display. Silent displays were used as catch trials: Participants did not have to respond to these stimuli. Reaction times were calculated as the interval between the sound’s onset and the spacebar pressure.

Apparatus, stimuli, and procedure: Perceptual evaluation of sounds. The visual stimulus and apparatus were identical to those used in Experiment 1. This part of the experiment was performed with a subset of the audiovisual displays used in Experiment 1. The four audiovisual displays where the sound’s onset was simultaneous with the discs’ touch were selected and concatenated in pairs. Four pairs were created: a ramped-tone display followed by a damped-tone display (and the time-reversed pair); a ramped-noise display followed by a damped-noise display (and the time-reversed pair). Participants observed three random sequences of display pairs. Within each sequence, each display pair was presented four times for a total of 16 pairs per sequence. The participant was asked to evaluate the sounds while looking at the motion display. After each display pair she or he was asked to report which of the two sounds was perceived as longer in duration (first sequence), as louder (second sequence), and as changing in loudness (third sequence). The order of the three questions was selected randomly for each participant.

Results

Reaction time. Participants’ responses to catch trials, reactions smaller than 120 ms (i.e., anticipations), and reactions greater than 500 ms (i.e., delayed responses) were excluded from the analysis. Excluded data were \(~3\%\) of the data collected. Reaction times were subjected to a 2 (sound timbres) × 2 (sound envelopes) ANOVA. Participants’ reactions were independent from the timbre of the sound, \(F(1, 13) = 2.58, p > .05\), and were faster for damped (\(M = 237.7\) ms, \(SD = 40.3\) ms) than for ramped (\(M = 252.6\) ms, \(SD = 35.7\) ms) sounds, \(F(1, 14) = 44.34, p < .0001\). In brief, the reaction to ramped sounds was, on average, 15 ms slower than that to damped sounds.

Perceptual evaluation of sounds. The proportion of ramped longer/louder/changing-in-loudness-more-than-damped responses was computed from individual data. These proportions were, respectively, 92\% (\(SD = 11\%\)), 35\% (\(SD = 27\%\)), and 95\% (\(SD = 11\%\)). Three one-sample \(t\) tests (\(H_0 = 50\%\)) revealed that ramped displays were perceived as longer, \(t(11) = 13.14, p = .0001\), and as more changing in loudness, \(t(11) = 12.13, p = .0001\), than damped sounds. However, ramped and damped sounds were judged as similarly loud, \(t(11) = -1.84, p < .05\).

Experiment 4

Method

Fourteen new participants with normal or corrected-to-normal vision and hearing participated in the experiment. They performed the two tasks of Experiment 3 (i.e., reaction time and perceptual evaluation of sounds), but the sounds were those used in Experiment 2.

Results

Reaction time. Responses to catch trials, anticipations, and delayed responses were excluded from the data analysis (\(~5\%\) of

\[^4\] We could not ask the original Neuhoff (1998, 2001) question (i.e., “Which of the two sounds has the greater change in loudness?”) because the sounds were too short in duration and the participants hardly perceived that damped sounds changed in sound pressure level over time. Damped sounds were perceived as rather impulsive. In contrast, participants did perceive that ramped rounds changed in sound pressure level over time.
data collected). The remaining data were subjected to a 2 (sound timbres) × 2 (sound envelopes) ANOVA. Mean reaction times were independent from the timbre of the sound (F < 1) and were identical for damped (M = 259.5 ms, SD = 39.4 ms) and ramped (M = 259.2 ms, SD = 49.8 ms) sounds (F < 1, statistical power p = .114).

Perceptual evaluation of sounds. The proportion of ramped longer, louder, changing–in-loudness-more-than-damped responses was computed from individual data. Observed proportions were, respectively, 91% (SD = 7%), 97% (SD = 3%), and 94% (SD = 6%). Ramp ed sounds were judged as longer, t(13) = 22.27, p < .0001, louder, t(13) = 57.51, p < .0001, and more changing in loudness, t(13) = 28.85, p < .0001, than damped sounds.

Discussion of Experiments 3 and 4

Experiments 3 and 4 were designed to control for the results of Experiments 1 and 2 and to assess the validity of the assumptions supporting the experiments. The results of Experiment 3 do not provide a net interpretation of the results of Experiment 1. They suggest that, overall, the ramped sounds of Experiment 1 did not capture attention as fast as the damped sounds of the same experiment. Moreover, they do not fully support the hypothesis that ramped sounds were perceptually more salient than damped sounds. In contrast, the results of Experiment 4 offer a net interpretation of the results gathered in Experiment 2. Ramped and damped sounds used in Experiment 2 captured attention equally fast. However, ramped sounds were perceptually more salient than damped sounds. In Experiment 4, in fact, amplified ramped sounds were judged invariably as longer and louder, and were characterized by a greater change in loudness than damped sounds. In other words, in Experiment 2, ramped and damped sounds were (at least) equally capable of subtracting attentional resources. However, in that experiment, damped displays gathered 80% of bounce responses, whereas ramped displays gathered 50% of bounce responses. Therefore, the attentional explanation by itself is not sufficient to explain the whole ABE.

Experiment 5

In Experiments 3 and 4 we investigated whether ramped and damped sounds have a similar capability to subtract attentional resources. In Experiments 5 and 6 we investigated whether ramped and damped sounds have a similar capability to change perception because of their similarity (or lack of similarity) to real sound events. We cannot exclude the fact that in Experiments 1 (in particular) and 2, ramped displays did not gather as many bounce responses as damped displays because ramped sounds were perceived as streaming-like sounds. If this was the case, ramped sounds would be unsuitable to induce the ABE. The acoustic pattern of ramped sounds indeed simulates the acoustic pattern of real events. For example, it simulates the acoustic pattern resulting from the motion of an approaching sound source (Neuhoff, 1998, 2001). But it is also the first half of a Doppler acoustic array (McBeath & Neuhoff, 2002; Neuhoff & McBeath, 1996), an event that can be thought of as streaming-like.

Let us assume that ramped sounds are perceived as streaming-like sounds. In Experiment 1, the number of streaming responses to silent displays was nearly at ceiling (i.e., more than 88% of streaming responses); therefore, the addition of the ramped sound to the discs’ motion could not add a further increment to the number of streaming responses. The results of Experiment 2 suggest that ramped sounds should not favor the perception of streaming because, in that experiment, ramped displays gathered fewer streaming responses than silent displays did. However, the ramped and damped sounds used in Experiment 2 were different (in average sound pressure level as well as in change in sound pressure level) from the ramped and damped sounds used in Experiment 1 (see Figure 1).

Experiments 5 and 6 were designed to find out whether ramped sounds could be perceived as streaming-like sounds and promote the perception of streaming. The rationale driving both experiments is the following. If ramped sounds are perceived as streaming-like sounds, they should resolve in favor of streaming the perception of an ambiguous display (i.e., a silent display that evokes an equal number of streaming and bouncing responses). The silent display used in Experiments 1 and 2 does not possess this characteristic because the streaming response largely prevails over the bouncing one. In Experiments 5 and 6 the discs’ motion was of two types. The first was identical to that used in the previous experiments. The second was similar; however, in this new motion display discs never overlapped completely (see Figure 3).

Brief discontinuities in the discs’ motion are known to increase the number of bounce responses to silent displays (Remijn & Ito, 2007; Sekuler & Sekuler, 1999). In the new motion display, discs started their motion, overlapped for a maximum of 70% of their diameter, and returned to their starting position (see Figure 3). We found this motion pattern particularly suitable for testing whether ramped sounds are streaming-like sounds: A pilot experiment revealed that the proportion of bounce responses to a silent version of this display was about 50%.

In Experiments 5 and 6 participants performed the tasks of Experiments 1 and 2. The discs’ motion could be either identical to that used in the previous experiments (i.e., complete overlap [CO]) or like that just described (i.e., partial overlap [PO]). In Experiment 5, CO and PO were accompanied by the ramped and damped tones used in Experiment 1. In Experiment 6, the ramped and damped tones used in Experiment 2 were used for the sounds.

Method

Twelve new participants with normal or corrected-to-normal vision and hearing participated in the experiment. The apparatus was identical to that used previously. In the current experiment, discs had two different motion patterns. The discs’ motion was either identical to that used in the previous experiments (i.e., CO display) or different (i.e., PO display). In the PO display, discs overlapped for a maximum of 70% of their diameter and then returned to their starting position. The PO display was obtained by

---

4 This hypothesis has been advanced by the attendants of all talks and conferences where we presented the current study, as well as by reviewers.
removing three frames from the original CO display. The 70% overlap was chosen on the basis of a pilot experiment. This overlap was evoked in fact an ambiguous percept with bounce and streaming responses almost equally occurring. Moreover, this display could not be easily discriminated from a CO display. In a pilot experiment we tested whether participants could tell the difference between PO and CO. We set a 2I-2AFC experiment. PO and CO displays were presented in pairs, and after each pair participants were asked to report which the CO display was. Participants’ accuracy was 79%—that is, just above a 75% discrimination threshold.

In the current experiment, PO and CO displays were accompanied by either the ramped (or damped) tone of Experiment 1, or by no sound, for a total of six different stimuli. Stimulation were presented 30 times each in random order. In all audiovisual displays, tones were switched on simultaneously with the discs’ touch. The participant’s task was identical to the task in Experiments 1 and 2.

Results

The proportions of bounce responses were calculated for each stimulus and participants. Proportions were subjected to a 2 (PO, CO display) × 3 (no sound, ramped sound, damped sound) ANOVA. PO displays gathered a higher number of bounce responses than CO displays, \(F(1, 11) = 42.02, p < .0001\). Moreover, the number of bounce responses gathered by silent, ramped, and damped displays was different, \(F(2, 22) = 25.89, p < .0001\). The two-way interaction was not significant, \(F(1, 11) = 25.32, p < .0001\). Two planned comparisons showed that ramped displays returned a number of bounce responses identical to silent displays \((F < 1)\), whereas damped displays gathered a higher number of bounce responses, \(F(1, 11) = 25.32, p < .0001\). Moreover, a one-sample \(t\) test \((H_0 = 50\%\) revealed that the silent PO display was ambiguous \((t < 1)\)—that is, it gathered an identical number of streaming and bouncing responses (see Figure 4).

Experiment 6

Method

Fourteen new participants with normal or corrected-to-normal vision and hearing participated in the experiment. They were all naive to the purpose of the study. Participants performed the task of Experiment 5; however, the sounds of the audiovisual displays were the tones used in Experiment 2.

Results

The proportions of bounce responses were subjected to a 2 (PO, CO display) × 3 (no sound, ramped sound, damped sound) ANOVA. PO displays collected more bounce responses than CO displays, \(F(1, 13) = 27.79, p < .0001\). Moreover, ramped, damped, and silent displays gathered a different number of bounce responses, \(F(2, 26) = 11.57, p < .0001\). The two-way interaction was not significant, \(F(2, 26) = 2.70, p > .05\). Two planned comparisons showed that ramped displays returned more bounce responses than silent displays both with PO, \(F(1, 13) = 6.10, p = .028\), and with CO motion, \(F(1, 13) = 6.28, p = .021\). Moreover, damped displays returned more bounce responses than ramped displays both with PO, \(F(1, 13) = 4.70, p = .049\), and with CO motion, \(F(1, 13) = 4.81, p = .047\). Finally, a one-sample \(t\) test \((H_0 = 50\%)\) revealed that the silent PO display was ambiguous \((t < 1)\)—that is, it gathered an identical number of streaming and bouncing responses (see Figure 4).

Discussion of Experiments 5 and 6

Experiments 5 and 6 investigated whether ramped sounds were perceived as streaming-like sounds—therefore, sounds intrinsically unsuitable to induce the ABE. We created an ambiguous display (i.e., PO) and investigated whether the addition of a ramped sound to this motion display increased the number of streaming responses in comparison to a silent version of it. In both Experiments 5 and 6 we failed to observe this result. In conclusion, ramped sounds do not favor the perception of streaming.

Experiments 5 and 6 show further results. First, they replicate findings of Experiment 1 and 2, with damped sounds that induce the ABE and ramped sounds that do not (Experiment 5) or do not as much (Experiment 6). Moreover, they confirm the literature findings, namely, that brief discontinuities in the discs’ motion increase the number of bounce responses observed with silent displays (e.g., Sekuler & Sekuler, 1999). The increment in the number of bounce responses observed passing from CO to PO displays can be easily explained. In PO displays discs never overlap completely, and this is compatible with a real bouncing event of solid objects. In PO displays discs overlap partially, such as two tennis balls hitting each other at high velocity: the balls would approach, touch, and compress one into the other, thus generating a two-dimensional percept of partial overlap. Because the discs never overlap completely, the visual system may be unwilling to interpret the PO as exclusively streaming.

Finally, in both Experiments 5 and 6, the ABE (i.e., CO display with the addition of a damped sound) was smaller than that observed in Experiments 1 and 2 (i.e., ~80% of bounce responses in Experiments 1 and 2 compared with ~50% bounce responses in

\[6\] Because of the three-frames subtraction, the duration of the PO display was 30-ms shorter than the original CO display (i.e., 1.370 ms). However, it is very unlikely that participants noticed such a difference: The difference in duration of the two displays is smaller than the duration discrimination threshold for visual events of such a duration (Westheimer, 1999).
Experiments 5 and 6). This reduction is certainly contextual: In Experiments 5 and 6 there were bounces (e.g., PO with the addition of a damped sound) that were perceived as more bouncing-like than the original ABE (i.e., CO with the addition of a damped sound).

General Discussion

In the current study we investigated the ABE (Sekuler et al., 1997) by manipulating the acoustic part of the stimulation. We took the streaming/bouncing display and added to it either a sound increasing in amplitude over time (i.e., ramped sound) or the same sound played backward in time (i.e., damped sound). These sounds were selected because they could disentangle the double origin of the ABE: the attentional one and the perceptual one. An effect of ramped sounds (and not of damped ones) should reveal the role of attention, whereas an effect of damped sounds (and not of ramped ones) should reveal the role of perception.

In the experiments we performed, we observed the ABE more often and more strongly with impact-similar (i.e., damped) sounds than with impact-dissimilar (i.e., ramped) sounds. The latter sounds either did not induce the ABE at all (Experiments 1 and 5) or, at maximum, resulted in equally occurring streaming and bouncing responses (Experiment 2). In particular, in Experiments 2 and 6, displays with damped sounds gathered more bounce responses than displays with ramped sounds, even though (a) the two sounds did not differ in their speed to capture the viewer’s attention and (b) ramped sounds were perceived as more salient than damped sounds (i.e., louder, longer, and characterized by a change in loudness; Experiment 4). Moreover, we investigated whether ramped sounds were perceived as streaming-like sounds and, thus, whether these sounds were intrinsically unsuitable to induce the ABE. The results of Experiments 5 and 6 failed to provide evidence in support of this hypothesis.

Overall, the results suggest that the effect of attention alone cannot explain the ABE. Stimulations that subtract attention from the motion display do increase the occurrence of bounce responses in comparison with conditions where the observer can pay full attention to it (e.g., Kawabe & Miura, 2006; Watanabe, 2001; Watanabe & Shimojo, 1998). However, here the reduction of attention from the motion display produced a maximum rate of bounce responses of ~50% (Figure 2, Experiment 2, ramped displays). This is the result one would expect on the basis of the attentional explanation: If the attentional resources that promote the streaming percept are unavailable, we would expect responses to be distributed equally between the two possible perceptual solutions—that is, streaming and bouncing.7

We believe that an explanation combining both attention and perception explains the origin of the ABE. In the current study, the reduction of attention from the discs’ motion inhibited the streaming response and made both responses (i.e., streaming and bouncing) occur equally (e.g., Figure 2, Experiment 2, ramped displays). However, impact-similar sounds (only) were able to turn the response into a stable bounce response (Figure 2, Experiment 2, damped displays). In our interpretation, the ABE has a double, additive origin. The first component is attentional: Sounds with a high intensity onset (e.g., both ramped and damped sounds of Experiment 2) subtract attentional resources and inhibit the integration of the local motion signals that promote the perception of streaming. This subtraction of attentional resources gives a first rise to the rate of bounce responses observed in audiovisual displays. The second component is perceptual and gives a further rise to the number of bounce responses. In the current study, the effect of this component could be observed with damped sounds but not with ramped ones. Damped sounds only were characterized by a waveform that was compatible with one of the two possible perceptual solutions of the motion display: the acoustic array of an elastic impact. The way this second component works is, however, unclear. It can either bias the observer’s response toward the bounce response or, alternatively, shape the multisensory image toward a percept that is maximally consistent across both modalities (e.g., Ernst & Banks, 2002).

7 We would like to add a further note on the possible distracting role of the sound. Recent evidence shows that the presentation of a sound stimulus does not necessarily distract from (thus, worsen) the performance of an observer engaged in a visual task. On the contrary, isolated, brief sounds presented simultaneously with visual targets have been found to improve both target detection and target discrimination (e.g., Lippert, Logothetis, & Kayser, 2007; Vroomen & de Gelder, 2000).
References


Received February 21, 2007
Revision received May 13, 2008
Accepted May 22, 2008