Research Article

WHEN SOUND AFFECTS VISION: Effects of Auditory Grouping on Visual Motion Perception

Katsumi Watanabe and Shinsuke Shimojo
California Institute of Technology-Pasadena

Abstract—*Two identical visual targets moving across each other can be perceived either to bounce off or to stream through each other. A brief sound at the moment the targets coincide biases perception toward bouncing. We found that this bounce-inducing effect was attenuated when other identical sounds (auditory flankers) were presented 300 ms before and after the simultaneous sound. The attenuation occurred only when the simultaneous sound and auditory flankers had similar acoustic characteristics and the simultaneous sound was not salient. These results suggest that there is an aspect of auditory-grouping (saliency-assigning) processes that is context-sensitive and can be utilized by the visual system for solving ambiguity. Furthermore, control experiments revealed that such auditory context did not affect the perceptual qualities of the simultaneous sound. Because the attenuation effect is not manifest in the perception of acoustic characteristics of individual sound elements, we conclude that it is a genuine cross-modal effect.*

Although there are numerous examples of visual influence on auditory perception (for reviews, see Calvert, Brammer, & Iversen, 1998; Driver & Spence, 1998), such as the McGurk effect (McGurk & MacDonald, 1976) and the ventriloquism effect (Howard & Templeton, 1966), it is hard to find a clear example of auditory influence on visual perception. R. Sekuler, Sekuler, and Lau (1997), however, recently devised a simple method to show a compelling effect of sound on visual perception. The basic idea is to introduce visual ambiguity, the solution of which is influenced by audiovisual interaction.

In a two-dimensional display, two identical visual targets moving across each other can be perceived either to bounce off or to stream through each other (Metzger, 1934; Michotte, 1946/1963). Despite the ambiguous nature of the visual stimulus, observers normally show a strong bias to see the streaming percept (Bertenthal, Banton, & Bradbury, 1993; Goldberg & Pomerantz, 1982; A.B. Sekuler & Sekuler, 1999; Watanabe & Shimojo, 1998, 2001). However, various factors have been reported to increase the relative frequency of the bouncing percept. These factors include a momentary pause of the targets (Bertenthal et al., 1993) and attentional distraction (Watanabe & Shimojo, 1998) at the timing of visual coincidence. Most intriguingly, R. Sekuler et al. (1997) showed that a brief sound, presented at the moment the targets coincide (hereafter referred to as the simultaneous sound), biases perception toward bouncing, and suggested that cross-modal interaction is involved in solving the streaming/bouncing motion ambiguity. This simple paradigm provided us with a chance to investigate the unexplored topic of auditory influence on visual event perception.

In this article, we show that the bounce-inducing effect is reduced when the simultaneous sound is preceded and followed by other identical sounds (auditory flankers) (Experiment 1), and that the attenuation of the bounce-inducing effect depends on auditory context and the saliency of the simultaneous sound (Experiments 2 and 3). Furthermore, we demonstrate that this auditory context does not alter perceptual properties of the simultaneous sound itself (Experiments 4–6).

Address correspondence to Katsumi Watanabe, Computation and Neural Systems, California Institute of Technology, Mail Code 139-74, Pasadena, CA 91125; e-mail: kw@caltech.edu.

EXPERIMENT 1: ATTENUATING THE BOUNCE-INDUCING EFFECT

Method

Observers

Seven people (20–48 years of age; including one of the authors, K.W.) with normal or corrected-to-normal vision and hearing participated. Except for the author, all were naive as to the purpose motivating the study.

Stimuli

Visual stimuli were displayed on a computer monitor, controlled by a Silicon Graphics Indigo2 workstation, in a dark room. A slow frame rate (20 Hz) was used to ensure synchrony between visual and auditory stimuli. A black fixation cross (0.01 cd/m^2, 0.35° in visual angle) was displayed against a gray background (8.5 cd/m^2). Two black disks (0.13° in diameter) appeared 1.47° above the fixation cross, initially separated by 3.27° (Fig. 1). Immediately after the onset, the disks moved laterally at 1.64°/s (frame-to-frame offset = 0.08°) toward each other, coincided, and continued moving until each reached the other's start point. An 1800-Hz tone burst was presented at (simultaneous-sound condition), 300 ms before (before-coincidence condition), or 300 ms after (after-coincidence condition) the visual coincidence, through a speaker in the computer. Additionally, no sound or three consecutive sounds were presented on some trials (no-sound and three-sound conditions, respectively). The duration of each single sound was 3 ms, and the sound pressure level was 58 dB at the observer's ear. There was background noise of about 53 dB.

Procedure

Observers viewed the display binocularly from a distance of 80 cm and reported their perception (i.e., streaming or bouncing) by pressing the appropriate button on the computer mouse. There were 20 trials in each of the five conditions, and the order of trials was random.

Results and Discussion

Sounds presented 300 ms before and 300 ms after the visual coincidence did not induce the bouncing percept (Fig. 2). Analyses of variance (ANOVAs) comparing the no-sound condition with the before-coincidence and after-coincidence conditions were not significant, $F(1, 6) = 2.95$, $p = .11$, and $F(1, 6) = 0$, $p = 1$, respectively. In contrast, the single sound presented at the visual coincidence enhanced

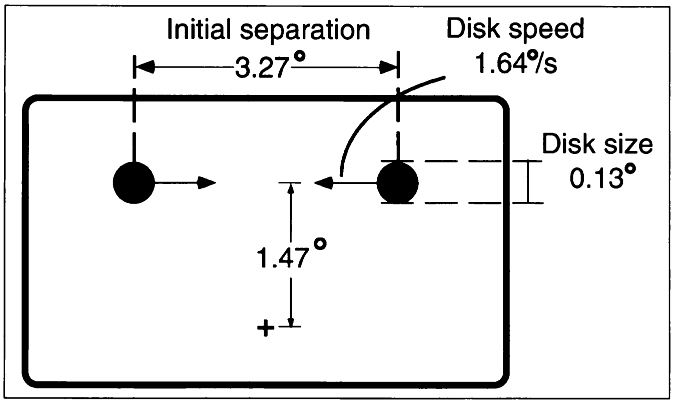

Fig. 1. Configuration of the visual stimulus used in the experiments. The background of the actual display was gray.

perception of bouncing, $F(1, 6) = 45.36$, $p < .001$ (no-sound condition vs. simultaneous-sound condition). When the simultaneous sound was flanked by other identical sounds, the bounce-inducing effect was present, $F(1, 6) = 4.39$, $p < .06$ (no-sound condition vs. three-sound condition), but attenuated significantly compared with the simultaneous-sound condition, $F(1, 6) = 5.88$, $p < .05$ (three-sound condition vs. simultaneous-sound condition).

This pattern of results indicates that the auditory flankers reduce the saliency of the simultaneous sound, which is critical for audiovisual interaction. That is, a single sound may be more distinct than a sound with auditory flankers. We hypothesize that repetitive sound elements may be spontaneously grouped into an auditory event because of their similar acoustic properties, and that, once grouped, the flanked sound element may lose its saliency for audiovisual interaction.

It is well known that perception of auditory elements is largely influenced by auditory context. For example, the tonal structure of a series of pure tones and its presentation rate affect the listener's ability to identify or discriminate the order of particular sound elements within the series (e.g., Bregman & Campbell, 1971; Miller & Heise, 1950). Sound elements with similar acoustic characteristics tend to group together and segregate from acoustically distant sound elements, and this grouping-segregation effect increases as the presentation rate increases (auditory streaming; Bregman, 1990; McAdams, 1984). Our account for the attenuation effect qualitatively resembles these auditory phenomena, although it is measured by cross-modal interaction in our display, and the integration time seems to be longer (about 300 ms and probably more; Watanabe & Shimojo, 1999).

In the next two experiments, we intended to examine the hypothesis that auditory grouping and saliency affect the attenuation of the

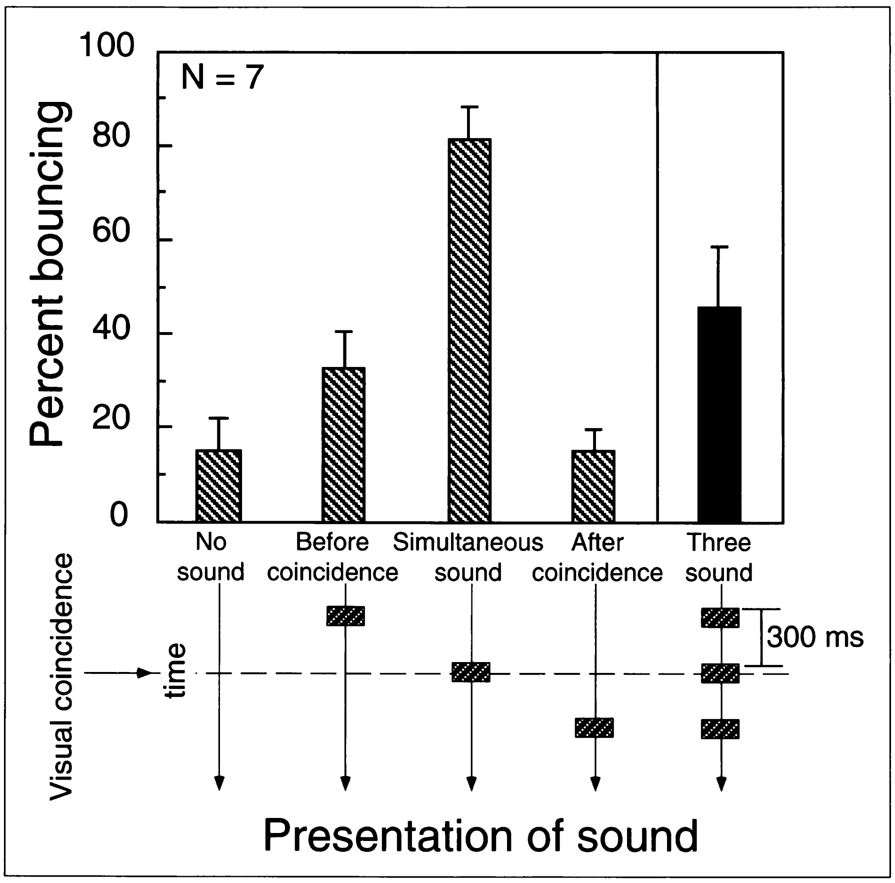

Fig. 2. Mean percentage of trials on which observers reported the bouncing percept in Experiment 1. Standard errors are shown. The schematic below the graph depicts the timing of the sound in each condition, with the rectangles representing the sounds.

bounce-inducing effect that we observed. We speculated that auditory grouping (and the attenuation effect) depends on similarity between the simultaneous sound and auditory flankers. Thus, if the simultaneous sound and auditory flankers had different acoustic characteristics, the simultaneous sound would "pop out" and the bounce-inducing effect would recover.

EXPERIMENT 2: RECOVERING THE BOUNCE-INDUCING EFFECT BY USING DIFFERENT SOUND FREQUENCIES

Method

Observers

The 7 observers from Experiment 1 and 2 additional naive observers participated.

Stimuli

The visual stimulus was identical to that in Experiment 1. There were two kinds of auditory conditions (Fig. 3). First, in the single-sound condition, a single sound was presented at the visual coincidence. The sound had the same acoustic profile as in Experiment 1 except that its frequency was either 900, 1800, or 2700 Hz. Trials in which no sound was presented (no-sound condition) were a control for the single-sound trials. Second, in the embedded-sound condition, seven sounds were presented successively with a sound onset asynchrony of 300 ms. The fourth sound (the embedded sound) was always presented at the time of the visual coincidence. The frequency of the embedded sound varied as in the single-sound condition, but the frequency of the auditory flankers was fixed at 1800 Hz. Note that when the frequency of the embedded sound was 1800 Hz, all seven sounds had the same frequency. Trials in which the embedded sound was omitted (sound-off condition) were a control for the embedded-sound trials.

Procedure

The procedure was the same as that of Experiment 1. There were 20 trials for each of the eight sound conditions (900-, 1800-, and 2700-Hz single-sound conditions; 900-, 1800-, and 2700-Hz embedded-sound conditions; no-sound condition; and sound-off condition). The order of trials was random.

Results and Discussion

The results are summarized in Figure 4. All the single sounds significantly enhanced the bouncing percept compared with the no-sound condition, $F(3, 24) = 84.4$, $p < .001$. With the single sounds, the bounce-inducing effect did not change as a function of sound frequency, $F(2, 16) = 1.63$, $p = .22$. Comparison of the embedded-sound condition and the sound-off condition showed that the bounce-inducing effect was also present in the embedded-sound condition, $F(3, 24) = 46.48$, $p < .001$. However, when the embedded sound had the same frequency as the auditory flankers, the bounce-inducing effect was significantly attenuated, $F(1, 8) = 47.21$, $p < .001$ (single-sound condition vs. 1800-Hz, embedded-sound condition). This attenuation did not occur when different frequencies were used for the embedded sound, $F(1, 8) = 1.06$, $p = .32$ (single-sound condition vs. 900-Hz, embedded-sound condition) and $F(1, 8) = 0.63$, $p = .44$ (single-sound condition vs. 2700-Hz, embedded-sound condition). These results are consistent with the auditory-grouping account for the attenuation of the bounce-inducing effect, because the attenuation was observed only when the embedded sound had the same frequency as the auditory flankers.

The sound-off condition did not lead to enhancement of the bouncing percept, a finding consistent with the results of R. Sekuler et al. (1997). This confirms the hypothesis that the saliency, not the mere presence nor the surprising nature, of the simultaneous sound is crucial for the bounce-inducing effect. In Experiment 3, we examined the role of auditory grouping and saliency further by changing the inten-

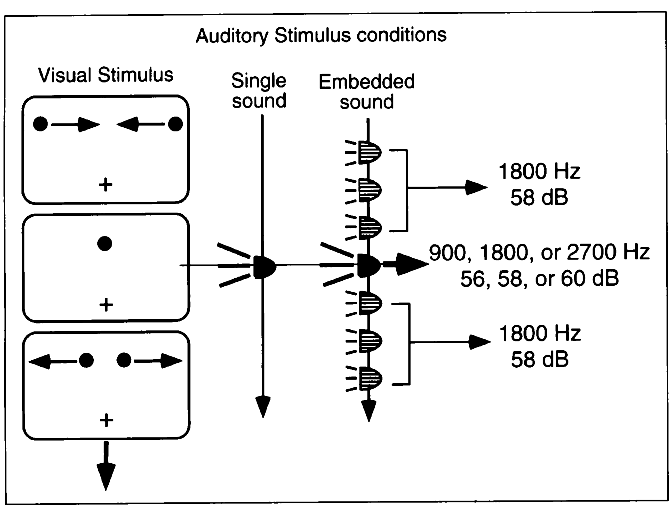

Fig. 3. Schematic diagram of the stimulus presentations in Experiments 2 and 3. In Experiment 2, the frequency of the single and embedded sounds was varied. In Experiment 3, the intensity of the single and embedded sounds was varied.

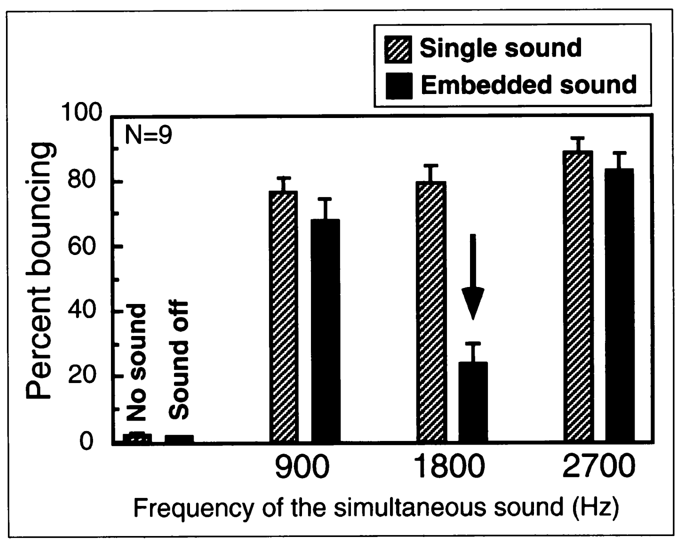

Fig. 4. Mean percentage of trials on which observers reported the bouncing percept in Experiment 2. The frequency of the auditory flankers was 1800 Hz. The arrow indicates the condition in which the attenuation effect was observed. Standard errors are shown.

sity of the embedded sound, as intensity differences are known to serve as important cues for segmenting auditory input into discrete perceptual events (e.g., Dowling, 1968; Schröger, Tervaniemi, Wolff, & Näätänen, 1996).

EXPERIMENT 3: RECOVERING THE BOUNCE-INDUCING EFFECT BY USING SOUND-INTENSITY DIFFERENCES

Method

Observers

The same 9 observers as in Experiment 2 participated.

Stimuli

The visual and auditory stimuli were almost the same as those used in Experiment 2. However, the frequency for all sounds was fixed at 1800 Hz, and the intensity of the single sound and embedded sound was varied (56, 58, or 60 dB; see Fig. 3). The intensity of the auditory flankers was 58 dB.

Procedure

The procedure was identical to that of Experiment 2.

Results and Discussion

The results are summarized in Figure 5. Again, all the single sounds produced the bounce-inducing effect, $F(3, 24) = 90.27$, $p < .001$. The bounce-inducing effect was still present with the embedded sounds, compared with the sound-off condition, $F(3, 24) = 45.02$, $p < .001$. However, the effect was not present in all the embedded-sound conditions. When the embedded sound had an intensity higher than that of the auditory flankers, the bounce-inducing effect was comparable to the effect with the single sound, $F(1, 8) = 2.49$, $p = .134$ (single-sound condition vs. 60-dB, embedded-sound condition). When the embedded sound had the same intensity as the auditory flankers or a lower intensity, the bounce-inducing effect was significantly reduced, $F(1, 8) = 47.02$, $p < .001$ (56-dB condition), and $F(1, 8) = 49.82$, $p < .001$ (58-dB condition). This absence of recovery with weaker sounds confirms that the bounce-inducing effect is a function of saliency of the simultaneous sound.

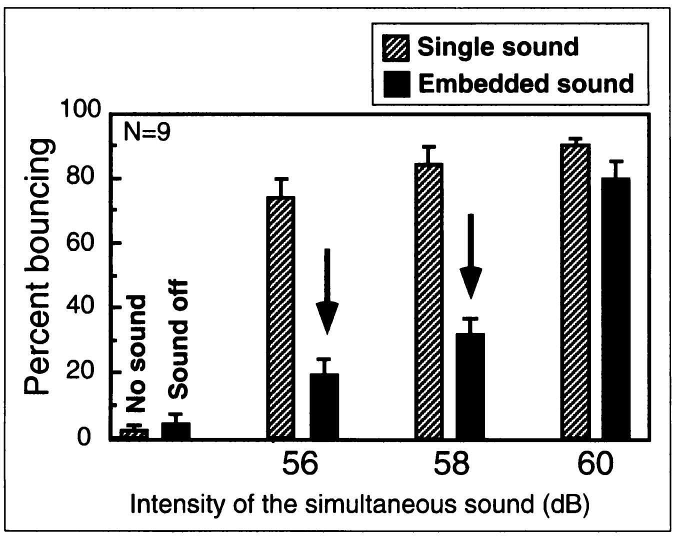

Fig. 5. Mean percentage of trials on which observers reported the bouncing percept in Experiment 3. The auditory flankers had an intensity of 58 dB. The arrows indicate the conditions in which the attenuation effect was observed. Standard errors are shown.

The bounce-inducing effect seemed to increase as a function of intensity. However, this dependency was not quite significant within the intensity range used, $F(2, 16) = 2.80$, $p = .08$. The small degree of intensity dependency appeared to exist even when the bounce-inducing effects were attenuated by the auditory flankers (56-dB vs. 58-dB embedded sounds), though the difference was not significant, $F(1, 8) = 2.74$, $p = .12$. Thus, it may be the case that both the auditory grouping with flankers and the intensity of the simultaneous sound itself are factors in determining the observer's perception.[1] These results of Experiment 3 may reflect the fact that in the natural environment, an increase in intensity against the background sound level may signal collision of objects, although an intensity decrease may not, because collision is almost always accompanied by an abrupt onset of sound.

The results of the first three experiments showed that the saliency of the simultaneous sound is critical for the bounce-inducing effect and that this saliency is determined through the auditory-grouping process. The next question we asked was, in what way is a sound salient? Did the observers hear the flanked sound as acoustically different from the single sound? Interestingly, when interviewed after Experiment 1, none of our observers reported that the simultaneous sound in the three-sound condition was heard differently from the simultaneous sound in the simultaneous-sound condition. It thus appeared that the flankers' effect of attenuating the bouncing percept, which was clearly conspicuous as a cross-modal effect, was not reflected in the conscious perception of acoustic characteristics of the sound elements.

Moreover, it is unlikely that the attenuation effect was caused by conventional auditory masking, which alters detection thresholds or the perceived characteristics of individual sounds (e.g., Lüscher & Zwislocki, 1947; Stevens & Davis, 1938). Auditory masking is an unlikely explanation because the effective range for auditory masking is no more than 250 ms proactively and retroactively (e.g., Massaro, 1970, 1975; Massaro, Cohen, & Idson, 1976; Viemeister & Plack, 1993), whereas we have observed significant attenuation of the bounce-inducing effect with much longer intervals (up to about 500 ms; Watanabe & Shimojo, 1999).

Nonetheless, there was the possibility that some audiovisual interactions might have changed the perceived acoustic properties of the simultaneous sound with auditory flankers, but that such changes escaped the spontaneous reports after the experiment. Our preliminary study had shown that although the pitch and timbre of the single simultaneous sound do not substantially alter the bounce-inducing effect, the effect becomes smaller as the sound intensity becomes lower,

1. A preliminary observation also confirmed the importance of saliency. When a long continuous sound was presented instead of a train of sounds, a situation that should reduce the saliency of the simultaneous sound maximally, the bounce-inducing effect was reduced down to the same level as in the no-sound condition, irrespective of sound intensity. This implies that the grouping factor has stronger influence than sound intensity on the audiovisual effect, with the parameters employed in the present study.

the sound duration becomes longer, or the asynchrony between the sound and the visual coincidence becomes larger (Watanabe, Scheier, Lewkowicz, & Shimojo, 1999). Therefore, we set out to examine whether the auditory flankers would alter these aspects of the simultaneous sound in the presence of the visual display. In the following control experiments, the visual collision was always displayed, even if it was irrelevant for the required task, and the 7 observers from Experiment 1 participated. We reasoned that if the attenuation of the bounce-inducing effect is a genuine cross-modal effect and not a side effect due to changes in auditory quality, the auditory flankers would not change the auditory perception of the simultaneous sound.

EXPERIMENT 4: DO FLANKERS AFFECT PERCEIVED LOUDNESS?

Experiment 4 was conducted to determine whether the simultaneous sound might be perceived as being less audible when it was accompanied by auditory flankers (Fig. 6a). Observers were asked to make judgments about auditory stimuli, not about visual events.

Method

Experiment 4a (within condition)

The auditory stimulus in this experiment was almost identical to that in the three-sound condition of Experiment 1. However, the intensity of the simultaneous sound was varied from 56 dB to 60 dB (in 0.5-dB steps), while the intensity of the auditory flankers was kept constant (58 dB). Observers compared the loudness of the simultaneous sound with the loudness of the auditory flankers. There were 20 trials with each intensity of the simultaneous sound, and the order of trials was random.

Experiment 4b (between condition)

The auditory stimuli used in the three-sound condition and the simultaneous-sound condition were presented successively in random order. The interval between displays was about 1 s. As in Experiment 4a, the intensity of the simultaneous sound in the three-sound condition was varied. The intensity of the single simultaneous sound was 58 dB. Observers compared the loudness of the simultaneous sound in the three-sound condition with the loudness of the simultaneous sound presented alone. Twenty trials were repeated for each intensity.

Results

Figure 6b shows the results of Experiment 4. A two-way ANOVA revealed that there was no significant difference between the subexperiments, $F(1, 6) = 0.51$, $p = .479$. In Experiment 4a, when the intensity of the simultaneous sound was identical to that of the auditory flankers (i.e., 58 dB), the mean percentage of judgments that "the simultaneous sound was louder (compared with the auditory flankers)" was 48.6% ($SD = 14.64$). This is not significantly different from chance, $t(7) = -0.26$, $p = .81$, two-tailed. Likewise, for the 58-dB simultaneous sounds in Experiment 4b, the mean percentage of judgments that "the simultaneous sound was louder (compared with the single sound)" was 51.4% ($SD = 19.52$). This percentage does not differ from chance either, $t(7) = 0.19$, $p = .85$. Thus, we found no evidence for a perceived change in the loudness of the simultaneous sound when accompanied by the auditory flankers.

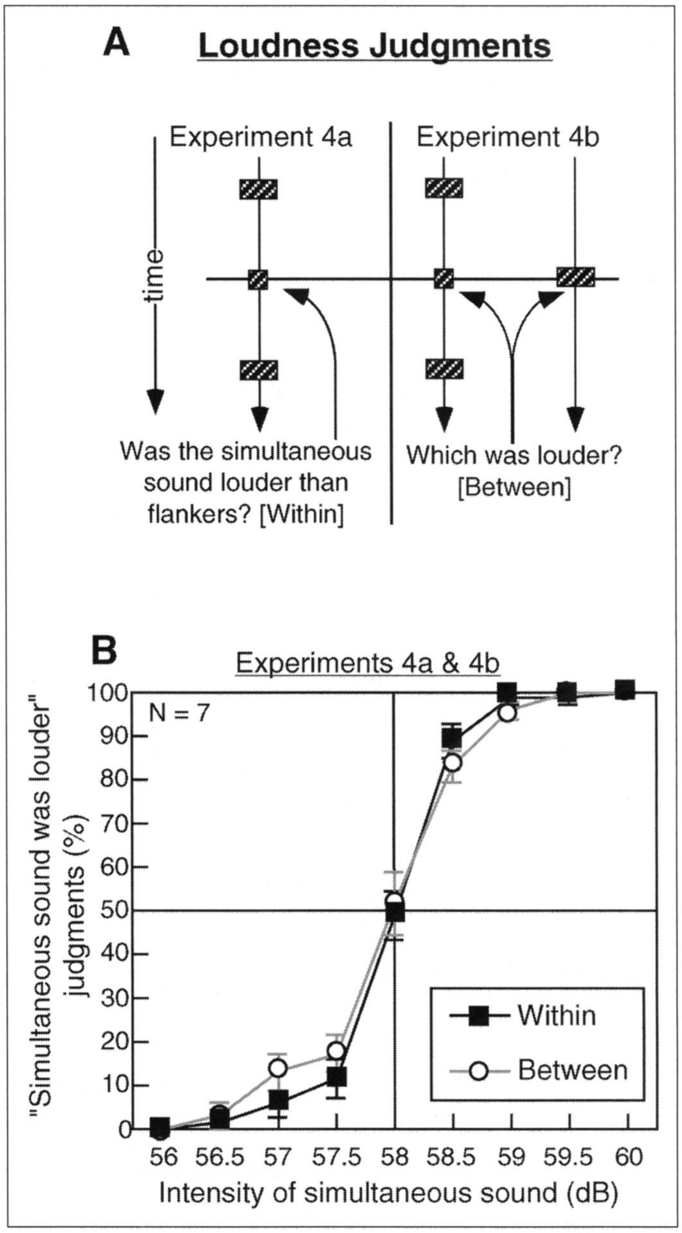

Fig. 6. Illustration of the task (a) and results (b) for Experiment 4. Observers judged the loudness of the simultaneous sound relative to auditory flankers (Experiment 4a) and relative to another single sound (Experiment 4b). The graph shows the percentage of judgments that the simultaneous sound was louder, as a function of the intensity of the simultaneous sound. The horizontal line shows chance performance. The vertical line indicates the value of the comparison sound. Standard errors are shown.

EXPERIMENT 5: DO FLANKERS AFFECT PERCEIVED DURATION?

Experiment 5 was conducted to determine whether the simultaneous sound might be perceived as lasting longer when it was accompanied by auditory flankers (Fig. 7a).

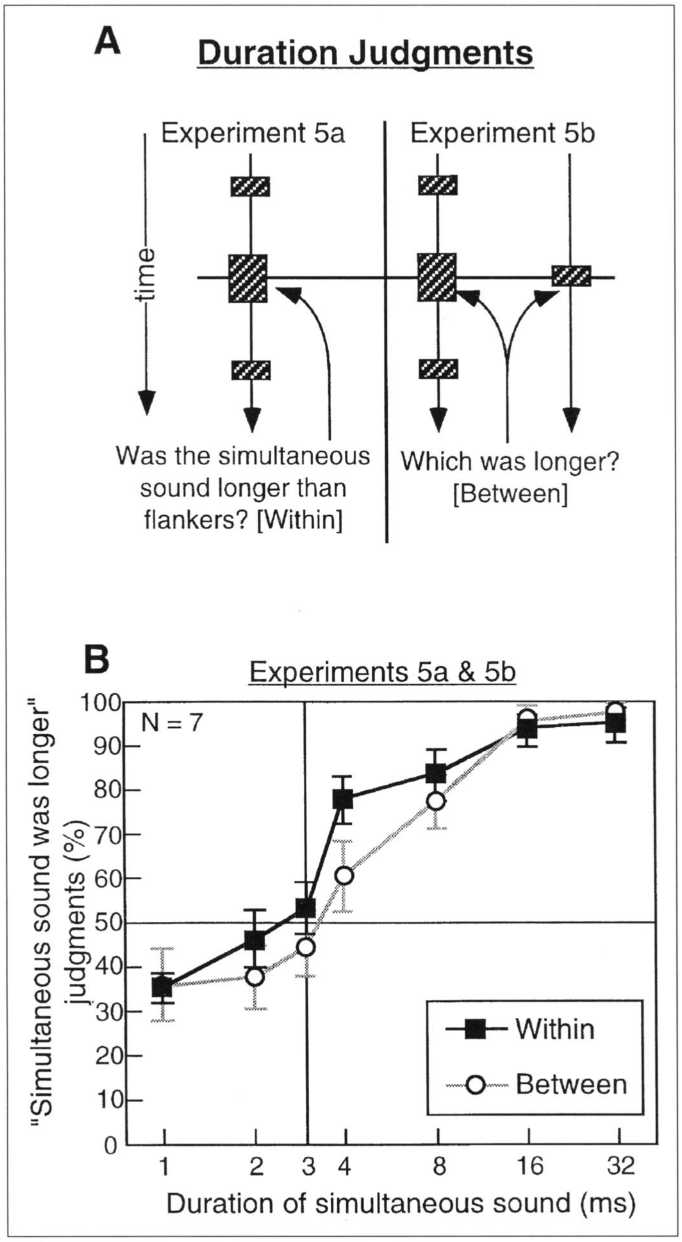

Fig. 7. Illustration of the task (a) and results (b) for Experiment 5. Observers judged the duration of the simultaneous sound relative to auditory flankers (Experiment 5a) and relative to another single sound (Experiment 5b). The graph shows the percentage of judgments that the simultaneous sound was longer, as a function of the log of the duration of the simultaneous sound. The horizontal line shows chance performance. The vertical line indicates the value of the comparison sound. Standard errors are shown.

Method

Experiment 5a (within condition)

The auditory stimulus and procedure were almost identical to those of Experiment 4a. However, the intensity of the simultaneous sound was kept at 58 dB, and the duration of the sound was varied from 1 ms to 32 ms. Observers reported whether the simultaneous sound was longer or shorter than the auditory flankers. There were 20 trials with each duration of the simultaneous sound, and the order of trials was random.

Experiment 5b (between condition)

The auditory stimulus and procedure were almost identical to those of Experiment 4b, but, as in Experiment 5a, the duration of the simultaneous sound in the three-sound condition was varied. Observers compared the duration of the simultaneous sound in the three-sound condition flanked by the other sounds with that of the simultaneous sound presented alone. For each duration, 20 trials were repeated.

Results

The results of Experiment 5 are shown in Figure 7b. The difference between the subexperiments failed to reach significance, $F(1, 6) = 0.77$, $p = .38$. In Experiment 5a, when the simultaneous sound and the auditory flankers had the same duration (3 ms), the observers reported that the simultaneous sound was longer than the auditory flankers in 52.9% of the trials on average (SD = 14.96). In Experiment 5b, the observers reported that the 3-ms simultaneous sound accompanied by flankers was longer than the 3-ms isolated sound in 44.3% of the trials on average (SD = 17.18). Neither of these percentages differs from chance, $t(7) = 0.51$, $p = .63$, for the within condition, and $t(7) = -0.88$, $p = .41$, for the between condition. Hence, it was not evident that the auditory flankers altered the perceived duration of the simultaneous sound.

EXPERIMENT 6: DO FLANKERS AFFECT PERCEIVED TIMING?

Experiment 6 was conducted to determine whether the auditory flankers induced uncertainty as to the timing of the sound relative to the visual coincidence (Fig. 8a).

Method

Experiment 6a (single-sound condition)

First, we presented the single sound around the time of the visual coincidence (from 200 ms before to 200 ms after, in 50-ms steps) and asked observers to judge whether the single sound appeared before or after the coincidence. There were 20 trials with each asynchrony, and the order of trials was random.

Experiment 6b (three-sound condition)

Next, we added auditory flankers to the stimulus in Experiment 6a. The flankers were presented 300 ms before and 300 ms after the visual coincidence. Observers judged whether the single sound (the second sound) appeared before or after the visual coincidence. Twenty trials were repeated for each asynchrony.

Results

Figure 8b shows the results of Experiment 6. The two psychometric curves, representing results with and without the auditory flankers, are similar. No difference was found between these two conditions, $F(1, 6) = 0.008$, $p = .93$. Thus, the perceived synchrony between the sound and the visual coincidence was not affected by the auditory flankers.

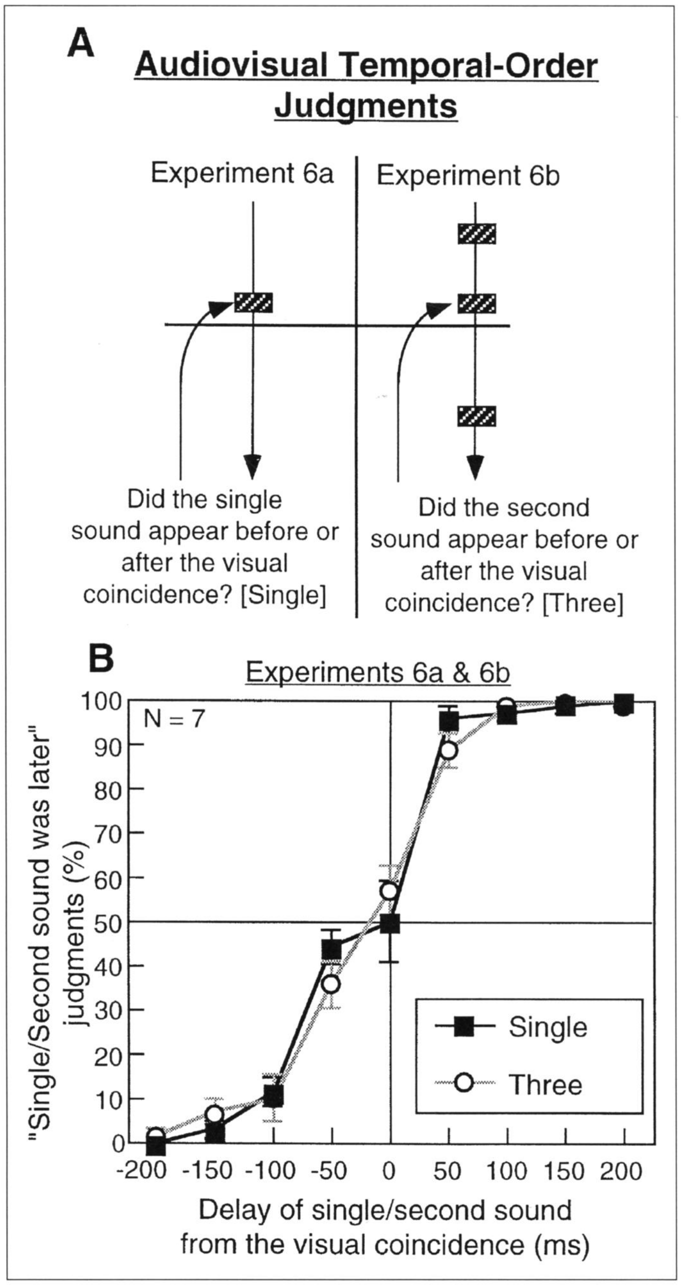

Fig. 8. Illustration of the task (a) and results (b) for Experiment 6. Observers judged whether a single sound (Experiment 6a) or the second sound in a three-sound sequence (Experiment 6b) came before or after the visual coincidence. The graph shows the percentage of judgments that the single or second sound was later than the visual coincidence, as a function of the delay of the sound from the visual coincidence. The horizontal line shows chance performance. The vertical line indicates the value of the comparison sound. Standard errors are shown.

Discussion

The three control experiments (Experiments 4–6) excluded the possibility that the auditory flankers altered the loudness, the duration, or the timing of the simultaneous sound, which in turn would affect the bounce-inducing effect. To reduce the bounce-inducing effect to the same level as in the three-sound condition of Experiment 1, according to our previous study (Watanabe et al., 1999), a single sound would have to be physically attenuated by more than 20 dB, lengthened by more than 200 ms, or desynchronized with the coincidence by about 150 ms. The changes in acoustic perception observed in Experiments 4 through 6, if any, were well below those levels of change. Therefore, we conclude that the attenuation of the bounce-inducing effect with auditory flankers is a cross-modal effect, not a side effect due to a change in auditory perception.

GENERAL DISCUSSION

We replicated the bounce-inducing effect (R. Sekuler et al., 1997); the single sound presented at the visual coincidence enhanced perception of bouncing. The new findings from this study are as follows: First, when the simultaneous sound is preceded and followed by other identical sounds (with 300-ms intervals), this same sound does not produce the bounce-inducing effect in the same magnitude (Experiment 1). Second, the attenuation of the bounce-inducing effect is a function of auditory context and saliency of the simultaneous sound (Experiments 2 and 3). Third, the auditory flankers do not affect the perception of acoustic properties of the simultaneous sound (Experiments 4–6). These results suggest that there is an aspect of auditory-grouping (saliency-assigning) processes that is context-sensitive and can be utilized by the visual system for solving ambiguity, but is nonetheless not manifest in the perception of the main acoustic characteristics of individual sound elements.

Cross-Modal Perception as "Inverse Physics" Problems

The dependence of the bounce-inducing effect on auditory context may reflect the likelihood of the combination of a physical collision and a transient sound in the natural environment. Geometrically, a point in a three-dimensional space can project onto a two-dimensional surface (such as the retina) in infinite ways, which results in the intrinsic ambiguity in vision and made Marr (1982) regard vision as a set of "inverse optics" problems. The facts that physical objects are seldom aligned at the same depth plane and that momentum makes objects move in the same direction as in the past may provide constraints to the visual system's interpretation. Consequently, given a two-dimensional image that two moving objects coincide, the visual system may be unwilling to interpret it as a physical collision in a three-dimensional world (leading to the dominance of the streaming percept; A.B. Sekuler & Sekuler, 1999; Sumi, 1995; Watanabe & Shimojo, 1998, 2001).

When physical objects actually collide, however, a transient sound often occurs (Gaver, 1993a, 1993b), and it is not probable that an independent event causes a sound at the very moment of a visual coincidence. So, detecting such an accidental audiovisual coupling, the brain should take the bouncing interpretation of the visual event (the bounce-inducing effect with a simultaneous sound). Nonetheless, if such a simultaneous sound and background sound have similar acoustic profiles, the physical cause of the sound should not be attributed to the visual coincidence (the attenuation effect). Despite the fact that the simultaneous sound alone implies a bouncing event, it would be highly accidental for the sound caused by the visual coincidence to resemble the background noise. In this case, the likelihood of auditory signals belonging together supersedes the likelihood of one element from that group being caused by a collision.

Although these operations can be regarded as a set of complex heuristics, they are functionally akin to rules used in Bayesian inference modeling in vision (i.e., solving inverse optics by computing conditional probabilities; Freeman, 1994; Knill, Kersten, & Yuille, 1996; Nakayama & Shimojo, 1992; Richards, 1988). The difference between visual and audiovisual perception is that the problem of audiovisual perception should be considered a problem of "inverse physics," not just inverse optics. The brain has to consider not only optics, but also physics and its projections to optics and acoustics. Consistent with the view that cross-modal perception involves computing conditional probabilities to solve problems of inverse physics, the bounce-inducing effect becomes less frequent but does not disappear altogether when there are auditory flankers, suggesting that the bounce-inducing effect and the attenuation effect are probabilistic.[2]

Dissociation Between Cross-Modal Saliency and Perception of Auditory Properties

Intriguingly, the processing of the auditory system does not result in the simultaneous sound with auditory flankers being perceptually different from the isolated sound. Hence, the bounce-inducing effect and its attenuation with auditory flankers should be considered a genuine audiovisual effect, not an audiovisual effect that results from auditory effects. We think that what is altered by auditory flankers may be the saliency of the simultaneous sound, and that the visual system or the audiovisual system may use such saliency and depend less on the perception of individual sound elements when solving motion ambiguity. The present study thus provides a clear example of dissociation between functional saliency and perceptual awareness. In other words, some cross-modal effects can occur without changing perceptual elements in the critical modality.

CONCLUSION

By using an ambiguous audiovisual display, the present study demonstrated the influence of auditory context on visual motion perception and the dissociation between perceptual properties of sound and its saliency. Further investigations will provide new insights into interactions between auditory and visual perception and into how the brain in general combines numerous physical constraints within and between modalities to construct the unity of the perceptual world.

Acknowledgments—We thank Allison Sekuler and Peter Grossenbacher for valuable comments. This work was supported by a grant from the Sloan Foundation Fellowship for Theoretical Neurobiology to K.W.

REFERENCES

Bertenthal, B.I., Banton, T., & Bradbury, A. (1993). Directional bias in the perception of translating patterns. *Perception*, *22*, 193–207.
Bregman, A.S. (1990). *Auditory scene analysis: The perceptual organization of sound.* Cambridge, MA: MIT Press.
Bregman, A.S., & Campbell, J. (1971). Primary auditory stream segregation and perception of order in rapid sequences of tones. *Journal of Experimental Psychology*, *89*, 244–249.
Calvert, G.A., Brammer, M.J., & Iversen, S.D. (1998). Crossmodal identification. *Trends in Cognitive Sciences*, *2*, 247–253.
Dowling, W.J. (1968). Rhythmic fission and perceptual organization. *Journal of the Acoustical Society of America*, *44*, 369.
Driver, J., & Spence, C. (1998). Attention and the cross modal construction of space. *Trends in Cognitive Sciences*, *2*, 254–262.
Freeman, W.T. (1994). The generic viewpoint assumption in a framework for visual perception. *Nature*, *368*, 542–545.
Gaver, W.W. (1993a). How in the world do we hear? Exploration in ecological acoustics. *Ecological Psychology*, *5*, 285–313.
Gaver, W.W. (1993b). What in the world do we hear? An ecological approach to auditory event perception. *Ecological Psychology*, *5*, 1–29.
Goldberg, D.M., & Pomerantz, J.R. (1982). Models of illusory pausing and sticking. *Journal of Experimental Psychology: Human Perception and Performance*, *8*, 547–561.
Howard, I.P., & Templeton, W.B. (1966). *Human spatial orientation.* London: Wiley.
Knill, D.C., Kersten, D., & Yuille, A. (1996). Introduction: A Bayesian formulation of visual perception. In D.C. Knill & W. Richards (Eds.), *Perception as Bayesian inference* (pp. 1–21). Cambridge, England: Cambridge University Press.
Lüscher, E., & Zwislocki, J. (1947). The decay of sensation and the reminder of adaptation after short pure tone impulses on the ear. *Acta Otolaryngologia*, *35*, 428–446.
Marr, D. (1982). Vision: A computational investigation into the human representation and processing of visual information. New York: W.H. Freeman.
Massaro, D.W. (1970). Retroactive interference in short-term recognition memory for pitch. *Journal of Experimental Psychology*, *83*, 32–39.
Massaro, D.W. (1975). Backward recognition masking. *Journal of the Acoustical Society of America*, *58*, 1059–1065.
Massaro, D.W., Cohen, M.C., & Idson, W.L. (1976). Recognition masking of auditory lateralization and pitch judgments. *Journal of the Acoustical Society of America*, *59*, 434–441.
McAdams, S. (1984). The auditory image: A metaphor for musical and psychological research on auditory organization. In W.P. Crozier & A.J. Chapman (Eds.), *Cognitive process in the perception of art* (pp. 289–323). Amsterdam: North-Holland.
McGurk, H., & MacDonald, J.W. (1976). Hearing lips and seeing voices. *Nature*, *264*, 746–748.
Metzger, W. (1934). Beobachtungen über phänomenale Identität. *Psychologische Forschung*, *19*, 1–60.
Michotte, A. (1963). *The perception of causality* (T.R. Miles & E. Miles, Trans.). New York: Basic Books. (Original work published 1946)
Miller, G.A., & Heise, G.A. (1950). The trill threshold. *Journal of the Acoustical Society of America*, *22*, 637–638.
Nakayama, K., & Shimojo, S. (1992). Experiencing and perceiving visual surfaces. *Science*, *257*, 1357–1363.
Richards, W. (1988). The approach. In W. Richards (Ed.), *Natural computation* (pp. 3–17). Cambridge, MA: MIT Press.
Schröger, E., Tervaniemi, M., Wolff, C., & Näätänen, R.N. (1996). Preattentive periodicity detection in auditory patterns as governed by time and intensity information. *Cognitive Brain Research*, *4*, 145–148.
Sekuler, A.B., & Sekuler, R. (1999). Collisions between moving visual targets: What controls alternative ways of seeing an ambiguous display? *Perception*, *28*, 415–432.
Sekuler, R., Sekuler, A.B., & Lau, R. (1997). Sound alters visual motion perception. *Nature*, *385*, 308.
Stevens, S.S., & Davis, H. (1938). *Hearing: Its psychology and physiology.* New York: Wiley.
Sumi, S. (1995). Bounce effect seen in opposite motions. *Japanese Psychological Research*, *37*, 195–200.
Viemeister, N.F., & Plack, C.J. (1993). Time analysis. In W.A. Yost, A.N. Popper, & R.R. Fay (Eds.), *Human psychophysics* (pp. 116–154). New York: Springer.
Watanabe, K., Scheier, C., Lewkowicz, D.J., & Shimojo, S. (1999, October). *Temporal window for audiovisual interaction revealed with an ambiguous motion display.* Poster presented at the Experimental Psychology Society Workshop on Crossmodal Attention and Multisensory Integration, Oxford, England.
Watanabe, K., & Shimojo, S. (1998). Attentional modulation in perception of visual motion events. *Perception*, *27*, 1041–1054.
Watanabe, K., & Shimojo S. (1999, March). *Temporal window for audiovisual integration revealed with bistable motion display.* Poster presented at the Joint Symposium on Frontiers in Computational Neuroscience, Elat, Israel.
Watanabe, K., & Shimojo, S. (2001). Postcoincidence trajectory duration affects motion event perception. *Perception & Psychophysics*, *63*, 16–28.

(Received 6/2/99; Revision accepted 6/6/00)

2. We (Watanabe & Shimojo, 2001) have shown that if observers are given the option to respond that they are unsure whether the disks stream or bounce, the frequency of such responses is quite low. This indicates that observers perceive either clear streaming or clear bouncing, though their perception can be probabilistic across trials.