Audiovisual integration: An investigation of the 'streaming-bouncing' phenomenon

Remijn, Gerard Bastiaan
Department of Visual Communication Design, Kyushu University

Ito, Hiroyuki
Department of Visual Communication Design, Kyushu University

Nakajima, Yoshitaka
Department of Acoustic Design, Kyushu University

出版情報：Journal of Physiological Anthropology and Applied Human Science. 23 (6), pp.243-247, 2004-11-01. 日本生理人類学会
バージョン：
権利関係：

Journal of
PHYSIOLOGICAL
ANTHROPOLOGY
and Applied Human Science

Audiovisual Integration: An Investigation of the 'Streaming-bouncing' Phenomenon

Gerard B. Remijn[1)], Hiroyuki Ito[1)] and Yoshitaka Nakajima[2)]

1) Department of Visual Communication Design, Kyushu University
2) Department of Acoustic Design, Kyushu University

Abstract Temporal aspects of the perceptual integration of audiovisual information were investigated by utilizing the visual 'streaming-bouncing' phenomenon. When two identical visual objects move towards each other, coincide, and then move away from each other, the objects can either be seen as streaming past one another or bouncing off each other. Although the streaming percept is dominant, the bouncing percept can be induced by presenting an auditory stimulus during the visual coincidence of the moving objects. Here we show that the bounce-inducing effect of the auditory stimulus is paramount when its onset and offset occur in temporal proximity of the onset and offset of the period of visual coincidence of the moving objects. When the duration of the auditory stimulus exceeded this period, visual bouncing disappears. Implications for a temporal window of audiovisual integration and the design of effective audiovisual warning signals are discussed. *J Physiol Anthropol Appl Human Sci 23(6): 243–247, 2004* http://www.jstage.jst.go.jp/browse/jpa

Keywords: audiovisual integration, visual motion perception, streaming, bouncing

Introduction

Many warning signals both inside and outside the home simultaneously stimulate the visual and auditory senses. The use of simultaneous auditory and visual warning information improves the localization of the signal (Doyle and Snowden, 2001), the speed with which we react to the signal (Schroger and Widmann, 1998), and is beneficial for those with a loss in hearing and/or eyesight (Alain and Woods, 1999). Most authors suggest that simultaneous audiovisual warning information somehow enhances and directs selective attention to the warning signals, compared with visual or auditory warning information presented alone. Some basic questions regarding the temporal aspects of the perceptual integration of audiovisual information, however, are still largely unanswered. In the present paper we wish to address a number of questions related to the temporal window of audiovisual integration, of possible importance for the design of warning signals in general.

The paradigm investigated here proceeds from a visual display consisting of two identical visual objects that move directly towards each other, then coincide, and move away from each other. Usually the objects are seen as streaming past one another, as they physically do, but occasionally they can be seen as bouncing off each other (Berthenthal et al., 1993; Sekuler and Sekuler, 1999). In this 'bouncing' percept, the objects perceptually move in a direction opposite to their original direction of movement when they coincide. Although the appearance of the 'streaming' percept is generally dominant, a 'bouncing' percept can be induced when an external, transient stimulus is presented along with the moving objects at the moment of their visual coincidence (Sekuler et al., 1997; Watanabe and Shimojo, 2001). The external stimulus can be a short visual flash, but also a brief tactile (somatosensory) or auditory stimulus, showing that cross-modal perceptual integration can be involved in solving the streaming/bouncing ambiguity.

Shimojo et al. (2001) have shown that the effectiveness with which an external stimulus can induce visual bouncing depends on the timing of the presentation of the stimulus relative to the coincidence of the moving objects. A brief sound, for example, induces visual bouncing most effectively when it is presented about 50 ms before the moving objects coincide. Even when more asynchrony exists between the sound and the moment of visual coincidence, bouncing can be induced. The data of Shimojo et al. (2001) showed that a rather large temporal window of audiovisual integration exists, from about 250 ms before the visual coincidence of the moving objects to 150 ms after the visual coincidence, in which a brief auditory stimulus can induce visual bouncing. However, since all research on the effect of auditory stimuli on visual bouncing was done with a single short click of 20 ms or less presented in the temporal window of audiovisual integration (Sekuler et al., 1997; Watanabe and Shimojo, 2001), we wish to further investigate the bounce-inducing effects of sounds in the present study.

The purpose of the present study is two-fold. First, we aimed to investigate whether the duration of a sound influences its bounce-inducing capacity in a streaming-bouncing display. In other words, can audiovisual integration also occur when a part of a sound (i.e., its onset or offset) falls outside of the suggested window of audiovisual integration? Second, we intended to measure the bounce-inducing capacity of auditory stimuli consisting of more than one sound presented within the suggested audiovisual window of integration. Watanabe and Shimojo (2001) have shown that a sound loses its bounce-inducing capacity when it is flanked by other sounds with the same duration and frequency, so that a repetitive stream of identical sounds can be perceived. In their experiments, however, an interonset-interval of 300 ms was used. As a result, the onset and offset of the string of sounds occurred outside the audiovisual window of integration. Moreover, only a single sound was presented in the audiovisual window of the integration at each presentation.

In our experiment, we used a streaming-bouncing display with bounce-inducing auditory stimuli with variations in their duration, and the number of sounds. In the multiple-sound stimuli, the sounds are separated by an interonset-interval that is considerably shorter than that used by Watanabe and Shimojo (2001). By using these auditory stimuli, more than one sound can fall within the suggested audiovisual window of integration. The onset of all auditory stimuli relative to the moment of exact coincidence of the moving objects is systematically changed in seven steps.

Method

Participants

Five male participants joined the experiment. They were students and researchers of visual or auditory perception. They had normal or corrected-to-normal vision and normal hearing.

Apparatus

The visual stimuli were generated and controlled by a personal computer (Sotec PV2240M) and displayed on a computer screen (Mitsubishi RDF 19X) in a dark room. Each participant viewed the display on the computer screen binocularly from a distance of 50 cm, with a chin-and-forehead rest steadying the participant's head, and the center of the computer screen at eye level. The auditory stimuli were generated by the same personal computer, and binaurally presented through headphones (Audio-Technica ATC-HA7USB) at a peak sound pressure level of 68 dBA on average (Fast-peak). The levels were measured with a sound level meter (Brüel and Kjær 2209), and an artificial ear (Brüel and Kjær 4153) mounted with a microphone (Aco 7013). The temporal tuning between the movement of the visual stimuli and the onset of the auditory stimuli was calibrated with the use of an oscilloscope (Agilent MegaZoom 54621A).

Stimuli

The stimuli consisted of two white squares (65.5 cd/m^2, 1.0°×1.0° in visual angle), presented on a visual display in which a gray fixation cross (19.7 cd/m^2, 0.84°×0.84° in visual angle) was set against a black background (1.38 cd/m^2). The squares appeared 2.51° in visual angle above the fixation cross, separated by 16.74° in visual angle. Initially, the squares were stationary, and then moved laterally towards each other with a speed of 7.55°/s, coincided, and then continued moving until each square reached the other's starting point. The squares were stationary for 1782 ms and in motion for 2218 ms, so that each stimulus lasted 4 s in total.

The movement of the visual objects was accompanied by one of five auditory stimuli. The first, second and third auditory stimulus each consisted of a single tone of 1500 Hz with a rise and fall time of 10 ms, with cosine-shaped ramps. The tone had a duration of 50 ms, 150 ms, and 250 ms, respectively, in these three stimuli. In the fourth auditory stimulus, the 50-ms tone was presented twice, separated by a silence of 50 ms. The total duration of this stimulus was 150 ms. In the fifth auditory stimulus, the 50-ms tone was presented three times, separated by two silences of 50 ms. The total duration of this stimulus was 250 ms. In a sixth control condition, the visual stimulus was presented without any sound.

Seven variations were made in the temporal tuning of the visual stimulus and the auditory stimuli. The auditory stimuli started 300, 200, or 100 ms before, simultaneously with (0 ms), or 100, 200, or 300 ms after the moment at which the moving objects exactly coincided on the visual display. Combining these seven variations in temporal tuning with the five auditory stimuli and the silent control condition, resulted in a total of 42 stimuli.

Procedure

The task of the participant was a 2AFC-task in which he had to judge whether the two white squares on the visual display were streaming or bouncing. Before starting the experiment, verbal and visual explanations of streaming and bouncing percepts were given to the participant. After the participant indicated that he understood the task, the experiment was started. The experiment consisted of five sessions. Each session consisted of 84 randomized trials (two randomized series of the 42 stimuli). Each stimulus was therefore judged twice in each session, and ten times in total by each participant. A trial was started by clicking a 'start'-pane on a computer screen. This was followed by the presentation of the stimulus. After this, the participant could indicate his percept by clicking a 'streaming' or a 'bouncing'-pane on the computer screen. Each of the five sessions started with three warm-up trials, randomly selected from the 84 stimuli in the session. The experiment took 75 minutes in total.

Results

The seven control stimuli, in which the visual display was presented without an auditory stimulus, rendered bouncing percepts in 4–12% of the total of 50 judgments per stimulus (five participants × ten repetitions). The percentages of bouncing percepts in stimuli consisting of the visual display accompanied by one of the five auditory stimuli are presented in Figure 1 and Figure 2. The percentages are displayed against each of the seven values of temporal tuning (−300 ms through +300 ms) that represent the onset of the auditory stimulus relative to the moment of exact coincidence of the moving objects.

Figure 1 shows the percentage of bouncing percepts seen in visual displays accompanied by an auditory stimulus consisting of one, two, or three sounds of 50 ms, respectively. The results of the 21 stimuli displayed in Figure 1 were analyzed as follows. The proportions of bouncing percepts for each participant were normalized by an arcsine transformation (Zar, 1996), and subjected to an analysis of variance (ANOVA) in a two-way, within-subject design. The independent variables were temporal tuning (seven levels; −300 ms to +300 ms) and number of sounds (three levels; one, two, or three 50-ms sounds). Two significant effects were found. The main effect of temporal tuning was significant [$F(6,84)=14.95$, $p<0.01$], and the interaction between number of sounds and temporal tuning was significant [$F(12,84)=2.17$, $p<0.05$]. The figure shows that when an auditory stimulus was presented along with the moving objects, the percentage of bouncing increased at certain values of temporal tuning, compared with the percentage of bouncing percepts in the control conditions without sound (not included in the figure). The significant interaction effect shows that the effect of temporal tuning depended on the number of sounds presented with the moving objects. As can be seen by the asymmetric bell-shaped curve in Figure 1, the bounce-inducing effect of the auditory stimulus was strong when the onset of the auditory stimulus appeared 100 ms before or in synchrony with the moment of exact coincidence of the moving objects. When multiple sounds were presented in the auditory stimulus, the percentages of bouncing percepts at temporal tuning values of −100 ms and 0 ms grew closer to those obtained at other values of temporal tuning. Since the duration of the auditory stimulus increased with the number of sounds it contained, however, the percentages of bouncing percepts could have been easily tempered because of the increase in the duration of the auditory stimulus.

Figure 2 shows that this was indeed the case. Figure 2 shows the percentage of bouncing percepts when the moving objects were accompanied by an auditory stimulus consisting of a single sound with a duration of 50, 150, or 250 ms, respectively. The results of the 21 stimuli displayed in Figure 2 were also subjected to a two-way ANOVA, in the way described above. The factors were temporal tuning (seven levels: −300 ms through +300 ms) and sound duration (three levels: 50, 150, and 250 ms). Three significant effects were found. The main effects of temporal tuning [$F(6,84)=19.95$, $p<0.01$] and sound duration [$F(2,84)=4.49$, $p<0.05$], and their interaction effect were significant [$F(12,84)=2.28$, $p<0.05$]. An asymmetric bell-shaped curve can also be seen in Figure 2, and from the significant interaction effect it becomes clear that temporal tuning facilitates visual bouncing especially

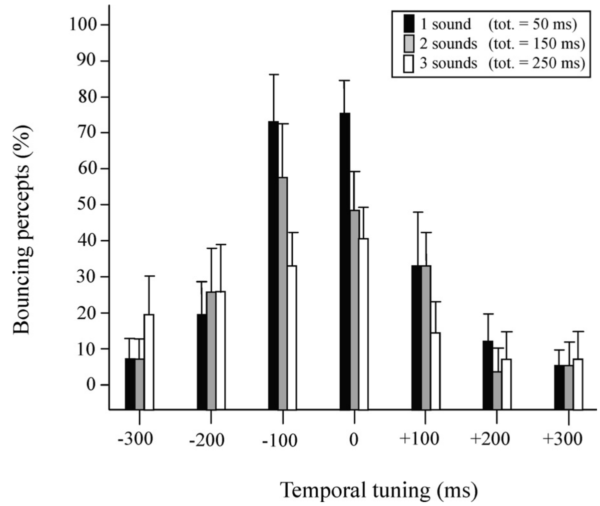

Fig. 1 Obtained percentage of bouncing percepts ($N=5\times10$ repetitions). Error bars indicate the standard error of the mean. The visual displays were accompanied by an auditory stimulus consisting of a single sound of 50 ms, two sounds of 50 ms separated by a silence of 50 ms, and three sounds of 50 ms separated by two silences of 50 ms. The percentage of bouncing percepts are indicated for each of the seven variations in temporal tuning. The temporal tuning value of 0 ms represents the moment of exact coincidence of the moving objects.

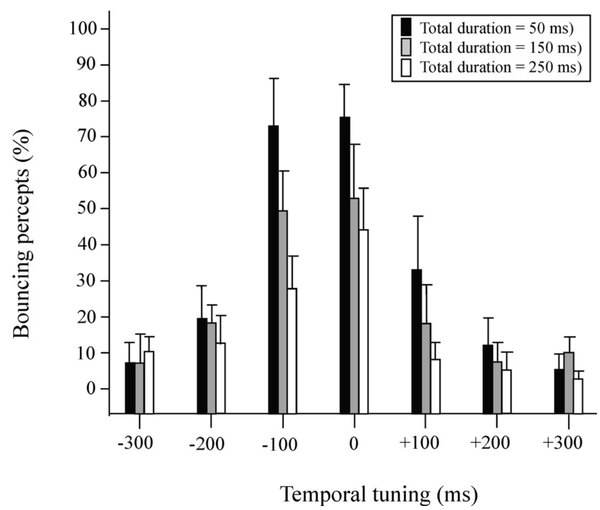

Fig. 2 Obtained percentage of bouncing percepts ($N=5\times10$ repetitions). Error bars indicate the standard error of the mean. The visual displays were accompanied by an auditory stimulus consisting of a single sound of 50 ms (replicated in Figure 1), a single sound of 150 ms, and a single sound of 250 ms. The percentage of bouncing percepts are indicated for each of the seven variations in temporal tuning. The temporal tuning value of 0 ms represents the moment of exact coincidence of the moving objects.

well when its value is −100 ms or 0 ms and when the auditory stimulus is short.

In summary, the results of the experiment show that the two moving objects were perceived as bouncing off each other when a relatively short auditory stimulus was presented 100 ms before or in synchrony with the moment of exact coincidence of the objects. When the duration of the auditory stimulus increased, its bounce-inducing effect was still relatively good at temporal tuning values of −100 ms or 0 ms, yet gradually decreased.

Discussion

The present results show that it is important that at least the onset of the auditory stimulus, but preferably both the onset and offset, occurs in close temporal proximity to the moment of exact coincidence of the moving objects. When only the offset of the auditory stimulus occurred in temporal proximity of the moment of visual coincidence, as for example in cases where a 150-ms sound started 200 ms before the moment of visual coincidence, the bounce-inducing effect of the sound was negligible. Sekuler et al. (1999) also found a superior effect of the onset of an auditory stimulus in facilitating bouncing percepts. The results further show that the bounce-inducing effect of the auditory stimulus decreased when its duration became longer, even when its onset was in close proximity to the moment of exact coincidence of the moving objects. This occurred when the auditory stimulus consisted of multiple sounds, but also when it consisted of a single sound. Rather than an effect of number of sounds, the total duration of the auditory stimulus thus is the determining factor in inducing visual bouncing.

The effect of duration of the auditory stimulus on visual bouncing may have certain implications for the derivation of a window of audiovisual integration from our data, similar to Shimojo et al. (2001). With regard to the window of audiovisual integration, though, two points need to be discussed. Both points are related to each other, and concern the width of the window of audiovisual integration, and the question of what exactly the visual 'event' is that can be integrated with the auditory stimulus in a streaming-bouncing display. Concerning the latter question, we assume that a perceptually significant change occurs in the visual display when the moving objects touch each other, overlap each other, and then move away from each other. For a while, only one object, that changes in size, is seen instead of two objects. This 'period of coincidence' of the moving objects may be interpreted as the visual 'event' that can be integrated with the auditory stimulus in a streaming-bouncing display. In our experiment, the period of coincidence lasted for 133 ms, from −66.5 to +66.5 ms, with 0 ms as the moment of exact coincidence.

Second, concerning the width of the temporal window of audiovisual integration, the results show that the bounce-inducing effect of an auditory stimulus was the highest at temporal tuning values of −100 ms and 0 ms. Reasonable bouncing was obtained at the temporal tuning value of +100 ms. A range from −100 to +100 ms corresponds quite well with the period of visual coincidence. The results thus suggest a window of audiovisual integration ranging from −100 to +100 ms around the moment of exact visual coincidence, skewed to the left.

The results show that when the auditory stimulus was shorter than the period of visual coincidence, audiovisual integration occurred. However, when the auditory stimulus exceeded this period, audiovisual integration deteriorated. Therefore, we can speculate from the present data that the width of the temporal window of audiovisual integration is determined by the period of visual coincidence, the proposed visual 'event' in the streaming-bouncing display. If so, then a change in the speed of the visual motion may shorten or lengthen the window of the audiovisual integration. More research with streaming-bouncing displays with variations in the speed of the moving visual objects, and finer variations in sound duration and temporal tuning is necessary, however, to confirm the matter.

Conclusion

In this study, we used a paradigm in which a change in the perceived motion of two moving objects was induced by sound. From our experimental results, we obtained a measure of the temporal window of integration for audiovisual information that may be important for the design of effective warning signals. Our results suggest that for effective audiovisual warning signals, it is essential that the onset of the auditory event occurs in synchrony, or at least in close temporal proximity, with the onset of the visual event. This implies that alternating auditory and visual events, even when generated in the same spatial location, are less effective as audiovisual warning signals. Furthermore, when started synchronously, it seems recommendable that the duration of the auditory event does not exceed that of the visual event. Our results show that when the auditory event is shorter in duration than the accompanying visual event, audiovisual integration can occur. However, audiovisual integration deteriorates when the duration of the auditory event exceeds that of the visual event.

In view of the increase in the population of elderly people in many countries, influencing the number of people with a loss in hearing and/or eyesight, the design of effective audiovisual warning signals on smoke detectors, dashboard-interfaces in automobiles, water-cookers and other household appliances, etc. becomes increasingly important. Since aging also can harm the ability to selectively attend to a stimulus, because of losses in the ability to inhibit the processing of irrelevant stimuli ('inhibitory deficit hypothesis', Hasher et al., 1991), temporal synchronization of auditory and visual events, in accordance with the above-described suggestions, seems essential for the design of effective audiovisual warning signals.

Acknowledgements The present study was supported by the 'Center of Excellence' (COE) program of the Japanese government, entitled 'Design of Artificial Environments on the Basis of Human Sensibility', Kyushu University. We thank Kazuo Ueda for his advice and help.

References

Alain C, Woods, DL (1999) Age-related changes in processing auditory stimuli during visual attention: evidence for deficits in inhibitory control and sensory memory. Psychol of Aging 14: 507–519

Berthenthal BI, Banton T, Bradbury A (1993) Directional bias in the perception of translating patterns. Perception 22: 193–207

Doyle MC, Snowden, RJ (2001) Identification of visual stimuli is improved by accompanying auditory stimuli: the role of eye movements and sound location. Perception 30: 795–810

Hasher L, Stolzfus ER, Zacks RT, Rypma B (1991) Age and inhibition. J Exp Psychol (LMC) 17: 163–169

Schroger E, Widmann A (1998) Speeded responses to audiovisual signal changes result from bimodal integration. Psychophysiol 35: 755–759

Sekuler AB, Sekuler R (1999) Collisions between moving visual targets: what controls alternative ways of seeing an ambiguous display? Perception 28: 415–432

Sekuler R, Sekuler AB, Lau R (1997) Sound alters visual motion perception. Nature 385: 308

Shimojo S, Scheier C, Nijhawan R, Shams L, Kamitani Y, Watanabe K (2001) Beyond perceptual modality: auditory effects on visual perception. Acous Science Technol 22: 61–67

Watanabe K, Shimojo S (2001) Effects of auditory grouping on visual motion perception. Psychol Science 12: 109–116

Zar JH (1996) Biostatistical analysis, 3rd ed., Prentice Hall Inc., Englewood Cliffs, NJ

Received: September 6, 2004
Accepted: September 17, 2004
Correspondence to: G. B. Remijn, Kyushu University, Faculty of Design, Department of Visual Communication Design, 4–9–1 Shiobaru, Minami-ku, Fukuoka 815–8540, Japan
Phone: +81–92–553–4508
Fax: +81–92–553–4508
e-mail: remijn99@hotmail.com / ito@design.kyushu-u.ac.jp