CROSSMODAL INTERACTION IN HUMANS

Thesis by

Katsumi Watanabe

In Partial Fulfillment of the Requirements

for the Degree of Doctor of Philosophy

(COMPUTATION AND NEURAL SYSTEMS)

California Institute of Technology

Pasadena, California

2001

Acknowledgment

I would like to thank the committee, Professors John Allman, Richard Andersen, Christof Koch, and Mark Konish, for their valuable time and expertise. I would like to thank all persons, inside and outside the lab with whom I have shared wonderful times at CalTech. There are many way to say thank you, but I could not find an appropriate way to express my special thanks to Professor Shinsuke Shimojo: Thanks Shin!

This work was supported by the Sloan Foundation for Theoretical Neurobiology.

Abstract

Since different senses are attuned to the different aspects of the environment, crossmodal interactions can reduce perceptual ambiguity which may result from relying on a single sensory modality. While many studies have shown that visual information affects perception in the other modalities, little is known about how auditory and tactile information affect visual perception. The present thesis explores how the human perceptual system synthesizes visual event perception by using an ambiguous visual motion display: 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. Extensive psychophysical investigations on this bounce-inducing effect show that: (1) A salient sensory transient biases visual perception toward bouncing, irrespective of the modality in which such a transient is presented. (2) The magnitude of the bounce-inducing effect depends on both the saliency of a sensory transient and the timing of a sensory transient relative to the visual coincidence. (3) The magnitude of the bounce-inducing effect increases as the sensory saliency of bounce-inducing transients increases. (4) The interaction window during which a sensory transient can bias visual perception is fixed in time. (5) Auditory and tactile transients have much larger temporal interaction windows than visual transients. (6) Auditory and tactile transients presented before the visual coincidence tend to have a stronger bounce-inducing effect than those presented after the visual coincidence; visual transients do not show this asymmetry. (7) The lack of attention to the visual moving targets in the streaming/bouncing motion display increases the likelihood of the bouncing percept. Based on these results, I conclude that the streaming percept arises when amodal attentional resource is available for moving stimuli around the moment of the visual coincidence and that the bouncing percept results from the lack of attentional resource. The role of salient sensory transients in the bounce-inducing effect is to distract attention from the coinciding moving stimuli. Thus, this thesis shows that dynamics of attentional resource allocation, not simply associative learning, can be the cause of amodal event perception in humans.

Contents

List of Figures

Chapter I: Introduction

Background

Crossmodal interaction

Most animals have several different senses through which information about the environment is obtained. Each sense is attuned to a specific form and range of energy so that we can sense the different aspects of the environment. Despite this fact, the environment appears as a unitary percept, suggesting that information from different senses interact with each other to form the integrated representation of objects and events. Crossmodal integration or crossmodal interactions of different sensory inputs have behavioral advantages by, at least, two means: (1) by overcoming a specific sensory deprivation (e.g., auditory and tactile senses in darkness) and (2) by reducing perceptual ambiguity which may result from relying on a single sensory modality (for example, tactile and olfactory senses prevent us from eating a plastic replica of food). This thesis concerns mainly the latter, namely, the reduction of perceptual ambiguity by crossmodal interaction. The particular paradigm employed is to measure the efficacy of timed sensory transients in disambiguating visual motion perception (Sekuler et al., 1997).

Many studies have investigated the behavioral outcome of crossmodal interaction (for reviews, see Stein & Meredith, 1993; Driver & Spence, 1998; Calvert et al., 1998).

Crossmodal interaction makes the detection of, and orientation to, stimuli easy (Stein et al., 1988; 1989). It also affects localization of the stimuli (Held, 1955; Pick et al., 1969; Choe et al.,1975; Bermant & Welch, 1976; Thurlow & Rosenthal, 1976; Shelton & Searle, 1980; Welch & Warren, 1980; Bertelson & Radeau, 1981; Warren et al., 1981; Bertelson, 1997). Perceptual effects of crossmodal interaction in stimulus localization are generally termed the "ventriloquism effect" (Howard & Templeton, 1966). When the ventriloquist speaks without noticeable movement of the lips, normal people inevitably perceive the puppet talking, despite our knowledge of the muteness of the puppet (Figure 1.1).

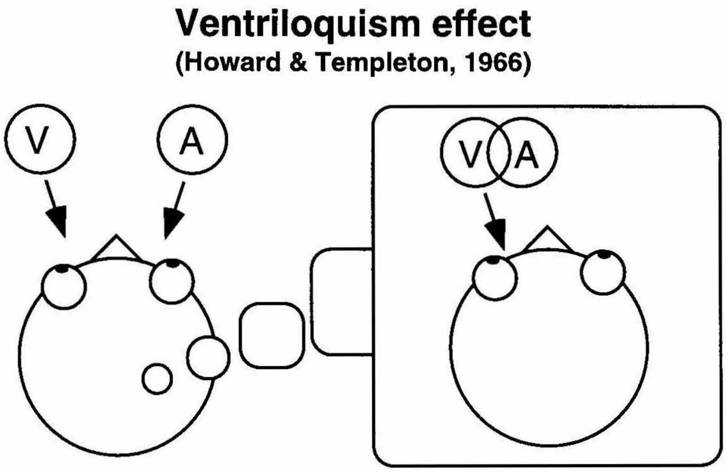

FIGURE 1.1 VENTRILOQUISM EFFECT.
When an auditory stimulus is presented at the same time as but at different location from a visual stimulus, the location of the sound is perceived to be dragged toward the location of the visual stimulus.

Crossmodal interaction is not restricted to localization judgments. Stimulus categorization and identification (event perception) can also be affected by crossmodal interaction (for reviews, see Calvert et al., 1998). The best (and so far only one) example is the McGurk effect (McGurk & MacDonald, 1976). When the human face articulates

/ga/ and the auditory stimulus is /ba/, many people typically hear /da/ (Figure 1.2). The McGurk effect has been demonstrated to provide information about the timing of articulation and information about the class of produced phoneme (MacDonald & McGurk, 1978; Green & Kuhl, 1989; Green et al., 1991; Sekiyama & Tohkura, 1991; Massaro & Cohen, 1993; Munhall et al., 1996). By doing so, it reduces perceptual ambiguity and helps identify linguistic information (Sumby & Polack, 1954; Reisberg, 1978; Summerfield, 1979, 1987; Massaro & Cohen, 1990; Driver, 1996). Additionally, such audiovisual interaction has been shown to extend to judgments of non-speech stimuli; the perception of sound of a cello can be affected by the visual perception of the cellist's movements of the hand (plucks or bows) (Saldaña & Rosenblum, 1993).

Vision dominates?

In most crossmodal interactions, the visual modality has been demonstrated to predominate (e.g., Gibson, 1933; Hay et al., 1965; Easton & Moran, 1978). For instance, whereas there are numerous examples of visual influence on auditory perception (Driver & Spence, 1998; Calvert et al., 1998), it is hard to find a strong example of auditory influence on visual perception. Only a few studies have reported the situations where audition affects vision (e.g., Shipley, 1964; Welch et al., 1986; Stein et al., 1996; Vroomen & de Gelder, 2000), but the magnitude of effects is far from being highly effective in a natural condition.

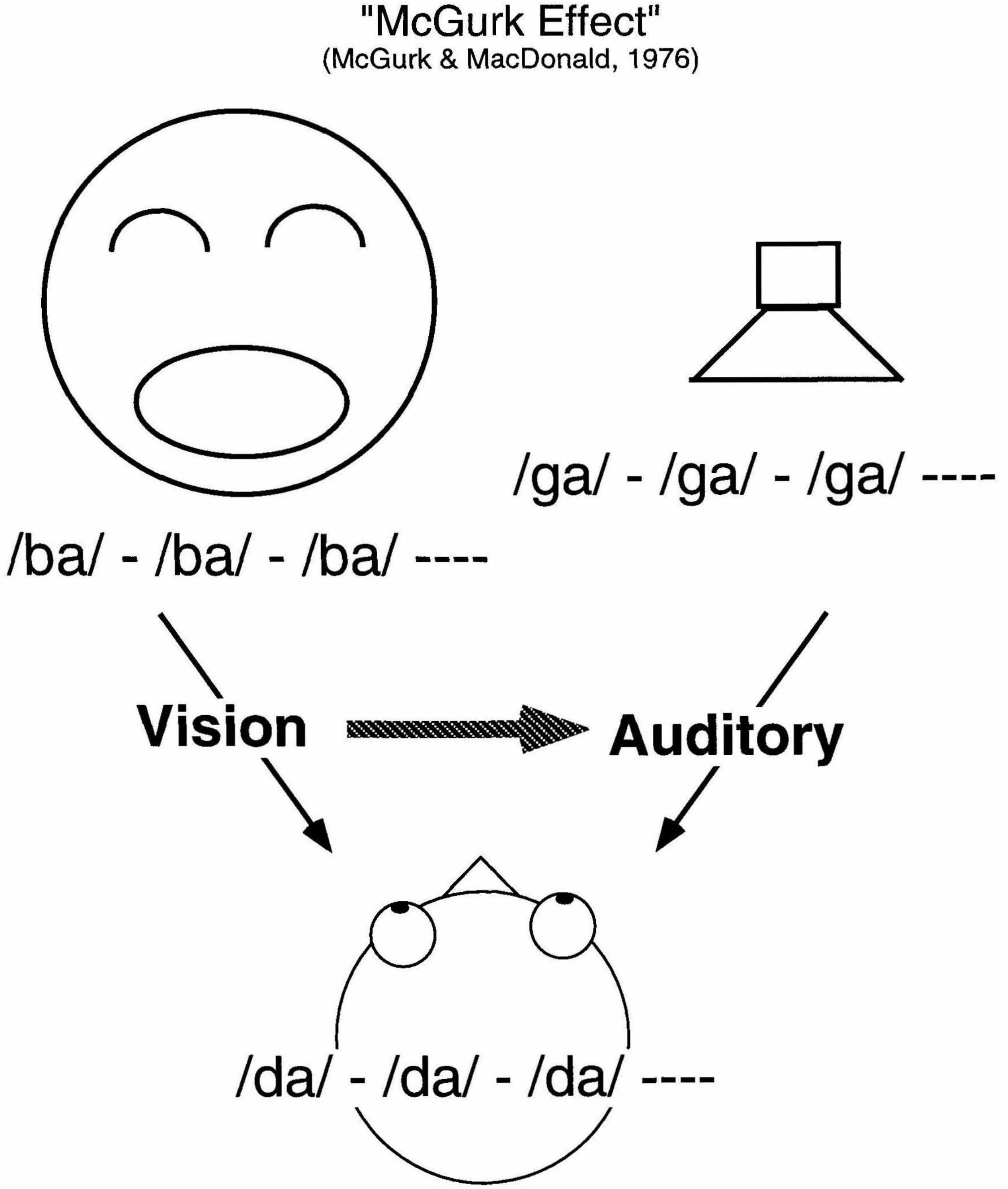

FIGURE 1.2 MCGURK EFFECT.

Faced with a conflict between visual and auditory speech information, observers typically hear sounds (voices) altered by the visual stimulus (talking head). For instance, a simultaneous presentation of visual "ba" and auditory "ga" leads to auditory perception of "da."

Researchers have speculated why vision dominates. Rock (1966) explained that the visual dominance in crossmodal interaction is due to vision's innate superiority in spatial information. Welch and Warren (1980) expanded this hypothesis and suggested that the modality with the highest precision with regard to the required task tends to be dominant (modality precision hypothesis), and vision is good at localization. The modality precision hypothesis has been advocated on the basis of indirect evidence. Studies of the ventriloquism effect, namely, the localization task under the influence of another modality, have outnumbered those of the other effects. Results have revealed, not surprisingly, that vision dominates.

On the other hand, the examples of auditory effects on visual perception mentioned above do not involve a localization task. Stein et al. (1996) showed that the detection of a near-threshold visual stimulus can be enhanced by the simultaneous presentation of sound. Vroomen and de Gelder (2000) have recently demonstrated that the detection of a visual shape improves when a deviant sound is presented at the same time. Particularly suggestive with respect to the focus of this thesis is "auditory-flutter-driven visual-flicker" (Shipley, 1964; Welch et al., 1986): The perception of visual flicker rate can be biased toward the auditory flutter rate. This suggests that, when the observer's task involves temporal perception, audition may dominate and affect vision.

It is well known that audition has a better temporal resolution than vision. Human temporal resolution in audition is estimated to be about 1-2 ms (Plomp, 1964; Ronken, 1970; Buunen & Valkenburg, 1979; Henning & Geskell, 1981; Forrest & Green, 1987), whereas human temporal resolution in vision is about 20 ms (critical flicker fusion = 50

Hz: Kelly, 1971ab, 1972). Therefore, the modality precision hypothesis predicts that the auditory effect on visual perception occurs in a non-spatial task that includes a temporal decision. The results of Shipley (1964) and Vroomen and de Gelder (2000) partially support this idea.

Bounce-inducing effect

Researchers have struggled to find a more dramatic auditory effect on visual perception for a long time, because the magnitude of the known effects is not so compelling (e.g., at most, 10% increase in the correct direction percentage from chance level; Vroomen & de Gelder, 2000). This lack of a psychophysical paradigm can be a disaster for the other fields of neuroscience because it limits the direction of crossmodal research. However, Sekuler et al. (1997) have recently devised a simple method to show the compelling effect of sound on visual perception. The basic idea is to introduce visual ambiguity, the solution of which is influenced by the timing of 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. This visual display was first described by Gestalt psychologist Metzger (1934), then later revisited by Michotte (1963). The ambiguous motion condition happens if the two targets are small and the motion is smooth so that observers cannot use the other cues (such as the overlap or occlusion of the targets) to interpret the visual event (Watanabe & Shimojo, 2000)[1].

Despite this ambiguous nature of the visual stimulus, observers normally show a strong bias toward seeing the streaming percept (Goldberg & Pomerantz, 1982; Bertenthal et al., 1993; Sekuler & Sekuler, 1999). However, various factors have been reported to increase the relative frequency of the bouncing percept, such as a momentary pause of the targets at the visual coincidence (Bertenthal et al., 1993). Most intriguingly, Sekuler et al. (1997) have shown that a brief sound, presented at the moment the targets coincide (the simultaneous sound, hereafter), biases perception toward bouncing (Figure 1.3). They have suggested that crossmodal interaction is involved in solving the streaming/bouncing visual motion ambiguity.

The success of Sekuler's display requires: (1) that the visual stimulus is *ambiguous*, (2) that the possible two interpretations are *mutually exclusive*, (3) that the solution of crossmodal interaction depends on the relative *timing* between auditory and visual stimuli, and (4) that it uses a *simple* auditory click and does not require complex audiovisual stimuli such as a talking head in the McGurk effect.

The streaming/bouncing ambiguous motion display provides us with a chance to investigate the unexplored topic of auditory influence on visual event perception.

---

[1] In other words, the display forces observers to use clues other than visual. This is the key feature of the streaming/bouncing motion display.

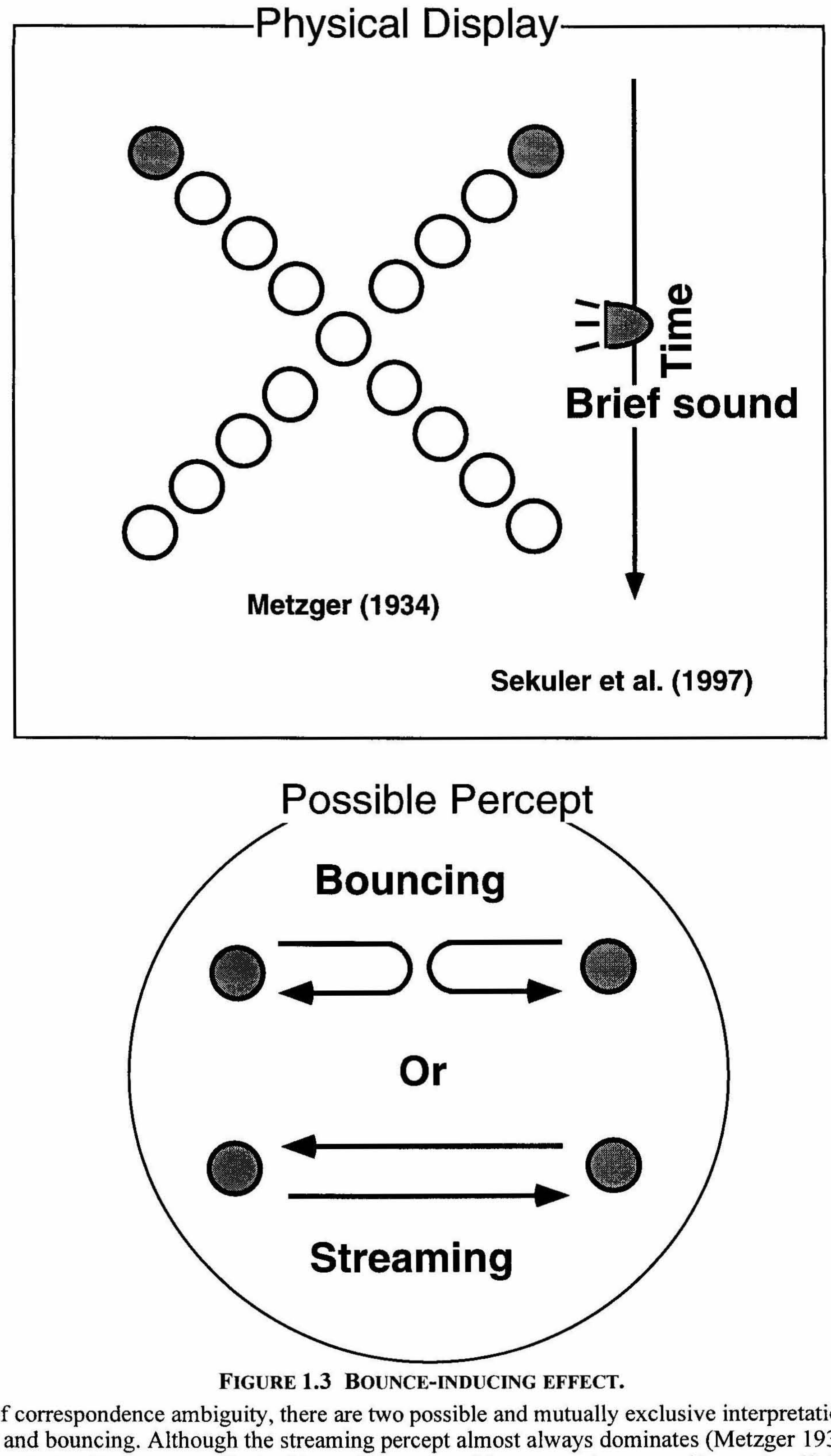

FIGURE 1.3 BOUNCE-INDUCING EFFECT.

Because of correspondence ambiguity, there are two possible and mutually exclusive interpretations: streaming and bouncing. Although the streaming percept almost always dominates (Metzger 1934), a sound presented at the moment of coincidence is known to bias perception toward bouncing (Sekuler et al., 1997).

Moreover, as we will see, the simplicity of the display leaves open the possibility of investigations on the temporal window of crossmodal interaction and on the underlying mechanism of event perception in general.

Outline of the Thesis

In this thesis, I employ the bounce-inducing effect as the basic paradigm and investigate the nature of various transients on the streaming/bouncing ambiguous motion display in order to explore how the human perceptual system synthesizes the unified percept of the sensory world (Figure 1.4).

In the experiments of Chapter II, the bounce-inducing effect is used to determine the temporal window of audiovisual event perception. The temporal window of the bounce-inducing effect turns out to be about 400 ms, which is about the same size as the temporal range reported in the McGurk effect, suggesting a common mechanism in both the bounce-inducing effect and the McGurk effect despite that the directions of crossmodal influence are opposite.

In Chapter III, I explore the effects of auditory context on the bounce-inducing effect. The bounce-inducing effect by a simultaneous sound can be attenuated when the sound is embedded in other identical sounds (the attenuation effect). Subsequent experiments show that the bounce-inducing effect can be recovered if the simultaneous sound "pops out" from the embedding sound sequence. These results suggest that the presence of the simultaneous sound is not sufficient, and that the saliency of auditory transients is *important* for the bounce-inducing effect.

In Chapter IV, I extend the bounce-inducing effect to tactile-visual and visual-visual interactions and report that any sensory transient presented around the moment of the visual coincidence biases visual perception toward bouncing (amodal saliency). Furthermore, by changing the intensity and the timing of auditory, tactile, and visual transients, it is shown that the bounce-inducing effect increases as the intensity (or saliency) of the sensory transient increases (saliency dependency), but the temporal window for the bounce-inducing effect is fixed in time (temporal gain modulation).

Motivated by the results of amodal saliency dependency, I propose the attention hypothesis for the bounce-inducing effect. The attention hypothesis states that the streaming percept requires attentional resource, presumably to enhance "motion recruitment" processes (i.e., tendency of motion as continuing to occur in the same direction as in the past), and that auditory, tactile, and visual transients automatically capture attention. Attentional capture reduces attentional recourse available to motion processing, which in turn leads to the enhancement of the bouncing percept. I devise experiments to test the attention hypothesis and obtain results that are consistent with the hypothesis (Chapter V).

Finally, in the General Discussion (Chapter VI), I summarize the empirical findings and suggest further directions of research.

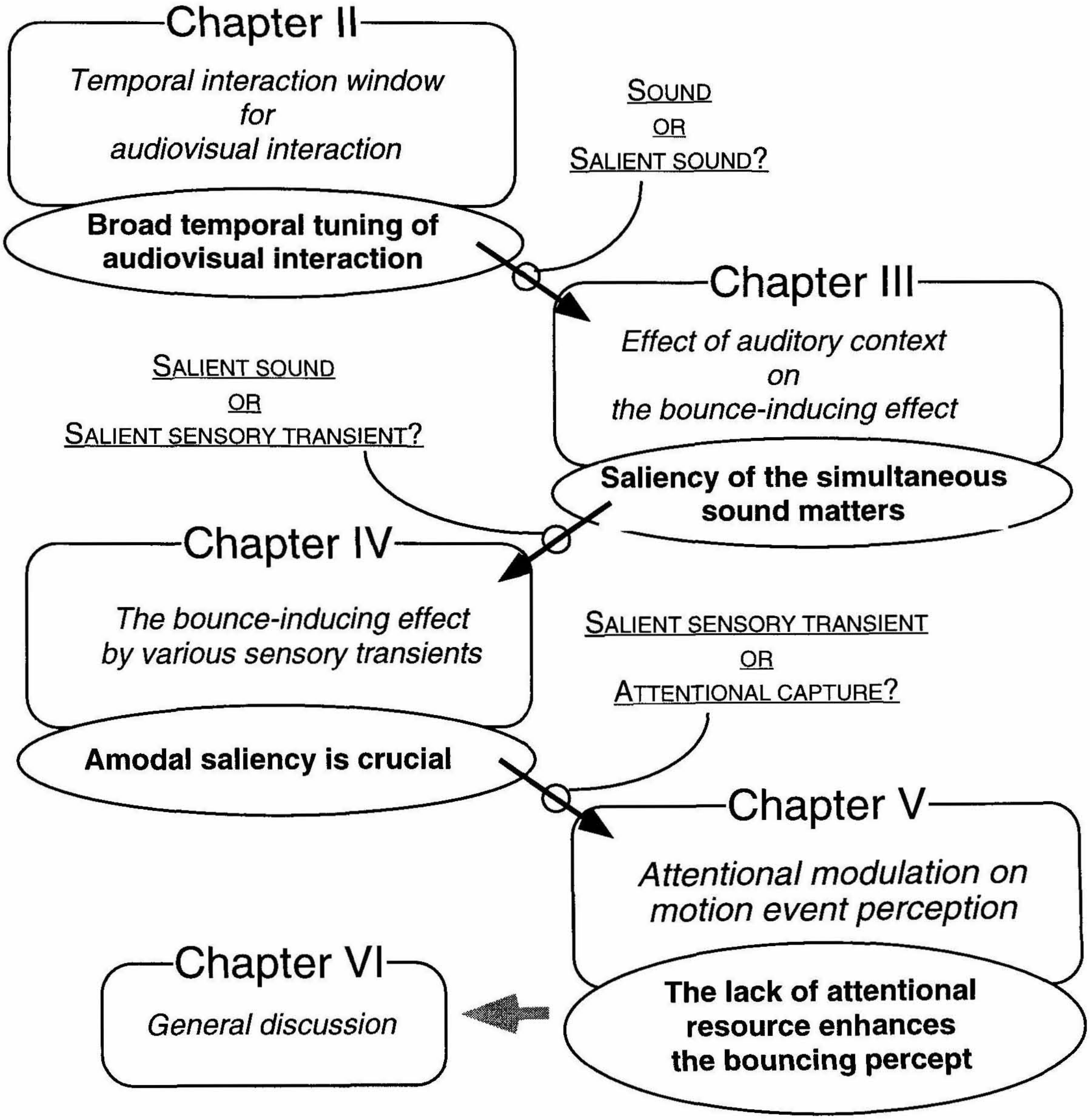

FIGURE 1.4 OVERVIEW OF THE THESIS.

Chapter II: The temporal window of audiovisual event perception

Introduction

Two identical visual targets moving across each other with equal and constant speed can be perceived either to bounce off or to stream through each other (Metzger, 1934; Michotte, 1963; Goldberg & Pomerantz, 1982). Sekuler et al. (1997) have shown that a brief sound which is roughly synchronized with the visual coincidence biases perception toward bouncing and suggested that crossmodal interaction is involved in solving the streaming/bouncing motion ambiguity. Furthermore, they reported that the sound enhances the perception of bouncing (though the effect was reduced) even when the sound was presented 150 ms before or after the visual coincidence. This observation raises an important question: what is the temporal window of audiovisual interaction?

The question of the temporal window of audiovisual interaction is one of the main questions in crossmodal interaction. For audiovisual localization (the ventriloquism effect), the temporal proximity between auditory and visual stimuli has to be within 200 ms (Thomas, 1941; Jack and Thurlow, 1973; Choe et al., 1975; Welch and Warren, 1980;

Radeau & Bertelson, 1977; Radeau & Bertelson, 1987)[2]. For audiovisual event perception (e.g., the McGurk effect), researchers all agree that the temporal proximity is important, but disagree about the extent to which different sensory inputs need to be close in time (up to 200 ms of audiovisual asynchrony; Massaro et al., 1996; Munhall et al., 1996; Bertelson et al., 1997).

The paramount reason for this disagreement is the lack of an adequate audiovisual display. The only available display for audiovisual event perception has been the McGurk effect or its variants. Although the McGurk effect is a robust phenomenon, the effect inherently depends on the complex temporal structure of auditory (speech) and visual (facial movement) stimuli. These complex dynamics have many dimensions other than time, which would alter the magnitude of the McGurk effect. Moreover, the McGurk effect is known to be influenced by cultural and linguistic factors (Sekiyama & Tohkura, 1991; Sekiyama, 1997). Thus, it is a formidable task to separate the temporal interaction window from such dimensions.

In contrast to these complications, the bounce-inducing effect has an advantage in that the timings of auditory and visual events are well defined. Therefore, it is best suited for the investigation of audiovisual temporal window. Additionally, the bounce-inducing effect involves an auditory effect on visual perception, whereas the McGurk effect involves a visual effect on auditory perception. By using the bounce-inducing effect, I

---

[2] A recent review (Calvert et al., 1998) suggested that for the ventriloquism effect, the temporal proximity between auditory and visual stimuli is a strict necessity. They concluded that even a slight temporal discrepancy completely abolishes audiovisual effect. My extensive literature study, including their citations (Jack & Thurlow, 1973; Choe et al., 1975; Welch & Warren, 1980), refuted it.

expected to see the commonalties and differences between these effects that result from the different crossmodal effect directions.

Experiment 1: Temporal window of the bounce-inducing effect

The purpose of Experiment 1 was to replicate the basic bounce-inducing effect reported by Sekuler et al. (1997) and to employ it to find the temporal window of audiovisual interaction.

Method

Observers

Nine observers voluntarily took part in the experiment. Seven of them were naive as to the purpose of the study. One of the non-naive observers was the author. Their ages ranged from 20 to 43 years. All observers had normal or corrected-to-normal visual and auditory senses and experienced no severe difficulty with the tasks. Informed consent was obtained from all observers before the experiment.

Stimuli

Visual stimuli were displayed on a Sony color monitor, controlled by a Silicon Graphics Indigo2 work station, in an otherwise dark environment. The frame-rate was 20 Hz. This relatively low frame-rate was chosen for enforcing a precise audiovisual synchronization. A black fixation cross (0.01 cd/m^2, 0.35 deg in visual angle) was continuously displayed against a gray background (8.5 cd/m^2) throughout a session (Figure 2.1). At the beginning of each trial, two black disks (0.13 deg in diameter) appeared at opposite sides of the frame, initially separated by about 3.27 deg. The disks moved laterally toward each other at a constant speed of 1.64 deg/s with no inter-stimulus interval between frames. The frame-to-frame spatial offset was about 0.065 deg. The two disks coincided at 1.47 deg above the fixation cross and then continued to move to the other's start position. The duration of a total sequence was about 2 s. Because of the disk size and the frame-to-frame spatial offset, the two disks never overlapped except at the coincidence point.

In addition to the visual sequence, auditory stimuli were delivered through a computer's built-in speaker, which was placed about 11 deg below the fixation cross. The sound was a digitized 1.8 kHz tone burst with a duration of 3 ms (virtually zero rise time and approximately 1.5 ms fall time). The highest sound pressure level was 58 dB (sound pressure level; SPL) at the observer's ear (measured with an Edmund scientific Model YF-20 sound level meter). The sound occurred at various times relative to the visual coincidence (50-, 100-, 150-, 200-, 250-, 450-, 650-, 850-, 1050-ms before, after, or at the same time as the visual coincidence). No sound was presented for one-twentieth of the trials. There was background noise of about 53 dB (SPL).

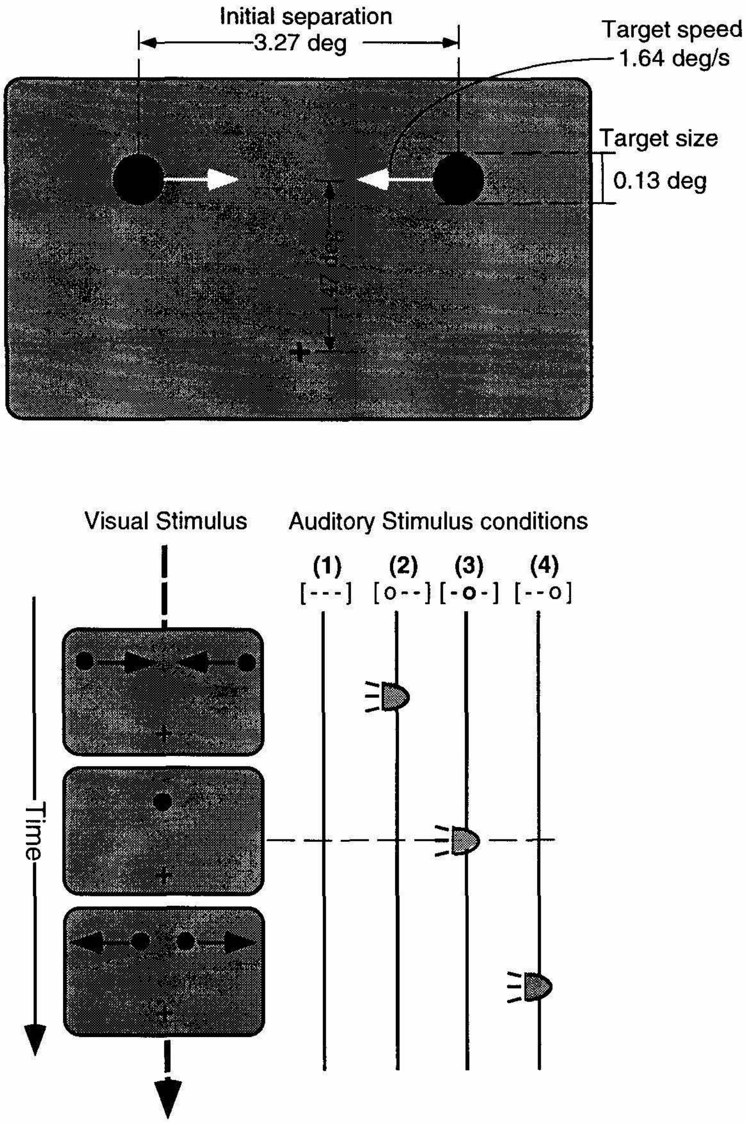

FIGURE 2.1 VISUAL DISPLAY AND EXPERIMENTAL PARADIGM OF EXPERIMENT 1 AND EXPERIMENT 2. Top: Two identical visual targets moved toward each other, coincided, and continued to move to the other's start position. Bottom: A tone burst was presented at various timings relative to the visual coincidence.

Procedure

Observers viewed the stimulus display binocularly and binaurally from a distance of 80 cm while staring at the fixation cross throughout each trial. The two disks appeared, immediately started to move, coincided, continued to move, and disappeared at the other's start position. Observers were asked to report whether the two disks appeared to "bounce off" (reversing their directions of motion) or "stream through" (no reversal of motion direction) each other by pressing one of the mouse buttons. Immediately after the observer's response, another trial began. For each asynchrony between the sound and the visual coincidence (20 conditions, including no-sound condition), 20 trials were shown in random order (the method of constant stimuli). A full experiment was divided into four sessions. Each session consisted of 100 trials.

Results

Figure 2.2 shows the averaged percentage of bouncing responses across the nine observers as a function of asynchrony of the sound relative to the visual coincidence. Compared with the no sound condition (mean = 1.25 %, standard deviation = 2.50), the presentation of the single sound at the moment of the visual coincidence (mean = 78.03 %, standard deviation = 12.08) clearly enhanced the perception of bouncing (ANOVA with repeated measures, $F(1,8) = 349.3$, $p < 0.001$). When the single sound occurred before the visual coincidence (the left half of Figure 2.2), the bounce-inducing effect decreased as the audiovisual asynchrony increased ($F(8,64) = 39.28$, $p < 0.001$). Similarly, the sound presented after the visual coincidence enhanced the bouncing percept when it was presented soon after the visual coincidence (the right half of Figure 2.2; $F(8,64) = 54.83$, $p < 0.001$). However, the decay of the bounce-inducing effect as a function of audiovisual asynchrony differed significantly between the pre- and the post-coincidence sounds ($F(2,16) = 6.446, p < 0.001$). A series of planned comparisons showed that the sound presented 250-ms before and the sound presented 150-ms after the visual coincidence were still effective to induce the bouncing percept (See Appendix A, Table 1).

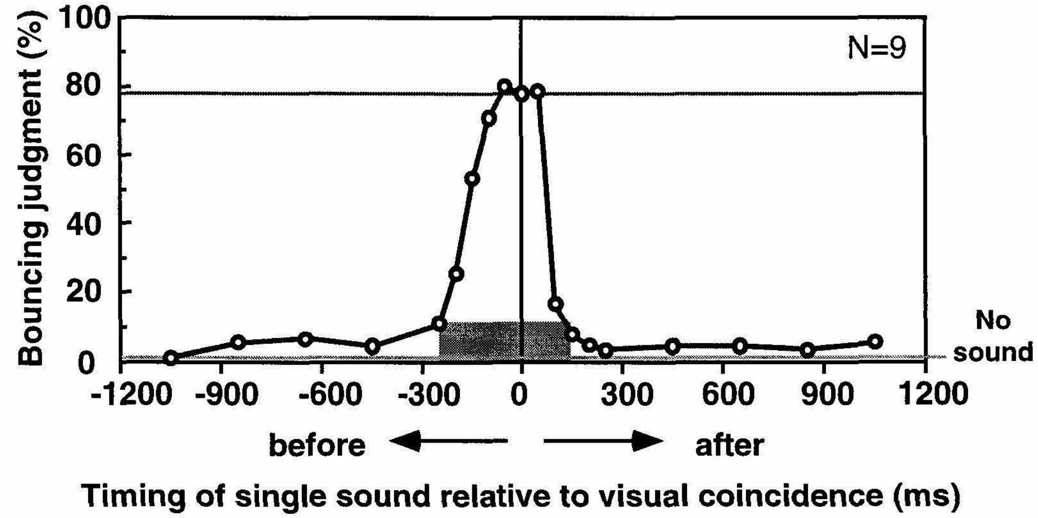

FIGURE 2.2 RESULTS OF EXPERIMENT 1.
The averaged percentage of bouncing judgment is plotted against the sound timing relative to the visual coincidence. The gray region represents the temporal window of the bounce-inducing effect, within which the enhancement of the bouncing percept was significant ($p < 0.05$).

Discussion

The results of Experiment 1 confirmed Sekuler's basic results of the bounce-inducing effect (Sekuler et al., 1997). The single sound presented at or near the time of the visual coincidence promoted the perception of bouncing. Generally, as asynchrony between the sound and the visual coincidence increased, the frequency of the bouncing report decreased. Approximately 250 ms prior to and 150 ms after the visual coincidence, the single sound had its effect on visual motion perception.

The results are a clear demonstration of the interaction window during which auditory information affects visual perception. A similar tolerance for asynchrony has been reported in the McGurk effect (the temporal window of visual effect on auditory perception), which is up to about 200 ms of visual advance and delay relative to auditory stimulus (Radeau & Bertelson, 1977; McGrath & Summerfield, 1985; Massaro et al., 1996; Munhall et al., 1996). However, the reported temporal ranges vary significantly among studies. This is probably because, for complex stimuli such as speech, it may be very difficult to define audiovisual synchrony; speech stimuli can differ along so many different perceptual or informational dimensions that the estimated thresholds for audiovisual asynchrony may differ considerably. The present experiment has overcome this complication by the virtue of the simplicity of the experimental paradigm and has showed that the tolerance for audiovisual asynchrony in the bounce-inducing effect is comparable to the longest interaction range of the McGurk effect (Massaro et al., 1996).

Another significance of the results of Experiment 1 is that, despite that fact that the directions of crossmodal influence are opposite, the audiovisual interaction range observed in the bounce-inducing effect (auditory effect on visual perception) is close to the temporal interaction range in the McGurk effect (visual effect on auditory perception). This suggests that the broad temporal interaction window may be a relatively common feature among audiovisual event perception.

Experiment 2: Is the audiovisual temporal window time- or space-dependent?

I have stated that the bounce-inducing effect depends on the temporal window of audiovisual interaction. However, since the speed of the visual targets was fixed in Experiment 1, one could argue that the results are equivocal in terms of whether the observed window is temporal or spatial. This is because, if the results were plotted as a function of distance from the coincidence point, as the speed of the targets was fixed, I could come up with a *spatial* interaction window for the bounce-inducing effect as well as a temporal interaction window (Figure 2.3). The aim of Experiment 2 was to address this issue.

In order to tear apart the spatial and temporal factors, the speed of the targets was doubled in Experiment 2, while using the otherwise same display and the same observers. If the audiovisual interaction window observed in Experiment 1 is fixed in space, it should shrink when it is plotted against time. In contrast, if the bounce-inducing effect is in fact mediated by a fixed temporal window, it should not be affected by the speed change of the visual target (when it is plotted against time).

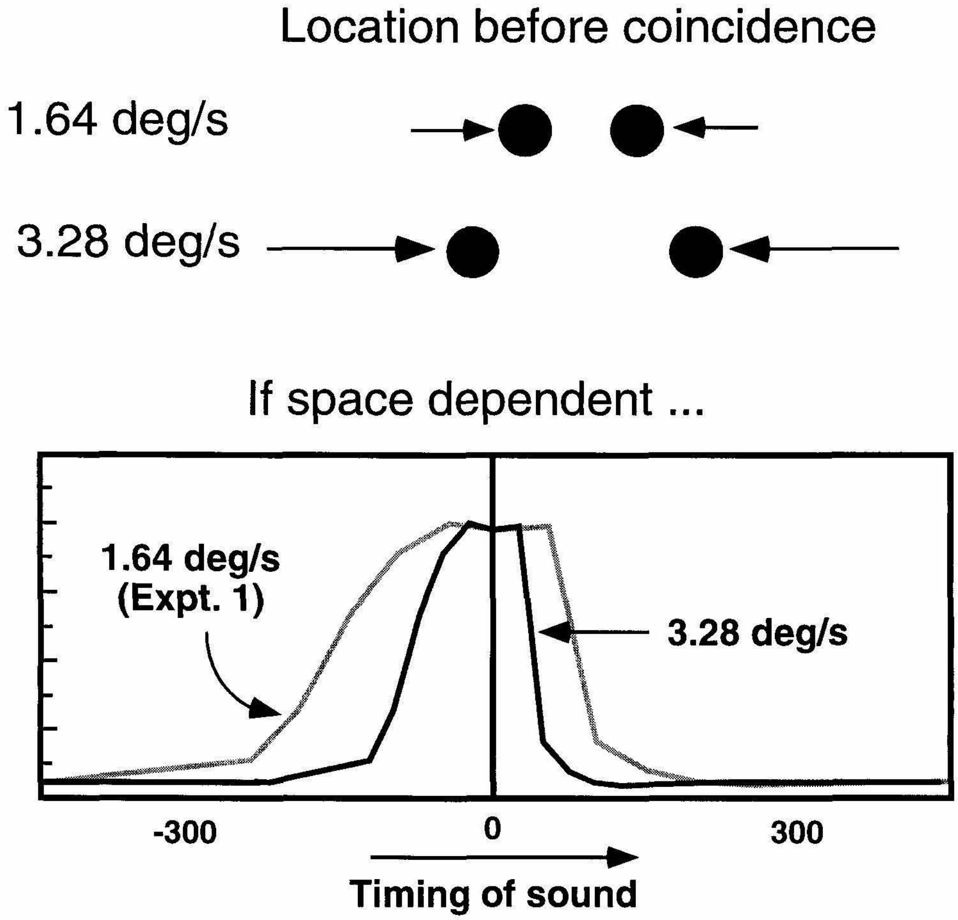

FIGURE 2.3 PREDICTION FOR DIFFERENT TARGET SPEEDS.

Top: Distance from the visual coincidence point at a given time for two speeds. Bottom: If a spatial interaction window mediates the bounce-inducing effect, the interaction window as a function of time would shrink with the doubled speed of the visual targets.

Method

Observers, stimuli, and experimental procedure were identical to those of Experiment 1, except that the frame-to-frame spatial offset of the moving disks was doubled (about 0.13 deg), resulting in a target speed of 3.28 deg/s.

Results

The results of Experiment 2 are presented in Figure 2.4. The results of Experiment 1 are shown as a gray-shaded region for comparison purposes. The curves of the two experiments nearly overlap each other, when they are plotted against the sound timing relative to the visual coincidence. A two-way ANOVA failed to find a significance between the results of Experiment 1 and those of Experiment 2 ($F(1,8) = 1.447, p = 0.230$) while indicating a significant main effect of audiovisual asynchrony ($F(18,144) = 109, p < 0.001$). Post-hoc statistical comparisons indicated that, in Experiment 2, as in Experiment 1, the sound presented in the range from 250-ms before to 150-ms after the visual coincidence significantly enhanced the perception of bouncing (see Appendix A, Table 2).

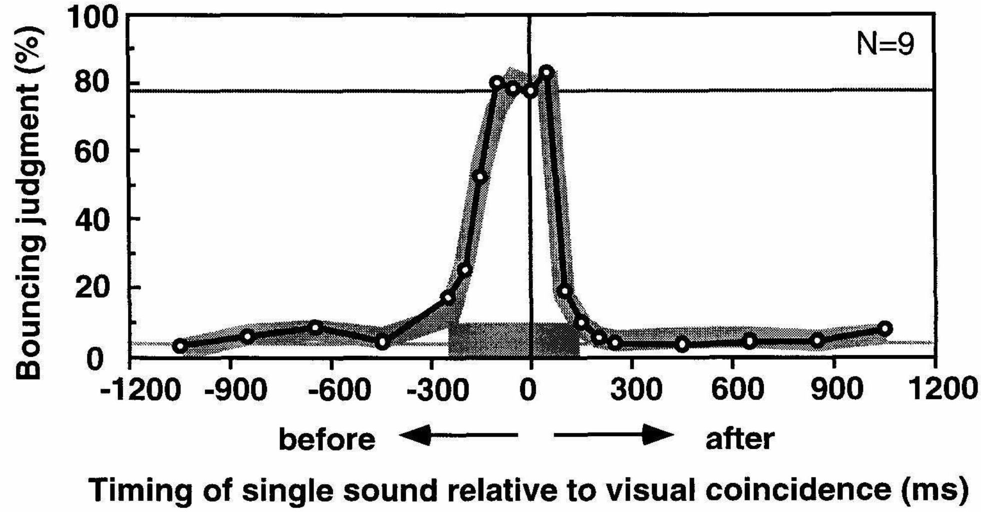

FIGURE 2.4 RESULTS OF EXPERIMENT 2.

The average percentage of bouncing judgment is plotted against the sound timing relative to the visual coincidence. The gray region represents the temporal window of the bounce-inducing effect, within which the enhancement of the bouncing percept was significant ($p < 0.05$). For comparison purpose, the results from Experiment 1 are also plotted as the thick gray line.

Discussion

Since the frame-rate was fixed, the doubled speed of the moving disks in Experiment 2 resulted in a doubled distance from the coincidence point at a given time, compared with those in Experiment 1. If the audiovisual interaction window of the bounce-inducing effect depended on the distance between the disks, the width of the temporal window in Experiment 2 should have been narrower than that in Experiment 1 (Figure 2.3). Quite to the contrary, the results indicate that the frequency of the bouncing percept increases as an almost identical function of audiovisual asynchrony in Experiment 1 and Experiment 2. Therefore, I conclude that time, rather than distance, is the determining factor for the audiovisual interaction in the bounce-inducing effect.

The results of Experiment 1 and Experiment 2 together imply that the temporal window for the bounce-inducing effect may be fixed. Also notable in both experiments is the curious asymmetry of the temporal interaction window. The sound presented before the visual coincidence tended to be more effective in enhancing the bouncing percept. The issues of the inflexible temporal window and temporal asymmetry will be examined and discussed in Chapter IV.

Experiment 3: Is the bounce-inducing effect a perceptual or cognitive effect? Temporal window versus Temporal uncertainty

Is the bounce-inducing effect a perceptual or cognitive effect? Although it is very difficult to draw a clear line separating perception from cognition (e.g., Bruner, 1957), it is clear that a part or an aspect of perception occurs regardless of the observer's conscious knowledge of real stimuli (cognitive impenetrability: Pylyshyn, 1984, 1999). Most geometrical perceptual illusions do not disappear when observers know the physical characteristics of the stimuli (Gregory, 1997). For example, measuring two lines to be precisely equal length does not make them look equal when arrowheads are added to them to form the Müller-Lyer illusion (Figure 2.5).

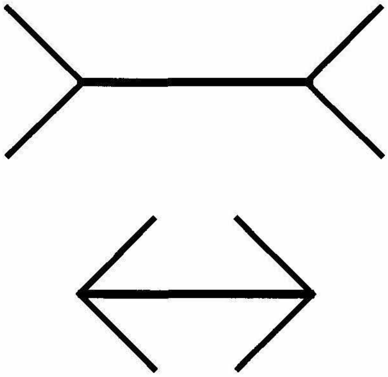

FIGURE 2.5 MÜLLER-LYER ILLUSION.
The length of a line appears longer when the outward "fins" are attached at the both ends of the line (top) than when the inward "fins" are attached (bottom). Knowing that the physical lengths of the lines are identical does not eliminate the illusion.

The possible cognitive factor in the bounce-inducing effect may be related to the uncertainty in temporal order between the sound and the visual coincidence. Observers might simply report the bouncing percept whenever they are uncertain about the temporal order between the sound and the visual coincidence (i.e., when they think two events happen simultaneously), and report the streaming percept otherwise. If this temporal uncertainty is the main cause of the bounce-inducing effect, the temporal order judgment between the sound and the visual coincidence would correlate closely with the bounce-inducing effect. Experiment 3 examined and rejected this possibility.

Method

Observers

Seven observers from the previous experiments (including the author) participated.

Stimuli

The visual and auditory stimuli were almost the same as those of Experiment 1. However, the timing of the sound presentation varied only from 450-ms before to 450-ms after the visual coincidence.

Procedure

For 14 conditions (13 different audiovisual asynchrony and no-sound condition), 20 trials were repeated (280 trials). There were two tasks that observers performed. Bounce judgment: As in the previous experiments, observers were asked to report their percept by pressing the mouse button. Temporal order judgment: The stimuli were the same as in

Bounce judgment. But, observers were instructed to judge whether the sound occurred before or after the visual coincidence in a 2-alternative forced-choice paradigm. A full experiment was composed of four sessions (two for Bounce judgment, two for Temporal order judgment). Each session consisted of 140 trials.

Results

The top panel of Figure 2.6 depicts the results for Bounce judgment in Experiment 3. The results again replicated the results of Experiment 1 ($F(6,72) = 15.88, p < 0.001$, see also Appendix A, Table 3). The results of the temporal-order judgment (bottom panel of Figure 2.6) revealed that the observers were quite sensitive to the audiovisual asynchrony employed in the present set-up. They reliably detected an audiovisual asynchrony greater than or equal to 100 ms (see Appendix A, Table 4). It is worth noting that the temporal-order judgment results showed no temporal asymmetry whereas the bounce judgment results did. These results were also replicated with a three-alternative forced-choice method (bouncing, streaming, or unsure; 3 observers; Figure 2.7).

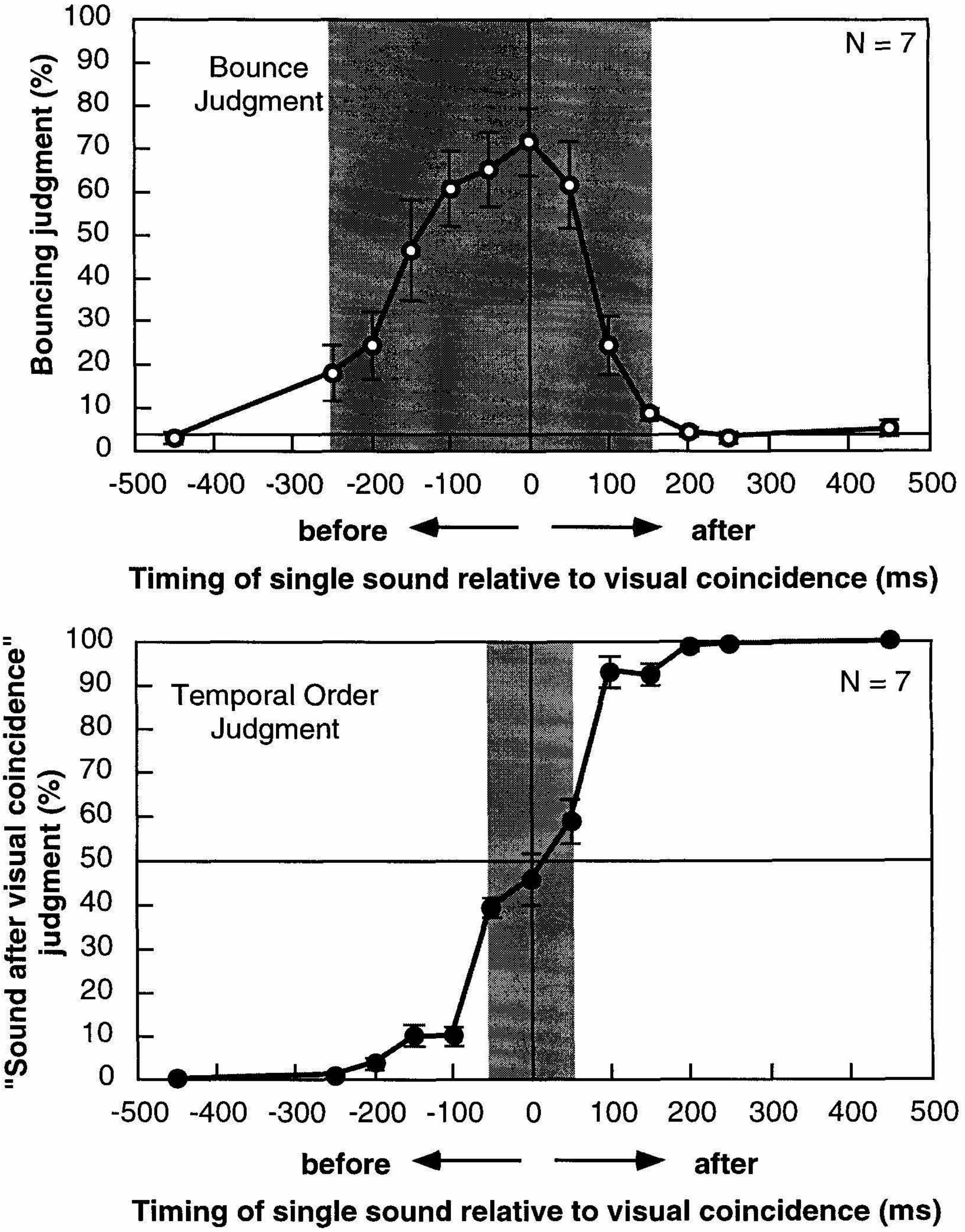

FIGURE 2.6 RESULTS OF EXPERIMENT 3 (1).

Top: Bouncing judgment. The averaged percentage of bouncing judgment is plotted against the sound timing relative to the visual coincidence. The gray region represents the temporal window of the bounce-inducing effect, within which the enhancement of the bouncing percept was significant ($p < 0.05$). Bottom: Temporal order judgment: Observers were quite sensitive to the audiovisual asynchrony.

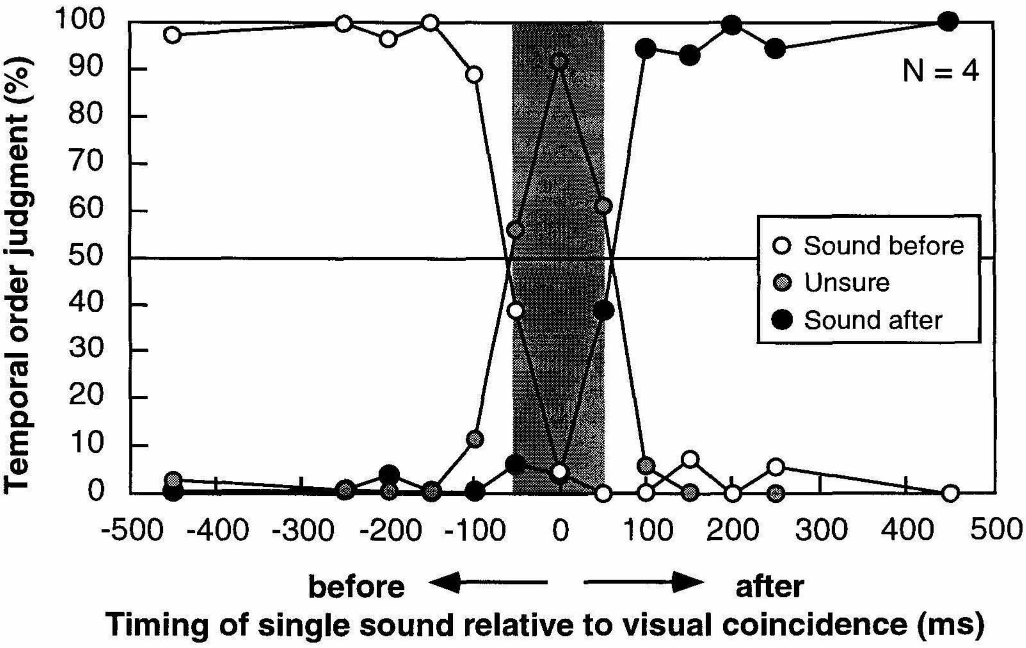

FIGURE 2.7 RESULTS OF EXPERIMENT 3 (2).
Temporal order judgment with a 3 alternative forced choice procedure.

Discussion

The results of Experiment 3 indicated that the observers were able to notice an audiovisual asynchrony of about 100 ms with certainty. Nonetheless, with a 100-ms asynchrony, the same observers reported the bouncing percept more frequently than with larger asynchronies or without sound. Furthermore, no temporal asymmetry was observed in the temporal order judgment, which was clear in the bouncing judgment. These results clearly eliminate the possibility that cognitive penetration due to audiovisual temporal uncertainty is the main factor for the bounce-inducing effect. In other words, the bounce-inducing effect is cognitively impenetrable.

When crossmodal displays are complex and/or meaningful, observers sometimes do not notice large temporal discrepancies. For example, Dixon and Spitz (1980) presented a film in which a hammer was hitting a peg and asked observers to either advance or delay the hitting sound until they perceived asynchrony. Their observers advanced the sound by 75 ms and delayed it by 180 ms. When the sound track and image of a videorecording of a talker were used, observers advanced the sound by 150 ms and delayed it by 250 ms. In contrast, for simple transient stimuli, such as a click and a flash of light, practiced observers can be very sensitive to difference in intermodal timing. An asynchrony as small as 20 ms can be detected (Hirsh & Sherrick, 1961). The temporal order judgment performance in Experiment 3 was obviously better than the results of Dixon and Spitz (1980) and relatively close to those of Hirsh and Sherrick (1961).

The most important point in Experiment 3 is the discrepancy between the audiovisual temporal uncertainty and the audiovisual temporal interaction window in the bounce-inducing effect. Several studies have shown the independence of crossmodal effect from perceptions of stimuli in each modality (Green et al., 1991; Green & Gerdeman, 1995). For instance, Green et al. (1991) found that knowing whether the information from two modalities corresponded was not a precondition for perception of the McGurk effect. Observers viewed stimuli composed of faces and voices of different genders. When male faces were combined with female voices and vice versa, observers showed no decrease in the magnitude of the McGurk effect even though it was clear that the genders of the faces and voices were incompatible. The resistance or immunity to knowledge about intramodal stimuli may be another shared characteristics of audiovisual event perception.

Experiment 4: The spatial window of the bounce-inducing effect

The aim of Experiment 4 was to obtain some insights on the spatial tuning of auditory effect on visual perception, by using the bounce-inducing effect. Although the main theme of the present thesis is the temporal window for event perception and underlying mechanisms, it may be informative to examine the spatial tuning, too.

The spatial interaction windows of crossmodal localization and crossmodal event perception are known to be different (e.g., Calvert et al., 1998). The spatial window for the ventriloquism effect was estimated as about 10 degree in visual angle (Bermant & Welch, 1976). In contrast, the McGurk effect can be found without loss of magnitude with a spatial separation up to about 40 deg (Witkin et al., 1952; Jackson, 1953; Jack & Thurlow, 1973; Fisher & Pylyshyn, 1994; Jones & Munhall, 1997; Bertelson, 1998).

The broad temporal window observed in the previous experiments suggests some commonalties between the bounce-inducing effect and the McGurk effect. If this is the case, a broad spatial interaction window as in the McGurk effect could be expected for the bounce-inducing effect. Therefore, measuring the spatial interaction range might add further evidence to the idea that there are some general rules for audiovisual event perception.

Moreover, again note that the bounce-inducing effect is an auditory effect on visual perception. The spatial interaction window of auditory effect on visual perception has not been investigated. The results of Experiment 4 would provide the first demonstration of it.

Method

Observers

Three observers participated. One observer was the author and two observers were naive but all had participated in previous experiments of the bounce-inducing effect.

Stimuli

Visual stimuli were displayed on a Nokia color monitor (25 Hz frame-rate), controlled by a Power Macintosh 8600/200, in a dark room. Two identical visual disks moved in an 'X' shaped path; one target moved the left-down direction, and the other the right-down direction (see Figure 2.8). Observers looked at the visual display from a distance of 120 cm. In order to compensate for the increased observation distance compared with the previous experiments, the physical size of the visual display was magnified by a factor of 1.5. Before the experimental sessions, the downward shift of the fixation cross from the visual coincidence was adjusted for each observer so that, in about 90% of trials, the bouncing percept would occur when the sound was presented[3]. This adjustment was needed because of the fewer number of observers and an inevitable increase in variability of results. The other visual parameters were the same as those of Experiment 1.

---

[3] It is known that bounce-inducing effect is enhanced as the eccentricity of the visual coincidence is increased (Bertenthal et al., 1993).

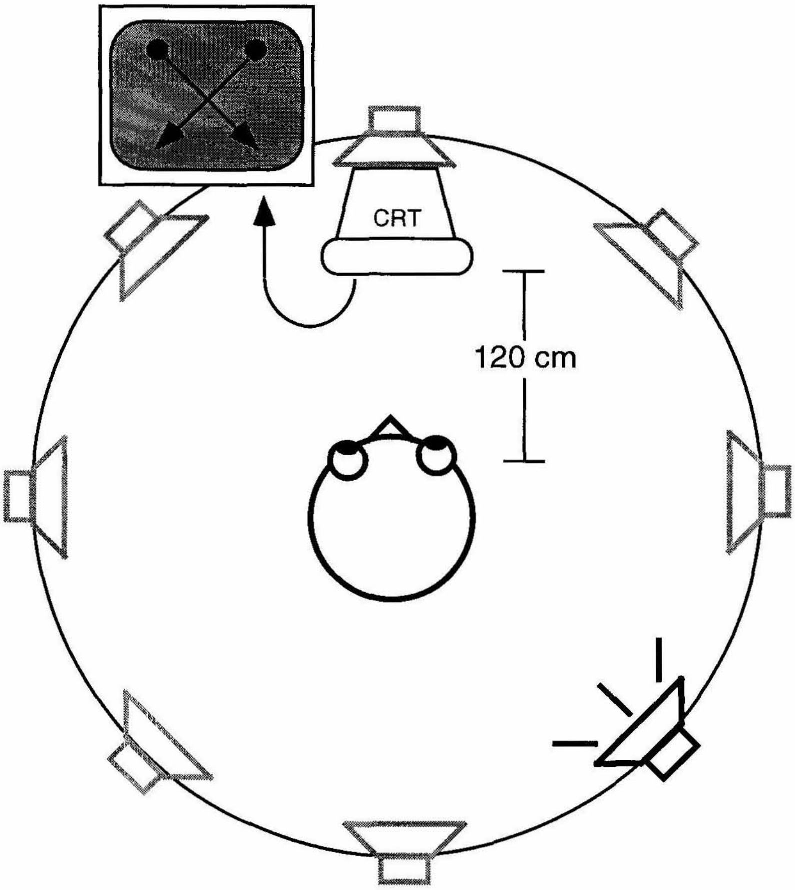

FIGURE 2.8 SET-UP OF EXPERIMENT 4 (SPATIAL WINDOW OF AUDIOVISUAL INTERACTION).

Auditory stimuli were delivered through a loud speaker (Harman/Kardon, model HK193), which was located 10-deg below the observer's eye-level. The possible locations of the speaker are depicted in Figure 2.8. To avoid possible complications due to speaker-specific characteristics, the same speaker was used for all locations. Therefore, the location of the speaker was fixed within a session, but varied randomly from session to session. The sound (58 dB at the observer's ears, 1800 Hz) was always presented at the moment of the visual coincidence. Otherwise, the properties of the sound were set as close as possible to those of Experiment 1.

Procedure

For each speaker location (14 locations), 40 trials were repeated. A full experiment was divided into 14 sessions, during which the location of the speaker was fixed.

Results

The results of Experiment 4 are shown in Figure 2.9. The bounce-inducing effect survived the audiovisual source separation up to about 30-45 degree of visual angle, thus showing only a very crude spatial tuning.

Discussion

The results of Experiment 4 suggest that the spatial interaction window of the bounce-inducing effect is quite similar to that of the McGurk effect (up to 30-45 deg). This might suggest a common mechanism of spatial interaction for the bounce-inducing effect (auditory effect on visual perception) and the McGurk effect (visual effect on auditory perception).

It is possible that the sound localization mechanism has influences on or interacts with the bounce-inducing effect in humans. A partial support for this idea comes from the small enhancement of the bouncing percept by the sound presented just behind the observer's head relative to elsewhere behind the observer (Figure 2.9). This may correspond with the relatively poor localization in the front-back direction in humans (e.g., Geissler, 1915; Stevens & Newman, 1936; but see Plenge, 1914; Blauert, 1996). Jack and Thurlow (1973) reported a similar result of a sound presented just behind the observer's head that produced a strong visual capture.

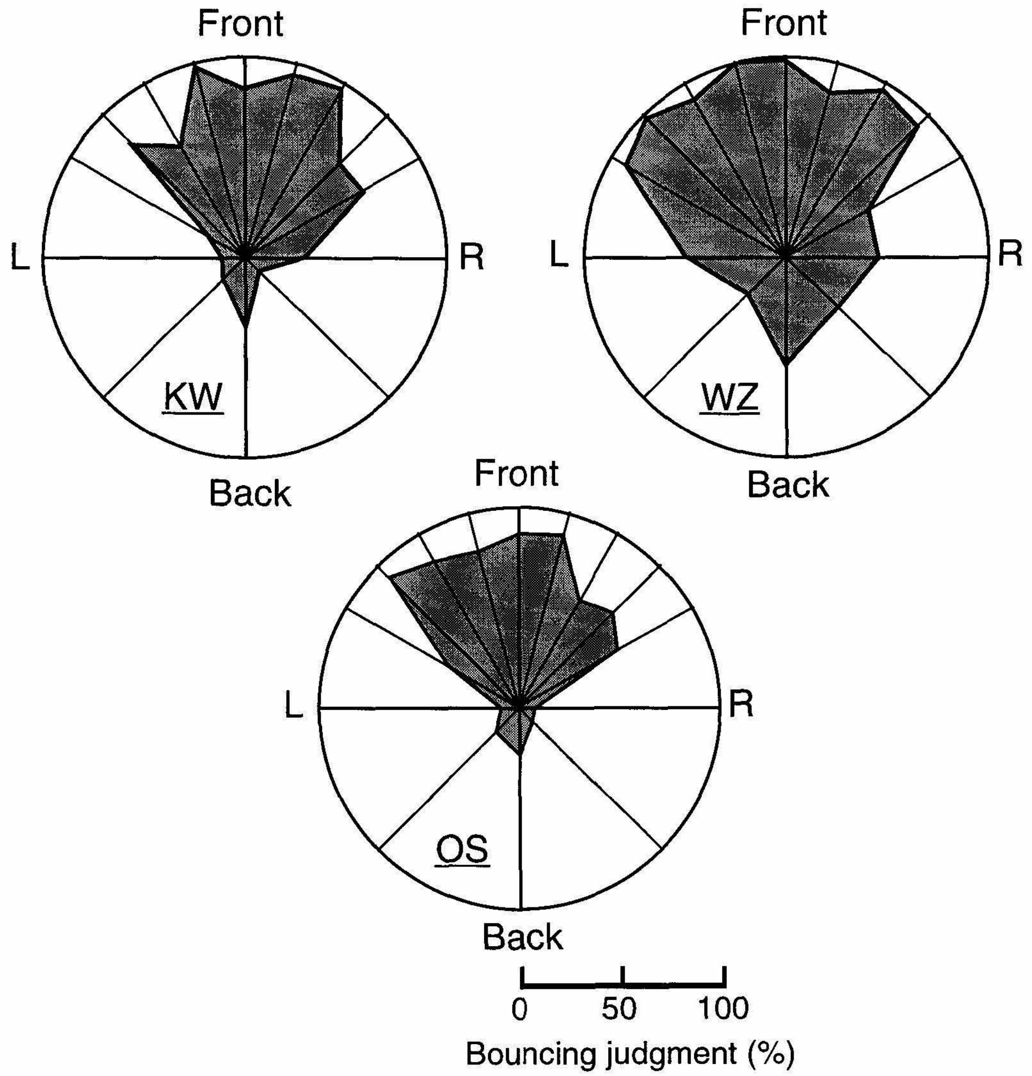

FIGURE 2.9 RESULTS OF EXPERIMENT 4.
The spatial tunings of the bounce-inducing effect from three observers are shown. There seems to be a wide spatial window for the bounce-inducing effect. Up to 30-45 deg of visual angle, the bounce-inducing effect occurred relatively unattenuated.

Additionally, the spatial interaction window observed in Experiment 4 may be a product of interactions between audiovisual localization (the ventriloquism effect) and audiovisual event perception (the bounce-inducing effect). Driver (1996) has shown that a sound that is subjectively mislocalized via the ventriloquism effect strongly affects the McGurk effect and suggests interactions between audiovisual localization and audiovisual identification.

Considering the possible interactions among auditory localization, audiovisual localization, and audiovisual event perception, it would be interesting to systematically change acoustic characteristics that influence auditory localization and to see whether these parameters affect the spatial window of the bounce-inducing effect. Humans have two ears on each side of the head. This means that a sound not emanating from a source on the midline reaches the two ears at different times (interaural time difference; ITD) and with different intensities (interaural intensity difference; IID) (e.g., Middlebrooks & Green, 1991; King & Carlile, 1995). The sound localization performance in the horizontal plane depends critically on acoustic characteristics of target sound such as tone frequency (Berger, 1981). Sound localization is accurate for low frequencies less than 1000 Hz and for very high frequencies over 10000 Hz. Localization errors are highest for the middle frequencies between 2000 and 4000 Hz (Stevens & Newman, 1936). This dip in sound localization performance at the middle frequencies is due to the transition between the ITD mechanism and the IID mechanism[4].

Would the spatial window for the bounce-inducing effect be different for different tone frequencies? This will be an interesting and plausible future investigation because it will provide clues to how auditory localization, audiovisual localization, and audiovisual event perception interact. However, as the main focus of the present thesis is the temporal aspect of crossmodal interaction, I will not discuss the spatial aspect of crossmodal interaction in detail.

General Discussion

Empirical findings

The results of Experiments 1 and Experiments 2 showed that the bounce-inducing effect is mediated by the temporal interaction window, the width of which is about 400 ms (250 ms before to 150-ms after the visual coincidence). Experiment 3 demonstrated that the temporal window for the bounce-inducing effect is much wider than the window of audiovisual temporal uncertainty, and confirmed that the bounce-inducing effect is a perceptual phenomenon. Finally, Experiment 4 showed that the bounce-inducing effect

---

[4] In humans, a low-frequency tone up to 1,500 Hz can be localized by using ITD. For higher frequencies, localization based on ITD is precluded by the phase ambiguity, which results from the use of phase-looked impulses for the measurement of ITD. By contrast, IID is more reliable for localization of high frequency tones (over 4000 Hz) because the head produces an acoustic shadow for high frequency tones (not for low frequency tones). Neither binaural cue works well in the transition region (2000-4000 Hz).

can be observed even when the spatial separation between the visual coincidence and the simultaneous sound was up to 30-45 degrees of visual angle.

Theoretical implications

The results described in Chapter II are the first demonstrations of the temporal and spatial window within which auditory information affects visual perception. Although the relationship between the affecting and affected modalities is reversed, there are several similarities between the bounce-inducing effect and the McGurk effect.

One similarity is the relatively broad temporal range for audiovisual interaction. In the bounce-inducing effect, the single sound presented 250-ms before and 150-ms after the visual coincidence significantly enhances the bouncing percept. In the McGurk effect, observers were strongly influenced by visual stimuli presented up to about 200 ms before or after the auditory stimulus (Campbell & Dodd, 1980; Munhall et al., 1996; Massaro & Cohen; 1993). Another similarity between the bounce-inducing effect and the McGurk effect can also be found in the tolerance for spatial separation of the auditory and visual information sources. The McGurk effect is just as pronounced when the mismatching auditory and visual stimuli come from separate spatial locations (up to about 40 deg), as when they share the same (Witkin et al., 1952; Jackson, 1953; Jack & Thurlow, 1973; Fisher & Pylyshyn, 1994; Jones & Munhall, 1997; Bertelson, 1998). This seems to be the case for the bounce-inducing effect, too.

The broad spatial/temporal interaction window may be a common characteristic of audiovisual event perception. Recently, several researchers have suggested that principles of crossmodal interaction discovered in one phenomenon, such as speech perception (the McGurk effect), may extend to many other crossmodal phenomena and therefore reflect general architectural constraints (Driver & Spence, 1998; Massaro, 1999; Calvert et al., 2000). The results and implications of experiments in Chapter II are consistent with this view, because the auditory effect on visual perception (the bounce-inducing effect) and the visual effect on auditory perception (the McGurk effect) have a similar spatio-temporal interaction window. The bounce-inducing effect should become an excellent tool to investigate audiovisual interaction, especially because of the simplicity of this experimental paradigm.

Possible neurophysiological correlates

The neural site(s) for crossmodal interaction is still under intensive debate. Several neural structures (e.g., superior colliculus, parietal lobe, superior temporal sulcus, putamen, and regions in the frontal lobe) have been reported to contain neurons that respond to multiple modalities (e.g., Andersen et al., 1993; Stein & Meredith, 1993; Wallace & Stein, 1994; Graziano & Gross, 1996; Wilkinson et al., 1996; Andersen et al., 1997; Graziano & Gross, 1998). Although most of these crossmodal regions have been examined in monkeys (Streicher & Ettlinger, 1987), homologous regions have been suggested in humans (Mesulam, 1992, 1994), including the superior colliculus, posterior parietal cortex, regions within lateral temporal cortex extending into the depths of the superior temporal sulcus, putamen, and parts of prefrontal cortex.

In order to perceive a unified event from two or more sensory inputs, the brain must detect common features between them. Temporal and spatial proximity should be major factors for any crossmodal integration (Radeau, 1994). As to this co-registration principle, the most well-established crossmodal site is the deeper layers of the superior colliculus (SC), where neural responses are enhanced by bimodal stimulation when a stimulation occurs at a common location, and suppressed when a stimulation occurs at different locations (Sprague & Meikle, 1965; Schneider, 1969; Casagrande et al., 1972; Goodale & Murison, 1975; Stein, 1984; Sparkes, 1986; Wallace et al., 1993; for review see Stein & Meredith 1993). The temporal window during which crossmodal interaction occurs in the SC neurons is surprisingly long (Meredith et al., 1987; Wallace & Stein, 1994). Although typical SC neurons tolerate bimodal stimulus onset asynchronies up to 100-200 ms, some SC neurons can have a temporal window of more than 500 ms. The relatively long temporal interaction window might imply that the bounce-inducing effect is partly mediated at the level of the SC.

However, it seems that the SC alone cannot explain the bounce-inducing effect. For example, the relative insensitivity of SC neurons to spatial separation of auditory and visual inputs is not large enough to account for the spatial separation tolerance of the bounce-inducing effect (and the McGurk effect). Also, there is a possibility that the bounce-inducing effect may interact with sound localization. Although the representation of acoustic space in the mammalian brain is not well understood (Middlebrooks &

Pettigrew, 1981; Imig et al., 1990; Rajan et al., 1990; Middlebrooks et al., 1994; Brugge et al., 1996), auditory cortex is clearly necessary to localize sounds in the contralateral hemisphere (Neff et al., 1956; Heffner & Masterton, 1975; Heffner, 1978; Kavanagh & Kelly, 1987; Jenkins & Merzenich, 1984; Heffner & Heffner, 1990). Therefore, there is evidence, though indirect, for the involvement of cortical processes in the bounce-inducing effect.

In fact, it has been shown that SC neurons receive sensory feedforward and feedback signals from many areas of visual, auditory, and somatosensory cortex (Edwards et al., 1979; Huerta & Harting, 1984; Stein & Meredith, 1991). In the cat cortex, the anterior ectosylvian sulcus (AES), composed of an auditory area (Clarey & Irvine, 1986), a somatosensory area (Clemo & Stein, 1982) and a visual area (Mucke et al., 1982; Olson & Graybiel, 1987), has been reported to play an important role in crossmodal interaction. The AES has many projections to the deep layers of the SC (Stein et al., 1993; Wallace et al., 1993), and when the AES is temporarily deactivated, crossmodal interaction in SC neurons diminishes (Wallace & Stein, 1994), while, importantly, unimodal responses of these neurons are not affected. Thus some of the crossmodal interactions in the SC are under the control of cortical functions. Furthermore, Stein & Wallace (1996) reported that the sensory receptive fields of cortical crossmodal neurons are much larger than SC crossmodal neurons, which may explain the tolerance of the bounce-inducing effect for audiovisual spatial disparity. Recent studies in humans support the cortical involvement of crossmodal interaction by suggesting that the parietal lobe is important for the ventriloquism effect (Soroker et al., 1995) and for the McGurk effect (Driver, 1996).

Thus, the involvement of cortical processes in the bounce-inducing effect seems highly likely.

Summary

Chapter II described experiments that are intended to reveal the spatial-temporal interaction window for the bounce-inducing effect by sound. The results suggest that there is a rather broad interaction window both in temporal and spatial domains. The size of these windows resembles those for the McGurk effect. The fact that the direction of crossmodal influences are opposite between the bounce-inducing effect (audition affects vision) and McGurk effect (vision affects audition) implies a common mechanism for audiovisual event perception in general.

Chapter III: The effect of auditory context on audiovisual event perception

Introduction

The experiments in Chapter III mainly address the effect of auditory context on the bounce-inducing effect. This issue is important because in a natural environment, a sound never occurs in isolation. Therefore, in order for audiovisual interaction to be significantly functional, it should deal with auditory context.

Contextual effects of auditory perception have been extensively studied and described in Bregman's book, "*Auditory scene analysis*" (Bregman, 1990). Typically, auditory stream formation or auditory stream segregation is demonstrated by alternating pure tones of high and low frequencies (Figure 3.1). When the frequency separation of high and low tones is small and the alternation rate is low, a single sequence which fluctuates in frequency is perceived. When the frequency separation is large and the alternation rate is high, two auditory streams without frequency fluctuation are perceived; one stream is high in frequency, and the other is low. Once an auditory stream is formed, each sound element is no longer perceived as an individual sound event and therefore the transient nature (or saliency) of each sound element is diminished.

The task of the auditory system is to detect, to localize, and to identify relevant sound sources in various contexts. The human auditory system organizes information about auditory events, depending on the events' similarity to other sounds, on their spatial location, and on their temporal structure. The most useful organizational principle is to create auditory streams or auditory objects in which sound elements with similar physical characteristics (e.g., frequency, temporal regularity, tempo, spatial location, intensity, timbre, changes in spatial envelope, etc.) are presumed to originate from the same source (McAdams & Bregman, 1979; McAdams, 1984; Bregman, 1990).

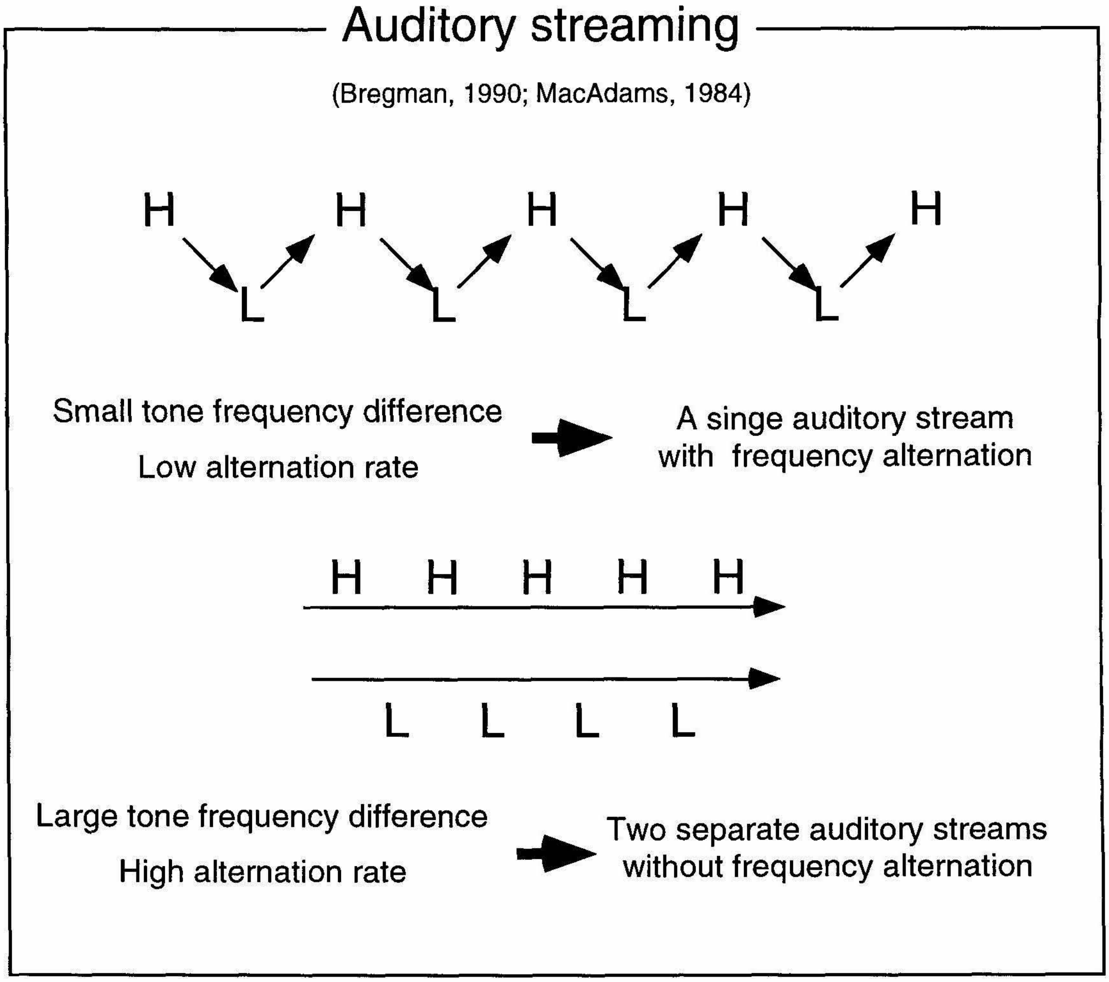

FIGURE 3.1 AUDITORY STREAMING (BREGMAN, 1990).
When the frequency separation of the high and low tones is small and the alternation rate is low, a single coherent sequence which fluctuates in frequency is perceived. When the frequency separation is large and the alternation rate is high, two auditory streams without frequency fluctuation are perceived.

In order to investigate the effects of auditory context on the bounce-inducing effect, a simultaneous sound (that was synchronized with the visual coincidence) was presented with additional preceding and/or following sounds. By changing the interval (Experiment 5), frequency difference (Experiment 6), and intensity difference (Experiment 7) between the simultaneous sound and additional sounds, the experiments were expected to reveal how auditory context affects the audiovisual interaction involved in the bounce-inducing effect.

Experiment 5: Attenuating the bounce-inducing effect

In Experiment 5, I examined whether the bounce-inducing effect caused by the simultaneous sound would be influenced by an additional preceding sound and/or an additional following sound. What if, for example, a three-sound sequence is presented and the second sound is synchronized with the visual coincidence? There would be three possible outcomes: (A) The bounce-inducing effect would not change because only synchrony between a sound and the visual coincidence is important. (B) The three-sound sequence might enhance the bounce-inducing effect because the amount of integrated energy from auditory stimulation might be more than a single sound. (C) The three sounds would form a 'sound stream' (Bregman, 1990) and thus the second sound would lose its transient nature. If the bounce-inducing effect requires such a transient signal in an auditory scene, then the effect may actually decrease.

Method

Observers

Nine observers took part in Experiment 5. Seven of the observers (including the author) had participated in Experiment 1 (Chapter II).

Stimuli

The visual stimuli were identical to those of Experiment 1. The sound had the same acoustic properties as in Experiment 1. However, there were 3 types of auditory events (Figure 2.2): (1) An additional sound was presented before the visual coincidence with the simultaneous sound. (2) An additional sound was presented after the visual coincidence with the simultaneous sound. (3) Two additional sounds (auditory flankers) were presented before and after the visual coincidence with the simultaneous sound. For each auditory event condition, the sound-onset asynchrony (SOA) was varied (50, 100, 150, 200, 250, 450, 650, 850, or 1050 ms). In control conditions, only the single sound or no sound was presented. Note that the simultaneous sound was always present except for the no-sound condition.

Procedure

The experimental procedures were identical to those of Experiment 1. For each of the 9 SOAs and three types of auditory events, and for the single-sound and no-sound conditions, 20 trials were presented randomly (580 trials). A full experiment was divided into 4 sessions. Each session consisted of 145 trials.

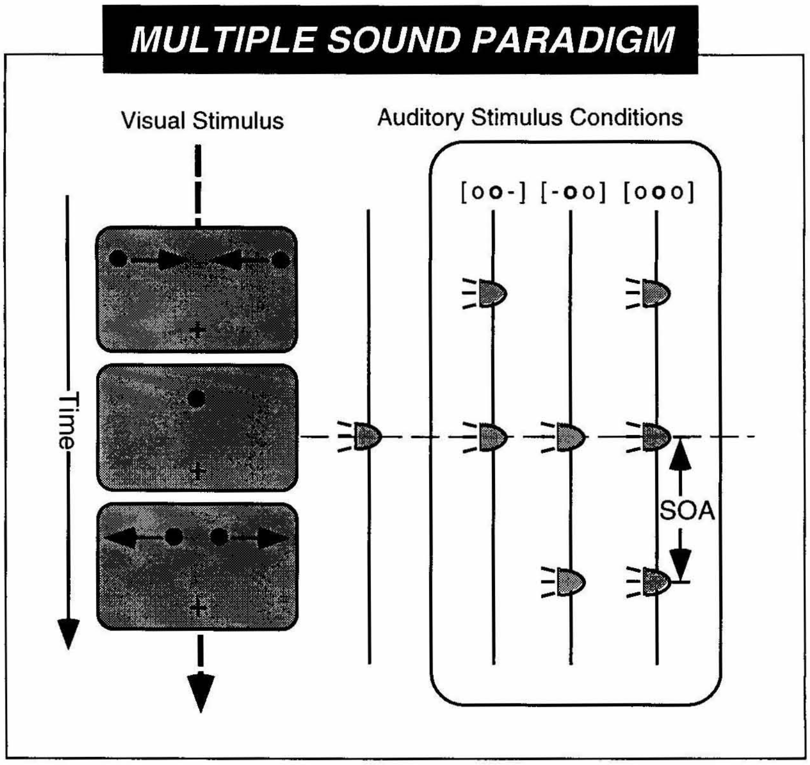

FIGURE 3.2 EXPERIMENTAL PARADIGM OF EXPERIMENT 5.
The simultaneous sound was preceded and/or followed by additional identical sounds.

Results

Figure 3.3 shows the mean percentage of bouncing responses pooled across all the observers as a function of asynchrony of the additional sound relative to the simultaneous sound (and the visual coincidence). The bouncing percentage for the single-sound and the no-sound conditions are represented as horizontal lines.

When an additional sound was presented before the simultaneous sound, a significant modulation of the bounce-inducing effect as a function of SOA was observed (left half of

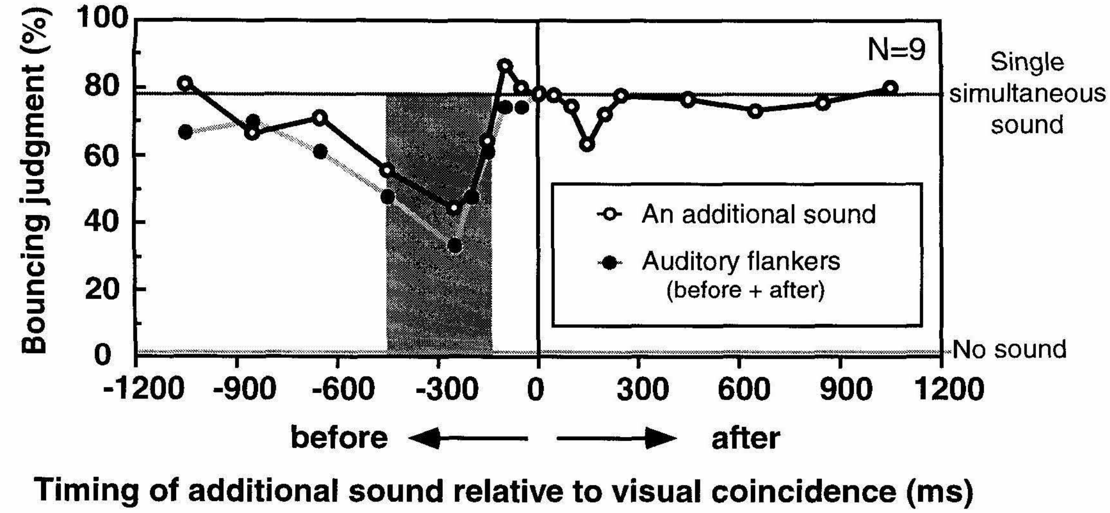

FIGURE 3.3 RESULTS OF EXPERIMENT 5.

The averaged percentage of bouncing judgment is plotted against the additional sound timing relative to the visual coincidence. Note that the visual coincidence was always accompanied with the simultaneous sound. Black curves with open circles represent the data from the preceding (left half) and following (right half) sound conditions. The gray curve with filled circles represents the data from the three sound condition (auditory flankers, left-half). The gray region shows the time interval for which the attenuation effect was significant ($p < 0.05$).

Figure 3.3, black curve; $F(8,64) = 3.29, p < 0.001$). The bounce-inducing effect with short SOAs (50 -150 ms) was as prominent as in the single-sound condition, but significantly attenuated with moderate SOAs (200 - 450 ms; the attenuation effect), and then recovered again with longer SOAs (650 -1050 ms) (see also Appendix A, Table 4).

With an additional sound presented after the visual coincidence (right half of Figure 3.3, black curve), the bounce-inducing effect appeared to be slightly attenuated at the 150-ms SOA ($F(1,8) = 4.36, p < 0.06$), but not at the other SOAs ($F(8,64) = 0.61, p = 0.77$; see also Appendix A, Table 4).

When the simultaneous sound was flanked by the additional sounds (i.e., when both the pre- and post-coincidence sounds were presented), the results resembled those with a pre-coincidence sound (left half of Figure 3.3 gray curve, also see Appendix A, Table 5), although a small enhancement of the attenuation effect was observed, compared with the results of the preceding-sound condition ($F(1,8) = 3.77, p < 0.06$).

Discussion

The results of Experiment 5 showed that the bounce-inducing effect can be attenuated by an additional sound presented before the visual coincidence, or by auditory flankers (the attenuation effect). The attenuation effect was maximal with sound-onset asynchronies (SOAs) of 250-300 ms and was detectable with an SOA of up to 450 ms. It is important to note that a single sound presented 250-300 ms before the visual coincidence by itself has no bounce-inducing effect (Experiment 1). Nevertheless, a physically identical sound modulated the effect of the simultaneous sound on visual motion perception. Thus, the attenuation effect could not be due to a simple energy summation or probability summation in the perceptual judgment. Moreover, it is unlikely that the attenuation effect was caused by conventional auditory masking, by which detection thresholds and/or the perceived characteristics of individual sounds are altered (e.g., Stevens & Davis, 1938; Lüscher & Zwislocki, 1947; Irwin & Zwislocki, 1971; Zwislocki & Sokolich, 1974). The effective range for auditory masking is no more than 250 ms proactively and retroactively (e.g., Wright, 1964; Massaro, 1970, 1975; Fastl, 1976; Massaro et al., 1976; Kallman & Morris, 1984; Viemeister & Plack, 1993), whereas

I observed significant attenuation of the bounce-inducing effect with much longer intervals (up to 450 ms)[5].

While the attenuation effect by a preceding sound is clear, the attenuation effect by a following sound is not so obvious. A single following sound produced a small attenuation effect with an SOA of 150 ms. Plus, auditory flankers (a preceding and following sound) led to a slightly stronger attenuation effect than did a single preceding sound. Therefore, the attenuation effect might exist even with a following sound, and the possibility that a following sound can affect the efficacy of the simultaneous sound to produce the bounce-inducing effect is quite interesting. However, further research is necessary because, statistically speaking, the effect of a following sound was not significant ($p < 0.06$).

The pattern of results for Experiment 5 can be interpreted in that the preceding sound (and maybe the following sound) reduces the transient nature, or *saliency* of the simultaneous sound, which is critical for audiovisual interaction. That is, repetitive sound elements may be spontaneously grouped into an auditory stream because of their similar acoustic properties (McAdams & Bregman, 1979; McAdams, 1984; Bregman, 1990), and, once grouped, the simultaneous sound loses its saliency for audiovisual interaction (unless the simultaneous sound is the first element of the auditory stream).

My account for the attenuation effect qualitatively resembles auditory streaming phenomena (Bregman & Campbell, 1971; Bregman, 1990). The differences are that the

---

[5] A set of control experiments has also shown that the bounce-inducing effect does not alter the auditory perception of the simultaneous sound when it is flanked by additional identical sounds (Appendix B).

effect was measured via crossmodal interaction in my display, and that the auditory integration time seems to be longer (up to 400 ms) than that measured in typical auditory streaming experiments; auditory streaming cannot be observed with a 400-ms sound interval (Anstis & Saida, 1985).

Experiment 6: Recovering the bounce-inducing effect with sound frequency differences

The main finding of Experiment 5 is the attenuation effect; the efficacy of the simultaneous sound to induce the bouncing percept is attenuated by the presence of additional sounds. My working hypothesis here is that auditory grouping reduces the saliency of the simultaneous sound that is critical for audiovisual interaction. Experiment 6 was conducted to examine this auditory grouping and saliency hypothesis. Auditory grouping depends on similarity among sound elements (e.g., Bregman, 1990). Therefore, if the simultaneous sound had acoustic characteristics distinct from auditory flankers, it would "pop out" and the bounce-inducing effect would recover. In Experiment 6, I changed the tone frequency of the simultaneous sound which was embedded in a sequence of additional background sounds.

Method

Observers

The same 9 observers as in Experiment 5 participated.

Stimuli

The visual stimulus was identical to that of Experiment 5. There were 2 types of auditory stimulus conditions (Figure 3.4). (1) Single sound: 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. As a control condition for the single-sound trials, no sound was presented (no-sound condition). (2) Embedded sound: Seven sounds were presented successively with an SOA of 300 ms. The fourth sound was always presented at the moment of the visual coincidence. The frequency of the embedded sound varied as in the single sound conditions, whereas those of the other sound (auditory flankers) were fixed at 1800 Hz. Note that when the frequency of the embedded sound was 1800 Hz, all seven sounds had the same frequency. In a control condition for the embedded-sound trials, the simultaneous sound was omitted (sound-off condition).

Procedure

The procedures were the same as those of Experiment 1. For eight sound conditions (900, 1800, and 2700-Hz single-sound conditions [3] + 900, 1800, and 2700-Hz embedded-sound conditions [3] + no-sound and sound-off conditions [2]) 20 trials were repeated randomly.

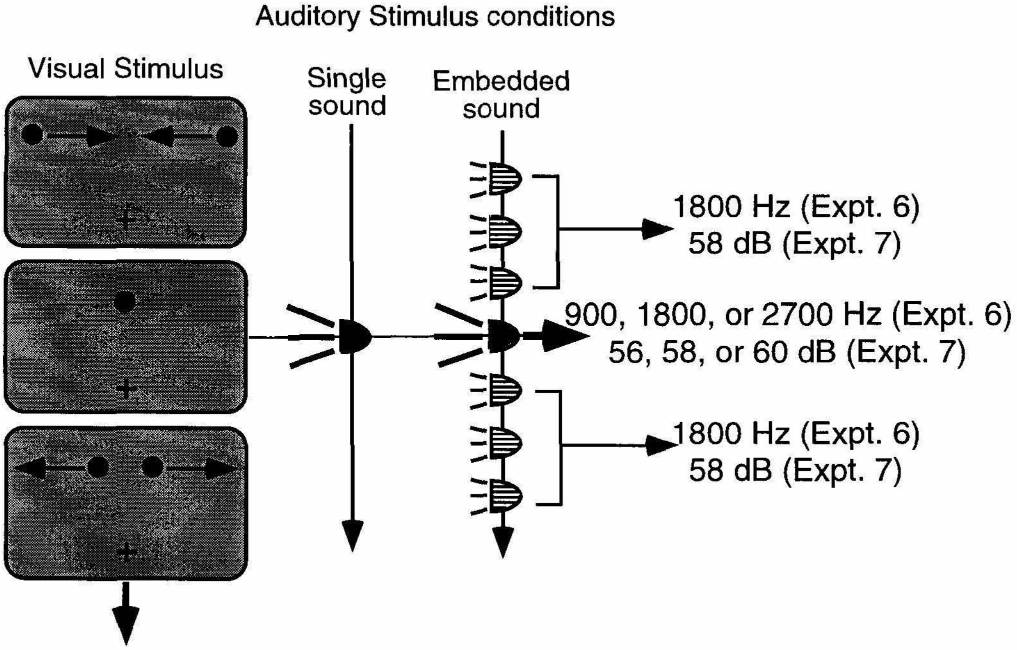

FIGURE 3.4 SCHEMATIC OF THE STIMULUS PRESENTATIONS IN EXPERIMENT 6 AND EXPERIMENT 7. In Experiment 6, the frequency of the single/embedded sound was varied. In Experiment 7, the intensity of the single/embedded sound was varied.

Results

All the single simultaneous sounds significantly enhanced the bouncing percept, compared with the no-sound condition (Figure 3.5, $F(3,24) = 84.4, p < 0.001$). For the single sound conditions, the bounce-inducing effect did not change as a function of sound frequency ($F(2,16) = 1.63, p = 0.22$). With the embedded sounds, the bounce-inducing effect was still observed, compared with the sound-off condition ($F(3,24) = 46.48, p < 0.001$). However, when the embedded sound had the same tone frequency as the background sounds, the bounce-inducing effect was significantly attenuated (single vs.

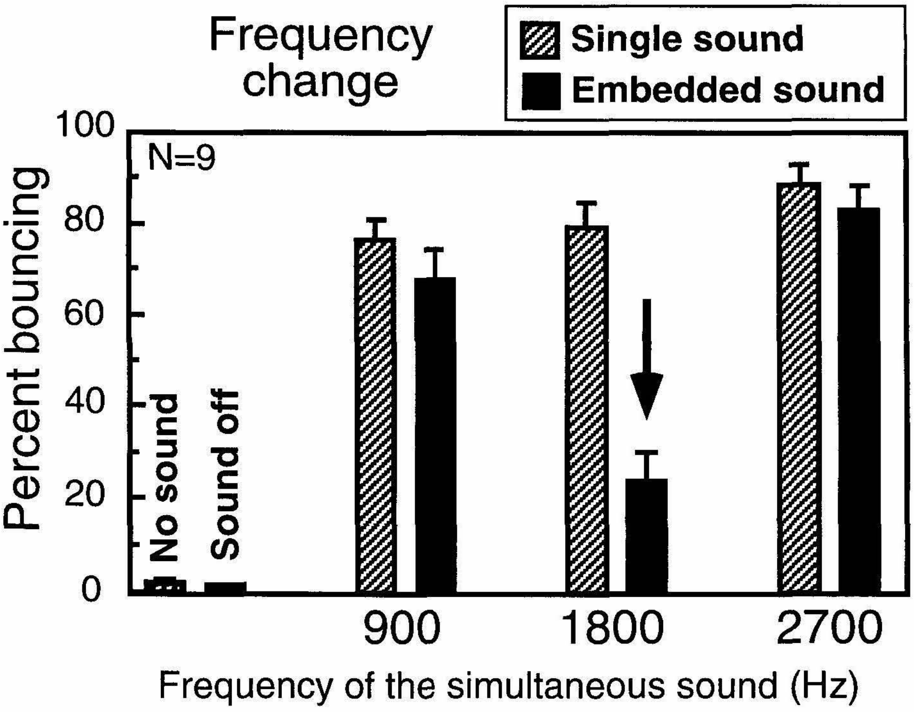

FIGURE 3.5 RESULTS OF EXPERIMENT 6.

Standard errors are shown. The frequency of the auditory flankers was 1800 Hz. The arrow indicates the condition in which the attenuation effect was observed. The bounce-inducing effect was attenuated only when the simultaneous sound had the same frequency as the auditory flankers.

embedded, 1800 Hz, $F(1,8) = 47.21, p < 0.001$). The attenuation effect did not occur when different frequencies were used for the embedded simultaneous sound (single vs. embedded, 900 Hz, $F(1,8) = 1.06, p = 0.32$; 2700 Hz, $F(1,8) = 0.63, p = 0.44$).

Discussion

The results of Experiment 6 are consistent with the auditory grouping and saliency account for the attenuation effect, since the attenuation was observed only when the embedded sound had the same tone frequency as the auditory flankers. Interestingly, the sound-off condition did not lead to the enhancement of the bouncing percept. This confirms the hypothesis that neither the mere presence nor the surprising nature, but the *saliency* (strong stimulus × surprising nature) of the simultaneous sound, is crucial for the bounce-inducing effect. The following experiment further examined the auditory grouping and saliency hypothesis by changing the intensity of the embedded sound, as intensity differences are known to serve as an important cue for segmenting the auditory input into discrete perceptual events (e.g., Dowling, 1968; Schröger et al., 1996).

Experiment 7: Recovering the bounce-inducing effect with sound intensity differences

The saliency hypothesis is partially supported by the fact that the sound-off condition did not result in the recovery of the bounce-inducing effect. The aim of Experiment 7 was to further examine the saliency hypothesis in the bounce-inducing effect. The specific prediction was: If the saliency of the simultaneous sound is really important, the intensity increase of the simultaneous sound relative to auditory flankers would recover the bounce-inducing effect because it is more salient than the background sounds. But, this may not be true for the intensity decrease.

Method

Observers

The same 9 observers as in Experiment 6 participated.

Stimuli

The visual and auditory stimuli were almost the same as those used in Experiment 6. 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 Figure 3.4). The intensities of the auditory flankers were 58 dB.

Procedure

The procedures were identical to those of Experiment 6.

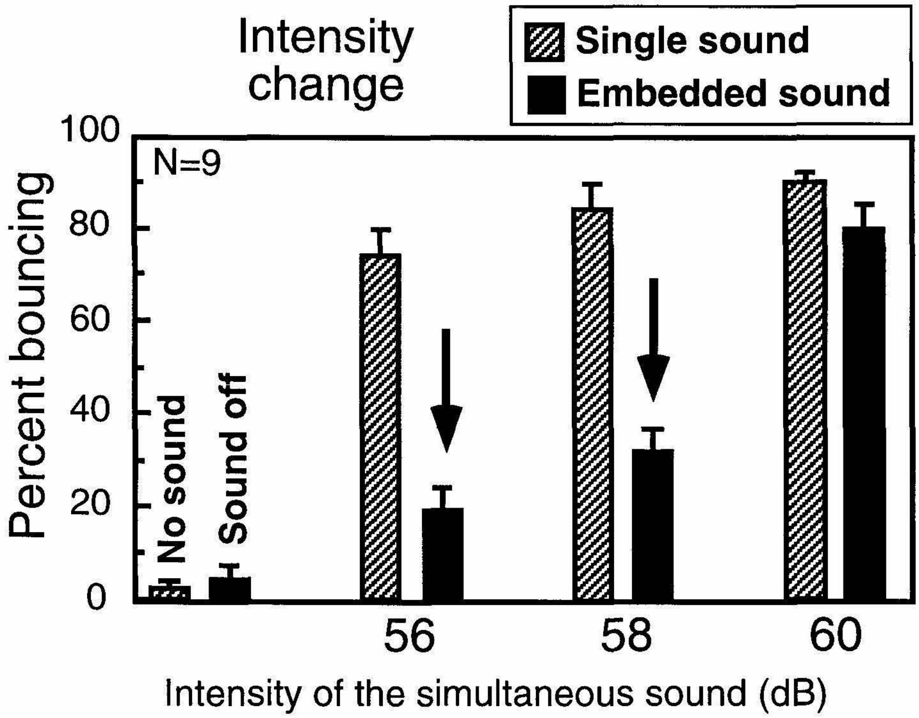

FIGURE 3.6 RESULTS OF EXPERIMENT 7.

Standard errors are shown. The auditory flankers had 58 dB intensity. The arrows indicate the conditions in which the attenuation effect was observed. The bounce-inducing effect was reduced when the simultaneous sound had the same intensity (58 dB) as or lower intensity (56 dB) than the auditory flankers.

Results

Again, all the single simultaneous sounds produced the bounce-inducing effect (Figure 3.6, $F(3,24) = 90.27, p < 0.001$). The bounce-inducing effect was still present with the embedded simultaneous sounds, compared with the sound-off condition ($F(3,24) = 45.02$, $p < 0.001$). However, only when the embedded sound had a higher intensity than the auditory flankers was the bounce-inducing effect comparable to the effect with the single sound (single vs. embedded, 60 dB, $F(1,8) = 2.49, p = 0.134$). With the same or the lower intensity, the bounce-inducing effect was significantly reduced (56 dB, $F(1,8) = 47.02, p < 0.001$; 58 dB, $F(1,8) = 49.82, p < 0.001$).

Discussion

The absence of recovery with weaker sounds strengthens the hypothesis that the bounce-inducing effect is a function of *saliency* of the simultaneous sound. Intuitively, the results of Experiment 7 may reflect the fact that, in a natural environment, the increase in intensity against the background sound level may signal collision of objects, whereas the intensity decrease may not.

Experiment 8: Auditory threshold vs. audiovisual threshold

Experiments 5-7 have shown that the saliency of the simultaneous sound is crucial for audiovisual interaction involved in the bounce-inducing effect. But, to what degree should the simultaneous sound be salient for the bounce-inducing effect?

The auditory differences (frequency and intensity) introduced in the previous experiments were always suprathreshould for the human auditory system. That is, the observers heard the simultaneous sound being distinct from auditory flankers. It is natural to assume that the bounce-inducing effect would be attenuated as the acoustic difference between the simultaneous sound and auditory flankers is reduced. However, the relationship between the detection threshold for auditory deviation and the magnitude of the bounce-inducing effect had not been examined.

In Experiment 8, firstly, I obtained a psychometric function for auditory detection of a deviant sound as a function of tone frequency difference between the simultaneous sound and auditory flankers (Figure 3.7). Secondly, I measured the bounce-inducing effect as a function of tone frequency difference and drew a psychometric function for the audiovisual interaction. By normalizing and comparing those two psychometric functions, the experiment would reveal the relationship between auditory processing and audiovisual interaction processing for the bounce-inducing effect.

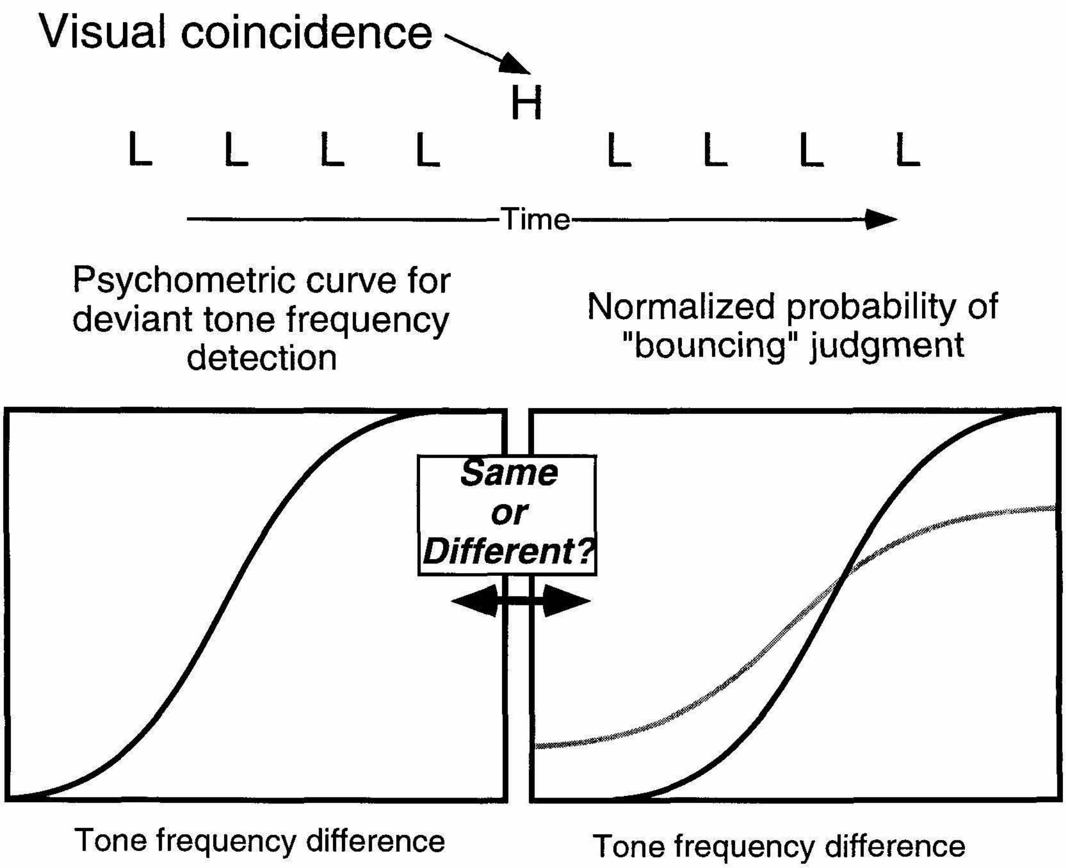

FIGURE 3.7 SCHEMATIC OF THE PARADIGM OF EXPERIMENT 8.
Sound sequence used in Experiment 8 (top). For each observer, a psychometric function for auditory detection of a deviant sound and a psychometric function for the bounce-inducing effect were independently obtained, normalized, and compared (bottom).

Method

Observers

Five people (23 to 48 years of age; including the author) participated. All had normal vision and hearing.

Stimuli

The visual stimulus was the same as that of Experiment 4 (the spatial-tuning experiment in Chapter II). There were always seven consecutive sounds with an SOA of

300 ms, with the fourth sound being synchronized with the visual coincidence. In one half of trials, the frequency of the simultaneous sound was randomly varied between 1810 Hz and 1900 Hz (10 Hz step). In the other half of trials, the frequency of the simultaneous sound was 1800 Hz (i.e., no tone-frequency difference). Those trials were intermixed in order to avoid a bias in the observer's judgment criterion. The other parameters of auditory stimuli were identical to those of the previous experiments.

Procedure

In separate sessions, observers performed two different tasks. (1) Bounce judgment: As in the previous experiments, observers reported their percept (streaming or bouncing) by pressing the mouse buttons. (2) Deviant sound detection: Observers were asked to close their eyes and judge whether the fourth sound in the sound sequence had higher frequency than the other sounds. There were 400 trials with a deviant sound (10 frequencies × 40 repeat) and 400 trials with no deviant sound. A full experiment was conducted in 4 separate sessions with 200 trials each.

Results

The results of Experiment 8 are shown in Figure 3.8. After normalization, data from the bounce judgment and deviant sound detection experiments were fit with a logistic function,

$$y = C \cdot \left(\frac{(ax+b)+1}{\sqrt{(ax+b)^2+1}} \right)$$

for each observer and for each task-condition to obtain psychometric curves (Finney, 1971). Here, a and b are two parameters for a logistic function, and C is a normalization factor. The psychometric curves for deviant sound detection and those for the bouncing percentages are quite similar. A two-way ANOVA on the pooled results among observers indicated that there was no difference between these two psychometric functions (frequency difference, $F(10,40) = 107, p < 0.001$, deviant sound detection vs. normalized percentage bounce, $F(1,4) = 0.015, p = 0.90$; interaction, $F(10,40) = 1.216, p = 0.29$).

Fifty-percentile thresholds for deviant sound detection and those for bouncing percentages were estimated from the derived psychometric functions. The difference between these "thresholds" did not reach a significant level ($F(1,4) = 0.092, p = 0.77$).

Discussion

The bounce-inducing effect appears to closely correlate with the detection of a salient auditory stimulus. In other words, there seems to be little information loss at the transition from the auditory processing stage to the audiovisual processing stage. These results suggest that the audiovisual interaction involved in the bounce-inducing effect should have a high functional significance in a natural condition.

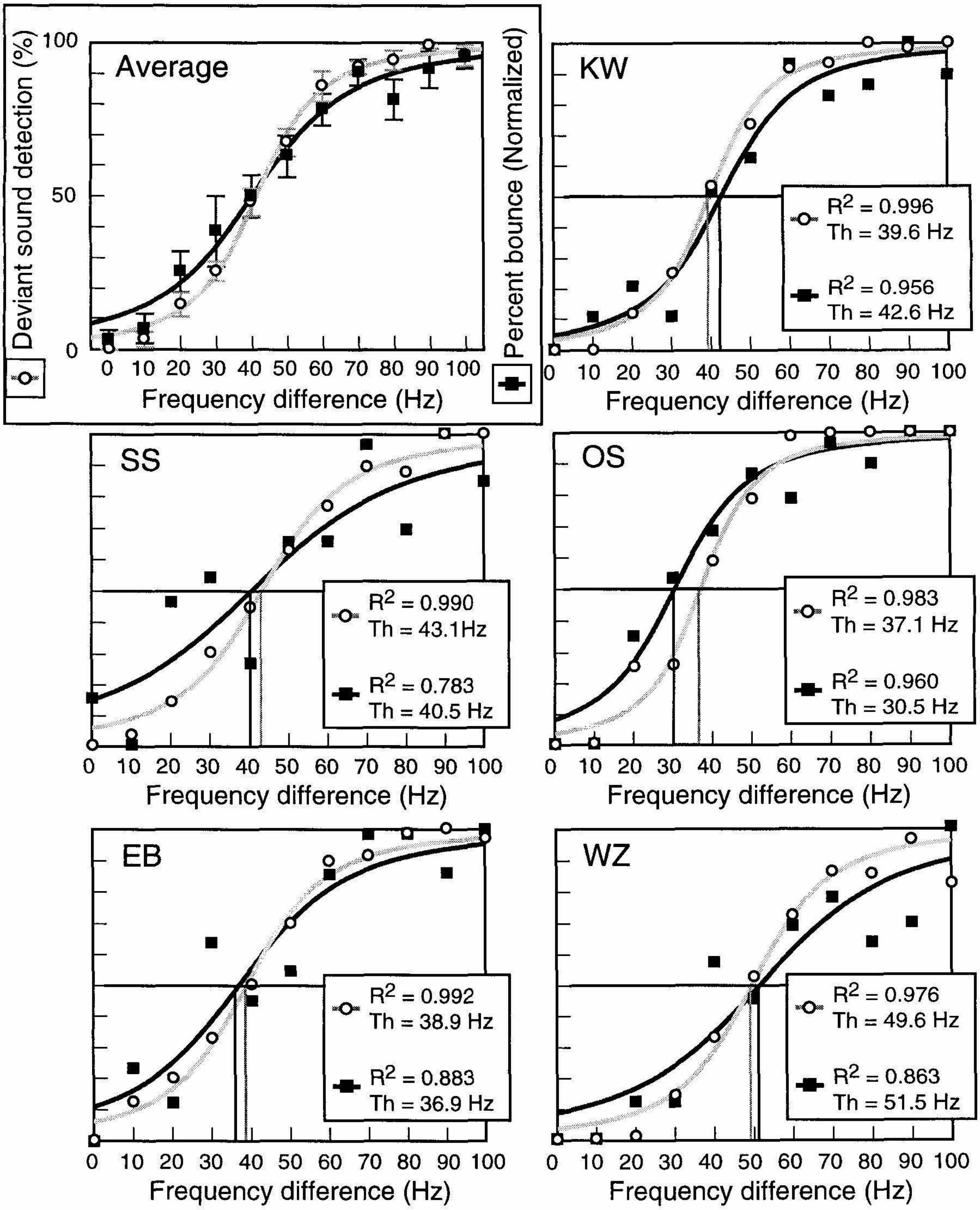

FIGURE 3.8 RESULTS OF EXPERIMENT 8.

Open circles show the results of the auditory detection experiment, and black squares represent the results of the bounce judgment. The plot at the top-left is the data averaged for all observers.

Experiment 9: The fixed time-window of audiovisual interaction revisited

In the experiments in Chapter II, I have confirmed that the bounce-inducing effect depends on the auditory context in which the simultaneous sound occurs. The results have converged to suggest that the saliency of the simultaneous sound is the determining factor of the bounce-inducing effect. Besides these theoretical implications, I happened to notice that the attenuation effect would provide an excellent way to investigate the nature of the temporal window for audiovisual interaction.

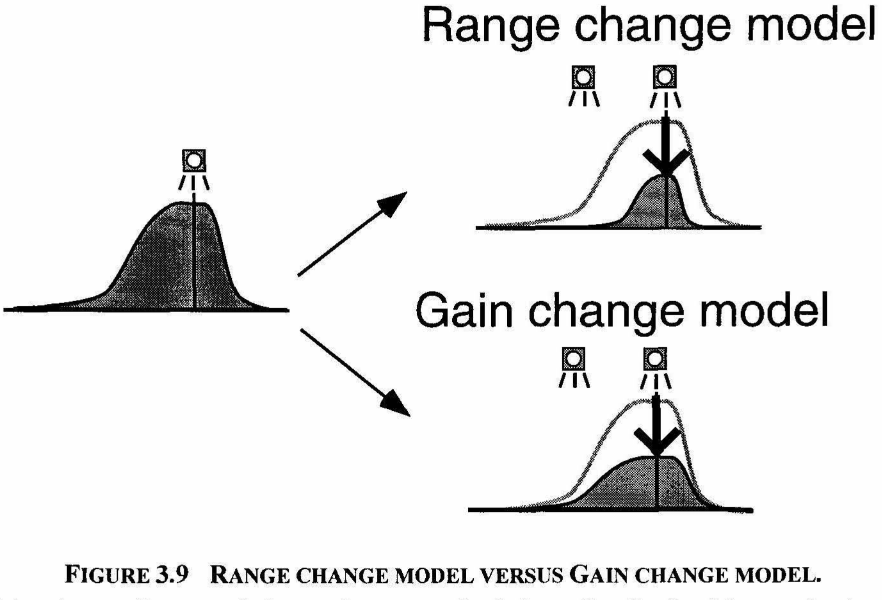

FIGURE 3.9 RANGE CHANGE MODEL VERSUS GAIN CHANGE MODEL.
Does an additional preceding sound change the temporal window of audiovisual interaction by changing the range without changing the slope (range change model) or by changing the slope without changing the interaction range (gain change model)?

A preceding sound reduces the bounce-inducing effect of the simultaneous sound (Experiment 5). The attenuation effect can be described as a downward shift of the 0-ms asynchrony point of the interaction window (downward arrows in Figure 3.9). What would happen for the range and the slope of other parts of the interaction window? Does a preceding sound change the temporal range of audiovisual window without changing the slope (range change model)? Or, does the slope of the window change while the range remain constant (gain change model)? In Experiment 9, I investigated the modulation of the range and shape of the audiovisual interaction window due to the attenuation effect.

Method

Observers

In order to compare results with those of Experiment 1 (Chapter I), the same 9 observers from Experiment 1 took part in Experiment 9.

Stimuli

Two sounds were presented successively with a 300-ms interval. The visual coincidence happened at the first or the second sound, or lagged by 50 - 300 ms from the second sound (Figure 3.10). In control trials, a single sound or no sound was presented. The other stimulus parameters were identical to those of Experiment 1.

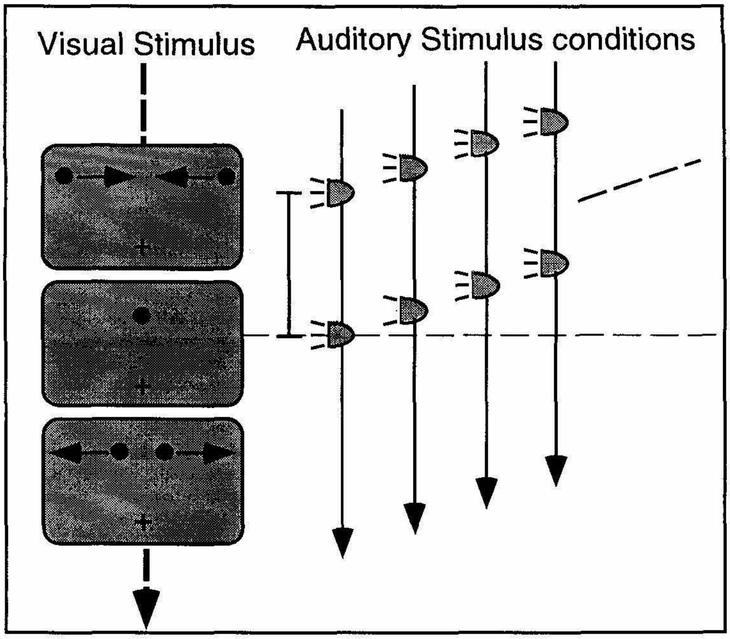

FIGURE 3.10 EXPERIMENTAL PARADIGM OF EXPERIMENT 9.
An additional sound was presented 300-ms before the second sound. The second sound preceded the visual coincidence by various times.

Procedure

The procedures were the same as those of Experiment 1. For each of 10 different sound conditions, 20 trials were presented in random order (200 trials). A full experiment was divided into 2 sessions. Each session consisted of 100 trials.

Results and Discussion

Figure 3.11 shows the averaged percentage of bouncing judgments across all observers. The straight lines represent the data from the two baseline conditions (the single-sound and no-sound conditions). A part of the results of Experiment 1 (when the

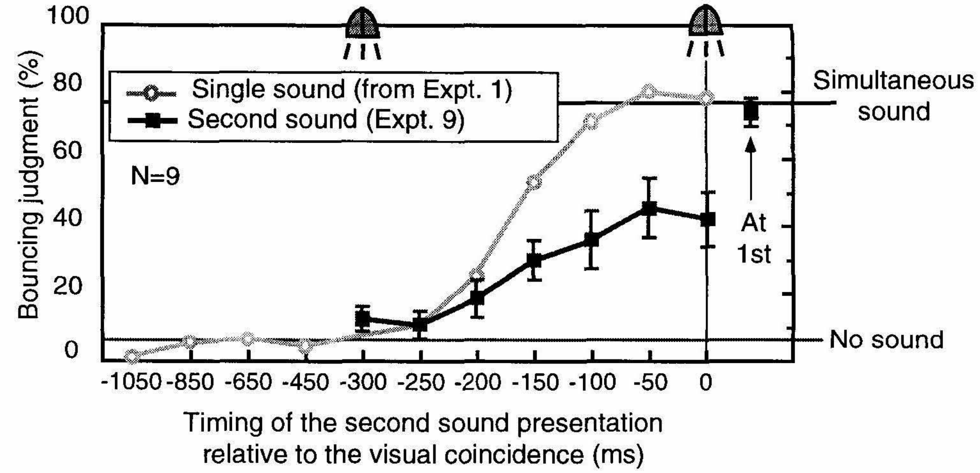

FIGURE 3.11 RESULTS OF EXPERIMENT 9.

The averaged percentage of bouncing judgment is plotted against the second timing relative to the visual coincidence (black curves with filled squares). A part of the results of Experiment 1 (single sound preceding the visual coincidence) was also plotted for comparison purpose (gray curves with open squares). With the additional preceding sound, the attenuation effect was evident, but the temporal interaction window did not seem to change. When the first sound of the two sounds was synchronized (At 1st), the bounce-inducing effect occurred in the same magnitude as the single simultaneous sound.

sound was presented before the visual coincidence) is also plotted for comparison purposes. When the first sound of the sound pair was synchronized with the visual coincidence, it produced the bounce-inducing effect of the same magnitude as the single sound (At 1st vs. single sound, $F(1,8) = 0.161, p = 0.69$). As expected, the bouncing percept was less frequent when the second sound was synchronized with the visual coincidence than when the single sound was synchronized (the attenuation effect at 0-ms SOA, $F(1,8) = 12.28, p < 0.005$). The frequency of bouncing judgments decreased as the second sound was presented further before the visual coincidence ($F(6,48) = 4.426$, $p <$ 0.001). The significant bounce-inducing effect persisted up to 150-ms asynchrony (see Appendix A, Table 6).

This tolerance for the auditory-ahead asynchrony (up to 150-ms auditory advance) is similar to that of the single sound in Experiment 1. Thus, the attenuation effect made the slope of the temporal window shallower *without* changing the temporal interaction range (at least for the case of a preceding sound). These results further strengthen the hypothesis that the temporal window for the bounce-inducing effect is fixed in time, which had been implied in Experiment 2 (different speed experiment). This issue will be addressed in Chapter IV in detail.

General Discussion

Empirical findings

In the experiments in Chapter III, in order to investigate the effect of auditory context on the bounce-inducing effect, I utilized the reduction of the transient nature of auditory elements under the condition auditory streams form. In Experiment 5, the simultaneous sound was preceded, followed, or flanked by identical sounds. The results showed that the bounce-inducing effect is reduced when the simultaneous sound is preceded by another identical sound or flanked by two identical sounds (the attenuation effect). Furthermore, the experiment revealed that there is an optimal time-interval (about 250-300 ms) between sound elements for the attenuation effect. Subsequently, Experiment 6 and Experiment 7 demonstrated that the attenuation of the bounce-inducing effect depends on the saliency of the simultaneous sound. In Experiment 8, I addressed the relationship between the detectability of a sound variation in an auditory sequence and the bounce-inducing effect and found an almost-perfect correlation between the auditory threshold and the audiovisual threshold measured by the bounce-inducing effect. Finally, I utilized the attenuation effect to examine the hypothesis that the temporal window for audiovisual interaction is fixed in time and obtained supporting evidence (Experiment 9).

Theoretical Implications

The task of the auditory system is to determine the origin of a sound and interpret it. A sound almost always occurs within auditory contexts which include independent and overlapping sound sources. Therefore, this task should be performed by picking up relevant information in sound scenes (cocktail-party problem: Cherry, 1953; Broadbent, 1954; Wood & Cowan, 1995). Bregman (1990) suggested that solving the cocktail-party problem requires perceptually grouping distinct sounds.

The results of experiments in Chapter III suggest that audiovisual interaction (the bounce-inducing effect) is sensitive to auditory grouping (i.e., the attenuation effect) and is dependent on the *saliency* of the simultaneous sound (i.e., the recovery from the attenuation effect). The near-perfect correlation between the auditory threshold and the audiovisual threshold (measured by the bounce-inducing effect) suggests that the bounce-inducing effect may play a highly significant role in everyday situations; such a high correlation is not expected to happen if auditory grouping and the bounce-inducing effect are independent phenomena.

The auditory grouping observed in the experiments of Chapter III is similar to Bregman's auditory streaming effect. The degree of auditory grouping (measured by the attenuation effect) depends on the similarity between the simultaneous sound and auditory flankers, and the flanked sound seems to become less salient than an isolated sound. However, the attenuation effect is optimal at the SOA of 250-300 ms and the effect can be detected for even longer SOAs (up to 450 ms). The auditory streaming effect becomes weak or completely disappears at this rate of sound element alternation (Anstis & Saida, 1985). Thus, the auditory grouping processing for audiovisual event perception may have different characteristics from the auditory grouping processing for auditory perception.

Possible neural correlates of the attenuation effect

The auditory cortex has been recognized as playing an important role in the processing of auditory stimulus sequences. Creutzfeldt et al. (1980) suggested that cortical neurons process the time structure of transient auditory signals. In the auditory cortex, after the presentation of a tone, the neural response to a following tone can be inhibited, facilitated, or shifted in latency (Creutzfeldt et al., 1980; Langner, 1992). Duration of forward inhibition depends strongly on the frequency and intensity of the first tone of the stimulus pair. Also, the auditory cortex contains "combination-sensitive" neurons which respond to specific combinations of tone interval, tone frequency, and tone intensity (Suga et al., 1978; Margoliash, 1983; Suga, 1989; Doupe & Konishi, 1991; Margoliash &

Fortune, 1992; Lewicki & Konishi, 1995; Lewicki, 1996; Doupe, 1997). In subcortical stages of the auditory pathway, however, duration of forward inhibition is relatively independent of the neuron's characteristic frequency.

The temporal window for neuronal auditory interaction seems to increase for higher levels of auditory processing. Neurons at subcortical stages of the auditory pathway generally respond to a much higher repetition rate (e.g., Langner, 1992). In contrast, a number of studies have shown that cortical auditory neurons respond in a time-locked fashion to repetition rates of up to only several elements per second (Creutzfeldt et al., 1980; Eggermont, 1991; Phillips et al., 1989; Brosch & Schreiner, 1997; but see also de Ribaupierre et al., 1972). Typically, at higher repetition rates, neurons only respond to the first element of the sequence but only weakly (or not at all) to the following elements. Hocherman and Gilat (1981) found that two-thirds of the neurons in the auditory cortex can exhibit reduced responses if the inter-stimulus interval is less than 1.6 seconds. These cortical neurons may be responsible for rhythmic grouping of tone sequences and stream segregation and integration of sequences (Bregman, 1990; Brosch & Schreiner, 1997; also see Rogers & Bregman, 1993; Wang, 1996; McCabe & Denham, 1997; Rose & Moore, 1997), which in turn may produce the bounce-inducing effect. Thus, the attenuation effect should involve processing in the auditory cortex.

Relation to human electrophysiology

In human electroencephanography (EEG) and magnetoencephanography (MEG) studies, the organization of the auditory input to neural representations has been extensively investigated using mismatch negativity (MMN). This negative component of the event-related potential is elicited mainly from the auditory association cortex (Alan et al., 1998) when some physical pattern of a sound deviates from a repetitive standard sound, such as tones differing in frequency, intensity, or duration (Nordby et al., 1988; Saarinen et al., 1992; Schröger et al., 1992; Alho et al., 1993; Näätänen et al., 1993b; Alain et al., 1994; Schröger, 1994; Tervaniemi et al., 1994; Schröger et al., 1995; Winkler & Schröger, 1995; Schröger et al., 1996; Alain et al., 1998). These MMNs are evoked even when participants do not actively listen to the sounds but perform some other task, such as attention-demanding visual tracking (Näätänen, 1992; Näätänen et al., 1993a). Therefore, MMN elicitation by a sound change is thought to reflect preattentive perceptual chunking and periodicity detection. These characteristics of MMN appear to be consistent with those of the attenuation effect.

However, in recent studies (Yabe et al., 1997; Winkler et al., 1998; Yabe et al, 1998), MMN was elicited by random stimulus omission in a repetitive auditory sequence only when the constant stimulus onset asynchrony (SOA) was less than 150 ms. On the basis of behavioral and electrophysiological evidence, Näätänen and colleagues (Näätänen, 1990; Näätänen, 1992; Loveless & Hari, 1993) suggested that during the early stages of auditory information processing, the constant flow of input is processed in about 200 ms segments, termed the "temporal window of auditory integration." This temporal window for MMN (about 200 ms) is much shorter than the temporal interaction range for the attenuation effect (up to 450 ms). Based on many MMN studies, it is suggested that two subsequent phases for auditory sensory memory exist: the perceptual (shorter than 200 ms) and synthesized (longer than 200 ms) auditory stores. The short auditory store integrates incoming stimulation encountered within about 200 ms while extracting its basic acoustic features, and may be related to the MMN generation (Massaro, 1975; Cowan, 1984, 1988; Winkler & Schröger, 1995). The other long, synthesized auditory store may be responsible for the attenuation effect.

Summary

The experiments described in Chapter III suggest that audiovisual interaction involved in the bounce-inducing effect may depend on auditory context and the saliency of the simultaneous sound. Based on known physiological studies, it is inferred that auditory cortex, or higher levels of neural functional structure, is important for the contextual modulation of the bounce-inducing effect (the attenuation effect).

Chapter IV: Temporal gain modulation of the bounce-inducing effect

Introduction

An intuitive account for the bounce-inducing effect may be that the simultaneous sound increases the probability of seeing a bouncing event because people have experienced such synchronized events many times (the audiovisual associative learning account). The audiovisual associative learning account, however, is not totally satisfactory. For example, Experiment 3 in Chapter II showed that the bounce-inducing effect occurs even if the audiovisual asynchrony is easy to detect. How could associative learning occur when observers know that two events in different modalities are independent?

Based on the results so far, the bounce-inducing effect by sound can be described as follows: A salient auditory transient around the moment of the visual coincidence biases visual perception toward bouncing. In the present chapter, I examine whether the bounce-inducing effect is restricted to audiovisual interaction. If a sensory transient in modalities other than auditory produces the bounce-inducing effect, the conclusion would be more general: A salient sensory transient around the moment of the visual coincidence biases visual perception toward bouncing (the amodal saliency account).

Experiment 10: Effects of auditory, tactile, and visual sensory transients on visual motion event perception

Contrary to intuition (the audiovisual associative learning account), the amodal saliency account predicts that any sensory transient is capable of biasing visual perception toward bouncing. To test this, three different sensory transients (auditory, tactile, and visual) were presented at the moment of the visual coincidence.

Method

Observers

Four people (23 to 29 years age; including the author) participated.

Stimuli

The streaming/bouncing motion display was the same as that used in Experiment 4 (Chapter II, an 'X' shape trajectory). Three types of sensory transients were presented at the moment of the visual coincidence. (1) Auditory transient: A 200 Hz tone burst was presented for 40 ms through a speaker (model TD87-1; see Figure 4.1 for the experimental set-up). The sound pressure level was 58 dB. (2) Tactile transient: A 200 Hz vibration was applied to the observer's left index-finger. This was achieved by asking observers to put their fingers on the speaker (accordingly the duration of vibration was 40 ms). The intensity was determined for each observer by multiplying the individual observer's threshold for the vibratory stimulus by four. In order to prevent observers

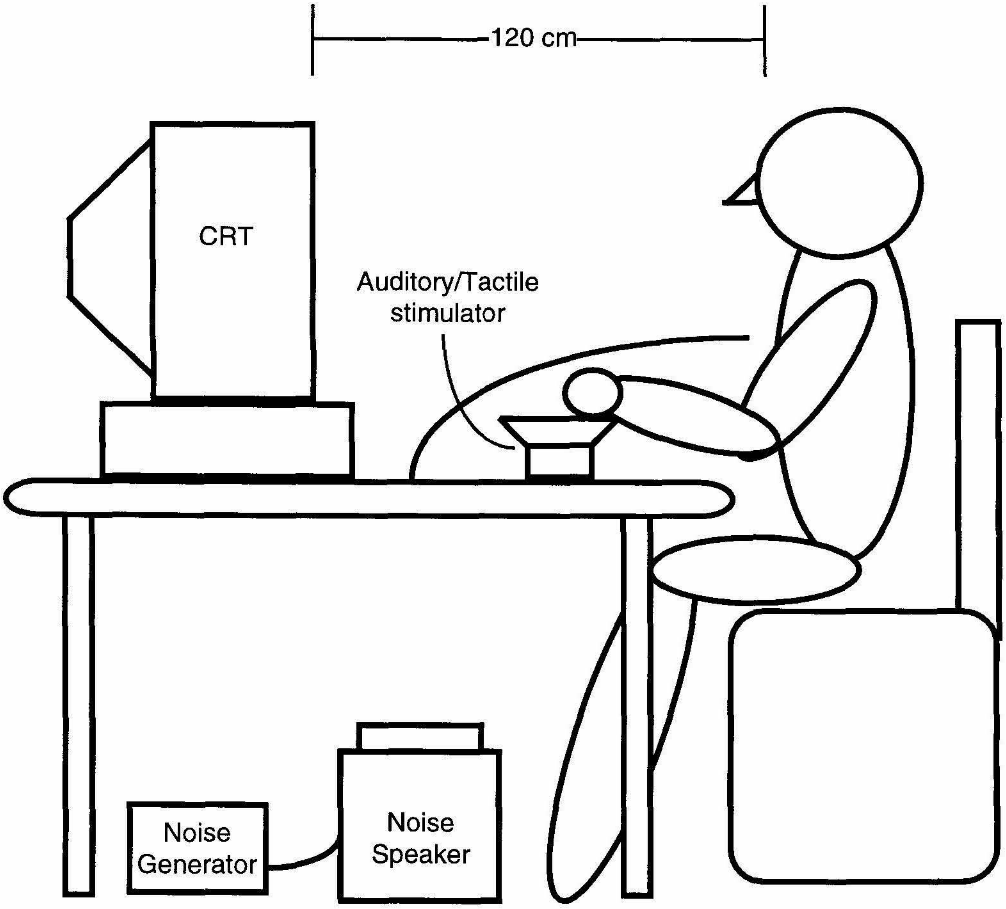

FIGURE 4.1 EXPERIMENTAL SET-UP FOR THE EXPERIMENTS IN CHAPTER IV.
The auditory and tactile stimuli were delivered by using a speaker. For Experiments 10 & 11 (auditory experiments), observers heard the sound via the speaker. For Experiments 10 & 12 (tactile experiments), observers touched the speaker which was placed under a cover while auditory noise completely masked the sound produced by the speaker. For Experiments 10 & 13 (visual experiments), the speaker was not used, but a visual circle was presented as a sensory transient.

---

from hearing the sound produced by the speaker, a loud white noise (about 90 dB) was continuously presented through another speaker (Bose, model 111AD), which completely masked the sound produced by the speaker. Observers also used ear plugs (about 32 dB noise reduction). (3) Visual transient: A black ring (0.01 cd/m^2, 0.5 deg in diameter) was presented at the location of the visual coincidence for 1 frame (= 40 ms).

Procedure. The experimental procedures were the same as those in Experiment 1. The auditory, tactile, and visual experiments were conducted in separate sessions. In each session, the sensory transient was presented in 40 out of 80 trials.

Results

The results of Experiment 10 are presented in Figure 4.2. All the sensory transients enhanced the bouncing percept (Auditory, $F(1,3) = 22.42, p < 0.01$; Tactile, $F(1,3) = 18.38, p < 0.01$; Visual, $F(1,3) = 63.84, p < 0.001$).

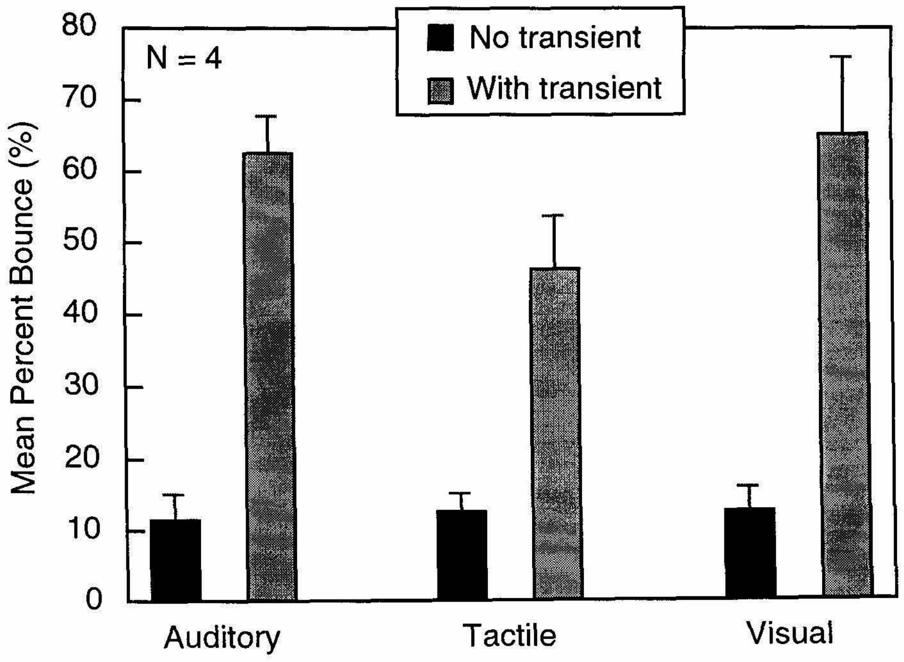

FIGURE 4.2 RESULTS OF EXPERIMENT 10.
All the three sensory (auditory, tactile, and visual) transients produced the bouncing-effect.

Discussion

The results of Experiment 10 have shown that, in order for the bounce-inducing effect to occur, a sensory transient need not be presented in the auditory modality. This clearly contradicts the simple associative learning account based on auditory-visual coupling of physical bouncing events and supports the amodal saliency account for the bounce-inducing effect. It is not *modal* because the sensory transients in other than the visual modality produce the bounce-inducing effect (in the sense that it is not restricted within the visual modality), and it is not always *crossmodal* because the visual transient produces the bounce-inducing effect. The amodal saliency account requires a common mechanism that mediates the disambiguation of the streaming/bouncing motion display. The possible underlying mechanism will be briefly mentioned in the end of this chapter's general discussion and extended in the following chapter.

At this point, my tentative conclusion is that a salient sensory transient around the moment of the visual coincidence biases perception toward bouncing. However, the temporal window of the bounce-inducing effect has been examined only in the audiovisual interaction (Chapter II). Therefore, the "around the moment" part in the above statement warrants investigation for the bounce-inducing effect by visual and tactile transients. One purpose of the following experiments (Experiments 11-13) is to investigate the temporal interaction window of the bounce-inducing effect by various sensory transients.

Additionally, the following experiments were designed to examine the effect of saliency on the temporal interaction window. The results of Experiment 2 have suggested that the interaction window of the bounce-inducing effect (by sound) may be fixed in time. Partially supporting this idea, the results of Experiment 9 have demonstrated that the attenuation effect does not change the temporal range of audiovisual interaction. I will call this inflexibility of temporal window *temporal gain modulation* because only the effect gain is altered while the temporal interaction range is unchanged. An interesting possibility is that the change in saliency of a sensory transient in any modality affects the magnitude of the bounce-inducing effect but does not affect the temporal interaction range.

The idea of temporal gain modulation resembles that of other gain modulations, which have been found in many areas in the non-human brain such as *spatial gain modulation* in the monkey parietal cortex (Andersen et al., 1985; Andersen & Zipser., 1988; Zipser & Andersen., 1988; Andersen, 1989; Andersen et al., 1990; Andersen, 1995), *orientation gain modulation* in the monkey extrastriate area V4 by attention (Connor et al., 1996; Connor et al., 1997; McAdams & Maunsell, 1999), *motion direction gain modulation* of the monkey middle temporal visual area by attention (Treue & Trujillo, 1999), and *crossmodal gain modulation* in the cat superior colliculus by bisensory stimulation (Stein & Meredith, 1993) and in the monkey premotor cortex and putamen (Duhamel et al., 1991; Fogassi et al., 1992; Colby & Duhamel, 1993; Graziano & Gross, 1993; Graziano et al., 1994). Shared with these gain modulations is the inflexibility of response ranges in the stimulus dimension (retinal location, orientation, motion direction, etc.). For example, the receptive field of a crossmodal neuron in the cat superior colliculus does not change with input from the other modality, but only the magnitude of cell activity changes (Stein & Meredith, 1993).

In the following experiments, I systematically changed the intensity, which correlates with saliency, and the timing of sensory transients which were presented at the auditory, tactile, or visual modality. By doing so, I intended to test the validity of the *temporal gain modulation* model. Specifically, the change in intensity of a sensory transient would change only the magnitude of the bounce-inducing effect without changing the temporal interaction range. If this is so, it would add an interesting case of gain modulation in which the stimulus dimension is a *relative time* between sensory events.

Experiment 11: Effects of intensity of auditory transients

The purpose of Experiment 11 was to directly test the temporal gain modulation model for the bounce-inducing effect by sound.

Method

Observers

Eight observers participated in the experiment (age range 22-29 years, including the author). Four observers, including the author, had taken part in Experiment 10.

Stimuli

The stimuli were almost the same as those of the auditory session of Experiment 10. However, the auditory stimulus occurred at various times relative to the visual coincidence (40-, 80-, 160-, 320-, 640-, 1280-ms before, after, or at the same time as the visual coincidence) and had three possible intensities 53, 55, and 58 dB (SPL). The background sound noise level was 53 dB.

Procedure

There were 13 conditions with different audiovisual asynchronies and a no-sound condition. For each condition 20 trials were repeated in a single session (a total of 280 trials). Within each session, the sound intensity was fixed. Otherwise, the experimental procedures were the same as those of Experiment 1.

Results

Figure 4.3 shows the increase in the bouncing percentage from the bouncing percentage of the no-sound condition, averaged over all eight observers. The magnitude of the bounce-inducing effect depended on both audiovisual asynchrony and sound intensity. There were significant effects of audiovisual asynchrony (two-way ANOVA, $F(12,96) = 52.49, p < 0.001$) and of sound intensity ($F(2,16) = 16.82, p < 0.001$). The interaction between these factors was also significant ($F(24,192) = 2.11, p < 0.01$). A series of planned comparisons indicated that the significant bounce-inducing effect disappeared when the sound occurred 320 ms before and 160 ms after the visual coincidence for all the sound intensities (Appendix A, Table 7).

To gain more insights on the validity of the temporal gain modulation model, the results were approximated with a Gaussian function without any parameter constraint:

$$y = C \cdot \exp\left(-\frac{1}{2\sigma^2}(x - M)^2\right)$$

Here, M donates the mean, and σ donates the standard deviation. C is a multiplicative constant. The standard deviations of the approximated Gaussian functions are similar to each other (124±9 ms; Figure 4.4), suggesting that the temporal window of the audiovisual interaction is, in fact, fixed in time. The means are about -63±12 ms. Both the standard deviations and the means are not significantly different among the sound intensities used.

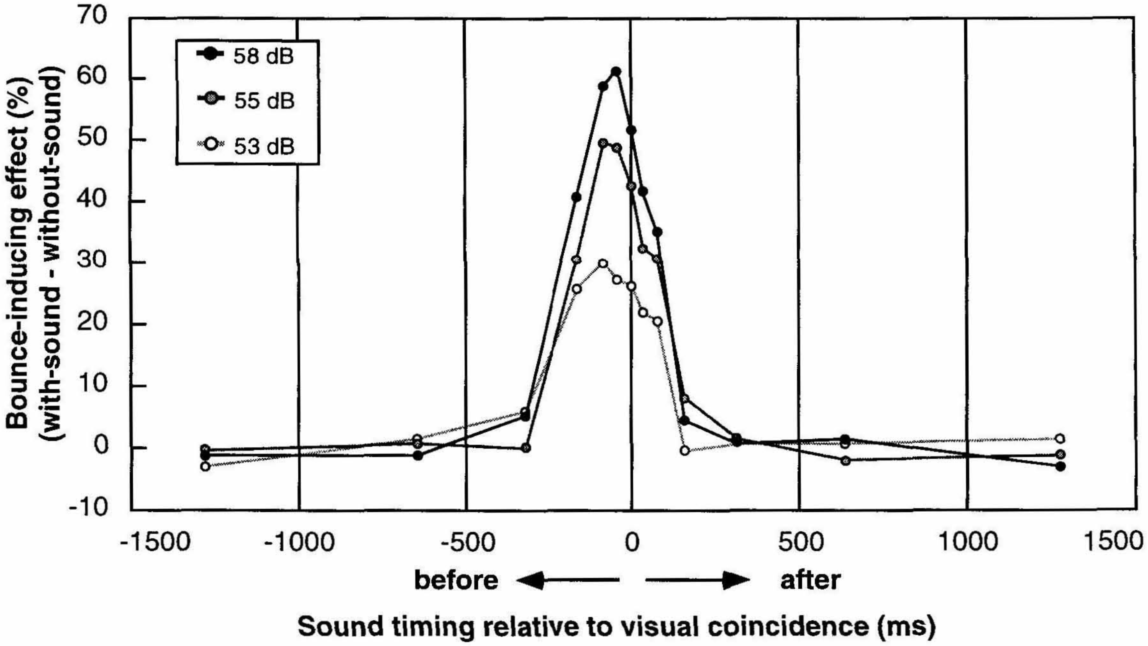

FIGURE 4.3 RESULTS OF EXPERIMENT 11 (AUDITORY).
Effect of sound intensity on the bounce-inducing effect. Background sound noise level was 53 dB.

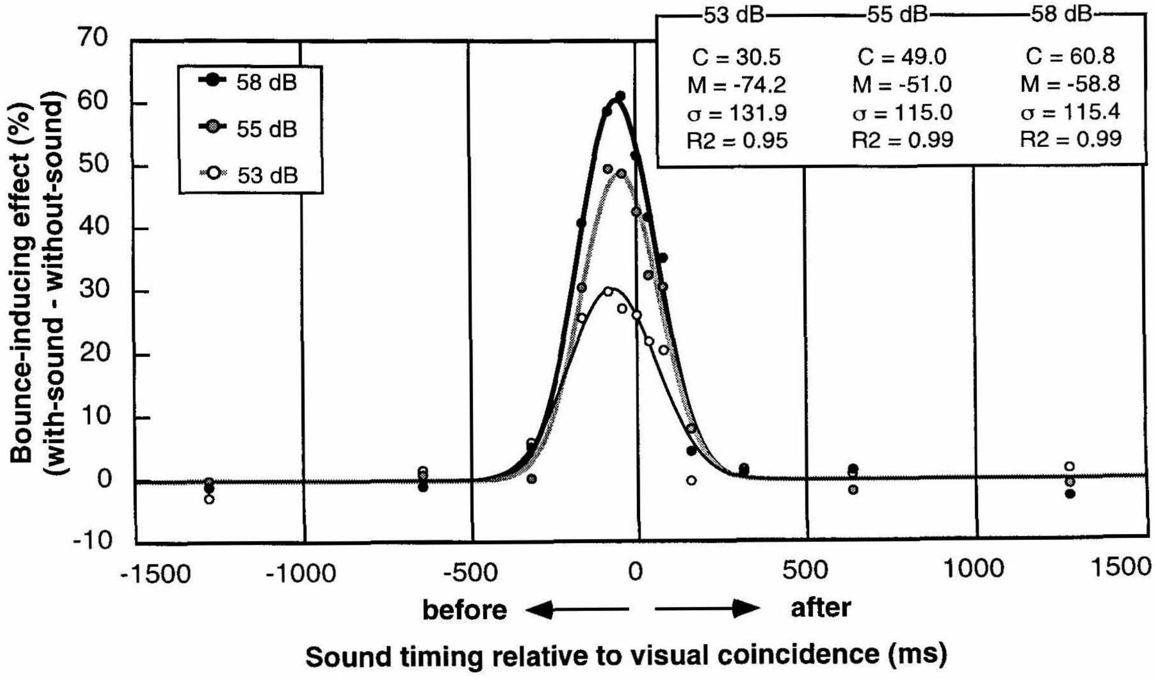

FIGURE 4.4 RESULTS OF EXPERIMENT 11 (AUDITORY): GAUSSIAN FIT OF FIGURE 4.3.
M donates the mean, and σ donates the standard deviation. C is a multiplicative constant.

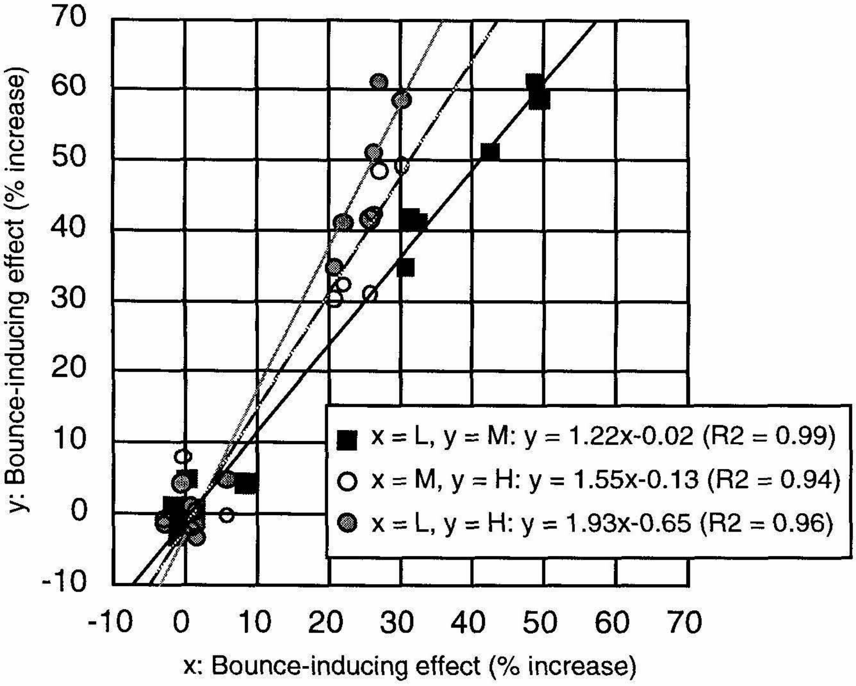

FIGURE 4.5 RESULTS OF EXPERIMENT 11 (AUDITORY): GAIN MODULATION FUNCTIONS.
In the inset "L," "M," and "H" donate the bounce-inducing effect by the low (53 dB), medium (55 dB), and high (58 dB) intensity sounds, respectively. For example, filled squares (x = L, y = M) represent the bounce-inducing effect by the medium intensity sound as a function of that by the low intensity sound.

Figure 4.5 replots the data from Figure 4.3, such that the bounce-inducing effects with the high (58 dB) and medium (55 dB) sound intensities are plotted against the bounce-inducing effect with the low sound intensity (53 dB). Fitting well with linear functions ($R^2 > 0.94$ for all plots) implies that the gain change due to sound intensity (saliency) were roughly proportional.

Discussion

Experiment 11 has provided a detailed look of the bounce-inducing effect by sound. The temporal range of audiovisual interaction measured in Experiment 11 is about 480 ms irrespective of sound intensity or saliency. Thus, the temporal gain modulation model seems adequate for the bounce-inducing effect by sound. Interestingly, the shift of the temporal window center from the 0-m asynchrony is also fixed. The sound presented about 60 ms before the visual coincidence produces the maximal bounce-inducing effect.

These results evoked a further interest in the possibility of temporal gain modulation in the bounce-inducing effect caused by tactile and visual transients. What is the temporal interaction window for these cases? Would be there temporal asymmetry of interaction window? If so, is it the same as in the bounce-inducing effect by sound? The next two experiments address these questions.

Experiment 12: Effects of intensity of tactile transients

In Experiment 12, I investigated (1) the temporal window of tactile-visual interaction, (2) temporal asymmetry of the tactile-visual interaction window, and (3) the validity of the temporal gain modulation model for the bounce-inducing effect by tactile transients.

Method

Observers, stimuli, and experimental procedures were almost the same as those of Experiment 11. However, instead of auditory stimuli, tactile (vibratory) stimuli were used as a bounce-inducing factor. The tactile stimuli (200 Hz, 40 ms) were presented in the same way as those in the tactile session of Experiment 10. The tactile intensity was determined for each observer, which was either two times, four times, or ten times larger than the individual observer's threshold for the vibratory stimulus. Within a session, the tactile intensity was fixed. In order to make sure that observers did not hear the sound at the highest tactile intensity, a control experiment was performed where the observers did not touch the speaker.

Results

The control experiment showed that without touching the speaker, no bounce-inducing effect was observed (data not shown). The results of Experiment 12 are shown in Figure 4.6 in the same manner as in Experiment 11. Similar to the results of Experiment 11, the bounce-inducing effect depended on both tactile-visual asynchrony and tactile intensity (tactile-visual asynchrony, $F(12,96) = 49.95, p < 0.001$; tactile intensity, $F(2,16) = 49.01$, $p < 0.001$; interaction, $F(24,192) = 6.68, p < 0.01$). However, the temporal interaction window of tactile-visual interaction was wider than that of audiovisual interaction (-640 ms to +40 ms; 680 ms, see Appendix A, Table 8). Also, the shape of the temporal interaction window was skewed toward the 0-ms asynchrony.

In order to fit the data of Experiment 12 with a Gaussian function, various transformations on the time axis were tested to correct the skewed distributions. It turned out that a square-root transformation, with the 0-ms asynchrony being the origin, was the best way to approximate the results of Experiment 12 (Figure 4.7). The standard deviations of the approximated Gaussian functions did not change among the three tactile

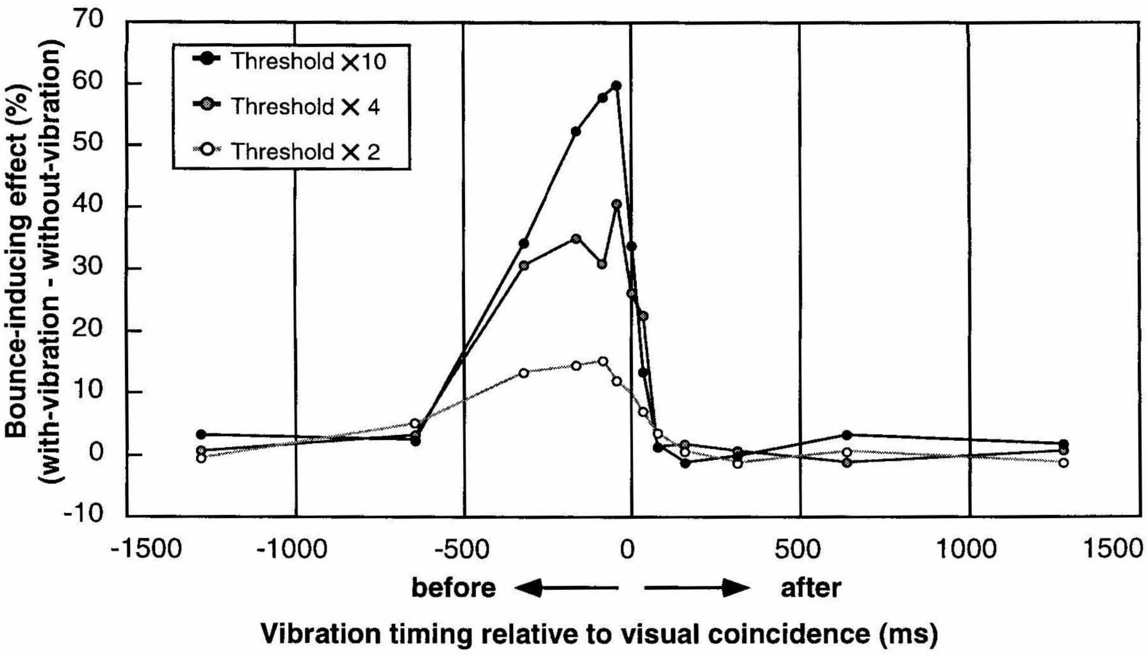

FIGURE 4.6 RESULTS OF EXPERIMENT 12 (TACTILE).
Effect of tactile intensity on the bounce-inducing effect.

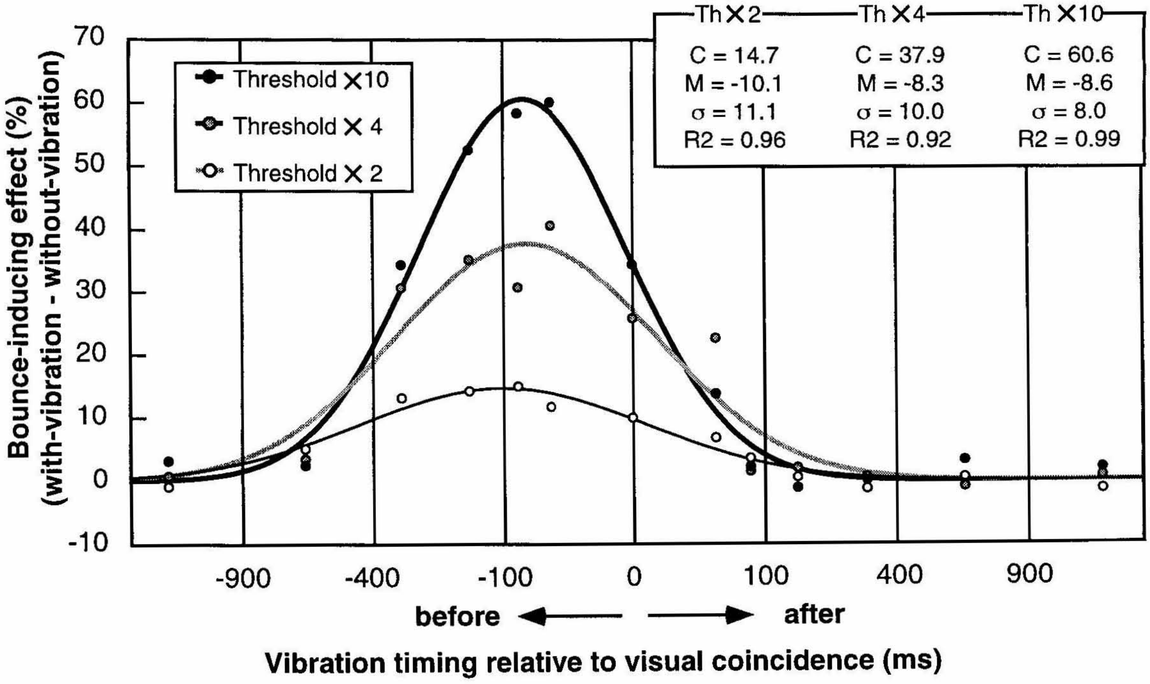

FIGURE 4.7 RESULTS OF EXPERIMENT 12 (TACTILE): GAUSSIAN FIT OF FIGURE 4.6.
M donates the mean, and σ donates the standard deviation. C is a multiplicative constant. Note that the time axis is square-transformed to fit a Gaussian function.

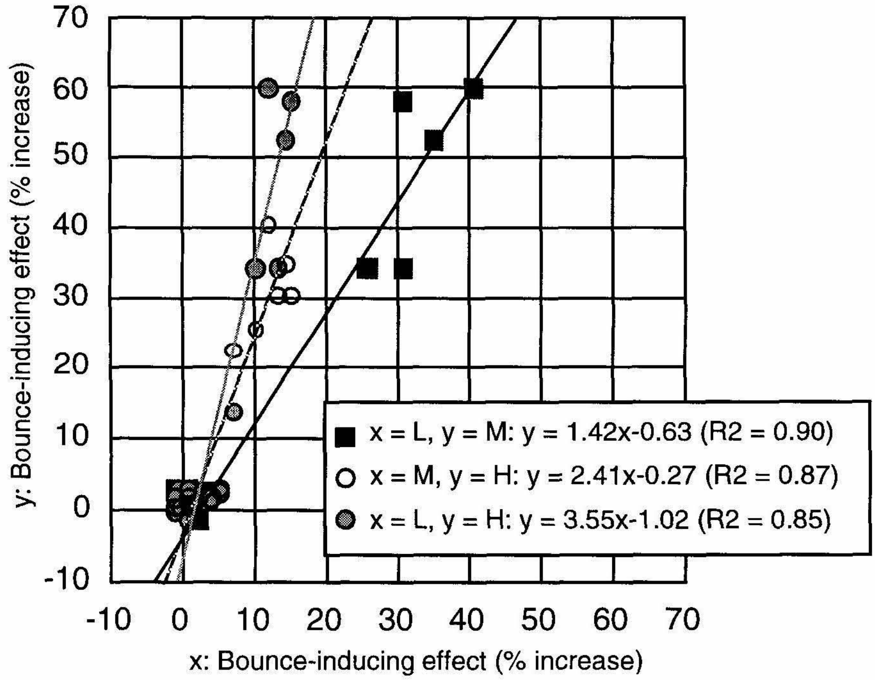

FIGURE 4.8 RESULTS OF EXPERIMENT 12 (TACTILE): GAIN MODULATION FUNCTIONS.
In the inset "L," "M," and "H" donate the bounce-inducing effect by the low (threshold × 2), medium (threshold × 4), and high (threshold × 4) intensity vibrations, respectively. For example, filled squares (x = L, y = M) represent the bounce-inducing effect by the medium intensity vibration as a function of that by the low intensity vibration.

---

intensities used (about 9.5±1.5 ms-square-root). The means are also nearly constant (about 9±1 ms-square-root). Figure 4.8 implies that the gain modulation in Experiment 12 can be approximated with linear functions ($R^2 > 0.85$ for all plots).

Discussion

The results of Experiment 12 are possibly the very first demonstration for the temporal interaction window of tactile-visual event perception. Significantly, most characteristics of the audiovisual interaction window were found in the tactile-visual interaction window. (1) The magnitude of the bounce-inducing effect is dependent on both the intensity (saliency) of the tactile stimulus and the temporal proximity between the tactile stimulation and the visual coincidence. (2) The temporal window of tactile-visual interaction is rather broad (about 680 ms). (3) The center of the temporal interaction window is shifted such that the tactile stimulus presented before the visual coincidence has a larger bounce-inducing effect than the tactile stimulus presented after the visual coincidence.

However, the differences between the bounce-inducing effect by auditory stimuli and that by tactile stimuli are also notable. The width of the temporal interaction window was larger for the tactile-visual interaction (680 ms; -640 ms to 40 ms) than for the audiovisual interaction (480 ms; -320 ms to 160 ms). The shift of the window center from the 0-ms asynchrony was also larger for the tactile-visual interaction (tactile-visual 81 (9^2) ms vs. audiovisual 63 ms). Moreover, in the audiovisual interaction, the interaction window was so symmetrical that a Gaussian function was able to fit the results without any transformation of the time axis. In contrast, the results of the tactile experiment required a square-root transformation of the time axis[6].

Despite these differences, the commonalties between the audiovisual and tactile-visual interactions in the bounce-inducing effect are striking. These commonalties definitely add further evidence that a salient sensory transient around the moment of visual motion ambiguity biases visual perception toward bouncing. However, in order to reach this

---

[6] This may suggest that the mapping from visual time to auditory time is linear, whereas visual time corresponds to the square of tactile time. The crossmodal correspondence problem in time is fascinating. Unfortunately, however, there has been no previous study that compared the representations of time between modalities. An independent set of new experiments is, therefore, necessary to address this issue.

conclusion, it is critical to examine the bounce-inducing effect by visual transients. Would temporal gain modulation occur even without crossmodal interaction?

Experiment 13: Effects of intensity of visual transients

Experiment 10 has shown that a visual transient could produce the bounce-inducing effect. In Experiment 13, I examined whether temporal gain modulation could be observed in the bounce-inducing effect by visual transients. This experiment is important because Experiment 11 and Experiment 12 involved crossmodal interactions. If the bounce-inducing effect by visual transients does not result in temporal gain modulation, temporal gain modulation may be something specific for crossmodal interaction. If temporal gain modulation is obtained in the following experiment, explaining the bounce-inducing effect should require a more general mechanism which functions *amodally*.

Method

Observers, stimuli, and experimental procedures were almost identical to those of Experiment 11. However, visual stimuli were presented as a bounce-inducing sensory transient. The visual stimuli were the same as those in the visual session of Experiment 12, except that the contrast of the flashed ring, which was fixed in a session, was either 0.6, 3.3, or 8.4 cd/m^2 on the gray (8.5 cd/m^2) background (respectively correspond to 7.9, 5.2, or 0.1 cd/m^2 in terms of absolute luminance).

Results

The results of Experiment 13 are presented in Figure 4.9. Both the asynchrony and the visual intensity had influences on the bounce-inducing effect (asynchrony, *F*(12,96) = 36.01, $p < 0.001$; intensity, *F*(2,16) = 15,78, $p < 0.001$; interaction, *F*(24,192) = 4.04, $p < 0.01$). A series of planned comparisons showed that the temporal window of visual-visual interaction is about 160 ms (-80 ms to + 80 ms; see also Appendix A, Table 9). A Gaussian function approximated the results of Experiment 14 excellently (Figure 4.10). The standard deviations were about 47±7 ms, and the means were around 6±0.5 ms. The differences in the standard deviation and the mean were not significant among the three visual contrasts used. Again, the gain modulation was approximately linear (Figure 4.11; $R^2 > 0.86$).

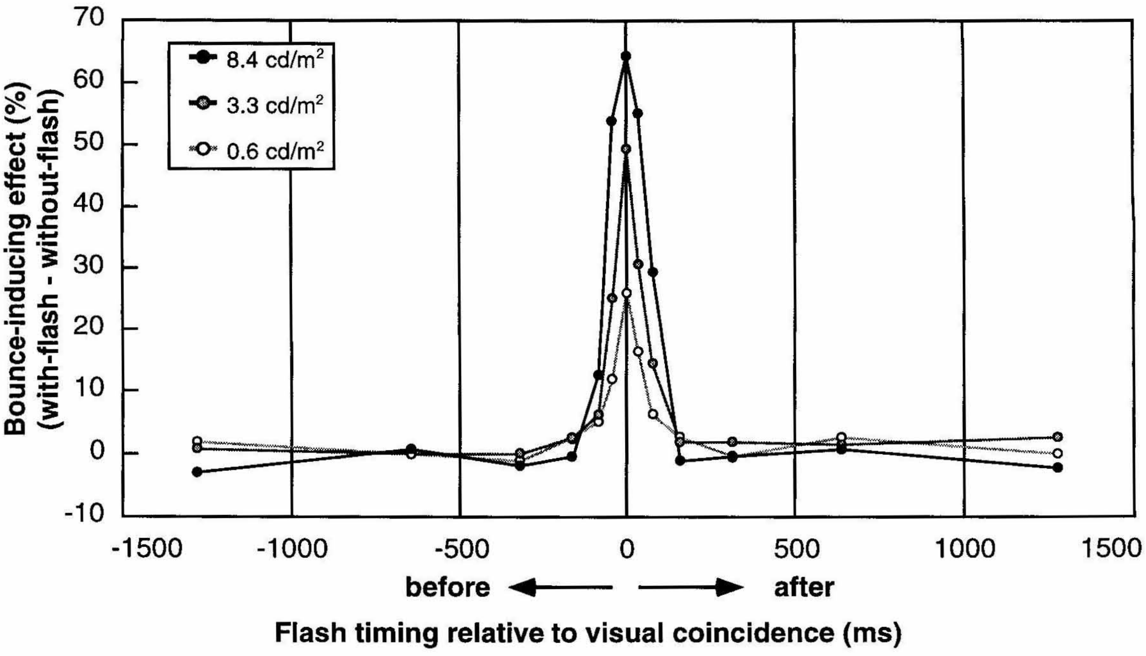

FIGURE 4.9 RESULTS OF EXPERIMENT 13 (VISUAL).
Effect of visual intensity on the bounce-inducing effect.

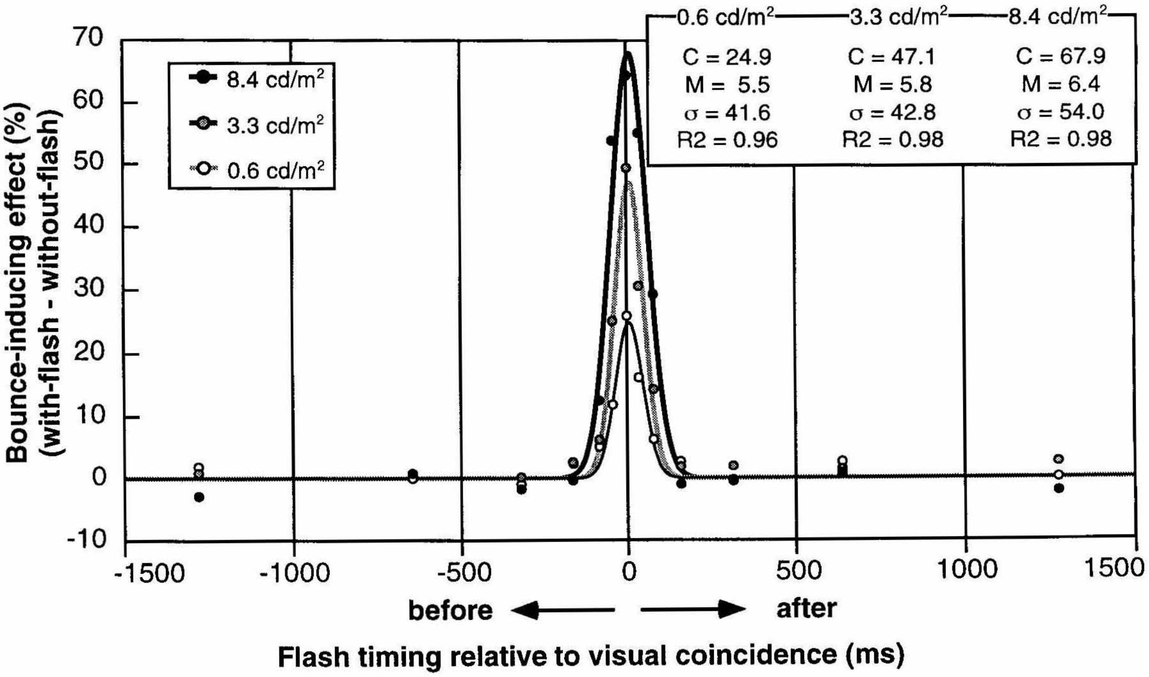

FIGURE 4.10 RESULTS OF EXPERIMENT 13 (VISUAL): GAUSSIAN FIT OF FIGURE 4.9.
M donates the mean, and σ donates the standard deviation. C is a multiplicative constant.

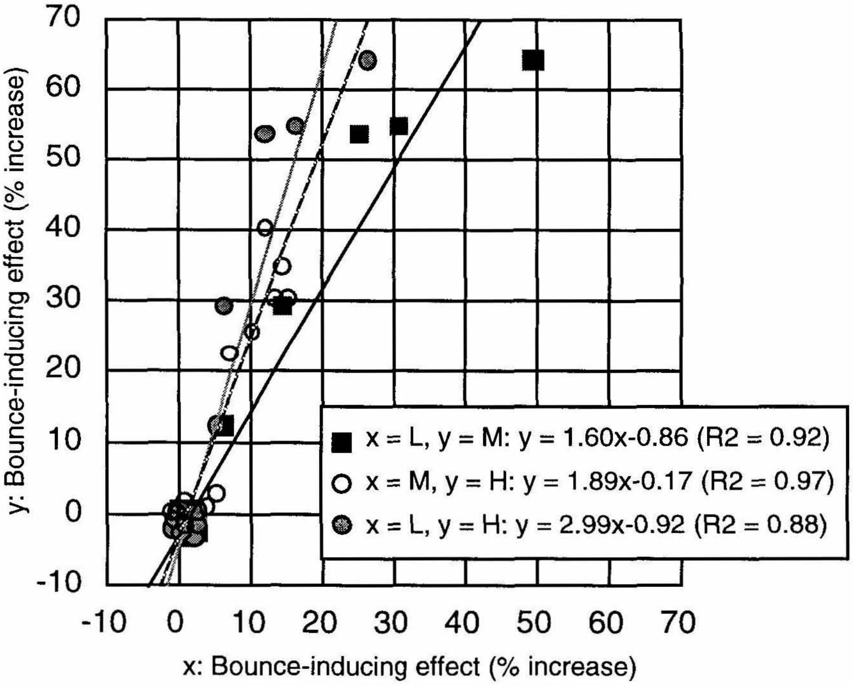

FIGURE 4.11 RESULTS OF EXPERIMENT 13 (VISUAL): GAIN MODULATION FUNCTIONS.
In the inset, "L," "M," and "H" donate the bounce-inducing effect by the low (0.6 cd/m^2), medium (3.3 cd/m^2), and high (8.4 cd/m^2) contrast visual flashes, respectively. For example, filled squares (x = L, y = M) represent the bounce-inducing effect by the medium contrast visual flash as a function of that by the low contrast visual flash.

Discussion

Similar to the bounce-inducing effect by sound and vibration, the bounce-inducing effect by visual transients depends on both the timing and intensity (saliency) of the visual transient. The temporal window of the visual-visual interaction shows a similar inflexibility to those of the audiovisual and tactile-visual interactions. Namely, temporal gain modulation is not restricted to crossmodal interaction.

The bounce-inducing effect by visual transients differs from those by auditory and tactile transients in that there is no temporal asymmetry of the interaction window. The approximated peak of the temporal window is at about a 6-ms delay of the visual transient, which is negligible. Moreover, the width of the visual-visual interaction window is much smaller than those of audiovisual and tactile-visual interactions.

General Discussion

Empirical findings

The experiments in Chapter IV lead to the conclusion that a salient sensory transient around the moment of the ambiguous motion display biases visual perception toward bouncing. Additionally, the transient's intensity (saliency) modulates the magnitude of the bounce-inducing effect without changing the temporal interaction range (*temporal gain modulation*: Experiments 11-13). The range and peak of temporal interaction has specific values for each type of bounce-inducing transient. The interaction range is about 480 ms for the auditory transient, about 680 ms for the tactile transient, and about 80 ms for the visual transient. The bounce-inducing effect is maximal when the auditory transient precedes the visual coincidence by about 60 ms, when the tactile transient precedes the visual coincidence by about 80 ms, and when the visual transient appears at the same time of the visual coincidence.

Bounce-inducing effect and amodal saliency

An important implication of the experiments in Chapter IV is that the bounce-inducing effect is a function of *amodal* saliency. Since any sensory transient, including a visual transient, produces the bounce-inducing effect, the saliency used in the bounce-inducing effect is neither modal nor crossmodal. This is not consistent with the audiovisual associative learning account. I will propose a new explanation for the bounce-inducing effect later in this chapter and in the next chapter.

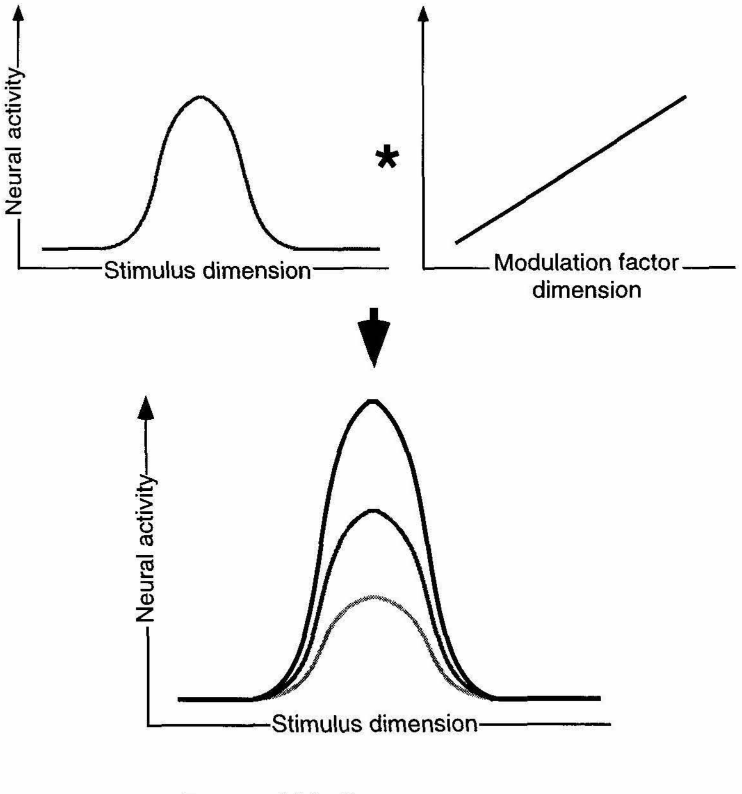

FIGURE 4.12 GAIN MODULATION.
Gain modulation is characterized by the response range (spatial receptive field, orientation tuning, etc.) and the gain modulation function (gaze direction, attention, etc.).

Temporal gain modulation

The temporal gain modulation found in the present experiments is the first gain modulation found in the time domain. It suggests that there is a fixed temporal range during which sensory transients interact with visual motion perception.

Gain modulation has been found in various regions of the brain. It is characterized by two functions (Figure 4.12): one that depends only on a stimulus dimension (retinal location, orientation, motion direction, etc.) and one that depends only on a modulation factor (gaze direction, attention, etc.). For example, gain modulated response in the monkey parietal neurons can be described as a product of the retinal receptive field and the gain field that is determined by gaze direction (Andersen & Mountcastle, 1983; Andersen et al., 1985; Andersen & Zipser., 1988; Zipser & Andersen, 1988; Andersen, 1989; Andersen et al., 1990; Andersen, 1995; Brotchie et al., 1995). The fact that gain modulation is ubiquitously found in the subcortical and cortical brain regions which combine two or more difference inputs (vision and proprioception: Andersen et al., 1985; Andersen & Zipser, 1988; Zipser & Andersen, 1988; Andersen, 1989; Andersen et al., 1990; Andersen, 1995, Brotchie et al., 1995; van Opstal et al., 1995; vision and touch: Fogassi et al., 1992; Colby & Duhamel, 1993; Graziano & Gross, 1993; Graziano et al., 1994; vision, audition, and touch: Stein & Meredith, 1993; vision and attention:

McAdams & Maunsell, 1999; Treue & Trujillo, 1999)[7] suggests that this type of nonlinear transformation is the general rule rather than the exception in the nervous system (Poggio, 1990). The present results add further evidence for the pervasiveness of gain modulation.

It is known to be very difficult to induce single neurons to exhibit mechanisms to carry out a product operation (Mel & Koch, 1990; Mel, 1992; Koch & Poggio, 1992; Mel, 1993). However, a group of neurons can exhibit gain modulation behavior if synaptic connections among neurons are recurrent and are adjusted so that neurons with overlapping response ranges excite each other, whereas those with nearby but separated response ranges inhibit each other[8]. When these neurons receive inputs representing the stimulus feature and modulation signals, the gain modulation property emerges (e.g., Salinas & Abott, 1996; Pouget & Sejnowski, 1997). Changes in the level of modulation input (e.g., eye proprioception input to the parietal neurons) results in changes in the neuronal activity without changing the response range (e.g., the visual receptive fields).

It should be stressed that topographic representation is not necessary for gain modulation to emerge in recurrent neural networks and, therefore, the dimension of response range can be any stimulus feature. Hence, it is possible that the brain represents

---

[7] There are also gain modulations between stimulus dimension in a single modality, all of which are found in vision. For example, contrast and spatial frequency (Dean, 1981; Holub & Morton-Gibson, 1981; Albrecht & Hamilton, 1982; Skottun et al., 1987; Geisler & Albrecht, 1997), contrast and spatial position (Geisler & Albrecht, 1997), contrast and orientation or direction of motion (Dean, 1981; Sclar & Freeman, 1982; Skottun et al., 1987), orientation and direction of motion (Geisler & Albrecht, 1997).

[8] It is worth noting that near-facilitation and far-inhibition is a common feature of recurrently connected cortical models (Ben-Yishai et al., 1995; Douglas et al., 1995; Somers et al., 1995; Stemmler et al., 1995; Zhang, 1996).

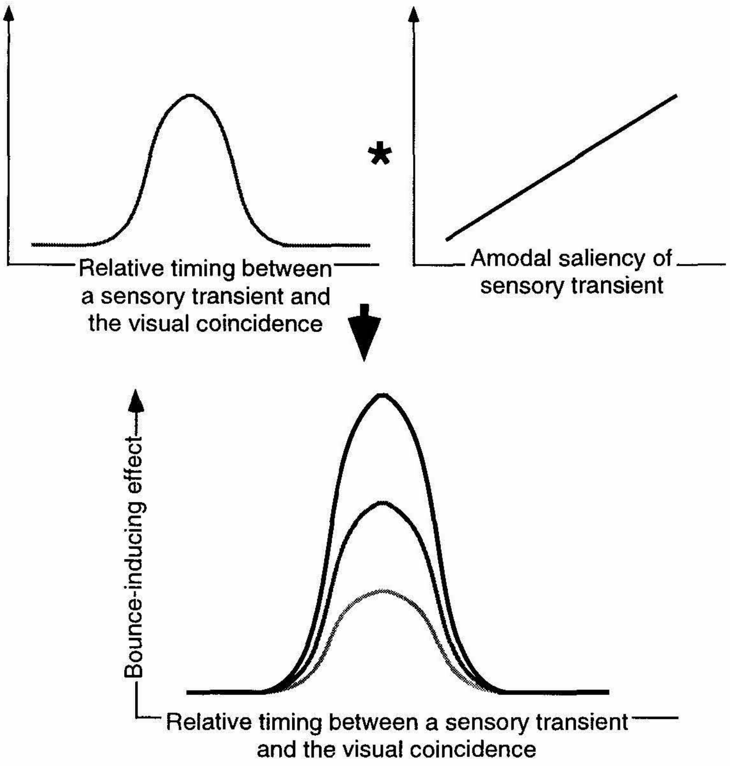

FIGURE 4.13 TEMPORAL GAIN MODULATION.

temporal gain modulation by mechanisms similar to those described above. The stimulus dimension is the relative timing between a sensory transient and the ambiguous visual coincidence, and the modulation input is the saliency of the sensory transient (Figure 4.13). In this regard, it would be interesting to see whether temporal gain modulation could be observed in neurophysiological experiments.

Temporal asymmetry in the bounce-inducing effect

In the bounce-inducing effects by sound and vibration, curious shifts of the temporal window center were observed. The peaks were shifted such that the transients presented before the visual coincidence produced a stronger bounce-inducing effect than those presented after the visual coincidence. The shifts were about 60 ms in the auditory-induced effect and about 80 ms in the tactile-induced effect respectively (see Figure 4.4 and Figure 4.7). Such a temporal asymmetry is virtually nonexistent in the bounce-inducing effect by visual transients.

The temporal asymmetry can be interpreted in that the visual coincidence must occur slightly after the auditory and tactile transients in order to achieve the maximal bounce-inducing effect (Figure 4.14, top). The first guess would be that this asymmetry may reflect a difference of intrinsic delay in information processing, and that the visual processing may be faster than the auditory and tactile processing. In order for a visual event and an auditory or tactile event *to collide in the brain*, the auditory or tactile event has to be presented before the visual event. This might also explain the absence of temporal asymmetry in the visual-visual interaction.

However, the differential latency account is not plausible because visual processing has been known to be slower than auditory and tactile processing on the order of 40-60 ms in humans (Robinson, 1934; Goldstone, 1968; Posner et al., 1976; Regan, 1989). Based on theses empirical facts, a visual event should happen before an auditory or tactile event *to collide in the brain* (Figure 4.14, bottom).

If the differential latency is not the cause of the temporal asymmetry, then what is? A clue comes from the McGurk effect, the visual effect on auditory perception. In the bounce-inducing effect, the auditory input leading the visual event has more effect than the auditory input following the visual event. Interestingly, in the McGurk effect, it has been noticed that visual information leading sound information has greater effect than

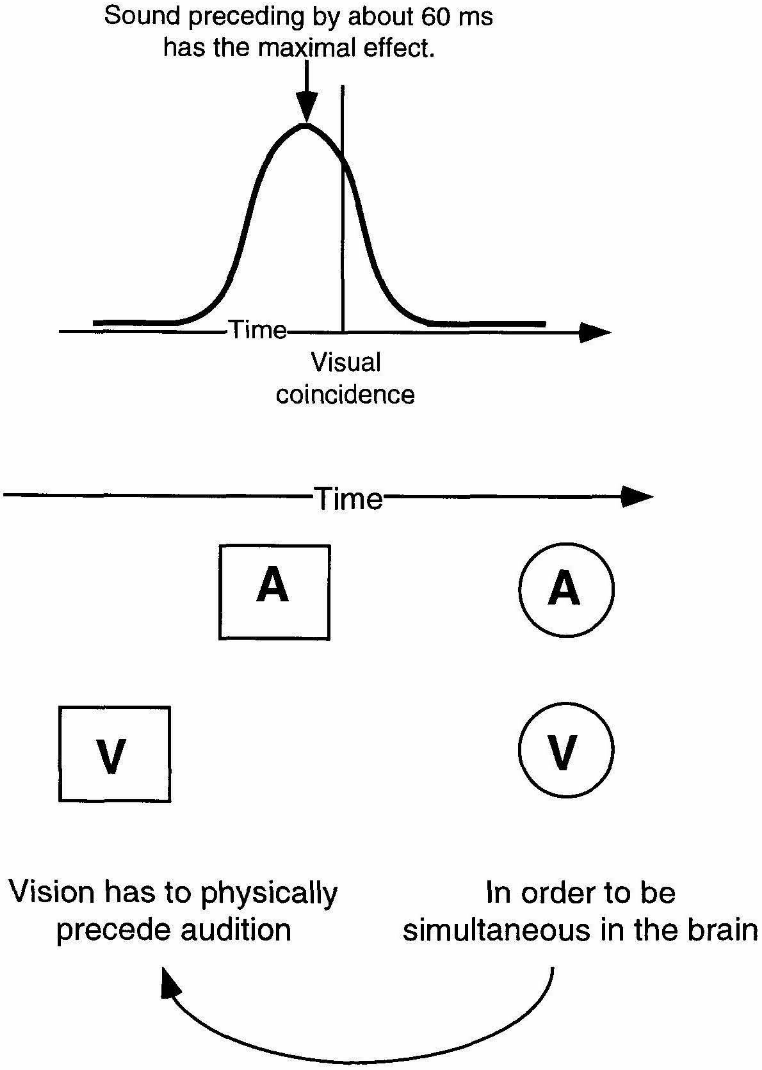

FIGURE 4.14 A PARADOX.

A paradox in the differential latency account for the temporal asymmetry in the bounce-inducing effect. Top: For example, the bounce-inducing effect is maximal when a sound precedes the visual coincidence. Bottom: However, the visual modality is known to be slower than the auditory modality.

visual information following sound information (e.g., Munhall et al., 1996), although researchers have not paid attention to the temporal asymmetry in the McGurk effect. The task in the bounce-inducing effect by sound is to categorize visual motion under the influence of a transient sound. The task in the McGurk effect is to identify auditory stimuli under the influence of visual stimuli. In both cases, the modulating event has larger effects when it leads the to-be-modulated event, irrespective of whether vision modulates auditory events or the other way around. This observation gave me an insight into a mechanism which may explain not only the temporal asymmetry but also the amodal nature of the bounce-inducing effect; that is *attention*. Briefly, I came up with the idea that the temporal asymmetry observed in the crossmodal bounce-inducing effects is due to the sluggishness of attentional shift between modalities, which motivated me to conduct experiments described in the following chapter.

Summary

Any salient sensory transient around the moment of the visual coincidence biases visual perception toward bouncing. The dependency of the bounce-inducing effect on the transient's intensity and on the relative timing between the transient and the visual coincidence can be described as *temporal gain modulation*. Such temporal gain modulation may be implemented in the brain by recurrent facilitation and inhibition among neurons that code the relative timing of perceptual events within or between modalities, in addition to a gain modulation input that is solely based on the sensory transient's saliency.

The auditory and tactile transients before the visual coincidence have larger bounce-inducing effect than those after the visual coincidence. Such a temporal asymmetry is not present in the bounce-inducing effect by visual transients. These temporal asymmetries require a more general framework to understand the bounce-inducing effect, beside simple crossmodal interactions. I propose the involvement of attention in the bounce-inducing effect.

Chapter V: Attentional modulation of motion event perception

Introduction

Attention account for the bounce-inducing effect

Until Sekuler et al. (1997) showed the involvement of crossmodal interaction, the streaming/bouncing ambiguous motion phenomenon has been investigated mostly with respect to local motion integration (Goldberg & Pomerantz, 1982; Bertenthal & Kramer, 1988; Bertenthal et al., 1993). For instance, Bertenthal et al. (1993) have suggested that temporal integration arising from cooperative interaction between local motion detectors is responsible for the dominant perception of streaming.

Temporal recruitment is the putative integration process across local motion detectors over time (Lappin & Bell, 1976; Nakayama & Silverman, 1984; McKee & Welch, 1985; Anstis & Ramachandran, 1986; Casco & Morgan, 1987; Bowne et al., 1989; Snowden & Braddick, 1989ab; 1991; Zanker, 1992; Watamaniuk et al., 1995). Several models involving cooperative interactions among a population of local motion detectors have been proposed (Williams & Sekuler, 1984; Chang & Julesz, 1984; Williams et al., 1986; Nawrot & Sekuler, 1990; Snowden, 1989; Snowden & Braddick, 1989a; Snippe & Koenderink, 1994; Grzywacz et al., 1995). They have successfully explained the directional bias of visual motion (motion as continuing to occur in the same direction as in the past; "motion inertia") as integration among passive local motion operators. However, it is obvious that temporal recruitment alone cannot explain the various phenomenology of the bounce-inducing effect such as crossmodal effects described so far in this thesis.

The results in Chapter IV have suggested that a common *amodal* mechanism is involved in disambiguating the streaming/bouncing motion display. Thus, to gain a more general understanding, a saliency-driven mechanism that involves both modal and crossmodal interaction has to be considered. Here, I propose that *attention* is the main factor in the bounce-inducing effect.

Generally, a sudden sensory event automatically attracts attention (attentional capture: Jonides & Yantis, 1988; Hillstrom & Yantis, 1994; Yantis & Egeth, 1994; Yantis & Jonides, 1996). The magnitude of attentional capture depends on the saliency of the sudden stimulus presentation. This fits very well with the saliency dependency of the bounce-inducing effect. Also, it is now clear that attention functions among two or more modalities as well as within a single modality (Driver & Spence, 1994; 1998; Spence & Driver, 1994, 1996), which may correspond with the amodal nature of the bounce-inducing effect. Moreover, most ambiguous motion perceptions are modulated (Ramachandran & Anstis, 1983; Gogel & Tietz, 1976; Gogel & MacCracken, 1979; Gogel & Sharkey, 1989; Chaudhuri 1990; Balz & Hock, 1997; for review Raymond, 2000), or even caused by attention (Cavanagh, 1992; Hikosaka et al., 1993ab; Lu & Sperling, 1995ab). Finally, there is evidence for the attention involvement in temporal recruitment of visual motion perception (Isaak & Fawcett, 1997). Since the streaming/bouncing motion display is ambiguous and may involve temporal motion recruitment, it is reasonable to expect it to be modulated by attention.

Furthermore, the temporal asymmetry of the interaction window may be explained by the attention account. The visual coincidence must occur slightly after the auditory and tactile transients in order to achieve the maximal bounce-inducing effect (the experiments in Chapter IV). The attention account explains that this temporal asymmetry may be due to the known time cost of attentional switching between two modalities (modality-shifting effect; Kristofferson, 1967; Hannes et al., 1968; LaBerge, 1973; Spence & Driver, 1997ab; Quinlan & Hill, 1999). Since the observer's task is to classify an ambiguous visual event, attention is usually on the moving stimulus in the visual modality (see Figure 5.1). An auditory or tactile transient attracts attention, but it takes time to shift attention between modalities. Hence, in order to produce the maximal attentional capture at the moment of the visual coincidence, an attention-capturing non-visual transient has to occur before the visual coincidence. The absence of the temporal asymmetry in the bounce-inducing effect by visual transients may be due to the absence of the modality-shifting effect, and/or the much faster attentional switching within the visual modality (Tsal, 1983; Weichselgartner & Sperling, 1987).

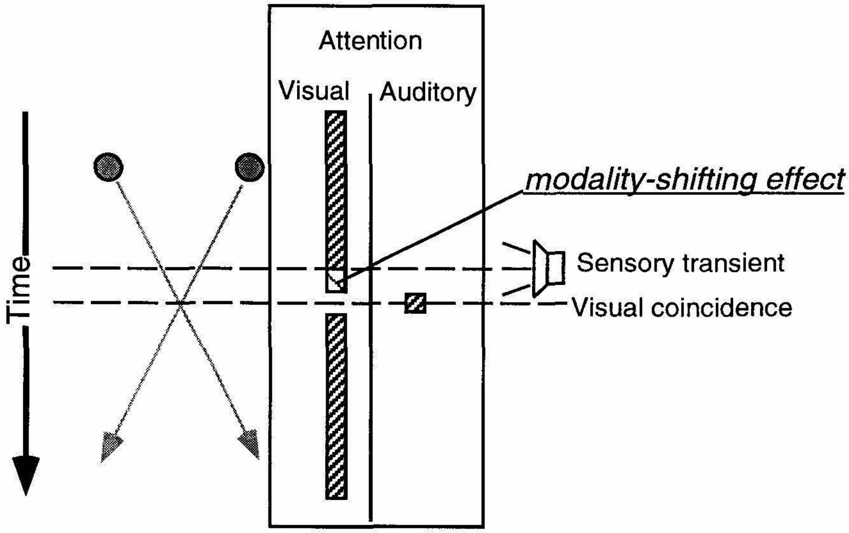

FIGURE 5.1 MODALITY SHIFTING EFFECT.
The attention account for the bounce-inducing effect suggests that the asymmetry of the temporal interaction window is due to the time-cost associated with crossmodal switching of attention. The sensory transient presented a little before the visual coincidence produces the maximal bounce-inducing effect, if it is not presented in the visual modality.

To sum up, the attention hypothesis for the bounce-inducing effect goes as follows: Temporal motion recruitment requires attention (Isaak & Fawcett, 1997). The dominance of the streaming percept reflects automatic deployment of attention to the moving stimuli (e.g., Pylyshyn, 1989, 1994). When a salient sensory transient is presented (whether modal or crossmodal; Driver & Spence, 1998; Spence et al., 1998), it automatically captures attention to the transient stimuli (attentional capture: Jonides & Yantis, 1988; Hillstrom & Yantis, 1994; Yantis & Egeth, 1994; Yantis & Jonides, 1996). When the temporal recruitment process is disrupted by this withdrawal of attention from the moving stimuli, the bouncing percept results. In short, the amount of attentional resource available to the moving stimuli at the moment of the visual coincidence determines the

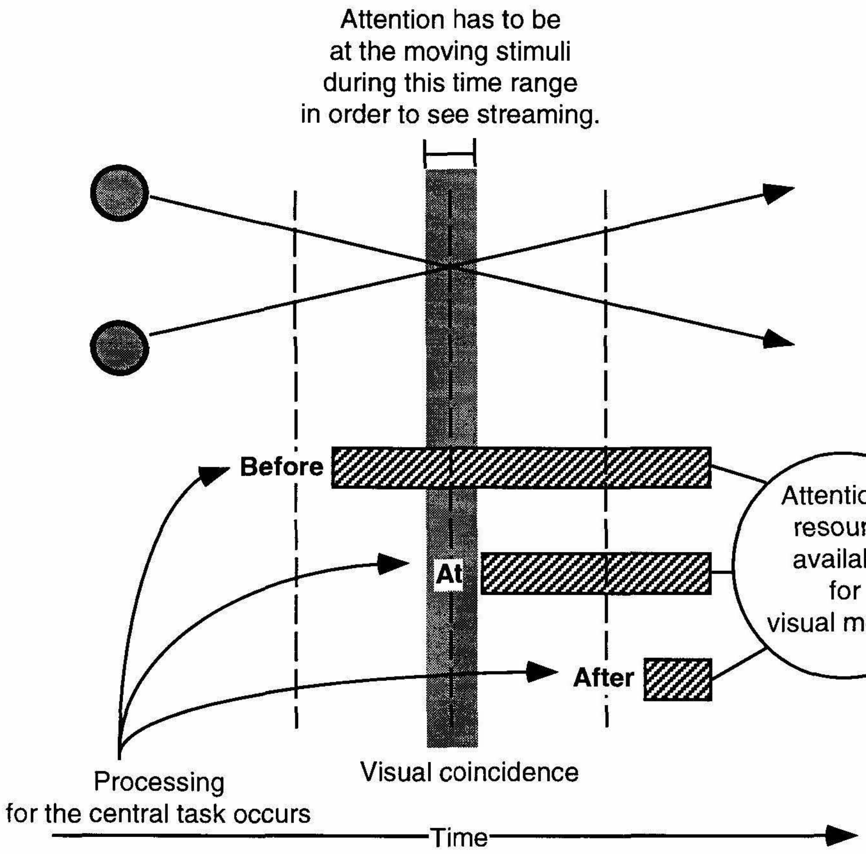

FIGURE 5.2 PREDICTION BASED ON THE ATTENTION ACCOUNT FOR THE BOUNCE-INDUCING EFFECT. Because the timing of the central task was unpredictable, observers had to attend the central task until it occurred.

---

relative frequency of the streaming/bouncing percept. In this chapter, I examine whether attention is the main determinant for the bounce-inducing effect.

Endogenous attentional modulation of motion perception

William James (James, 1890) distinguished between two forms of attention: active versus passive attention, respectively referred to as 'endogenous,' voluntary', or 'top-down' attention, and 'exogenous,' automatic, or 'bottom-up' attention (e.g., Yantis & Jonides, 1990). In the view of the attention account, the bounce-inducing effect by salient sensory transients occurs because of *exogenous* distraction of attention from visual motion stimuli. Effects of *endogenous* attention on the bounce-inducing effect, then, are clearly worth examining as the ultimate test of the attention account.

Endogenous components of attention can be investigated by the measurement of interference between concurrent tasks (Broadbent, 1958; 1982; Logan, 1978; Sagi & Julesz, 1986; Broadbent & Broadbent, 1987; Raymond et al., 1992; Braun & Sagi, 1990; Duncan, 1993; Braun, 1994; Duncan et al., 1994; Pashler, 1994; Bourke et al., 1996; Duncan et al., 1997). In the basic concurrent task paradigm, observers are asked to perform two tasks (Task A and Task B) at the same time, while primarily concentrating on one of the tasks (e.g., Task B). Task B is usually an attention-demanding task. In a separate session of the experiment observers are asked to perform only Task A, but the stimuli from Task B are also presented. Performance of Task A with the demand of Task B is compared to performance without it. Since the stimulus configuration is identical for both cases, if the demand of Task B has influence on the performance of Task A, Task A is generally thought to require attentional resources.

Inspired by these studies, I investigated whether an endogenous shift of visual attention would alter the relative frequency of streaming and bouncing percepts. I added an attention-demanding visual task on the fixation stimulus around the time of the visual coincidence. Since observers did not know when exactly this central task stimulus would happen, attention would be concentrated on the fixation stimulus until it occurred. Therefore, if the central task demand was imposed *at* or *after* the visual coincidence, attention would have to be focused on the fixation stimulus at the time of the visual coincidence. The attention account predicts that the bouncing percept would increase in such a condition because of less attention being available for processing motion perception (Figure 5.2). In contrast, if the stimulus for the central task occurred *before* the streaming/bouncing event, attention would be finished with the central task and already available for motion processing at the time of the visual coincidence. Thus, there would be no effect of the central task. To sum up then, asymmetry in the effect of the central task would be expected between trials where the central task occurred before the visual coincidence (no effect, or dominance of the streaming percept) and those where it occurred at or after the visual coincidence (increase in the bouncing percept).

Experiment 14: Effect of endogenous visual attention -- I. Temporal effect

Method

Observers

Six observers, including the author, took part in the experiment. All had normal or corrected-to-normal vision and experienced no severe difficulty with the tasks. Informed consent was obtained from all observers. Their participation was compensated at the rate of US$8 per hour. They were naive as to the purposes and hypotheses motivating the study, except the author.

Stimuli

Visual stimuli were displayed on a Sony color monitor (frame rate 72 Hz), controlled by a Silicon Graphics Indigo2 work station, in an otherwise dim environment. In order to facilitate fixation, a bull's eye (black and white, 0.58 deg in diameter; Figure 5.3) was continuously displayed on a random-dot background occupying a 20 × 20 deg area of visual angle (1.2% dot density with a luminance value of 10.2 cd/m^2; mean background luminance 0.87 cd/m^2). At the beginning of each trial, two white squares (30.67 cd/m^2, 0.2 deg in size) appeared at opposite sides of the display, initially separated by about 3.6 deg. The squares moved laterally toward the vertical center line of the display at the constant speed of 3.2 deg/s with no inter-stimulus interval between frames (frame-to-frame spatial offset = about 2.7 min per frame). The two squares coincided and then continued to move to the other's start position. The duration of a total sequence was 1.1 s. Around the time of the visual coincidence, a small spatial gap appeared for 13.9 ms in either the left or right side of the bull's eye. Before the experiment, the width of the gap was determined for each observer such that he or she was not able to discriminate the position of the gap without a firm fixation and focal attention on the bull's eye (average gap size 1.8 min). The gap was presented at five possible timings (0, 69, 139 ms before or after the visual coincidence). The eccentricity of the visual coincidence was fixed at 4.9 deg away from the fixation stimulus either in the lower visual field or in the upper visual field.

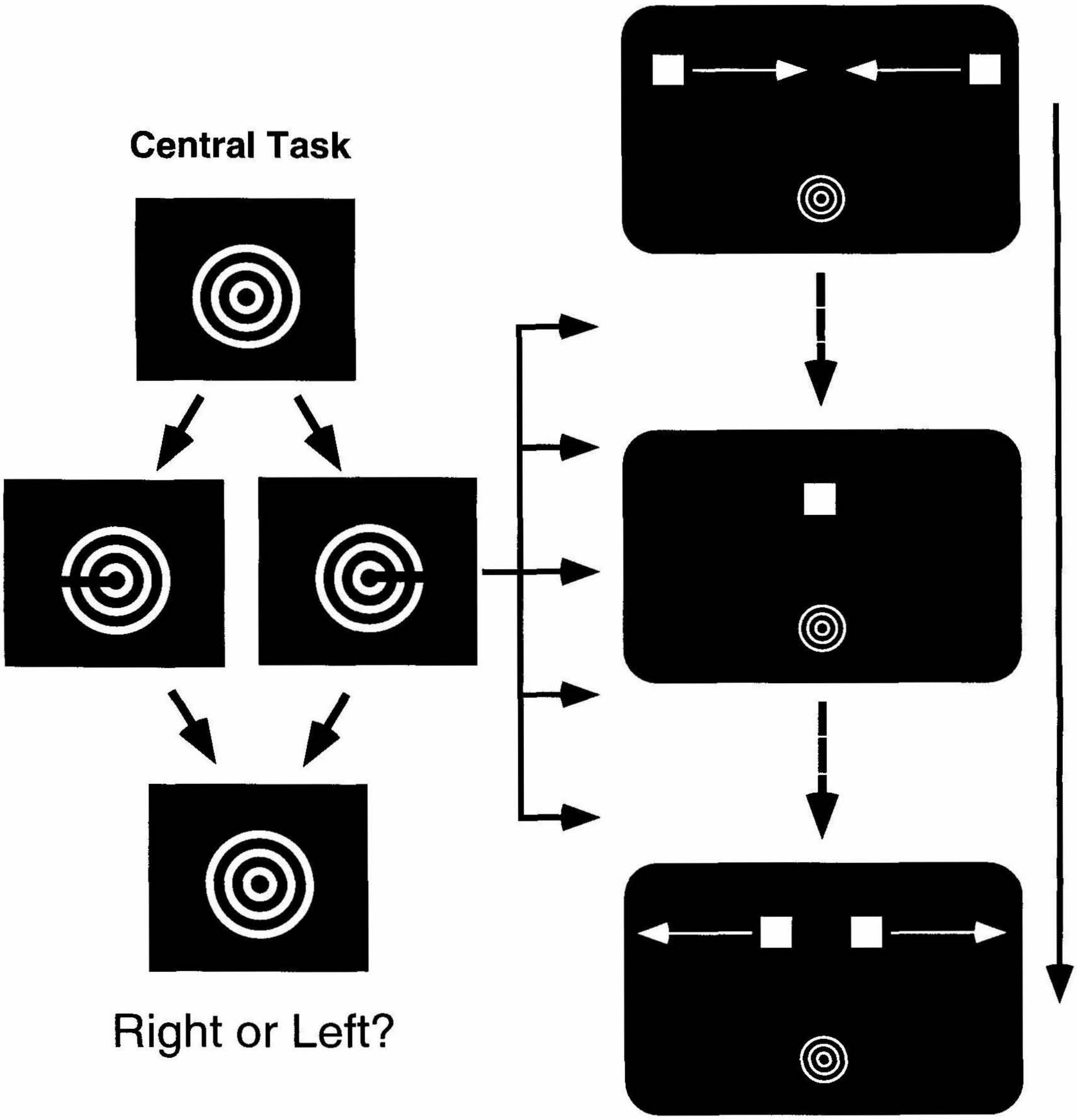

FIGURE 5.3 VISUAL STIMULI IN EXPERIMENTS 14 AND 15.
While observers viewed the streaming/bouncing motion display, either the left or right side of the bull's eye had a small spatial gap for 13.9 ms. The timing of the gap presentation was varied in Experiment 14. The location of the motion event was varied in Experiment 15.

Procedure

The observation distance was 70 cm. After each stimulus sequence, the display became black except for the bull's eye. Observers were asked to judge whether the two squares appeared to stream through or bounce off each another by pressing the mouse buttons accordingly. In one session, observers were also asked to report which side of the bull's eye had a gap and also asked to judge whether the two squares appeared to stream through or bounce off each other (with-task condition). They were told to primarily concentrate on the central task and informed that, if performance did not reach 90%, the results would be discarded. In the other session, observers were instructed to ignore the central task (but respond arbitrarily to it) and to report only the percept of streaming or bouncing (without-task condition). Note that the central task stimulus was presented regardless of whether the central task was required or not. The order of the two tasks was randomized among observers. Forty trials were shown randomly, for each of the 5 central task timings and 2 different visual fields (central task timing (5) × lower or upper visual field (2) × with or without central task (2) × repeat (40) = 800 trials). A full experiment was divided into 4 sessions (2 sessions of with-task condition, 2 sessions of without-task condition). Each session consisted of 200 trials.

Results

The mean performance on the central task was 96.6%. There was no statistical difference in the central task performance among the different timing conditions ($F(4,20) = 0.66, p > 0.05$). The results of the bouncing judgment in Experiment 14 are presented in Figure 5.4. When the central task was not required, there was no difference among the different central task timing conditions (gray curves in Figure 5.4). If it was required *before* the visual coincidence, no difference was found between the with-task and

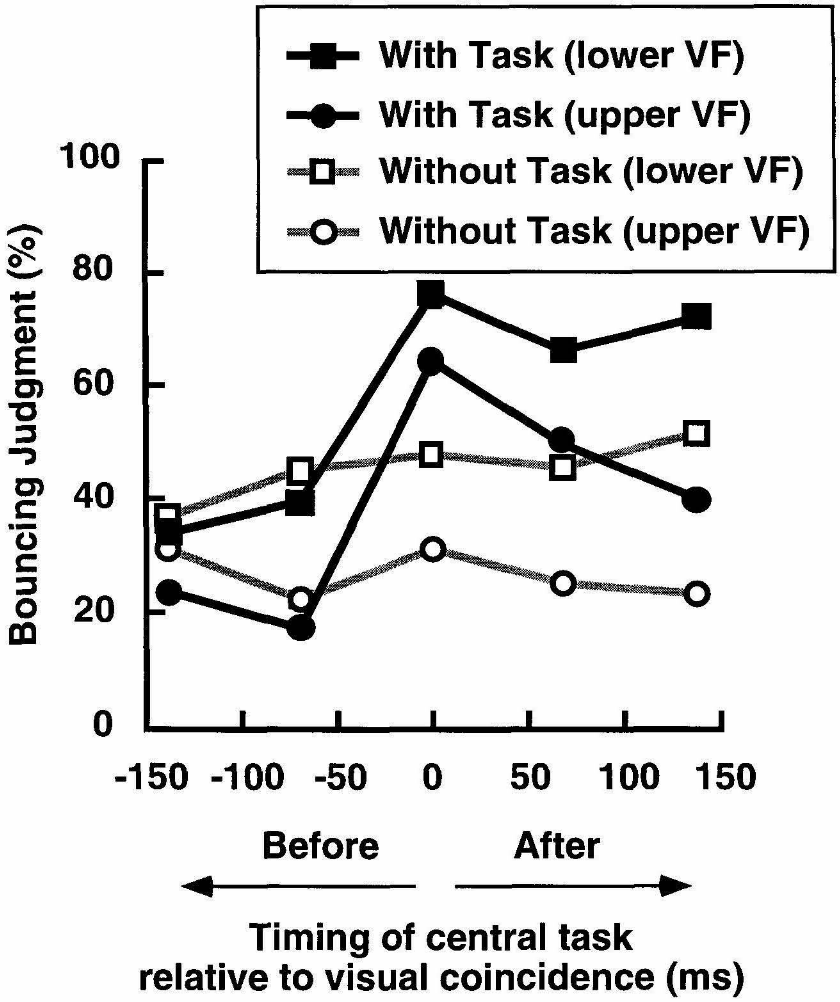

FIGURE 5.4 RESULTS OF EXPERIMENT 14 (TEMPORAL EFFECT OF ADDITIONAL TASK DEMAND). The vertical axis shows the averaged percentage of bouncing judgments. Gray lines indicate the conditions without the center task and black lines indicate the conditions with the center task. Circles and squares stand for the upper and lower visual field, respectively. Error bars were eliminated for viewing convenience. There were significant effects of the task demand, the visual field, and the center task timings ($p < 0.001$).

---

without-task conditions. In contrast, when the central task took place *at* or *after* the visual coincidence, the task demand increased the frequency of the bouncing percept. Both the effect of the task demand and that of the task timing were significant (task demand, $F(1,5) = 10.86, p < 0.001$; task timing, $F(4,20) = 8.58, p < 0.001$) as well as the interaction ($F(4,20) = 2.37; p < 0.01$). Additionally, the observers reported the bouncing percept more frequently when the visual event was presented in the lower visual field than when it was presented in the upper visual field ($F(1,5) = 3.54, p < 0.001$). I further tested the possibility that errors in the central task would correlate more with the streaming than bouncing percept, which is indeed a prediction based on the attentional modulation account. A further trial-basis analysis revealed that the bouncing percept tended to occur when the observers correctly responded to the central task and the streaming percept became frequent when the observers produced wrong responses (see Appendix A, Table 10, $\chi^2 = 4.25, p < 0.05$; Responses were pooled for all the observers).

Discussion

When the observers were not asked to perform the central task, the timing of the gap presentation had no effect. This means that the presentation of the small gap in the bull's eye did not by itself affect the perception of the streaming/bouncing motion display. However, when the observers performed the central task and the stimuli for the central task occurred *at* or *after* the visual coincidence, the frequency of the bouncing percept increased. In contrast, if the central task had been finished *before* the visual coincidence, the percept was not affected. This pattern of results is predicted, and can be interpreted in that when attention has to be concentrated on the central task at the moment of the visual coincidence ("at" and "after" conditions), less attention is available for motion processing. And, as expected from the attention hypothesis, the bouncing percept then becomes dominant (the "right-shoulder" in the black curves of Figure 5.4). Also, the overall positive correlation between the bouncing percept and correct responses to the central task and between the streaming percept and wrong responses (Appendix A, Table 10) imply that attention is less available for the central task when the streaming percept is dominant (and vice versa). Thus, the results of Experiment 14 confirm the involvement of attentional processes in the perceptual disambiguation of the streaming/bouncing motion display.

In addition to attentional modulation, an advantage of the upper visual field in producing the percept of the streaming event was observed accidentally. Since the visual field asymmetry is not the main issue here, it will be discussed in the general discussion of this chapter. Nevertheless, the visual field asymmetry brought up another issue: that of the spatial properties of attentional modulation. In the next experiment, in order to further characterize the attentional modulation involved here, I investigated the spatial aspect of attentional modulation of the bounce-inducing effect by manipulating the eccentricity of the motion event.

Experiment 15: Effect of endogenous visual attention -- II. Spatial effect

Bertenthal et al. (1993) informally reported that the dominant perception of streaming diminished when the fixation point was shifted more than 4.3 deg above or below the center of the visual coincidence in their experimental configuration. They reasoned that the density of local motion operators, which is thought to decrease as a function of eccentricity (e.g., Fredericksen et al., 1993; van de Grind et al., 1993), would be responsible for the eccentricity effect. If the eccentricity effect was governed only by the density of local motion detectors, it would not interact so much with attentional modulation. To investigate the eccentricity effect and its relation to attention, I varied the eccentricity of the visual coincidence, while applying again the concurrent task paradigm. If there was an interaction between the eccentricity effect and distraction of attention, I would argue for the importance of attention for the eccentricity effect.

Method

Observers

The same observers as in Experiment 14 participated in Experiment 15.

Stimuli

The stimuli were almost the same as those used in Experiment 14. However, the eccentricity of the visual coincidence location was varied among 10 possibilities (0.8, 2.9, 4.9, 6.9, 8.9 deg away from the fixation stimulus, in either the lower or upper visual field). The gap was always presented at the same time as the visual coincidence.

Procedure

Except for the fact that observers were informed of the different stimulus conditions, the procedures were the same as those of Experiment 14. For each of the five different eccentricities and the two visual fields, 40 trials were presented in random order [eccentricity (5) × lower or upper visual field (2) × with or without the central task (2) ×

repeat (40) = 800 trials]. A full experiment was divided into 4 sessions (2 sessions of with-task condition, 2 sessions of without-task condition). Each session consisted of 200 trials.

Results

The mean performance on the center task was 94.8%, and no significant difference among the different eccentricity conditions was found ($F(4,20) = 0.23, p > 0.05$). The averaged percentage of bouncing judgments across all observers is shown in Figure 5.5. When the central task was not required, the percentage of bouncing judgments increased with eccentricity ($F(4,20) = 13.00, p < 0.001$). When the central task was required, the overall bouncing percentage became higher ($F(1,5) = 28.97, p < 0.001$) but the dependency on eccentricity diminished. There was a significant interaction between task and eccentricity ($F(4,20) = 3.45, p < 0.001$). The significant advantage of the upper visual field in producing the streaming percept was found again ($F(1,5) = 7.60, p < 0.001$). No other interaction was significant.

Similar to the results of Experiment 14, a trial-basis analysis showed that there is a positive correlation between the bouncing percept and correct responses to the central task and between the streaming percept and wrong responses (see Appendix A, Table 11, $\chi^2 = 10.01, p < 0.01$).

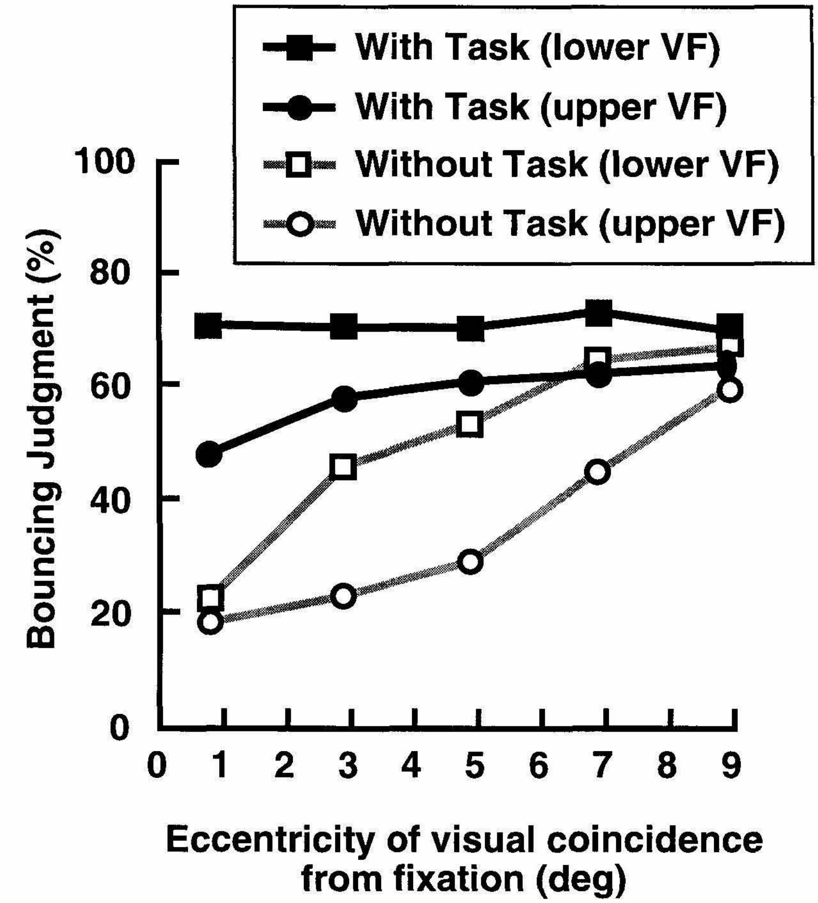

FIGURE 5.5 RESULTS OF EXPERIMENT 15 (SPATIAL EFFECT OF ADDITIONAL TASK DEMAND). The vertical axis shows the averaged percentage of bouncing judgments. Gray lines indicate the conditions without the center task and black lines indicate the conditions with the center task. Circles and squares stand for the upper and lower visual field, respectively. Error bars were eliminated for viewing convenience. There were significant effects of the task demand, the visual field, and eccentricity ($p < 0.001$).

Discussion

Again, the occurrence of the attention-demanding task at the moment of the visual coincidence facilitated the perception of bouncing and suppressed that of streaming. Furthermore, the bounce-inducing effect by attentional demand was more pronounced when the moving squares were near or in the foveal field and diminished when they appeared at more eccentric locations. Since this dependency on eccentricity was detectable only when attention was potentially available (i.e., with less distraction of attention), I argue that the eccentricity effect is due more to the passive distribution of spatial attention rather than to the distribution of local motion operators changing across eccentricities.

It is commonly assumed that the center of attention (or the highest availability of attentional resources) is at the fovea with a rather broadly spaced distribution, fading away with eccentricity (Sagi & Julesz, 1986; Balz & Hock, 1997), when a task is not required. However, once observers were asked to perform the central task in Experiment 15, attention might be concentrated on the fixation stimulus, and little attention was then available outside of the fovea. This is why, I think, the slope became flatter, which seems to be consistent with the fact that I did not observe a statistically significant effect of eccentricity when there was a central task demand.

The results of Experiments 14 and Experiment 15 (manipulation of endogenous attention), combined with those of Experiment 13 (visual flash experiment; presumably exogenous attentional capture), strongly suggest that attention facilitates perception of the streaming event and the lack of attention produces the bouncing percept.

Experiment 16: Effect of size of central task stimulus: A control experiment

Before coming to a conclusion about the attention account, I was curious about the absence of the effect of the central stimulus presentation itself. In Experiment 13 (Chapter IV), the abrupt presentation of the visual stimulus increased the bouncing percept. Nevertheless, when the center task was ignored, I did not observe an effect of the presentation of the central stimulus (gap) in Experiment 14. I hypothesized that the central stimulus in Experiment 14 and Experiment 15 might be so subtle that it could not capture attention automatically.

In this control experiment, I examined the effect of the size of the gap in the central task stimulus. A large salient gap at the time of the visual coincidence was expected to increase the frequency of the bouncing percept irrespective of the central task demand because of its salient (attention-capturing) nature. Another prediction was that the "right-shoulder" in the with-task condition in Experiment 14 would disappear and the graph would become symmetrical, since the large gap should make the central task much easier and free endogenous attention from the central task (as the large gap automatically attracts attention).

Method

Two observers (one the author, one naive) were tested with a larger gap size (0.29 deg in visual angle). This manipulation was expected to make the central task much easier and to induce attentional capture, if possible, through the salient nature of the stimulus change. The other stimulus and procedural conditions were the same as those of Experiment 14, except that the moving squares were presented only in the upper visual field.

Results and Discussion

The results are presented in Figure 5.6. Both observers made no error in the central task. The addition of this easy central task demand had little effect, and the perception of the bouncing event increased only when the gap was presented at the time of the visual coincidence. In other words, the results of Experiment 16 were similar to those of Experiment 13 (visual flash, exogenous attention distraction), not to those of Experiment 14 (central task, endogenous attention distraction).

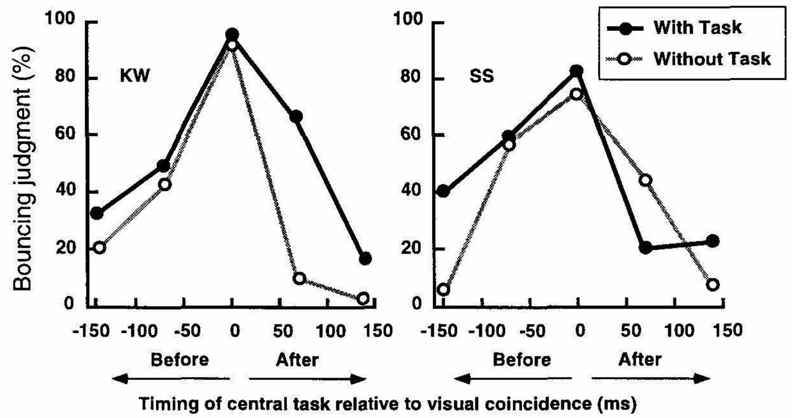

FIGURE 5.6 RESULTS OF EXPERIMENT 16 (WITH A LARGE GAP).
The vertical axis shows the percentage of bouncing judgments. Gray lines indicate the conditions without the center task and black lines indicate the conditions with the center task. Note the lack of the "right shoulder" compared with Figure 5.3.

When the central task was not required, the large gap at the fixation stimulus transiently suppressed the perception of streaming in Experiment 16, whereas the small gap in Experiment 14 did not. Thus, the absence of an increase in the bouncing percept in the without-task condition of Experiment 14 seems to be due to the much decreased saliency of the gap.

Moreover, this control experiment indirectly confirms the involvement of both exogenous and endogenous components of attention in the bounce-inducing effect. While the bouncing percept increased only when the central task demand occurred *at* or *after* the motion event in Experiment 14, this asymmetrical effect disappeared in Experiment 16. That is, Figure 5.6 is symmetrical, lacking the sustained "right shoulder" in the with-task condition in Figure 5.4. Therefore, the effect observed in Experiment 14 (the right sustained shoulder) should be interpreted as a component of the endogenous (top-down) distraction of attention, and the effect in Experiment 16 (and Experiment 13) should be a component of the exogenous (bottom-up) distraction of attention.

Experiment 17: Did observers just become uncertain? Another control experiment

The results of Experiments 14-16 undoubtedly demonstrate that attention modulates the perception of the streaming/bouncing motion display. However, one could argue that the lack of attention at the moment of the visual coincidence might increase the observer's uncertainty about the visual motion event. If this were the case, the central task demand might have simply brought the perceptual judgment close to chance level (50%). Although this is unlikely because the mean percentage of bouncing judgments at the "right shoulder" of Experiment 14 was well above chance (Figure 5.4), I conducted a control experiment to further confirm this point.

It has been known that a momentary pause of the visual targets at the coincidence point greatly enhances the bouncing percept; the vast majority of observers then perceive the bouncing percept more frequently than the streaming percept[9] (e.g., Bertenthal et al., 1993). In Experiment 17, I added a momentary pause at the visual coincidence, while observers performed the concurrent-task as in Experiment 14. If a lack of endogenous attention led to perceptual uncertainty, the central task demand would bring the perceptual judgment close to 50% and would produce a "left shoulder" pattern. That is, when the central task happened at or after the visual coincidence *with* the targets' pause at the visual coincidence, the frequency of the bouncing judgment would decrease.

Method

Two observers (one the author, one naive observer) participated in the experiment. The stimuli were nearly identical to those of Experiment 14 except that, in half of the

---

[9] This can be also explained by the attention account. The sudden pause of the visual targets captures attention and disrupts the temporal recruitment process.

trials, the visual target stopped at the moment of the coincidence for 3 frames. When the central-task stimulus was presented at the moment of the pause, it was presented at the second frame of 3 paused frames.

Results and Discussion

The results of Experiment 17 are shown in Figure 5.7. Without the pause, the "right shoulder" pattern was replicated for both observers. With the pause at the visual coincidence, the observers reported the bouncing percept most of the time. In contrast to the no-pause conditions, the with-pause conditions did not lead to the dependency on the central task timing.

The lack of timing dependency in the with-pause conditions is inconsistent with the uncertainty account because it should predict a "left shoulder" pattern. By contrast, the attention account can explain the present results in that the momentary pause at the visual coincidence captures exogenous attention (possibly in addition to the intrinsic effect of the pause to reset temporal recruitment of local motion signals) and therefore the bouncing percept predominates regardless of the timing of the central task. The additional manipulation of endogenous attention by the central task demand then makes no change in the perception.

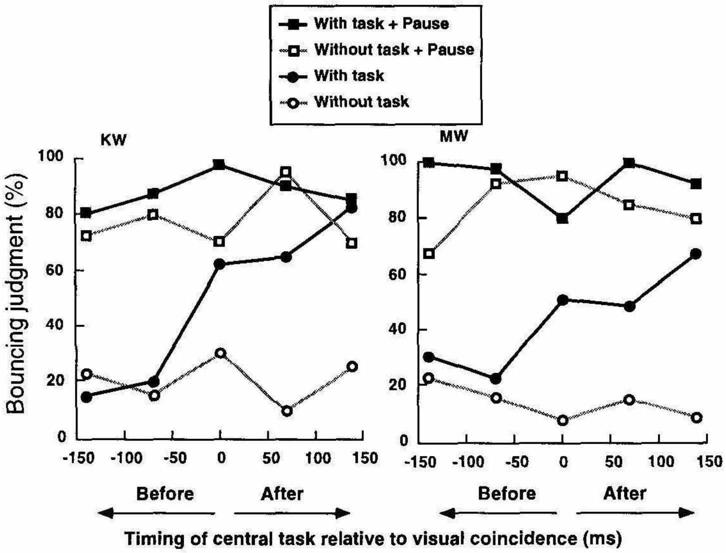

FIGURE 5.7 RESULTS OF EXPERIMENT 17 (WITH A PAUSE AT THE VISUAL COINCIDENCE).
The vertical axis shows the percentage of bouncing judgments. Gray lines indicate the conditions without the center task and black lines indicate the conditions with the center task. When there was a momentary pause at the visual coincidence, the two observers almost always perceived bouncing and the additional central task had no effect.

General Discussion

Attentional modulation of visual motion perception

Recent studies have demonstrated attentional modulation on motion perception by using psychophysical (Gogel & Tietz, 1976; Gogel & MacCracken, 1979; Gogel & Sharkey, 1989; Chaudhuri, 1990; Blakemore & Snowden, 1997; Balz & Hock, 1997; Isaak & Fawcett, 1997; Raymond et al., 1997; Treue & Husain 1997; Raymond 2000), electrophysiological (Treue & Maunsell, 1996; Treue & Husain, 1997; Valdes-Sosa et al., 1998; Treue & Trujillo, 1999), and functional imaging techniques (Beauchamp et al., 1997;

O'Craven et al., 1997; Büchel et al., 1998; Gandhi et al., 1999; Rees et al., 1999; Seidemann & Newsome, 1999).

Using the concurrent task paradigm, the experiments in Chapter V have demonstrated the effect of an endogenous attention shift on the perception of ambiguous motion. In short, the results suggest that (1) a dynamic attentional process is involved in perceptually disambiguating the streaming/bouncing motion display, and (2) attentional availability to visual motion contributes to the bias toward the perception of streaming, for distractions of attention increased the frequency of the bouncing percept.

If more attentional resources are available, attentional resources are automatically used for processing motion perception. This automaticity can be inferred because the observers were instructed not to intentionally track one of the targets with focal attention. In fact, if they were asked to choose and track one of the targets attentionally, the streaming percept became more dominant (informal observation). Yet, the possibility that involuntary tracking eye movements might cause the dominance of the streaming percept was examined and rejected by a control experiment (see Appendix C). Therefore, I also propose that (3) attentional mediation here may facilitate the streaming percept even without intentional tracking.

How exactly attention enhances the perception of streaming remains to be investigated. One possibility is that attention may directly enhance temporal recruitment. For instance, Isaak and Fawcett (1997) have measured the effect of attention on temporal recruitment. In their experiments, attention was controlled by manipulating spatial uncertainty. When a random-dot kinematogram was presented at a less probable location, or at a less attended location, temporal recruitment was slowed. So it may be possible that attention inhibits the bouncing percept and facilitates the streaming percept because it directly enhances the directional bias based on the past stimulation of local motion detectors. This interpretation fits well with the studies favoring integration of local motion signals (Bertenthal & Kramer, 1988; Bertenthal et al., 1993). However, since attention may mostly work on and be attached to perceived objects (Gibson & Egeth, 1994; Hikosaka et al., 1996), the other possibility of a passive form of attentional tracking, which may have a similar effect as active covert tracking (Cavanagh, 1992), cannot be ignored. In general, motion is thought to attract attention (James, 1890). Therefore, it is no wonder that the moving stimuli in the present experiments attract attention and an attentional process automatically tracks the moving objects. This latter view emphasizes representations of object identity more than local motion signals, in terms of the level of attentional resource allocation.

Whichever the case is, it is clear that the main determinant of the bounce-inducing effect is attention, and I conclude that attention may facilitate the intrinsic tendency of the visual system to perceive object motion as continuing to occur in the same direction as in the past ("motion inertia").

Attention as a limited resource and automaticity of attentional resource allocation

The attention account for the bounce-inducing effect is in accord with the idea that attention is a limited resource (Broadbent, 1985; Treisman, 1969; for review see

Kahneman & Treisman, 1984). The proposed automaticity of attentional allocation (default is for moving stimuli, then attentional-grabbing factors) resembles Lavie's theory of attention (Lavie & Tsal, 1994; Lavie, 1995). She has proposed that people cannot simply ignore a particular object or aspect of a stimulus; in order to ignore a stimulus we have to attend another task. A functional magnetic resonance experiment has recently presented supporting evidence (Rees et al., 1999). Observers performed an attention-demanding visual task while a task-irrelevant random-dot field moved in the same screen. Activation of area MT (visual motion area in humans) inversely correlated with the task load, but MT activation did not correlate with the observer's intention to ignore the random-dot field. So, attentional resource available to visual motion depends on how much a central task uses up the attentional resource. The results of experiments in Chapter V are in accord with this idea.

Visual field asymmetry

To my surprise, I found that the bouncing percept is more pronounced in the lower visual field than in the upper visual field (Experiments 14 & 15). Two basic accounts are possible for the visual field asymmetry. It could be caused either by asymmetry of attention distribution (higher attentional resource in the upper visual field), or by asymmetry of more bottom-up mechanisms, independent of attention (e.g., passive motion integration processes).

The attentional account for the visual field asymmetry is unlikely because the attentional asymmetry reported in the other studies predicts the opposite of my results. Higher attentional resolution has been reported in the lower visual field (He et al., 1996; Intriligator & Cavanagh, 1997), and it should decrease the frequency of the bouncing percept in the lower visual field. Furthermore, the known asymmetry of temporal resolution (higher temporal resolution in the lower visual field; Hylkema, 1942; Phillips, 1933; Skrandies, 1985; Tyler, 1987) is indeed consistent with the account based on the asymmetry in basic motion integration processing. This is because the frequency of the bouncing percept has been reported to increase if there is a pause when the targets coincide (Bertenthal et al., 1993; Sekuler et al., 1995). The greater temporal resolution might lead to a better detection of frame-to-frame temporal discontinuity in my computer-generated discrete display. Then, this may lead to the higher rate of the bouncing percept in the lower visual field.

Whatever the neural mechanisms underlying the visual field asymmetry are, this asymmetry in perceptual judgments may reflect the probability of bouncing events in the real environment. In the natural world, two (or more) objects cannot occupy the same place at the same time. Suppose that two objects move toward one another, coincide optically, then recede from each other. Optically coinciding objects in the lower field imply physical contact with higher probability than those in the upper visual field. This is because the lower visual field is usually occupied by the ground and thus optical coincidence is likely to reflect physical coincidence of two objects moving on the ground. On the other hand, the upper visual field is occupied by empty space so that it potentially has more depth than the lower visual field (Gibson, 1950). Consequently, the conditional probability of physical coincidence given optical coincidence would be lower.

Summary

Endogenous distraction of attention at the visual coincidence biases visual perception toward bouncing, despite the fact that there is no transient at the time of the visual coincidence. The results were interpreted in that attention may facilitate the intrinsic tendency of the visual system to perceive object motion as continuing to occur in the same direction as in the past. Thus, the amount of attentional resource assigned to the moving targets at the time of the visual coincidence seems to determine the streaming/bouncing percept.

Chapter VI: General Discussion

Summary of results

By using the streaming/bouncing ambiguous motion display, I have explored how the human perceptual system synthesizes visual perception under an ambiguous situation. The results are summarized as follows: (1) A salient sensory transient biases visual perception toward bouncing, irrespective of the modality in which a sensory transient is presented [amodal saliency]. (2) The magnitude of the bounce-inducing effect depends on both the saliency and timing (relative to the visual coincidence) of a sensory transient. (3) The more salient a sensory transient is, the stronger the bounce-inducing effect is [saliency dependency]. (4) However, the temporal window during which a sensory transient can bias visual motion perception is fixed [temporal gain modulation]. (5) Auditory, visual, and tactile transients have different temporal interaction windows. (6) Auditory and tactile transients presented before the visual coincidence tend to have stronger bounce-inducing effect than those presented after the visual coincidence; visual transients do not produce this temporal asymmetry. (7) The lack of visual attention to the visual moving targets in the streaming/bouncing motion display increases the bouncing percept and decrease the streaming percept. Based on these results, I conclude that the streaming percept arises when attentional resources are available for moving stimuli around the moment of the visual coincidence and the bouncing percept results from the lack of attentional resources. The role of the salient sensory transients is to distract attention from the coinciding visual motion stimuli. Many implications can be drawn from the present set of experiments, some of which will be discussed in the following.

Vision does not always dominate

Historically, the visual modality has been recognized as the primary sensory channel in humans (e.g., Rock, 1966). This is true in a sense; Losing vision has much more profound effect than losing the faculty in other modalities, and many examples of crossmodal interaction have shown the visual dominance over other modalities (Gibson, 1933; Hay et al., 1965; Easton & Moran, 1978; Driver & Spence, 1998; Calvert et al., 1998). However, as mentioned in Chapter I, the visual dominance may be overemphasized because the experimental paradigms of crossmodal interaction have involved mainly localization tasks. Therefore observers may rely more on visual information, with which location can be determined most correctly. The bounce-inducing effect in the streaming/bouncing motion display overturns this tradition, by introducing perceptual ambiguity in the visual stimulus. The present thesis is the first systematic investigation on the bounce-inducing effect.

Broad temporal tuning of crossmodal event perception

The temporal window during which a sensory transient influences visual motion perception is much broader for nonvisual-visual (crossmodal) interaction than visual-

visual (modal) interaction. The temporal interaction windows of the audiovisual and tactile-visual interactions are about 400-500 ms (Chapter II and Chapter IV) and 700 ms (Chapter IV), respectively. In contrast, a visual transient enhances the bouncing percept only when it is presented 80-ms before or after the visual coincidence (160 ms temporal window). Therefore, a broad temporal interaction window may be a general characteristic for crossmodal interaction (e.g., Stein and Meredith, 1993; Calvert et al., 1998). Because information processing speeds significantly differ among sensory modalities and depend on the modality to which attention is directed (Robinson, 1934; Goldstone, 1968; Posner et al., 1976; Regan, 1989), such a tolerance for asynchrony may have a functional significance in producing a unified perception of the sensory world. That is, it may allow two events occurring in different modalities to register *about* the same time in the system that synthesizes a unified perceptual world.

Dependence on context (the attenuation effect)

The bounce-inducing effect by sound depends on auditory context (the attenuation effect; Chapter III). The close correlation between auditory threshold for deviant sound detection and audiovisual threshold for the bounce-inducing effect (Experiment 8) suggests that the bounce-inducing effect by sound may have a strong functional significance in everyday environments in the sense that the auditory processing is the determinant for visual motion perception under circumstances where visual information is ambiguous.

This attenuation effect of the bounce-inducing effect may be a novel and useful method to study grouping effects that are crucial for event perception. The experiments in Chapter III have shown that auditory grouping for the bounce-inducing effect has a longer interaction range (up to 450 ms) than normal auditory grouping (Anstis & Saida, 1985; Bregman, 1990). This result intriguingly implies that a grouping mechanism that processes "audiovisual (amodal or crossmodal) continuity" and a grouping mechanism that processes "auditory (modal) continuity" may be different[10].

Further investigations about contextual effect on visual motion perception can be done by presenting repetitive tactile and visual stimuli with the streaming/bouncing motion display, similarly to the auditory sequence used in Chapter III. The results would provide important information as to whether this contextual effect (and dissociation between amodal/crossmodal and modal grouping processes) is specific for crossmodal interaction or common in defining sensory continuity.

Temporal gain modulation

Temporal gain modulation in the bounce-inducing effect has been observed with all the sensory transients used in this thesis (Chapter IV). As the intensity (or saliency) of a sensory transient increases, the bounce-inducing effect is enhanced without a change in

---

[10] This may be analogous to perception of music. While we can perceive each distinct note of melodies, we nonetheless still perceive the flow of melodies. The duration between the successive notes of melodies ranges from about 150 ms to 1 sec (Fraisse, 1963). However, interestingly, 150-290 ms and 300-900 ms tones make the fundamental rhythm in typical classical music (Fraisse, 1982). Although highly speculative, there might be a link between music perception and crossmodal continuity.

the temporal interaction range. In Chapter IV, I have argued that by taking the relative timing between sensory events as a stimulus dimension, temporal gain modulation may be implemented as facilitatory/inhibitory interaction among recurrently connected neurons (Salinas & Abott, 1996; Pouget & Sejnowski, 1997). This idea is in favor of the notion that gain modulation may be the general rule in the nervous systems (e.g., Poggio, 1990). In this regard, physiological experiments on the bounce-inducing effect in non-human primates will be interesting. First, the brain region where the bounce-inducing effect occurs should be determined (by functional imaging and/or neurophysiological methods), then the way in which neural activity would correspond with the change of perception should be examined. Note that the bounce-inducing effect is useful particularly for this purpose because it is difficult, if not impossible, to employ the McGurk effect as an audiovisual display in an experiment with non-human subjects.

Attention account for the bounce-inducing effect

At the first glance, the attention account for the bounce-inducing effect may sound counter-intuitive, given intuitive comprehension of the audiovisual associative learning account. However, it should be stressed that the attention account and the associative learning account are not mutually exclusive. Both may reflect and implement the likelihood of the combination of a physical collision and a transient event in the natural environment.

Geometrically, a point in a three-dimensional space can project onto a two-dimensional surface (such as the retina) in an infinite number of ways, which results in an 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 momentum makes physical objects move into the same direction as in the past may provide constraints to the visual system's interpretation and bias perception toward streaming as a default. Therefore, given a two-dimensional image where two moving objects coincide, the visual system may be unwilling to interpret it as a physical collision in a three-dimensional world (Sumi, 1995; Sekuler & Sekuler, 1999). When physical objects actually collide, however, a transient sound often occurs (Gaver, 1993ab) 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 adopt the bouncing interpretation (the bounce-inducing effect). Thus, streaming would be a more generic interpretation than bouncing without a sound, whereas bouncing would be more generic than streaming with a sound at the visual coincidence. Nonetheless, if such a simultaneous sound has a similar acoustic profile as the background sound, the physical cause of the sound should not be attributed to the visual coincidence. This is because, despite the fact that the simultaneous sound alone implies a bouncing event, it is highly improbable that the sound caused by the visual coincidence resembles the background noise (the attenuation effect). In this case, the likelihood of auditory signals belonging together supersedes the likelihood of one element from that group being caused by a collision. In a sense, the problem of general event perception (including crossmodal event perception) should be considered a problem of "inverse physics," not just inverse optics: The brain has to consider not only optics but also the physics of the world and to compute joint probabilities given sensory events within and between modalities. This is quite similar to the *generic view principle* that has been detailed in vision (Richards, 1988; Nakayama & Shimojo, 1992; Freeman, 1994; Knill et al., 1996).

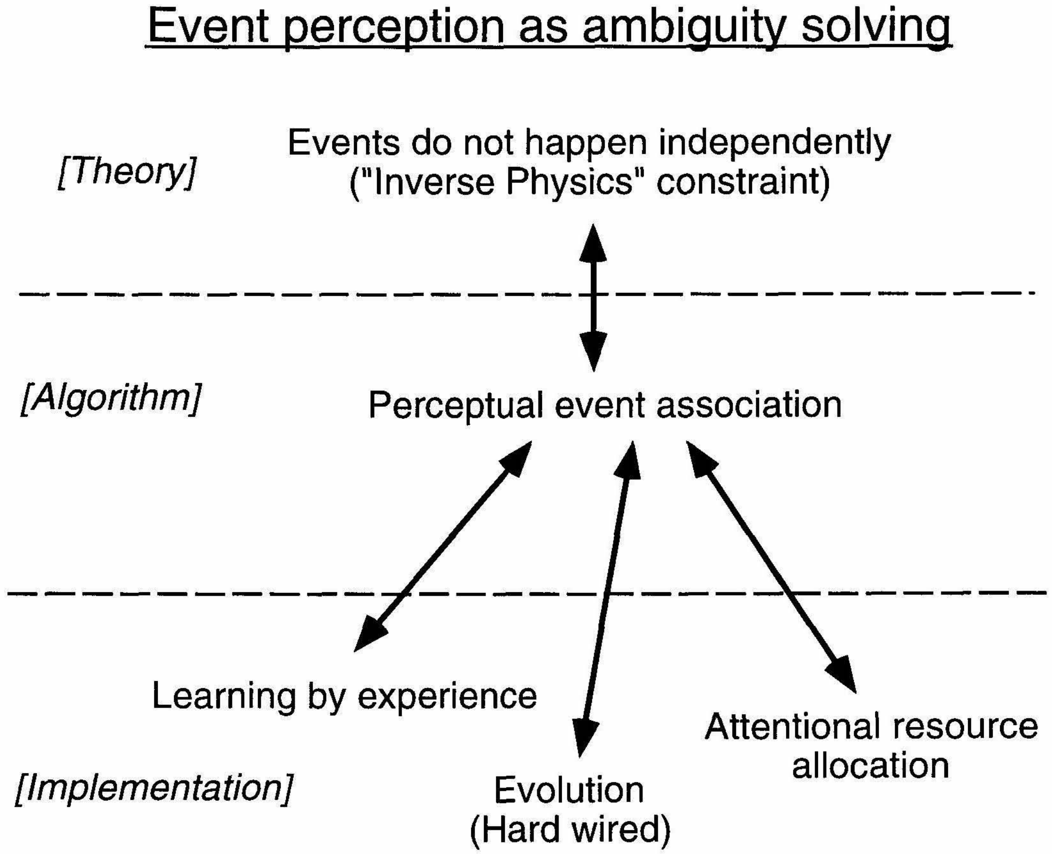

FIGURE 6.1 ATTENTION ACCOUNT AND ASSOCIATION LEARNING ACCOUNT.
Event perception is ambiguity solving. The fact that events in the world do not happen independently provides constraints in performing "inverse physics." Ambiguity solving is basically computing joint probabilities given sensory events within and between modalities (event association), which can be implemented in various ways.

This idea is in part compatible with the associative learning account. However, the associative learning account is a just-so theory that does not explain how such associations take place. Furthermore, the results of the present thesis cast doubt on the simplicity of the audiovisual associative learning account because all the transients (auditory, tactile, and visual) produced the bounce-inducing effect. If associative learning of sound-bouncing coupling was the main cause of the bounce-inducing effect, the bounce-inducing effect should have been restricted to audiovisual interaction. More importantly, associative learning based on experience is not an exclusive way to implement perceptual event coupling. As long as an algorithm can compute and represent joint probabilities given sensory events within and between modalities, any mechanism would suffice (Figure 6.1); facing the findings reported here, it is highly plausible to implement associative learning as a dynamic of attentional recourse allocation.

The attention account can capture most characteristics of the bounce-inducing effect by introducing the concepts of amodal saliency, attentional capture, and attentional modulation of visual motion perception (Figure 6.2). Also, there is additional evidence from functional imaging studies that transient sensory events in non-visual modalities deactivate visual responses in visual association areas (Kawashima et al., 1995; Downer et al., 2000). Additionally, another set of evidence has recently come from the developmental study of the bounce-inducing effect in our laboratory, which has demonstrated that at 6 months of age the bounce-inducing effect by sound becomes functional (Scheier et al., submitted). This age is thought to correspond to the period around which the attentional system of infants matures (Clohessy et al., 1991; Hood, 1993; Johnson & Tucker, 1996).

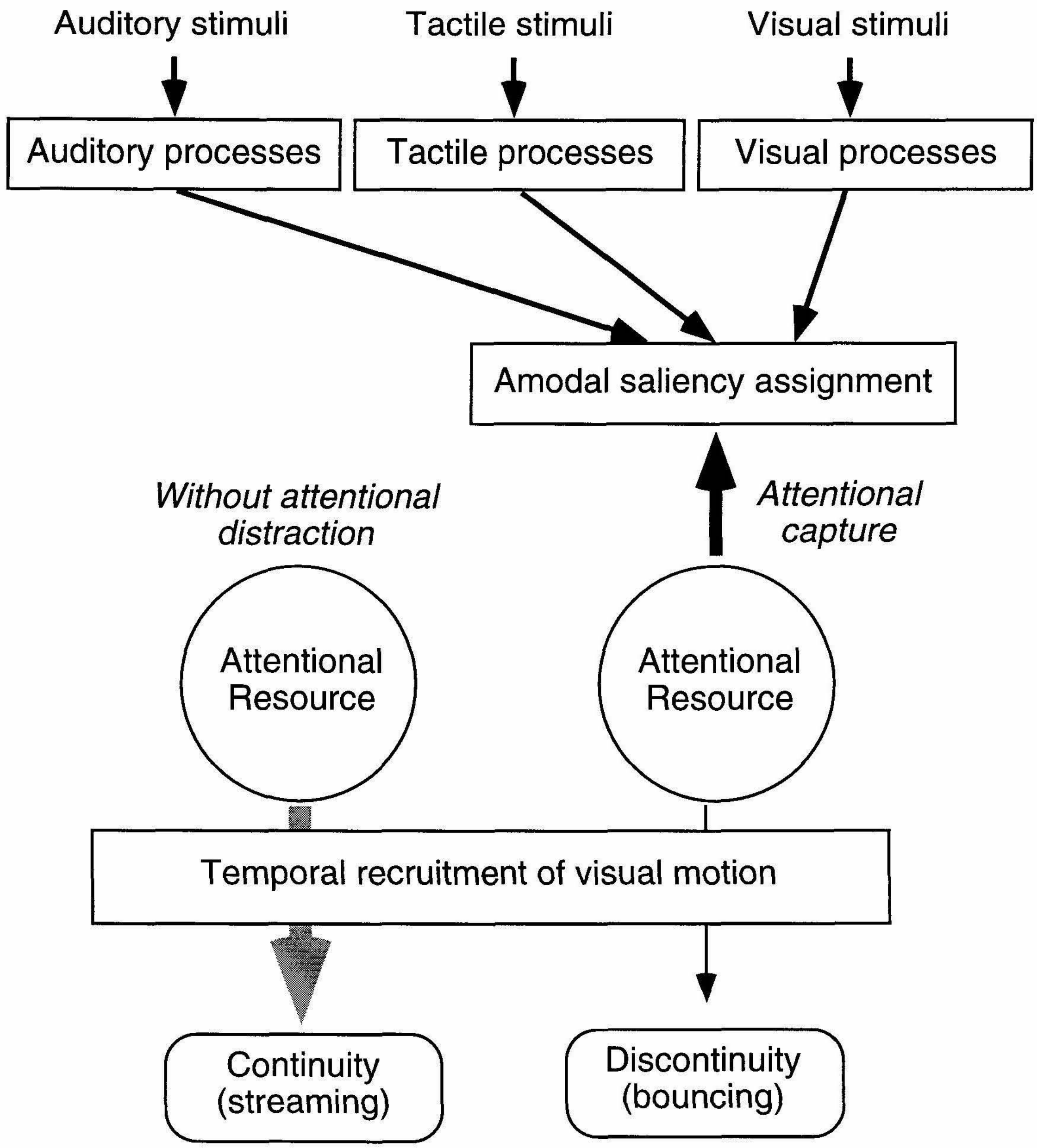

FIGURE 6.2 ATTENTION ACCOUNT FOR THE BOUNCE-INDUCING EFFECT.

Any salient sensory event decreases (via attentional capture) attentional resource available to temporal recruitment of visual motion, which leads to the enhancement of the bouncing percept.

Although the attention account has successfully explained the various phenomena in the bounce-inducing effect, more rigorous examinations of the attention account may be desirable. For example, I did not include an experiment where auditory endogenous attention is manipulated, partially because a large inter-observer difference was observed and the results were equivocal with respect to the attention account. I speculate that this inter-observer variance might reflect the difficulty, or variability in the observer's ability, to endogenously attend an auditory sequence while viewing a visual display. It has been shown that normal observers tended to allocate attentional resource to the visual modality, and there is a large inter-observer difference in endogenous attentional allocation to the auditory modality (Posner et al., 1976). In the future, I would like to devise an experiment to overcome this problem.

Attentional modulation of perceptual continuity

Finally, I would like to speculate about an extension of the attention account. Briefly, the attention account for the bounce-inducing effect suggests that attention may facilitate the intrinsic tendency of the visual system to perceive object motion as continuing to occur in the same direction as in the past. Is this restricted in the visual modality? Or, can it be said that attention tends to make the perceptual system interpret events as continuing to occur in the same way as in the past (the attention-for-continuity hypothesis)? For example, Bregman and his colleagues have described an auditory counterpart of the streaming/bouncing visual motion display (McAdams & Bregman, 1979; Ciocca & Bregman, 1987; Tougas & Bregman, 1985, 1990). The stimulus pattern is an X pattern consisting of two simultaneously gliding tones, one ascending frequency and the other descending frequency (Figure 6.3). The sequential grouping of the two tone glides can be based on the proximity principle (grouping on the basis of proximity in frequency) and the trajectory principle (grouping on the basis of a regular trajectory on frequency × time coordinates). The trajectory principle favors the streaming percept (or the crossing percept called by Bregman), but the auditory perception can be ambiguous, depending on the slope of tone gliding. It would be very interesting to see whether a visual flash roughly synchronized at the moment of the auditory coincidence biases auditory perception toward bouncing, and whether an endogenous shift of attention away from the auditory modality enhances the auditory bouncing percept. If so, it would

---

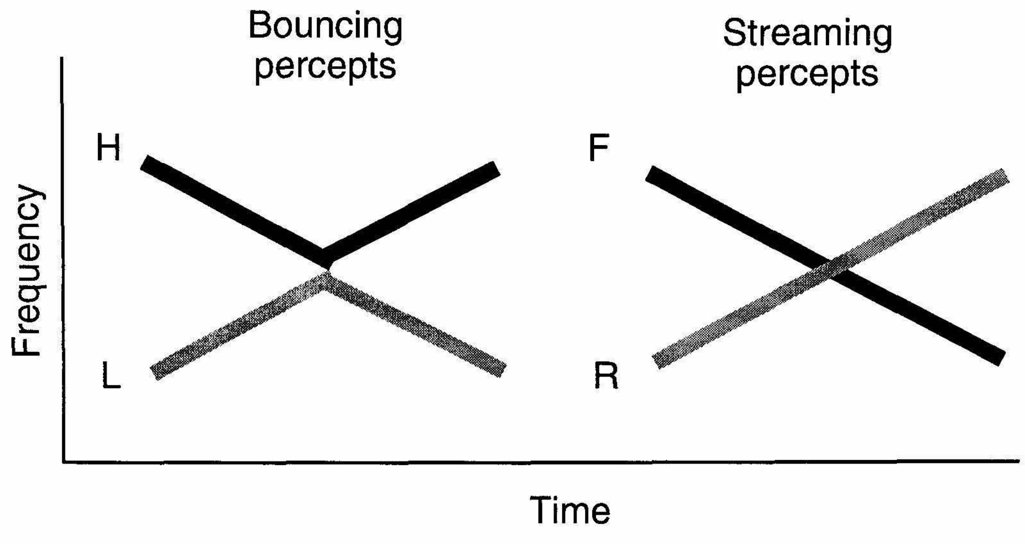

FIGURE 6.3 AUDITORY STREAMING/BOUNCING DISPLAY.

H = a high glide falling and then rising; L = a low glide rising and then falling; F = a falling glide; R = a rising glide. Modified from Tougas and Bregman (1990).

support a magnificent hypothesis that the general perceptual continuity of the sensory world is partly mediated and maintained by attentional mechanisms.

Concluding remarks

Additional measures of neuronal activity, such as functional imaging and single electrode physiology, will be necessary to directly resolve the issues discussed here; however, as I think I have succeeded in doing, psychophysical studies can also illuminate the most likely candidates for the underlying mechanism, define how such physiological studies can be designed, and suggest further interesting psychophysical experiments. In this respect, I believe the present thesis has advanced the knowledge of how the human brain synthesizes the unified perceptual world from different sensory channels.

Appendix A: Tables

In all tables, bold fonts indicate significance less than 0.05.

Table 1. Post-hoc statistical comparisons in Experiment 1

delay (ms)	F(1,8)	p
-1050	0.008016	0.92977
-850	1.545	0.23182
-650	4.742	0.04474
-450	1.578	0.22714
-250	**5.253**	**0.035801**
-200	**8.081**	**0.011757**
-150	**37.24**	**1.53E-05**
-100	**202.4**	**1.69E-10**
-50	**526.7**	**1.13E-13**
0	**349.3**	**2.7114E-13**
50	**282.5**	**1.37E-11**
100	**11.84**	**3.36E-03**
150	**5.107**	**3.81E-02**
200	1.578	0.22714
250	1.279	0.2748
450	0.5046	0.48771
650	0.8525	0.36957
850	0.7079	0.41254
1050	2.859	0.11026

Table 2. Post-hoc statistical comparisons in Experiment 2

delay (ms)	F(1,8)	p
-1050	0.1818	0.6755
-850	0.1447	0.7087
-650	1.609	0.22273
-450	0.008475	0.9278
-250	**7.171**	**0.016493**
-200	**7.976**	**0.012217**
-150	**35.7**	**1.9416e-05**
-100	**165.8**	**7.373e-10**
-50	**407.9**	**8.2234e-13**
0	**301.7**	**6.6472e-12**
50	**301.5**	**8.3671e-12**
100	**13.12**	**0.0022908**
150	**5.438**	**0.03309**
200	0.2837	0.60158
250	0.0241	0.87858
450	0.1085	0.74617
650	0.00938	0.92405
850	0.008475	0.9278
1050	1.535	0.23325

Table 3. Statistical analyses in Experiment 3:

Bounce-inducing effect vs. Temporal order judgment

	Bounce-inducing effect		Temporal Order judgment	
delay (ms)	F(1,6)	p	t(6)	p
-450	0.12	0.73503	inf.	0
-250	**5.106**	**0.043232**	-69	6.2341e-10
-200	**7.047**	**0.020996**	-32.5	5.6435e-08
-150	**13.85**	**0.0029213**	-16.395	3.2798e-06
-100	**42.76**	**2.7726e-05**	-18.33	1.6986e-06
-50	**51.36**	**1.1378e-05**	**-4.6663**	**0.50316**
0	**73.77**	**1.8033e-06**	**-0.71207**	**0.50316**
50	**33.7**	**8.4009e-05**	**1.6859**	**0.1428**
100	**9.704**	**0.0089358**	11.921	2.1101e-05
150	**6.125**	**0.029231**	17.526	2.2141e-06
200	0.1364	0.71836	52.673	3.1429e-09
250	0.12	0.73503	69	6.2341e-10
450	0.3636	0.55772	inf.	0

Table 4. Post-hoc statistical comparisons in Experiment 5:

Additional sound

delay (ms)	F(1,8)	p
-1050	0.1096	0.74486
-850	1.798	0.19874
-650	1.057	0.31922
-450	**5.056**	**0.038984**
-250	**11.29**	**0.003979**
-200	**8.529**	**0.010006**
-150	1.621	0.22109
-100	1.041	0.32286
-50	0.06433	0.80302
50	0.001612	0.96847
100	0.1249	0.7284
150	**4.356**	**0.053229**
200	0.4093	0.53139
250	0.00189	0.96586
450	0.03347	0.85714
650	0.5003	0.48954
850	0.1376	0.71552
1050	0.00887	0.76967

Table 5. Post-hoc statistical comparisons in Experiment 5: With auditory flankers

SOA (ms)	F(1,8)	p
-1050	2.341	0.14552
-850	0.8328	0.37501
-650	3.401	0.083756
-450	**15.24**	**0.001264**
-250	**32.93**	**3.05E-05**
-200	**7.22**	**0.016195**
-150	5.341	0.034484
-100	0.3564	0.55887
-50	0.132	0.72108

Table 6. Post-hoc statistical comparisons in Experiment 9:

Significance of the bounce-inducing effect compared with the no-sound condition

sound timing	F(1,8)	p
At the 1st sound	**240.1**	**4.6955E-11**
At the 2nd sound	**19.18**	**0.00046639**
50-ms after 2nd	**18.95**	**0.00049283**
100-ms after 2nd	**11.94**	**0.0032574**
150-ms after 2nd	**15.28**	**0.0012496**
200-ms after 2nd	4.944	0.040935
250-ms after 2nd	1.218	0.28607
300-ms after 2nd	2.472	0.13545

Table 7. Statistical results in Experiment 11:

Auditory transient [corresponding probabilities of *t*-test]

delay (ms)			
-1280	0.76396	0.55904	0.70083
-640	0.3873	0.4439	0.61557
-320	0.1815	0.5	0.21521
-160	**9.37E-05**	**4.27E-05**	**4.21E-05**
-80	**3.12E-05**	**2.29E-05**	**6.18E-06**
-40	**0.00011454**	**6.36E-06**	**3.79E-06**
0	**5.44E-05**	**0.00025423**	**3.10E-05**
40	**0.0035441**	**0.0016407**	**5.59E-05**
80	**0.00028376**	**1.86E-05**	**5.70E-05**
160	0.54098	0.10269	0.24246
320	0.44096	0.39244	0.42827
640	0.43096	0.77887	0.37031
1280	0.39694	0.60756	0.87523

Table 8. Statistical results in Experiment 12:

Tactile transient [corresponding probabilities of t-test]

delay (ms)			
-1280	0.59908	0.38134	0.090131
-640	**0.024934**	0.26038	0.13767
-320	**0.00036752**	**0.00027838**	**1.56E-06**
-160	**0.0081622**	**0.00013098**	**9.52E-07**
-80	**0.00076526**	**0.0029898**	**8.04E-08**
-40	**0.0050374**	**8.70E-05**	**1.02E-06**
0	**0.00056689**	**0.0023801**	**0.00013808**
40	0.11943	**9.03E-05**	**0.001908**
80	0.1108	0.29917	0.26405
160	0.43337	0.33111	0.67544
320	0.62568	0.43096	0.5
640	0.42827	0.65068	0.090131
1280	0.63901	0.40766	0.17531

Table 9. Statistical results in Experiment 13:

Visual transient [corresponding probabilities of *t*-test]

delay (ms)			
-1280	0.20075	0.38134	0.78262
-640	0.5	0.5	0.42827
-320	0.65068	0.5	0.68827
-160	0.17531	0.13767	0.56663
-80	0.24931	0.080207	**0.0039874**
-40	**0.082814**	**0.0041755**	**0.00015511**
0	**0.00036752**	**6.38E-05**	**4.79E-08**
40	**0.016197**	**0.0068687**	**0.00015349**
80	0.11268	0.018898	**0.0017093**
160	0.29089	0.30575	0.62202
320	0.58681	0.29186	0.57173
640	0.17531	0.34932	0.42525
1280	0.5	0.25824	0.76502

Table 10. Trial-basis analysis in Experiment 14:

Responses were pooled for all the observers and categorized as follows.

	Perceptual judgment		
Central-task Response	***Bouncing***	***Streaming***	***Total***
Correct	1173	1145	2318
Wrong	32	50	82
Total	1250	1195	2400

Table 11. Trial-basis analysis in Experiment 15:

Responses were pooled for all the observers and categorized as follows.

	Perceptual judgment		
Central-task Response	***Bouncing***	***Streaming***	***Total***
Correct	1414	861	2275
Wrong	60	65	125
Total	1474	926	2400

Appendix B: Control experiments of the attenuation effect

The involvement of auditory masking in the attenuation effect is very unlikely because the attenuation effect occurred with a time interval (over 400 ms) with which auditory masking would not occur (Wright, 1964; Massaro, 1970, 1975; Fastl, 1976; Massaro et al., 1976; Kallman & Morris, 1984; Viemeister & Plack, 1993). However, there is a possibility that the particular audiovisual display used in the present experiments made an unforeseen masking process functional. In the following set of control experiments, I eliminated even a slight possibility of auditory masking in the attenuation effect. The question is simple: Do observers hear the sound preceded and followed by additional sounds as being *acoustically* distinct from the single sound?

My preliminary observations had shown that, whereas the pitch and timbre of the single simultaneous sound do not substantially alter the bounce-inducing effect, the effect becomes smaller as (1) the sound intensity becomes lower, (2) the sound duration becomes longer, and/or (3) the asynchrony between the sound and the visual coincidence becomes larger (as in Experiment 1) (Watanabe et al., 1999). Therefore, I set out to examine whether the auditory flankers would alter these aspects of the simultaneous sound in the presence of the streaming/bouncing motion display.

Control experiment 1: Attenuating the bounce-inducing effect (replication)

The aim of Control experiment 1 was to replicate the attenuation effect so that I could compare the results with the same observers and the same audiovisual display.

Method

Observers

Seven people (20 to 48 years of age; including the author), 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

The visual stimuli were identical to those of Experiment 1. A 1800 Hz tone-burst was presented at (At), 300-ms before (Pre), or 300-ms after (Post) the visual coincidence, through a built-in speaker of the computer. Additionally, no sound (None) or three consecutive sounds (All-3; Pre+At+Post) were presented for one-fifth of the trials each. 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

The experimental procedures were the same as those in Experiment 1. For five sound conditions (None, Pre, At, Post, and All-3), 20 trials were repeated randomly in a single session (100 trials).

Results and Discussion

In Control experiment 1, the sounds presented 300-ms before and 300-ms after the visual coincidence did not induce the bouncing percept (Figure 7.1; None vs. Pre, $F(1,6) = 2.95, p = 0.11$; None vs. Post, $F(1,6) = 0, p = 1$). In contrast, the single sound presented at the visual coincidence enhanced perception of bouncing (None vs. At, $F(1,6) = 45.36, p < 0.001$). When this simultaneous sound was flanked by other identical sounds (All-3), the bounce-inducing effect was present (None vs. All-3; $F(1,6) = 4.39; p < 0.06$) but attenuated significantly (At vs. All-3; $F(1,6) = 5.88; p < 0.05$). All these results were consistent with the previous experiments. In the next three experiments, I examined whether the auditory perception of the simultaneous sound is altered when it is flanked by auditory flankers.

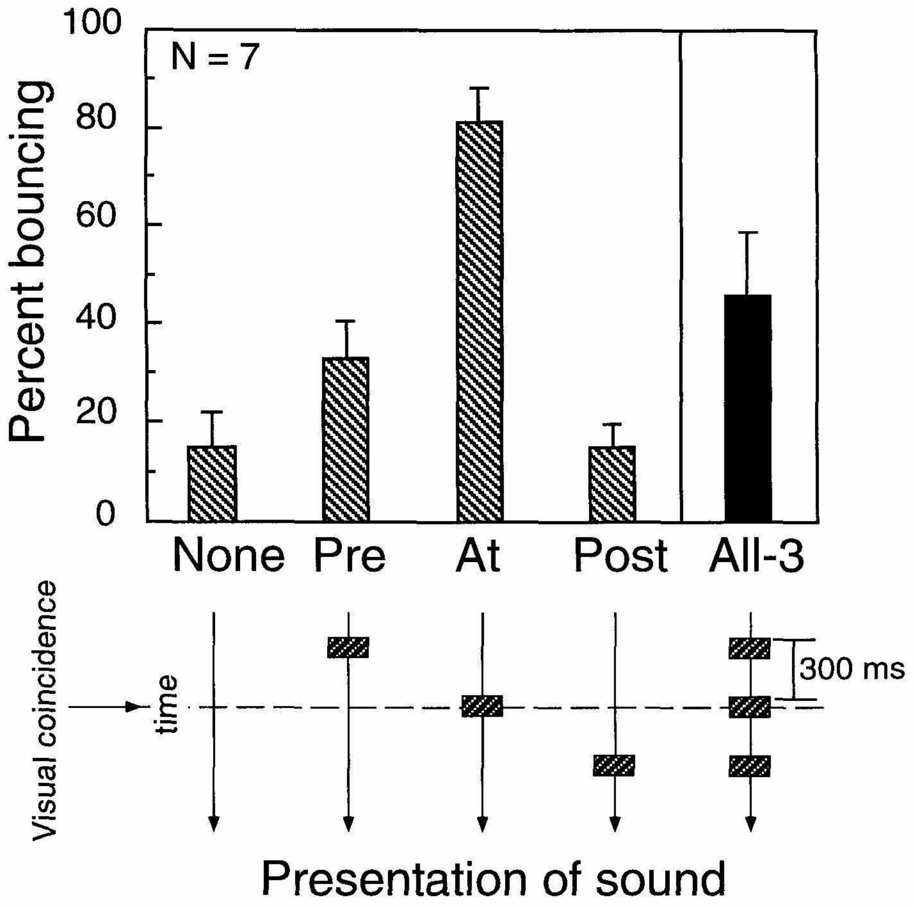

FIGURE 7.1 RESULTS OF CONTROL EXPERIMENT 1.
Standard errors are shown. The schematic below depicts the timing of the sound for each condition, with the rectangles representing the sounds.

Control experiment 2: Do flankers affect perceived loudness?

Control experiment 2 was conducted to determine whether the simultaneous sound might be perceived as being less audible with auditory flankers (Figure 7.2).

Method

Control experiment 2a

The stimulus was almost identical to that in the All-3 condition of Control experiment 1. However, the intensity of the simultaneous sound was varied from 56 dB to 60 dB (0.5 dB-step), while the intensities of the auditory flankers were kept constant (58 dB). Observers compared the loudness of the simultaneous sound with the loudness of the auditory flankers (Within condition; see Figure 7.2). There were 20 trials with each intensity and the order of trials was random.

Control experiment 2b

The stimuli used in the All-3 condition and the At condition were presented successively in random order. The interval between each display was about 1 s. As in Control experiment 2a, the intensity of the simultaneous sound in the All-3 condition was varied. The intensity of the single simultaneous sound was 58 dB. Observers compared the loudness of the simultaneous sound in the All-3 condition with the loudness of the simultaneous sound presented alone (Between condition). There were 20 trials with each intensity and the order of trials was random.

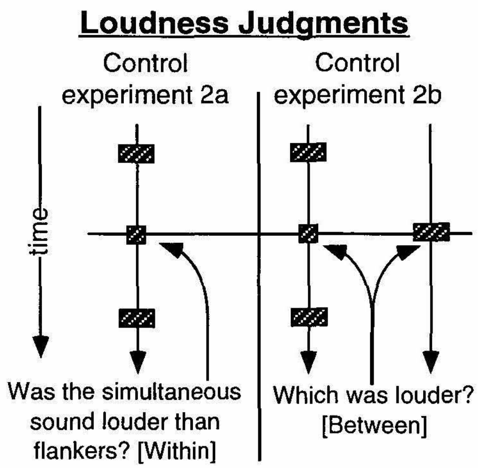

Figure 7.2 Stimulus presentation and task in Control experiment 2.
Loudness judgment of the simultaneous sound relative to the auditory flankers (Within condition; Control experiment 2a), and relative to the other single sound (Between condition; Control experiment 2b).

Results and Discussion

Figure 7.3 shows the results of Control experiment 2. A two-way ANOVA revealed that there was no significant difference between the Within and the Between conditions ($F(1,6) = 0.51, p = 0.479$). When the intensity of the simultaneous sound was identical to that of the auditory flankers (i.e., 58 dB), the mean percentage of "the simultaneous sound was louder (compared with the auditory flankers)" judgments in Control experiment 2a was 48.6% (SD = 14.64). This is not significantly different from chance (two-tailed *t*-test; $t(7) = -0.26, p = 0.81$). Likewise, the mean percentage of "the simultaneous sound was louder (compared with the single sound)" judgments in Control experiment 2b was 51.4% (SD = 19.52). This does not differ from chance, either ($t(7) = 0.19, p = 0.85$). Thus, I found no evidence for a perceived change in the loudness of the simultaneous sound with the auditory flankers.

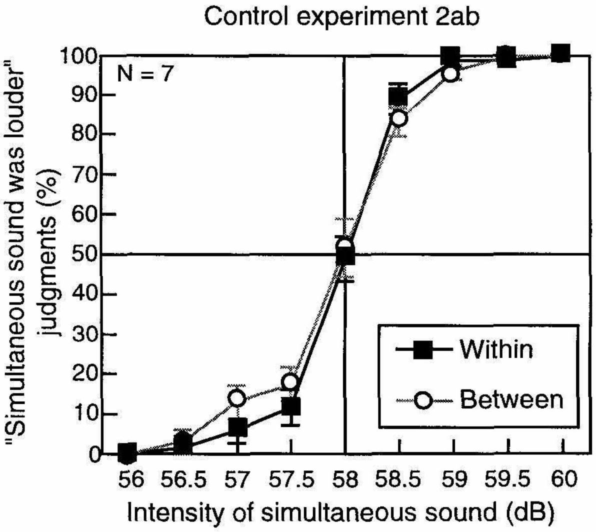

Figure 7.3 Results of Control experiment 2.

Percentage of "Simultaneous sound was louder" judgments were plotted against the intensity of the simultaneous sound. The vertical line indicates the value of the comparison sound. The horizontal line shows chance performance.

Control experiment 3: Do flankers affect perceived duration?

Control experiment 3 was conducted to determine whether the simultaneous sound might be perceived as lasting longer with auditory flankers (Figure 7.4).

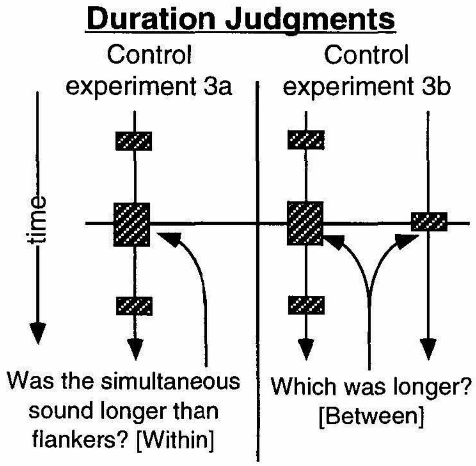

FIGURE 7.4 STIMULUS PRESENTATION AND TASK IN CONTROL EXPERIMENT 3. Duration judgment of the simultaneous sound relative to the auditory flankers (Within condition; Control experiment 3a), and relative to the other single sound (Between condition; Control experiment 3b).

Method

Control experiment 3a

The stimulus and procedure were almost identical to those of Control experiment 2a. However, the sound intensity was kept at 58 dB and the duration of the simultaneous sound was varied from 1 ms to 32 ms. Observers reported whether the simultaneous sound was longer or shorter than the auditory flankers (Within condition). There were 20 trials with each duration and the order of trials was random.

Control experiment 3b

The stimulus and procedure were almost identical to those of Control experiment 2b. But, as in Control experiment 3a, the duration of the simultaneous sound was varied. Observers compared the duration of the simultaneous sound flanked by the other sounds with that of the simultaneous sound presented alone (Between condition). There were 20 trials with each duration and that the order of trials was random.

Results and Discussion

The results of Control experiment 3 are shown in Figure 7.5. The difference between the Within and the Between conditions failed to reach significance (*F*(1,6) = 0.77, *p* = 0.38). If the simultaneous sound and the auditory flankers had the same duration (3 ms), the observers reported that the simultaneous sound was longer in 52.9% of the trials on average (SD = 14.96) when compared with the auditory flankers, and in 44.3% (SD = 17.18) when compared with the isolated simultaneous sound. Neither of these differs

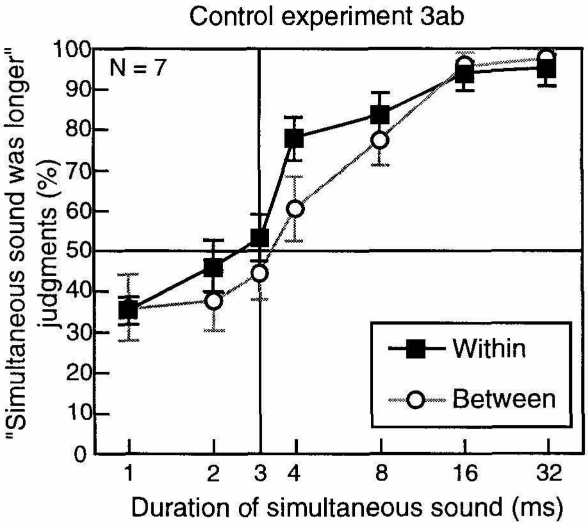

FIGURE 7.5 RESULTS OF CONTROL EXPERIMENT 3.

Percentage of "Simultaneous sound was longer" judgments were plotted against the log of duration of the simultaneous sound. The vertical line indicates the value of the comparison sound. The horizontal line shows chance performance.

from chance (Within, $t(7) = 0.51, p = 0.63$; Between, $t(7) = -0.88, p = 0.41$). Hence, it was not evident that the auditory flankers altered the perceived duration of the simultaneous sound.

Control experiment 4: Do flankers affect perceived timing?

Control experiment 4 was conducted to determine whether the auditory flankers might induce uncertainty as to the timing of the sound relative to the visual coincidence (Figure 7.6).

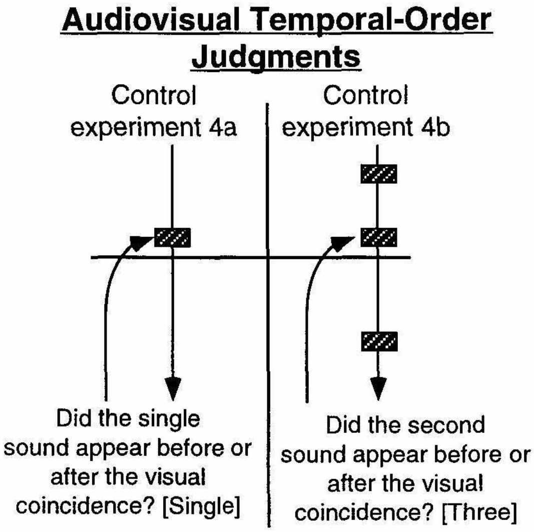

FIGURE 7.6 STIMULUS PRESENTATION AND TASK IN CONTROL EXPERIMENT 4.
Temporal order judgment between the visual coincidence and the single sound (Single condition; Control experiment 4a), and between the visual coincidence and the single-second sound in the three sound sequence (Three condition; Control experiment 4b).

Method

First, I presented the single sound around the time of the visual coincidence (from 200-ms before to 200-ms after; 50-ms step) and asked observers to judge whether the single sound appeared before or after the coincidence (the Single condition). For each asynchrony, 20 trials were repeated randomly. Next, in addition to the stimulus in the Single condition, I added the auditory flankers 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 (the Three condition). There were 20 trials with each audiovisual asynchrony and that the order of trials was random.

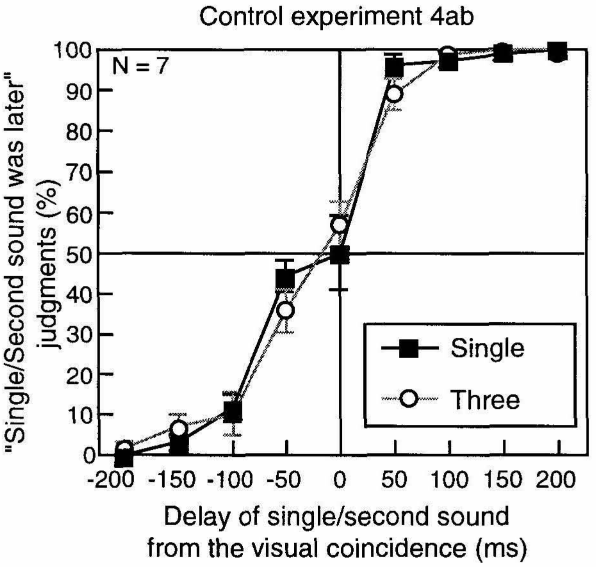

FIGURE 7.7 RESULTS OF CONTROL EXPERIMENT 4.

Percentage of "Single/Second sound was later than the visual coincidence" judgments were plotted against the delay of the single/second sound from the visual coincidence. The vertical line indicates the value of the comparison sound. The horizontal line shows chance performance.

Results and Discussion

Figure 7.7 shows the results of Control experiment 4. The two psychometric curves, representing results with (the Three condition) and without the auditory flankers (the Single condition), are similar. No difference was found between these two conditions (two-way ANOVA, $F(1,6) = 0.008$, $p = 0.93$). Thus, the perceived synchrony between the sound and the visual coincidence was not affected by the auditory flankers.

General discussion

The four control experiments 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. That is, there is no evidence for auditory masking to take place in the attenuation effect.

Besides this, the results are intriguing because it implies that there is an aspect of auditory-grouping (saliency-assigning) processes which is context-sensitive and can be utilized by the visual system for ambiguity solving, but is nonetheless not manifest in the perception of the main acoustic characteristics of individual sound elements. In other words, some crossmodal effects can occur without changing perceptual elements in the critical modality. The dissociation between intramodal perception and crossmodal perception would be interesting to investigate in the future.

Appendix C: A control experiment for eye-movement

The purpose of this control experiment was to examine the possibility that involuntary eye movements might cause the dominance of the streaming percept in the present study. I informally observed that when observers actively tracked one of the visual targets with their eyes, the streaming percept always occurred (see the general discussion section of Chapter V). This issue is particularly important for interpreting the results in Chapter V (attentional modulation of motion event perception). If the observer's eyes involuntary tracked one of the visual targets, then the results of attentional modulation might be caused by those involuntary eye movements of the observers. Thus, I decided to see if I could duplicate the results of Experiment 14 while monitoring eye movements to make sure that the observers do not move their eyes.

Method

Two observers participated in the eye-movement control experiment. The stimulus and procedure were almost identical to those of Experiment 14. However, the fixation stimulus was a white cross (0.8 deg) and the horizontal bar of the fixation cross was elongated to the left or the right by one pixel (1.2 min) for one frame (about 13.9 ms). The central task of the observer was to determine which part (left or right) of the fixation cross underwent elongation. Since the change (i.e., elongation) was so subtle that the observers had to keep fixating and attending the fixation cross. The central task stimulus was presented at five possible timings as in Experiment 14 (0, 69, 139 ms before or after the visual coincidence). The visual targets were presented only in the upper visual field.

Eye movement data were collected by means of a sclera reflection device (*Ober2, Permobil Meditech, Sweden*) at a sampling rate of 200 Hz. Only the movements of the right eye were stored for off-line analysis.

Results and Discussion

The results of the eye-movement control experiment are in Figure 8.1. For both observers the "right shoulder" pattern was replicated, confirming that the observers performed in the same way as in Experiment 14. The performance on the central task was 92.2% for KW and 91.8% for SS.

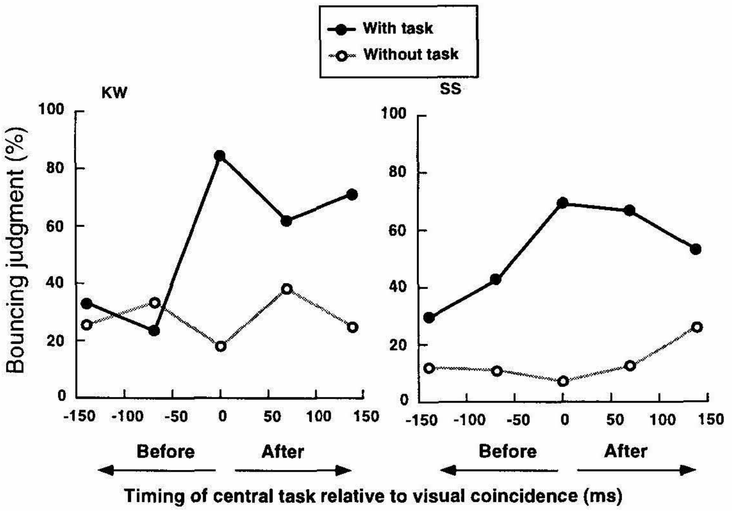

FIGURE 8.1 RESULTS OF EYE-MOVEMENT CONTROL EXPERIMENT.
The vertical axis shows the percentage of bouncing judgments. Gray lines indicate the conditions without the center task and black lines indicate the conditions with the center task. The "right shoulder" pattern was replicated.

Figure 8.2 shows the frequency distributions of horizontal eye positions. Eye position data were categorized based on the perceptual judgments that the observers made (i.e., streaming or bouncing). If the streaming percept were due to involuntary tracking of the visual targets, the gaze deviation should have been larger in trials where the observers reported the streaming percept than in trials where observers reported the bouncing percept. However, when the central task was required, the frequency distributions indicate that the behaviors of the eyes were quite the same irrespective of the observer's percept. The eyes of the observers seemed to deviate more from the center of the fixation cross when the central task was required, which may be interpreted in that the observers might actively inspect the fixation cross more frequently in order to perform the central task, but again no difference was found between eye movements in the 'bouncing' trials and the 'streaming' trials. Overall, the eye-movement control experiment confirmed that the observers did not track the visual targets during the streaming/bouncing judgment. Or, even if there were (very small; less than 1 deg) involuntary eye movements, those did not correlate with the perceptual judgments that the observers made.

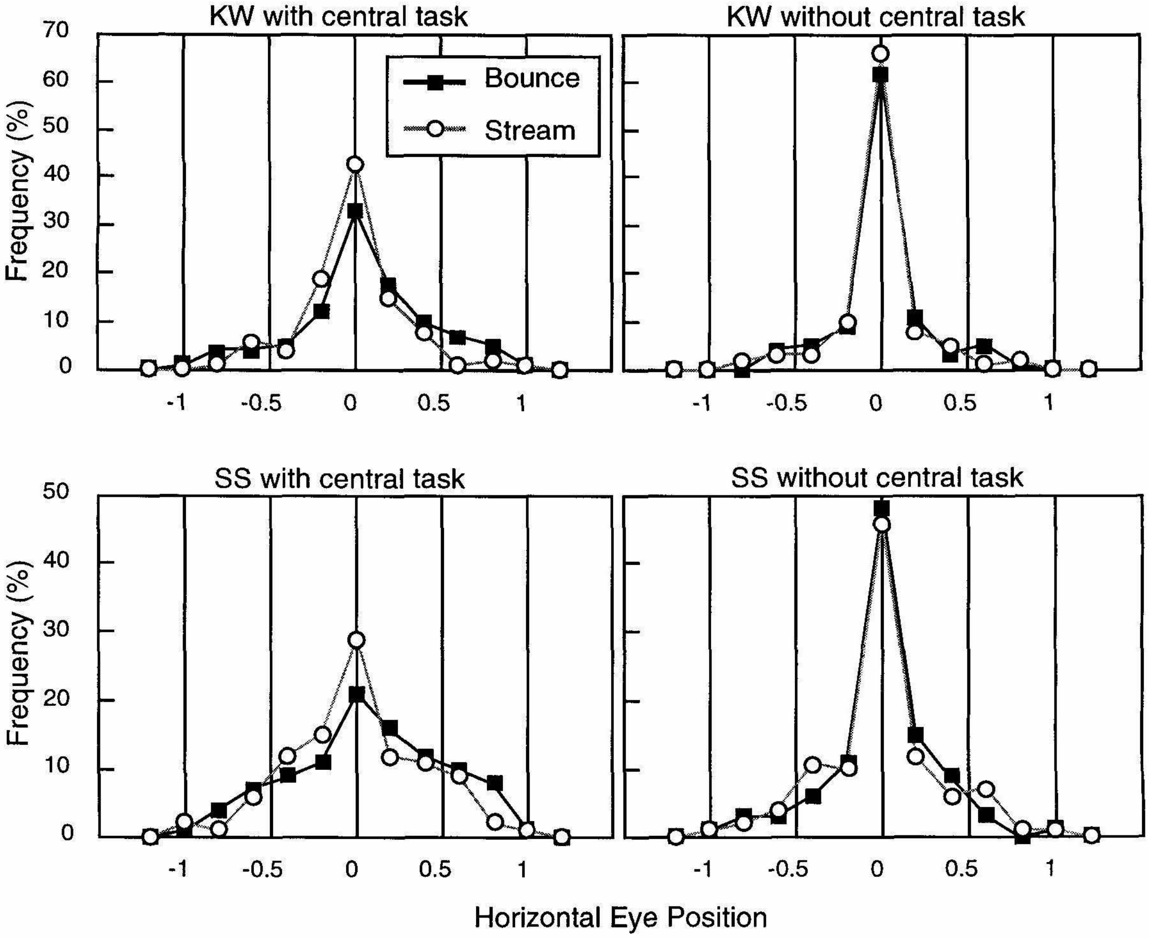

FIGURE 8.2 FREQUENCY DISTRIBUTIONS OF HORIZONTAL EYE POSITIONS.

The horizontal axis indicates eye position in visual angle. Negative values mean that the eyes were deviated toward the left of the fixation stimulus. The vertical axis shows the relative frequency of eye position. Gray lines indicate eye movements when the observers reported the bouncing percept and black lines indicate eye movements when the observers reported the streaming percept. Note that there is no difference between these conditions.

References

Alain, C., Woods, D. L., & Knight, R. T. (1998). A distributed cortical network for auditory sensory memory in humans. *Brain Research,* 812, 23-37.

Alain, C., Woods, D. L., & Ogawa, K. H. (1994). Brain indices of automated pattern processing. *NeuroReport*, 6, 140-144.

Albrecht, D. G., & Hamilton, D. B. (1982). Striate cortex of monkey and cat: contrast response function. *Journal of Neurophysiology*, 48, 217-237.

Alho, K., Houtilainen, M., Tiitinen, H., Ilmoniemi, R. J., Knuutila, J., & Näätänen, R. (1993). Memory-related processing of complex sound patterns in human auditory cortex: a MEG study. *NeuroReport*, 4, 391-394.

Andersen, R. A. (1989). Visual and eye movement functions of the posterior parietal cortex. *Annual Review in Neuroscience*, 12, 377-403.

Andersen, R. A. (1995). Encoding of intention and spatial location in the posterior parietal cortex. *Cerebral Cortex*, 5, 457-469.

Andersen, R. A., Bracewell, R., Barash, S., Gnadt, J., & Fogassi, L. (1990). Eye position effect on visual memory and saccade-related activity in areas LIP and 7a of macaque. *Journal of Neuroscience*, 10, 1176-1196.

Andersen, R. A., Essick, G., & Siegel, R. (1985). Encoding of spatial location by posterior parietal neurons. *Science*, 230, 456-458.

Andersen, R. A., & Mountcastle, V. B. (1983). The influence of the angle of gaze upon the excitability of the light-sensitive neurons of the posterior parietal cortex. *Journal of Neuroscience*, 3, 532-548.

Andersen, R. A., Snyder, L. H., Bradley, D. C., & Xing, J. (1997). Multimodal representation of space in the posterior parietal cortex and its use in planning movements. *Annual Review in Neuroscience*, 20, 303-330.

Andersen, R. A., Snyder, L. H., Li, C.-S., & Stricanne, B. (1993). Coordinate transformations in the representation of spatial information. *Current Opinion in Neurobiology,* 3, 171-176.

Andersen, R. A., & Zipser, D. (1988). The role of the posterior parietal cortex in coordinate transformations for visual-motor integration. *Canadian Journal of Physiology and Pharmacology*, 66, 488-501.

Anstis, S. M., & Ramachandran, V. S. (1986). Entrained path deflection in apparent motion. *Vision Research,* 26, 1731-1739.

Anstis, S. M., & Saida, S. (1985). Adaptation to auditory streaming of frequency-modulated tones. *Journal of Experimental Psychology: Human Perception and Performance*, 11, 257-261.

Balz, G. W., & Hock, H. S. (1997). The effect of attentional spread on spatial-resolution. *Vision Research,* 37, 1499-1510.

Beauchamp, M. S., Cox, R. W., & DeYoe, E. A. (1997) Graded effects of spatial and featural attention on human area MT and associated motion processing areas. *Journal of Neurophysiology,* 78, 516-520.

Ben-Yishai, R., Lev Bar-Or, R., & Sompolinsky, H. (1995). Theory of orientation tuning in visual cortex. *Proceedings of the National Academy of Science USA*, 92, 3844-3848.

Berger, E. H. (1981). Reexamination of the low-frequency (50-1000 Hz) normal threshold of hearing in free and diffuse sound field. *Journal of the Acoustical Society of America*, 70, 1635-1645.

Bermant, R. I., & Welch, R. B. (1976). The effect of degree of visual-auditory stimulus separation and eye position upon spatial interaction of vision and audition. *Perceptual and Motor Skills*, 43, 487-493.

Bertelson, P. (1997). Ventriloquism: a case of crossmodal perceptual grouping. In G. Ashersleben, T. Bachmann, & J. Müsseler (Eds.), *Cognitive Contributions to the Perception of Spatial and Temporal Events* (pp. 347-362). Amsterdam: Elsevier Science.

Bertelson, P. (1998). Starting from the ventriloquist: The perception of multimodal events. In M. Sabourin, F. I. M. Craik, & M. Robert (Eds.), *Advances in psychological science, Vol. 2: Biological and cognitive aspects* (pp. 419-639). Hove, UK: Psychology Press.

Bertelson, P., & Radeau, M. (1981). Crossmodal bias and perceptual fusion with auditory-visual spatial discordance. *Perception & Psychophysics*, 29, 578-584.

Bertenthal, B. I., Banton, T., & Bradbury, A. (1993). Directional bias in the perception of translating patterns, *Perception,* 22, 193-207.

Bertenthal, B. I., & Kramer, S. J. (1988). Perceptual constraints on the direction of moving surfaces. *Investigative Ophthalmology and Visual Science (Supplement),* 29, 250.

Blakemore, M. R., & Snowden, R. J. (1997). Attentional effects upon motion sensitivity. *Investigative Ophthalmology and Visual Science*, 38(4), S373.

Blauert, J. (1996). *Spatial Hearing: the Psychophysics of Human Sound Localization.* Cambridge, MA: MIT Press.

Bourke, P. A., Duncan, J., & Nimmo-Smith, I. (1996). A general factor involved in dual task performance decrement. *Quarterly Journal of Experimental Psychology*, 49, 525-545.

Bowne, S. F., McKee, S. P., & Glaser, D. A. (1989). Motion interference in speed discrimination. *Journal of the Optical Society of America*, 6, 1112-1121.

Braun, J. (1994). Visual search among items of different salience: removal of visual attention mimics a lesion in extrastriate area V4. *Journal of Neuroscience*, 14, 554-567.

Braun, J., & Sagi, D. (1990). Vision outside the focus of attention. *Perception & Psychophysics*, 48, 45-58.

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound.* Cambridge, MA: MIT Press.

Bregman, A. S., & Campbell, G. L. (1971). Primary auditory streaming segregation and perception of order in rapid sequences of tones. *Journal of Experimental Psychology*, 98, 244-249.

Broadbent, D. E. (1954). The role of auditory localization in attention and memory span. *Journal of Experimental Psychology*, 47, 191-196.

Broadbent, D. E. (1958). *Perception and Communication.* New York: Pergamon.

Broadbent, D. E. (1982). Task combination and the selective intake of information. *Acta Psychologica,* 50, 253-290.

Broadbent, D. E., & Broadbent, M. H. P. (1987). From detection to identification: response to multiple targets in rapid serial visual presentation. *Perception & Psychophysics*, 42, 105-113.

Brosch, M., & Schreiner, C. E. (1997). Time course of forward masking tuning curves in cat primary auditory cortex. *Journal of Neurophysiology*, 77, 923-943.

Brotchie, P. R., Andersen, R. A., Snyder, L. H., & Goodman, S. J. (1995). Head position signals used by parietal neurons to encode locations of visual-stimuli. *Nature*, 375, 232-235.

Brugge, J. F., Reale, R. A., & Hind, J. E. (1996). The structure of spatial receptive fields of neurons in primary auditory cortex of the cat. *Journal of Neuroscience,* 16, 4420-4437.

Bruner, J. S. (1957). On perceptual readiness. *Psychological Review*, 64, 123-152.

Büchel, C., Josephs, O., Rees, G., Turner, R., Frith, C. D., & Friston, K. J. (1998). The functional anatomy of attention to visual motion: A functional MRI study. *Brain*, 121, 1281-1294.

Buunen, T. J. F., & Valkenburg, D. A. (1979). Auditory detection of a single gap in noise. *Journal of the Acoustical Society of America*, 65, 534-537.

Calvert, G. A., Brammer, M. J., & Iversen, S. D. (1998). Crossmodal identification. *Trends in Cognitive Sciences*, 2, 247-253.

Calvert, G. A., Campbell, R., & Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Current Biology*, 10, 649-657.

Campbell, R., & Dodd, B. (1980). Hearing by eye. *Quarterly Journal of Experimental Psychology*, 32, 85-99.

Casagrande, V. A., Harting, J. K., Hall, W. C., & Diamond, I. T. (1972). Superior colliculus of the tree shrew: a structural and functional subdivision into superficial and deep layers. *Science*, 177, 444-447.

Casco, C., & Morgan, M. J. (1987). Detection of moving local density differences in dynamic random patterns by human observers. *Perception,* 16, 711-712.

Cavanagh, P. (1992) Attention-based motion perception. *Science*, 257, 1563-1565.

Chang, J. J., & Julesz, B. (1984). Cooperative phenomena in apparent movement perception of random-dot cinematograms. *Vision Research*, 24, 1781-1788.

Chaudhuri, A. (1990). Modulation of the motion aftereffect by selective attention. *Nature*, 344, 60-62.

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *Journal of the Acoustical Society of America*, 25, 975-979.

Choe, C. S., Welch, R. B., Guilford, R. M., & Juola, J. F. (1975). The 'ventriloquist effect': visual dominance or response bias? *Perception & Psychophysics*, 18, 55-60.

Ciocca, V., & Bregman, A. S. (1987). Perceived continuity of gliding and steady-state tones through interrupting noise. *Perception & Psychophysics*, 42, 476-484.

Clarey, J. C., & Irvine, D. R. F. (1986). Auditory response properties of neurons in the anterior ectosylvian sulcus of the cat. *Brain Research*, 386, 12-19.

Clemo, H. R., & Stein, B. E. (1982). Somatosensory cortex: a 'new' somatotopic representation. *Brain Research*, 235, 162-168.

Clohessy, A. B., Posner, M. I., Rothbart, M. K., & Vecera, S. P. (1991). The development of inhibition of return in early infancy. *Journal of Cognitive Neuroscience*, 3, 345-350.

Colby, C., & Duhamel, J. (1993). Ventral intraparietal area of the macaque: Anatomic location and visual response properties. *Journal of Neurophysiology*, 69, 902-914.

Connor, C. E., Gallant, J. L., Preddie, D. C., & Van Essen, D. C. (1996). Responses in area V4 depend on the spatial relationship between stimulus and attention. *Journal of Neurophysiology*, 75, 1306-1308.

Connor, C. E., Preddie, D. C., Gallant, J. L., & Van Essen, D. C. (1997). Spatial attention effects in macaque area V4. *Journal of Neuroscience*, 17, 3201-3214.

Cowan, N. (1984). On short and long auditory stores. *Psychological Bulletin*, 96, 341-370.

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological Bulletin*, 104, 163-191.

Creutzfeldt, O. D., Hellweg, F. C., & Schreiner, C. E. (1980). Thalamocortical transformation of responses to complex auditory stimuli. *Experimental Brain Research*, 39, 87-104.

Dean, A. F. (1981). The variability of discharge of simple cells in the cat striate cortex. *Experimental Brain Research,* 44, 437-440.

de Ribaupierre, F., Goldstein., M. H., & Yeni-Komshian, G. (1972). Cortical coding of repetitive acoustic pulses. *Brain Research*, 48, 205-225.

Dixon, N., & Spitz, L. (1980) The detection of audiovisual desynchrony. *Perception*, 9, 719-721.

Douglas, R. J., Koch, C., Mahowald, M., Martin, K. A. C., & Suarez, H. H. (1995). Recurrent excitation in neocortical circuits. *Science*, 269, 981-985.

Doupe, A. J. (1997). Song- and order-selective neurons in the songbird anterior forebrain and their emergence during vocal development. *Journal of Neuroscience*, 17, 1147-1167.

Doupe, A. J., & Konishi, M. (1991). Song-selectivity in auditory circuits in the vocal control system of the zebra finch. *Proceedings of the National Academy of Science USA*, 88, 11339-11343.

Dowling, W. J. (1968). Rhythmic fission and perceptual organization. *Journal of the Acoustical Society of America*, 44, 369-369.

Downer, J., Crawley, A. P., Mikulis, D. J., & Davis, K. D. (2000). A multimodal cortical network for the detection of changes in the sensory environment. *Nature Neuroscience*, 3, 277-283.

Driver, J. (1996). Enhancement of selective listening by illusory mislocation of speech sounds due to lip-reading. *Nature*, 381, 66-68.

Driver, J., & Spence, J. (1994). Spatial synergies between auditory- and visual-attention. In C. Umiltà, & M. Moscovitch (Eds.), *Attention and Performance XV* (pp. 311-331) Cambridge, MA: MIT Press.

Driver, J., & Spence, C. (1998). Attention and the cross modal construction of space. *Trends in Cognitive Sciences*, 2, 254-262.

Duhamel, J. R., Colby, C. L., & Goldberg, M. E. (1991). Congruent representations of visual and somatosensory space in single neurons of monkey ventral intraparietal cortex (area VIP). In J. Paillard (Ed.), *Brain and Space* (pp. 223-236). New York: Oxford University Press.

Duncan, J. (1993). Similarity between concurrent visual discriminations - dimensions and objects. *Perception & Psychophysics,* 54, 425-430.

Duncan, J., Ward, R., & Shapiro, K. (1994). Direct measurement of attentional dwell time in human vision. *Nature*, 369, 313-315.

Duncan, J., Martens, S., & Ward, R. (1997). Restricted attentional resource within but not between sensory modalities. *Nature*, 387, 808-810.

Easton, R. D., & Moran, P. W. (1978). A quantitative confirmation of visual capture of curvature. *Journal of General Psychology*, 98, 105-112.

Edwards, S. B., Ginsburgh, C. L., Henkel, C. K., & Stein, B. E. (1979). Sources of subcortical projections to the superior colliculus in the cat. *Journal of Comparative Neurology*, 184, 309-330.

Eggermont, J. J. (1991). Rate and synchronization measures for periodicity coding in primary auditory cortex. *Hearing Research*, 56, 153-167.

Fastl, H. (1976). Temporal masking effects: I. Critical band noise masker. *Acustica*, 36, 317-331.

Finney, D. J. (1971). *Probit analysis*. Cambridge, UK: Cambridge University Press.

Fisher, B. D., & Pylyshyn, Z. W. (1994). The cognitive architecture of bimodal event perception: a commentary and addendum to Radeau (1994). *Cahiers de Psychologie Cognitive*, 13, 92-96.

Fogassi, L., Gallese, V., di Pellegrino, G., Fadiga, L., Gentilucci, M., Luppino, G., Matelli, M., Pedotti, A., & Rizzolatti, G. (1992). Space coding by premotor cortex. *Experimental Brain Research*, 89, 686-690.

Forrest, T. G., & Green, D. M. (1987). Detection of partially filled gaps in noise and the temporal modulation transfer function. *Journal of the Acoustical Society of America*, 82, 1933-1943.

Fraisse, P. (1963). *The psychology of time* (J. Leith, Trans.) New York: Harper & Row.

Fraisse, P. (1982). Rhythm and tempo. In D. Deutsch (Ed.), *The psychology of music* (pp. 149-180), New York, Academic Press.

Fredericksen, R. E., Verstraten, F. A. J., & van de Grind, W. A. (1993) Spatiotemporal characteristics of human motion perception. *Vision Research,* 33, 1193-1205.

Freeman, W. T. (1994). The generic viewpoint assumption in a framework for visual perception. *Nature*, 368, 542-545.

Gandhi, S. P., Heeger, D. J., & Boynton, G. M. (1999). Spatial attention affects brain activity in human primary visual cortex. *Proceedings of the National Academy of Sciences USA*, 96, 3314-3319.

Gaver, W. W. (1993a). What in the world do we hear? An ecological approach to auditory event perception. *Ecological Psychology*, 5, 1-29.

Gaver, W. W. (1993b). How in the world do we hear? Exploration in ecological acoustics. *Ecological Psychology*, 5, 285-313.

Geisler, W. S., & Albrecht, D. G. (1997). Visual cortex neurons in monkeys and cats: detection, discrimination and identification. *Visual Neuroscience*, 14, 897-919.

Geissler, L. R. (1915). Sound localization under determined expectation. *American Journal of Psychology*, 26, 268-285.

Gibson, B. S., & Egeth, H. (1994). Inhibition of return to object-based and environment-based locations. *Perception & Psychophysics*, 55, 323-339.

Gibson, J. J. (1933). Adaptation, after-effect and contrast in the perception of curved lines. *Journal of Experimental Psychology*, 16, 1-31.

Gibson, J. J. (1950). *The perception of the visual world.* Boston: Houghton Mifflin.

Gogel, W. C., & MacCracken, P. J. (1979). Depth adjacency and induced motion. *Perceptual and Motor Skills*, 48, 343-350.

Gogel, W. C., & Sharkey, T. J. (1989). Measuring attention using induced motion. *Perception*, 18, 300-320.

Gogel, W. C., & Tietz, J. D. (1976). Adjacency and attention as determiners of perceived motion. *Vision Research*, 16, 839-845.

Goldberg, D. M., & Pomerantz, J. R. (1982). Models of illusory pausing and sticking. *Journal of Experimental Psychology: Human Perception and Performance*, 8, 547-561.

Goldstone, S. (1968). Reaction time to onset and termination of light and sounds. *Perceptual and Motor Skills*, 27, 1023-1029.

Goodale, M. A., & Murison, R. C. C. (1975). The effects of lesions of the superior colliculus on locomotor orientation and the orienting reflex in the rat. *Brain Research*, 88, 243-261.

Graziano, M., & Gross, C. (1993). A bimodal map of space - somatosensory receptive fields in the macaque putamen with corresponding visual receptive fields. *Experimental Brain Research*, 97, 96-109.

Graziano, M. S. A., & Gross, C. G. (1996) Multiple pathways for processing visual space. In T. Inui, & J. McClelland (Eds.), *Attention and Performance (Vol. 26): Information Integration in Perception and Communication* (pp. 181-207). Cambridge, MA: MIT Press.

Graziano, M. S. A., & Gross, C. G. (1998). Spatial maps for the control of movement. *Current Opinion in Neurobiology*, 8, 195-201.

Graziano, M. S. A., Yap, G., & Gross, C. (1994). Coding of visual space by premotor neurons. *Science*, 266, 1054-1057.

Green, K. P., & Gerdeman, A. (1995). Cross-modal discrepancy in coarticulation and the integration of speech information: The McGurk effect with mismatched vowels. *Journal of Experimental Psychology: Human Perception and Performance*, 21, 1409-1426.

Green, K. P., & Kuhl, K. P. (1989). The role of visual information in the processing of place and manner features in speech perception. *Perception & Psychophysics*, 45, 34-41.

Green, K. P., Kuhl, P. K., Meltzoff, A. N., & Stevens, E. B. (1991). Integrating speech information across talkers, gender, and sensory modality: Female faces and male voice in the McGurk effect. *Perception & Psychophysics*, 50, 524-536.

Gregory, R. L. (1997). Knowledge in perception and illusion. *Philosophical Transactions of the Royal Society of London B Biological Sciences*, 352, 1121-1127.

Grzywacz, N. M., Watamaniuk, S. N. J., & McKee, S. P. (1995). Temporal coherence theory for the detection and measurement visual motion. *Vision Research*, 35, 3183-3203.

Hannes, M., Sutton, S., & Zubin, J. (1968). Reaction time: Stimulus uncertainty with response certainty. *Journal of General Psychology*, 78, 165-181.

Hay, J. C., Pick, H. L. Jr., & Ikeda, K. (1965). Visual capture produced by prism spectacles. *Psychonomic Science*, 2, 215-216.

He, S., Cavanagh, P., & Intriligator, J. (1996). Attentional resolution and the locus of visual awareness. *Nature*, 383, 334-337.

Heffner, H. (1978). Effect of auditory cortex ablation on localization and discrimination of brief sounds. *Journal of Neurophysiology*, 41, 963-976.

Heffner, H., & Heffner, R. S. (1990). Effect of bilateral auditory cortex lesions on sound localization in Japanese macaques. *Journal of Neurophysiology*, 64, 915-931.

Heffner, H., & Masterton, B. (1975). Contribution of auditory cortex to sound localization in the monkey (Macaca mulatta). *Journal of Neurophysiology*, 38, 1340-1358.

Held, R. (1955). Shifts in binaural localization after prolonged exposures to atypical combinations of stimuli. *American Journal of Psychology*, 68, 526-548.

Henning, G. B., & Geskell, H. (1981). Monaural phase sensitivity with Ronken's paradigm. *Journal of the Acoustical Society of America*, 70, 1669-1673.

Hikosaka, O., Miyauchi, S., & Shimojo, S. (1993a). Focal visual attention produces illusory temporal order and motion sensation. *Vision Research*, 33, 1219-1240.

Hikosaka, O., Miyauchi, S., & Shimojo, S. (1993b). Visual-attention revealed by an illusion of motion. *Neuroscience Research*, 18, 11-18.

Hikosaka, O., Miyauchi, S., & Shimojo, S. (1996). Orienting spatial attention - its reflective, compensatory, and voluntary mechanisms. *Cognitive Brain Research*, 5, 1-9.

Hillstrom, A. P., & Yantis, S. (1994). Visual motion and attentional capture. *Perception & Psychophysics*, 55, 399-411.

Hirsh, I. J., & Sherrick, C. E. (1961). Perceived order in different sense modalities. *Journal of Experimental Psychology*, 62, 423-432.

Hocherman, S., & Gilat, E. (1981). Dependence of auditory cortex evoked unit activity on interstimulus interval in the cat. *Journal of Neurophysiology*, 45, 987-997.

Holub, R. A., & Morton-Gibson, M. (1981). Response of visual cortical neurons of the cat to moving sinusoidal gratings: Response-contrast functions and spatiotemporal interactions. *Journal of Neurophysiology*, 46, 1244-1259.

Hood, B. (1993). Inhibition of return produced by covert shifts of visual attention in 6-month old infants. *Infant Behavior and Development*, 16, 245-254.

Howard, I. P., & Templeton, W. B. (1966). *Human Spatial Orientation.* London: Wiley.

Huerta, M. F., & Harting, J. K. (1984). The mammalian superior colliculus: studies of its morphology and connections. In H. Vanegas (Ed.), *Comparative neurology of the optic tectum* (pp. 687-773). New York: Plenum Press.

Hylkema, B. S. (1942). Examination of the visual field by determining the fusion frequency. *Acta Ophthalmologica*, 20, 181-193.

Imig, T. J., Irons, W. A., & Samson, F. R. (1990). Single-unit selectivity to azimuthal direction and sound pressure level of noise bursts in cat high-frequency primary auditory cortex. *Journal of Neurophysiology*, 63, 1448-1466.

Intriligator, J., & Cavanagh, P. (1997). Individuation and the resolution of attention. *Investigative Ophthalmology and Visual Science*, 38(4), S688.

Irwin, R. J., & Zwislocki, J. J. (1971). Loudness effects in pairs of tone bursts. *Perception & Psychophysics*, 10, 198-192.

Isaak, M. I., & Fawcett, S. L. (1997). Attention facilitates sequential recruitment in the perception of motion. *Perception(Supplement)*, 26, 96.

Jack, C. E., & Thurlow, W. R. (1973). Effects of degree of visual association and angle of displacement on the 'ventriloquism' effect. *Perceptual and Motor Skills,* 37, 967-979.

Jackson, C. W. (1953). Visual factors in auditory localization. *Quarterly Journal of Experimental Psychology*, 5, 52-65.

James, W. (1890). *The principle of psychology*. New York: H. Holt.

Jenkins, W. M., & Merzenich, M. M. (1984). Role of cat primary auditory cortex for sound-localization behavior. *Journal of Neurophysiology*, 52, 819-847.

Johnson, M. K., & Tucker, L. A. (1996). The development and temporal dynamics of spatial orienting in infants. *Journal of Experimental Child Psychology*, 63, 171-188.

Jones, J. A., & Munhall, K. G. (1997). The effects of separating auditory and visual sources on audio-visual integration of speech. *Canadian Acoustics*, 2, 13-19.

Jonides, J., & Yantis, S. (1988). Uniqueness of abrupt visual onset in capturing attention. *Perception & Psychophysics*, 43, 346-354.

Kahneman, D., & Treisman, A. (1984). Changing views of attention and automaticity. In R. Parasuraman, & R. Davies (Eds.), *Varieties of Attention* (pp. 29-61). New York: Academic Press.

Kallman, H. J., & Morris, M. D. (1984). Duration perception and auditory masking. *Annals of the New York Academy of Sciences*, 423, 603-609.

Kavanagh, G. L., & Kelly, J. B. (1987). Contribution of auditory cortex to sound localization by the ferret (Mustela putorius). *Journal of Neurophysiology*, 57, 1746-1766.

Kawashima, R., O'Sullivan, B. T., & Roland, P. E. (1995). Positron-emission tomography studies of cross-modality inhibition in selective attentional tasks: closing the 'mind's eye'. *Proceedings of the National Academy of Science USA*, 92, 5969-5972.

Kelly, D. H. (1971a). Theory of flicker and transient responses. I. Uniform fields. *Journal of the Optical Society of America*, 61, 537-546.

Kelly, D. H. (1971b). Theory of flicker and transient responses. II. Counterphase gratings. *Journal of the Optical Society of America*, 61, 632-640.

Kelly, D. H. (1972). Adaptation effects on spatio-temporal sine-wave thresholds. *Vision Research*, 12, 89-101.

King, A. J., & Carlile, S. (1995). Neural coding for auditory space. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 279-293). Cambridge, MA: MIT Press.

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: Cambridge University Press.

Koch, C., & Poggio, T. (1992). Multiplying with Synapses and Neurons. In T. McKenna, J. L. Davis, & S. F. Zornetzer (Eds.), *Single Neuron Computation* (pp. 315-343). Boston: Academic Press.

Kristofferson, M. W. (1967). Shifting attention between modalities: A comparison of schizophrenics and normals. *Journal of Abnormal Psychology*, 72, 388-394.

LaBerge, D. (1973). Identification of two components of the time to switch attention: A test of a serial and a parallel model of attention. In S. Kornblum (Ed.), *Attention and Performance IV* (pp. 71-85). New York: Academic Press.

Langner, G. (1992). Periodicity coding in the auditory system. *Hearing Research*, 60, 115-142.

Lappin, J. S., & Bell, H. H. (1976). The detection of coherence in moving random-dot patterns. *Vision Research*, 16, 161-168.

Lavie, N., & Tsal, Y. (1994). Perceptual load as a major determinant of the locus of selection in visual-attention. *Perception & Psychophysics*, 56, 183-197.

Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology: Human Perception and Performance*, 21, 451-468.

Lewicki, M. S. (1996). Intracellular characterization of song-specific neurons in the zebra finch auditory forebrain. *Journal of Neuroscience*, 16, 5854-5863.

Lewicki, M. S., & Konishi, M. (1995). Mechanisms underlying the sensitivity of songbird forebrain neurons to temporal order. *Proceeding of the National Academy of Science USA*, 92, 5582-5586.

Logan, G. D. (1978). Attention in character-classification tasks: evidence for automaticity of component stage. *Journal of Experimental Psychology: General*, 107, 32-63.

Loveless, N. E., & Hari, R. (1993). Auditory evoked fields covary with perceptual grouping. *Biological Psychology*, 35, 1-15.

Lu, Z. L., & Sperling, G. (1995). Attention-generated apparent motion. *Nature*, 377, 237-239.

Lu, Z. L., & Sperling, G. (1995b) The functional architecture of human visual-motion perception. *Vision Research*, 35, 2697-2722.

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.

MacDonald, J., & McGurk, H. (1978). Visual influences on speech perception. *Perception & Psychophysics*, 24, 253-257.

Margoliash, D. (1983). Acoustic parameters underlying the responses of song specific neurons in the white-crowned sparrow. *Journal of Neuroscience*, 3, 1039-1057.

Margoliash, D., & Fortune, E. S. (1992). Temporal and harmonic combination-sensitive neurons in the zebra finch. *Journal of Neuroscience*, 12, 4309-4326.

Marr, D. (1982). *Vision: A computational Investigation into the Human Representation and Processing of Visual Information.* New York: W. H. Freeman and Company.

Massaro, D. W. (1970). Retroactive interference in short-term recognition memory for pitch. *Journal of Experimental Psychology*, 83, 32-39.

Massaro, D. W. (1975). *Experimental Psychology and Information Processing.* Chicago: Rand McNally.

Massaro, D. W. (1999). Speechreading: illusion or window into pattern recognition. *Trends in Cognitive Sciences*, 3, 310-317.

Massaro, D. W., & Cohen, M. M. (1990). Perception of synthesized audible and visible speech. *Psychological Science*, 1, 55-63.

Massaro, D. W., & Cohen, M. M. (1993). Perceiving asynchronous bimodal speech in consonant-vowel and vowel syllables. *Speech Communication*, 13, 127-134.

Massaro, D. W., Cohen, M .M., & Idson, W. L. (1976). Recognition masking of auditory lateralization and pitch judgments. *Journal of the Acoustical Society of America*, 59, 434-441.

Massaro, D. W., Cohen, M. M., & Smeele, P. M. (1996). Perception of asynchronous and conflicting visual and auditory speech. *Journal of the Acoustical Society of America*, 100, 1777-1786.

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.

McAdams, S., & Bregman, A. S. (1979). Hearing musical streams. *Computer Music Journal*, 3, 26-43.

McAdams, C. J., & Maunsell, J. H. R. (1999). Effects of attention on orientation-tuning functions of single neurons in Macaque Cortical Area V4. *Journal of Neuroscience*, 19, 431-441.

McCabe, S. L., & Denham, M. J. (1997). A model of auditory streaming. *Journal of the Acoustical Society of America*, 101, 1661-1721.

McGrath, M., & Summerfield, Q. (1985). Intermodal timing relations and audiovisual speech recognition by normal-hearing adults. *Journal of the Acoustical Society of America*, 77, 678-685.

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

McKee, S. P., & Welch, L. (1985). Sequential recruitment in the discrimination of velocity. *Journal of the Optical Society of America*, 2, 243-251.

Mel, B. W. (1992). NMDA-based pattern-discrimination in a modeled cortical neuron. *Neural Computation*, 4, 502-517.

Mel, B. W. (1993). Synaptic integration in an excitable dendritic tree. *Journal of Neurophysiology*, 70, 1086-1101.

Mel, B. W., & Koch, C. (1990). Sigma-pi learning: on radial basis functions and cortical associative learning. In D. S. Touretzky (Ed.), *Advances in Neural Information Processing Systems Vol. 2* (pp. 474-481). San Mateo, CA: Morgan Kaufmann.

Meredith, M. A., Nemitz, J. W., & Stein, B. E. (1987). Determinants of multisensory integration in superior colliculus neurons: I. Temporal factors. *Journal of Neuroscience*, 10, 3215-3229.

Mesulam, M-M. (1994). Neurocognitive networks and selectively distributed processing. *Revue Neuologique (Paris)*, 150, 564-569.

Mesulam. M-M., Hersh, L. B., Mash, D. C., & Geyla, C. (1992). Differential cholinergic innervation within functional subdivisions of the human cerebral cortex: a choline acetyltransferase study. *Journal of Comparative Neurology*, 318, 316-328.

Metzger, W. (1934). Beobachtungen über phänomenale Identität. *Psychologische Forschung*, 19, 1-49.

Michotte, A. (1963). *The perception of causality.* (T. R. Miles & E. Miles, Trans.). New York: Basic Books. (Original work published 1946).

Middlebrooks, J. C., Clock, A. E., Xu, L., & Green, D. M. (1994). A panoramic code for sound localization by cortical neurons. *Science*, 264, 842-844.

Middlebrooks, J. C., & Green, D. M. (1991). Sound localization by human listeners. *Annual Review of Psychology*, 42, 135-159.

Middlebrooks, J. C., & Pettigrew, J. D. (1981). Functional classes of neurons in primary auditory cortex of cat distinguished by sensitivity to sound location. *Journal of Neuroscience*, 1, 107-120.

Mucke, L., Norita, M., Benedek, G., & Creutzfeldt, O. (1982). Physiologic and anatomic investigation of a visual cortical area situated in the ventral bank of the anterior ectosylvian sulcus of the cat. *Experimental Brain Research*, 46, 1-11.

Munhall, K. G., Gribble, P., Sacco, L., & Ward, M. (1996). Temporal constraints on the McGurk effect. *Perception & Psychophysics*, 58, 351-362.

Näätänen, R. (1990) The role of attention in auditory information-processing as revealed by event-related potentials and other brain measures of cognitive function. *Behavioral and Brain Sciences*, 13, 201-288.

Näätänen, R. (1992). *Attention and Brain Function.* Hillsdale, NJ: Erlbaum.

Näätänen, R., Paavilainenm P., Tiitinen, H., Jiang, D., & Alho, K. (1993a). Attention and mismatch negativity. *Psychophysiology*, 30, 436-450.

Näätänen, R., Schröger, E., Karakas, S., Tervaniemi, M., & Paavilainen, M. (1993b). Development of a memory trace for a complex sound in the human brain. *NeuroReport*, 4, 503-506.

Nakayama, K., & Shimojo, S. (1992). Experiencing and perceiving visual surfaces. *Science*, 257, 1357-1363.

Nakayama, K., & Silverman, G. H. (1984). Temporal and spatial characteristics of the upper displacement limit for motion in random dots. *Vision Research*, 24, 293-299.

Nawrot, M., & Sekuler, R. (1990). Assimilation and contrast in motion perception: Explorations in cooperativity. *Vision Research*, 30, 1439-1451.

Neff, W. D., Fisher, J. F., Diamond, I. T., & Yela, M. (1956). Role of auditory cortex in discrimination requiring localization of sound in space. *Journal of Neurophysiology*, 19, 500-512.

Nordby, H., Roth, W. T., & Pfefferbaum, A. (1988). Event-related potentials to breaks in sequences of alternating pitches or interstimulus intervals. *Psychophysiology*, 25, 262-268.

O'Craven, K. M., Rosen, B. R., Kwong, K. K., Treisman, A., & Savoy, R. L. (1997). Voluntary attention modulates fMRI activity in human MT-MST. *Neuron*, 18, 591-598.

Olson, C. R., & Graybiel, A. M. (1987). Ectosylvian visual area of the cat: location, retinotopic organization, and connections. *Journal of Comparative Neurology*, 261, 277-294.

Pashler, H. (1994). Dual-task interference in simple tasks: data and theory. *Psychological Bulletin*, 116, 220-244.

Phillips, D. P., Hall, S. E., & Hollett, J. L. (1989). Repetition rate and signal level effects on neuronal responses to brief tone pulses in cat auditory cortex. *Journal of the Acoustical Society of America*, 85, 2537-2549.

Phillips, G. (1933). Perception of flicker in lesions of the visual pathways. *Brain*, 56, 464-478.

Pick, H. L., Jr., Warren, D. H., & Hay, J. C. (1969). Sensory conflict in judgments of spatial direction. *Perception & Psychophysics*, 6, 203-205.

Plenge, G. (1914). On the differences between localization and lateralization. *Journal of the Acoustic Society of America*, 56, 944-951.

Plomp, R. (1964). Rate of decay of auditory sensation. *Journal of the Acoustical Society of America*, 36, 277-282.

Poggio, T. (1990). A theory of how the brain might work. *Cold Spring harbor Symposium on Quantitative Biology*, 55, 899-910.

Posner, M. I., Nissen, M. J., & Klein, R. M. (1976). Visual dominance: An information-processing account of its origins and significance. *Psychological Review*, 83, 157-171.

Pouget, A., & Sejnowski, T. J. (1997). Spatial transformations in the parietal cortex using basis functions. *Journal of Cognitive Neuroscience*, 9, 222-237.

Pylyshyn, Z. W. (1984). *Computation and cognition: Toward a foundation for cognitive science.* Cambridge, MA: MIT Press.

Pylyshyn, Z. W. (1989). The role of location indexes in spatial perception: A sketch of the FINST spatial-indexing model. *Cognition*, 32, 65-97.

Pylyshyn, Z, W. (1994). Some primitive mechanisms of spatial attention. *Cognition*, 50, 363-384.

Pylyshyn, Z. W. (1999). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences*, 22, 341-423.

Quinlan, P. T., & Hill, N. I. (1999). Sequential effects in rudimentary auditory and visual tasks. *Perception & Psychophysics*, 61, 375-384.

Radeau, M. (1994). Auditory-visual spatial interaction and modularity. *Cahiers de Psychologie Cognitive*, 13, 3-15.

Radeau, M., & Bertelson, P. (1977). Adaptation to auditory-visual discordance and ventriloquism in semirealistic situations. *Perception & Psychophysics*, 22, 137-146.

Radeau, M., & Bertelson, P. (1987). Auditory-visual interaction and the timing of inputs: Thomas (1941) revisited. *Psychological Research*, 49, 17-22.

Rajan, R., Aitkin, L. M., Irvine, D. R. F., & McKay, J. (1990). Azimuthal sensitivity of neurons in primary auditory cortex of cats. I. Types of sensitivity and the effects of variations in stimulus parameters. *Journal of Neurophysiology*, 64, 872-887.

Ramachandran, V. S., & Anstis, S. M. (1983). Perceptual organization in moving patterns. *Nature*, 304, 529-531.

Raymond, J. E. (2000). Attentional modulation of visual motion perception. *Trends in Cognitive Sciences*, 4, 42-50.

Raymond, J. E., O'Donnell, H. L., & Tipper, S. (1997). Attention modulates global motion sensitivity. *Investigative Ophthalmology and Visual Science,* 38(4), S369.

Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporally suppression of visual processing in an RSVP task: an attentional blink? *Journal of Experimental Psychology: Human perception and performance,* 18, 849-860.

Rees, G., Frith, C. D., & Lavie, N. (1999). Modulating irrelevant motion perception by varying attentional load in an unrelated task. *Science*, 278, 1616-1619.

Regan, D. (1989). *Human brain electrophysiology: Evoked potentials and evoked magnetic fields in science and medicine.* New York: Elsevier.

Reisberg, D. (1978). Looking where you listen: visual cues and auditory attention. *Acta Psychologica*, 42, 331-341.

Richards, W. (1988). The approach. In W. Richards (Ed.), *Natural computation* (pp. 3-17). Cambridge, MA: MIT Press.

Robinson, E. S. (1934). Work of the integrated organism. In C. Murchison (Ed.), *Handbook of General Experimental Psychology* (pp. 571-650). New York: Rusell & Rusell.

Rock, I. (1966). *The nature of perceptual adaptation.* New York: Basic Books.

Rogers, W. L., & Bregman, A. S. (1993). An experimental evaluation of three theories of auditory stream segregation. *Perception & Psychophysics*, 53, 179-189.

Ronken, D. (1970). Monaural detection of a phase difference between clicks. *Journal of the Acoustical Society of America*, 70, 1091-1099.

Rose, M. R., & Moore, B. C. J. (1997). Perceptual grouping of tone sequences by normally hearing and hearing-impaired listeners. *Journal of the Acoustical Society of America*, 102, 1768-1778.

Saarinen, J., Paavilainen, P., Schröger, E., Tervaniemi, M., & Näätänen, R. (1992). Representation of abstract attributes of auditor-stimuli in the human brain. *NeuroReport*, 3, 1149-1151.

Sagi, D., & Julesz, B. (1986). Enhanced detection in the aperture of focal attention during simple discrimination tasks. *Nature*, 321, 693-695.

Saldaña, H. M., & Rosenblum, L. D. (1993). Visual influences on auditory pluck and bow judgments. *Perception & Psychophysics*, 54, 406-416.

Salinas, E., & Abbott, L. F. (1996). A model of multiplocative neural responses in parietal cortex. *Proceedings of the National Academy of Science USA*, 93, 11956-11961.

Scheier, C., Lewkowicz, D., & Shimojo, S. (submitted). Sound induces perceptual reorganization of an ambiguous motion display in human infants. *Psychological Science.*

Schneider, G. E. (1969). Two visual systems: brain mechanisms for localization and discrimination are dissociated by tectal and cortical lesions. *Science*, 163, 895-902.

Schröger, E. (1994) An event-related potential study of sensory representations of unfamiliar tonal patterns. *Psychophysiology*, 31, 175-181.

Schröger, E., Näätänen, R., & Paavilainen, P. (1992). Event-related potentials reveal how non-attended complex sound patterns are represented by the human brain. *Neuroscience Letters*, 146, 183-186.

Schröger E., Tervaniemi, M., & Näätänen, R. (1995). Time course of loudness in tone patterns is automatically represented by the human brain. *Neuroscience Letters*, 202, 117-120.

Schröger, E., Tervaniemi, M., Wolff, C., & Näätänen, R. (1996). Preattentive periodicity detection in auditory patterns as governed by time and intensity information. *Cognitive Brain Research*, 4, 145-148.

Sclar, G., & Freeman. R. D. (1982). Orientation selectivity in the cat's striate cortex is invariant with stimulus contrast. *Experimental Brain Research*, 46, 457- 461.

Seidemann, E., & Newsome, W. T. (1999). Effect of spatial attention on the responses of area MT neurons. *Journal of Neurophysiology*, 81, 1783-1794.

Sekiyama, K., & Tohkura, Y. (1991) McGurk effect in non-English listeners: Few visual effects for Japanese subjects hearing Japanese syllables of high auditory intelligibility. *Journal of the Acoustical Society of America*, 90, 1797-1805.

Sekiyama, K. (1997). Cultural and linguistic factors in audiovisual speech processing: The McGurk effect in Chinese subjects. *Perception & Psychophysics*, 59, 73-80.

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., & Brackett, T. (1995). When visual objects collide: repulsion and streaming. *Investigative Ophthalmology and Visual Science*, 36(4), S50.

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

Shelton, B. R., & Searle, C. L. (1980). The influence of vision on the absolute identification of sound-source position. *Perception & Psychophysics*, 28, 589-596.

Shipley, T. (1964). Auditory flutter-driving of visual flicker. *Science*, 145, 1328-1330.

Skottun, B. C., Ohzawa, I., Sclar, G., & Freeman, R. D. (1987). The effects of contrast on visual orientation and spatial frequency discrimination: a comparison of single cells and behavior. *Journal of Neurophysiology*, 57, 773-786.

Skrandies, W. (1985). Critical flicker fusion and double flash discrimination in different parts of the visual field. *International Journal of Neuroscience*, 25, 225-231.

Snippe, H. P., & Koenderink, J. J. (1994). Extraction of optical velocity by use of multi-input Reichardt detectors. *Journal of the Optical Society of America*, 11, 1222-1236.

Snowden, R. J. (1989). Motions in orthogonal directions are mutually suppressive. *Journal of the Optical Society of America*, 6, 1096-1101.

Snowden, R. J., & Braddick, O. J. (1989a). The combination of motion signals over time. *Vision Research*, 29, 1621-1630.

Snowden, R. J., & Braddick, O. J. (1989b). Extension of displacement limits in multiple-exposure sequences of apparent motion. *Vision Research*, 29, 1777-1787.

Snowden, R. J., & Braddick, O. J. (1991). The temporal integration and resolution of velocity signals. *Vision Research*, 31, 907-914.

Somers, D. C., Nelson, S. B., & Sur, M. (1995). Emergent model of orientation selectivity in cat visual cortical simple cells. *Journal of Neuroscience*, 15, 5448-5465.

Soroker, N., Calamaro, N., & Myslobodsky, M. S. (1995). Ventriloquist effect reinstates responsiveness to auditory stimuli in the 'ignored' space in patients with hemispatial neglect. *Journal of Clinical and Experimental Neuropsychology*, 17, 243-255.

Sparks, D. L. (1986). Translation of sensory signals into commands for control of saccadic eye movements: role of primate superior colliculus. *Physiological Reviews*, 66, 116-177.

Spence, C., & Driver, J. (1994). Covert spatial orienting in audition: Exogenous and endogenous mechanisms. *Journal of Experimental Psychology: Human Perception and Performance*, 20, 555-574.

Spence, C., & Driver, J. (1996). Audiovisual links in endogenous covert spatial attention. *Journal of Experimental Psychology: Human Perception and Performance*, 22, 1005-1030.

Spence, C., & Driver, J. (1997a). Audiovisual links in endogenous covert spatial orientation. *Perception & Psychophysics*, 59, 1-22.

Spence, C., & Driver, J. (1997b). On measuring selective attention to an expected sensory modality. *Perception & Psychophysics*, 59, 389-403.

Spence, C., Nicholls, M. E. R., Gillespie, N., & Driver, J. (1998). Crossmodal links in exogenous covert spatial orienting between touch, audition, and vision. *Perception & Psychophysics*, 60, 544-557.

Sprague, J. M., & Meikle, T. H. Jr. (1965). The role of the superior colliculus in visually guided behavior. *Experimental Neurology*, 11, 115-146.

Stein, B. E. (1984). Development of the superior colliculus. *Annual Review of Neuroscience*, 7, 95-125.

Stein, B. E., Huneycutt, W. S., & Meredith, M. A. (1988). Neurons and behaviour: the same rules of multisensory integration apply. *Brain Research*, 488, 355-358.

Stein, B. E., London, N., Wilkinson, L. K., & Price, D. D. (1996). Enhancement of perceived visual intensity by auditory stimuli: A psychophysical analysis. *Journal of Cognitive Neuroscience*, 8, 497-506.

Stein, B. E., & Meredith, M. A. (1991). Functional organization of the superior colliculus. In A. G. Leventhal (Ed.), *The neural basis of visual function* (pp. 85-110). Hampshire: Macmillan.

Stein, B. E., & Meredith, M. A. (1993). *The Merging of the Senses.* Cambridge, MA: MIT Press.

Stein, B. E., Meredith, M. A., Huneycutt, W. S., & McDade, L. (1989). Behavioural indices of multisensory integration: orientation to visual cues is affected by auditory stimuli. *Journal of Cognitive Neuroscience*, 1, 12-24.

Stein, B. E., Meredith, M. A., & Wallace, M. T. (1993). The visually responsive neuron and beyond: multisensory integration in cat and monkey. *Progress in Brain Research*, 95, 79-90.

Stein, B. E., & Wallace, M. T. (1996). Comparisons of crossmodality integration in midbrain and cortex. In M. Norita, T. Bando, & B. Stein (Eds.), *Progress in Brain Research Vol. 112* (pp. 289-299). Amsterdam: Elsevier Science.

Stemmler, M., Usher, M., & Niebur, E. (1995). Lateral interactions in primary visual cortex: A model bridging physiology and psychophysics. *Science*, 269, 1877-1880.

Stevens, S. S., & Davis, H. (1938). *Hearing. Its Psychology and Physiology.* New York: Wiley.

Stevens, S. S., & Newman, E. B. (1936). The localization of actual sources of sound. *American Journal of Psychology*, 48, 297-306.

Streicher, M., & Ettlinger, G. (1987). Crossmodal recognition of familiar and unfamiliar objects by the monkey: the effects of ablation of polysensory neocortex or of the amygdaloid complex. *Behavioral Brain Research*, 23, 95-107.

Suga, N. (1989). Principles of auditory information-processing derived from neuroethology. *Journal of Experimental Biology*, 146, 277-286.

Suga, N., O'Neill, W. E., & Manabe, T. (1978). Cortical neurons sensitive to combinations of information-bearing elements of biosonar signals in the mustache bat. *Science*, 200, 778-781.

Sumby, W. H., & Polack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America*, 26, 212-215.

Sumi, S. (1995). Bounce effect seen in opposite motions. *Japanese Psychological Research*, 37, 195-200.

Summerfield, A. Q. (1979). Use of visual information in phonetics perception. *Phonetica*, 36, 314-331.

Summerfield, A. Q. (1987). Some preliminaries to a comprehensive account of audio-visual speech perception. In B. Dodd & R. Campbell (Eds.), *Hearing by Eye: The Psychology of Lip-Reading* (pp. 3-51). London: Erlbaum.

Tervaniemi, M., Maury, S., & Näätänen, R. (1994). Neural representations of abstract stimulus features in the human brain as reflected by the mismatch negativity. *NeuroReport*, 5, 844-846.

Thomas, G. J. (1941). Experimental study of the influence of vision on sound localisation. *Journal of Experimental Psychology*, 28, 167-177.

Thurlow, W. R., & Rosenthal, T. M. (1976). Further study of existence regions for the "ventriloquism effect". *Journal of the American Audiology Society*, 1, 280-286.

Tougas, Y., & Bregman, A. S. (1985). The crossing of auditory streams. *Journal of Experimental Psychology: Human Perception and Performance*, 11, 788-789.

Tougas, Y., & Bregman, A. S. (1990). Auditory streaming and the continuity illusion. *Perception & Psychophysics*, 47, 121-126.

Treisman, A. M. (1969). Strategies and models of selective attention. *Psychological Review*, 76, 282-299.

Treue, S., & Husain, M. (1997). Temporal dynamics of sensory and attentional processing in motion perception. *Investigative Ophthalmology and Visual Science*, 38(4), S1151.

Treue, S., & Maunsell, J. H. R. (1996). Attentional modulation of visual-motion processing in cortical areas MT. *Nature*, 382, 539-541.

Treue, S., & Trujillo, J. C. M. (1999). Feature-based attention influences motion processing gain in macaque visual cortex. *Nature*, 399, 575-579.

Tsal, Y. (1983). Movement of attention across the visual field. *Journal of Experimental Psychology: Human Perception and Performance,* 9, 523-530.

Tyler, C. W. (1987). Analysis of visual modulation sensitivity III. Meridional variations in peripheral flicker sensitivity. *Journal of the Optical Society of America*, 4, 1612-1619.

Valdes-Sosa, M., Bobes, M. A., Rodriguez, V., & Pinilla, T. (1998). Switching attention without shifting the spotlight: Object-based attentional modulation of brain potentials. *Journal of Cognitive Neuroscience*, 10, 137-151.

van de Grind, W. A., van Doorn, A. J., & Koenderink, J. J. (1983). Detection of coherent movement in peripherally viewed random-dot patterns. *Journal of the Optical Society of America*, 73, 1674-1683.

van Opstal, A. J., Hepp, K., Suzuki, Y., & Henn, V. (1995). Influence of eye position on activity in monkey superior colliculus. *Journal of Neurophysiology*, 74, 1593-1610.

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.

Vroomen, J., & de Gelder, B. (2000). Sound enhances visual perception: Crossmodal effects of auditory organization on vision. *Journal of Experimental Psychology: Human Perception and Performance*, 26, 1583-1590.

Wallace, M. T., Meredith, M. A., & Stein, B. E. (1993). Converging influences from visual, auditory and somatosensory cortices onto output neurons of the superior colliculus. *Journal of Neurophysiology*, 69, 1797-1809.

Wallace, M. T., & Stein, B. E. (1994). Cross-modal synthesis in the midbrain depends on input from association cortex. *Journal of Neurophysiology*, 71, 429-432.

Wang, D. (1996). Primitive auditory segregation based on oscillatory correlation. *Cognitive Science*, 20, 409-456.

Warren, D. H., Welch, R. B., & McCarthy, T. J. (1981). The role of visual-auditory 'compellingness' in the ventriloquism effect: Implications for transitivity among the spatial senses. *Perception & Psychophysics*, 30, 557-564.

Watamaniuk, S. N. J., McKee, S. P., & Grzywacz, N. M. (1995). Detecting a trajectory embedded in random-direction motion noise. *Vision Research*, 35, 65-77.

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, UK.

Watanabe, K., & Shimojo, S. (2001). Postcoincidence trajectory duration affects motion event perception. *Perception and Psychophysics*, 63, 16-28.

Weichselgartner, E., & Sperling, G. (1987). Dynamics of automatic and controlled visual-attention. *Science*, 238, 778-780.

Welch, R. B., Duttonhurt, L. D., & Warren, D. H. (1986). Contributions of audition and vision to temporal rate perception. *Perception & Psychophysics*, 39, 294-300.

Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin*, 88, 638-667.

Wilkinson, L. K., Meredith, M. A., & Stein, B. E. (1996). The role of anterior ectosylvian cortex in cross-modality orientation and approach behavior. *Experimental Brain Research*, 112, 1-10.

Williams, D. W., Phillips, G., & Sekuler, R. (1986). Hysteresis in perception of motion direction as evidence for neural cooperativity. *Nature*, 324, 253-255.

Williams, D. W., & Sekuler, R. (1984). Coherent global motion percepts from stochastic local motions. *Vision Research*, 24, 55-62.

Winkler, I., Czigler I., Jaramillo, M., Paavilainen, P., & Näätänen, R. (1998). Temporal constraints of auditory event synthesis: evidence from ERPs. *NeuroReport*, 9, 495-499.

Winkler, I., & Schröger, E. (1995). Neural representation for the temporal structure of sound patterns. *NeuroReport*, 6, 690-694.

Witkin, H. A., Wapner, S., & Levelthal, T. (1952). Sound localization with conflicting visual and auditory cues. *Journal of Experimental Psychology*, 43, 58-67.

Wood, N. L., & Cowan, N. (1995). The cocktail party phenomenon revisited: Attention and memory in the classic selective listening procedure of Cherry (1953). *Journal of Experimental Psychology: General*, 124, 243-262.

Wright, H. N. (1964). Temporal summation and backward masking. *Journal of the Acoustical Society of America*, 36, 927-932.

Yabe, H., Tervaniemi, M., Reinikainen, K., & Näätänen, R. (1997). Temporal window of integration revealed by MMN to sound omission. *NeuroReport*, 8, 1971-1974.

Yabe, H., Tervaniemi, M., Sinkkonen, J., Huotilainen, M., Ilmoniemi, R. J., & Näätänen, R. (1998). Temporal window of integration of auditory information in the human brain. *Psychophysiology*, 35, 615-619.

Yantis, S., & Egeth, H. E. (1994). Visual salience and stimulus-driven attentional capture. *Investigative Ophthalmology & Visual Science*, 35(4), S1619.

Yantis, S., & Jonides, J. (1990). Abrupt visual onsets and selective attention: Voluntary versus automatic allocation. *Journal of Experimental Psychology: Human Perception & Performance*, 16, 121-134.

Yantis, S., & Jonides, J. (1996). Attentional capture by abrupt onsets - new perceptual objects or visual masking. *Journal of Experimental Psychology: Human Perception and Performance*, 22, 1505-1513.

Zanker, J. M. (1992). Noise thresholds of Fourier, drift-balanced and paradox theta motion. *Investigative Ophthalmology and Visual Science*, 33(4), S974.

Zhang, K. (1996). Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: A theory. *Journal of Neuroscience*, 16, 2112-2126.

Zipser, D., & Andersen, R. (1988). A back-propagation programmed network that simulates response properties of a subset of posterior parietal neurons. *Nature*, 331, 679-684.

Zwislocki, J. J., & Sokolich, W. G. (1974). On loudness enhancement of a tone burst by a preceding tone burst. *Perception & Psychophysics*, 16, 87-90.