Implicit auditory modulation on the temporal characteristics of perceptual alternation in visual competition

Kohske Takahashi

Research Center for Advanced Science and Technology, The University of Tokyo, Japan, & Japan Society for the Promotion of Science, Japan

Katsumi Watanabe

Research Center for Advanced Science and Technology, The University of Tokyo, Japan, Japan Science and Technology Agency, Japan, & National Institute of Advanced Science and Technology, Japan

Visual competition refers to the spontaneous change of the subjective perception of ambiguous visual patterns. We investigated how implicit and explicit auditory inputs affect the temporal characteristics of perceptual alternation and the interpretation bias in ambiguous visual patterns. Participants traced the perceived direction of apparent visual motion, while two task-irrelevant auditory tones were alternately presented. In the pre- and post-learning sessions, ambiguous apparent motion (seen as moving vertically or horizontally) was presented. In the learning session, disambiguated vertical and horizontal apparent motions were alternately presented and switched in synchronization with the changes in auditory tones. The results showed that the temporal intervals of perceptual alternation from the auditory switches were reduced after the participants experienced the synchronized audiovisual switches, even when the auditory switches were not consciously detectable. The magnitude of the effect was comparable for the implicit and explicit auditory switches. Neither explicit nor implicit auditory tones biased the interpretation of the ambiguous visual motion. These results suggest that auditory inputs implicitly affect the temporal characteristics of perceptual alternation after participants experience synchronized audiovisual events.

Keywords: visual competition, cross-modal interaction, auditory, implicit processing, learning

Citation: Takahashi, K., & Watanabe, K. (2010). Implicit auditory modulation on the temporal characteristics of perceptual alternation in visual competition. *Journal of Vision, 10*(4):11, 1–13, http://journalofvision.org/10/4/11/, doi:10.1167/10.4.11.

Introduction

When people look at ambiguous visual figures such as the Necker cube and the Rubin's vase, their visual interpretation of the figures spontaneously changes from one to another; this phenomenon has been termed visual competition (e.g., Blake & Logothetis, 2002; Kim & Black, 2005; Leopold & Logothetis, 1999; Sterzer & Kleinschmidt, 2007). Although perceptual alternation appears to occur spontaneously and stochastically, various internal and external factors are known to affect the temporal characteristics of perceptual alternation and bias the dominant interpretation of ambiguous stimuli (Blake, Sobel, & Gilroy, 2003; Kanai, Moradi, Shimojo, & Verstraten, 2005; Kornmeier, Hein, & Bach, 2009; Leopold, Wilke, Maier, & Logothetis, 2002; Maruya, Yang, & Blake, 2007). The inputs of sensory modalities other than vision also affect visual competition (Ando & Ashida, 2003; Blake, Sobel, & James, 2004; Bruno, Jacomuzzi, Bertamini, & Meyer, 2007; James & Blake, 2004; Sekuler, Sekuler, & Lau, 1997; van Ee, van Boxtel, Parker, & Alais, 2009; Watanabe & Shimojo, 2001a, 2001b). For example, when observers look at the Necker cube and touch a wire-frame cube with their hand, the interpretation of the ambiguous visual figure is biased so as to be congruent with the structure of the haptic object (Ando & Ashida, 2003; Bruno et al., 2007).

Does cross-modal modulation in visual competition occur via direct interaction among the different sensory modalities without specific intention or awareness, or is this modulation explicitly induced by top-down control? Research has shown that observers can intentionally change the frequency of perceptual alternation and bias the dominant interpretation of ambiguous visual patterns (Kornmeier et al., 2009; Meng & Tong, 2004; Toppino, 2003; Tsal & Kolbet, 1985); therefore, cross-modal modulation might reflect a top-down control of synchronization between visual perception and other modalities of sensory perception. However, it has also been shown that subthreshold sensory inputs can influence various types of behavioral performance and brain activity (Hoshiyama,

doi: 10.1167/10.4.11 Received July 31, 2009; published April 27, 2010 ISSN 1534-7362 © ARVO

Okamoto, & Kakigi, 2007; Sasaki et al., 2008; Tsushima, Sasaki, & Watanabe, 2006). If undetectable auditory or haptic inputs influence visual competition, it would imply that cross-modal interaction in visual competition can be implicit.

Previous research on cross-modal modulation on visual competition has mainly focused on the dominant interpretation of ambiguous visual patterns with regard to cross-modal congruency (Ando & Ashida, 2003; Blake et al., 2004; Bruno et al., 2007; James & Blake, 2004; Sekuler et al., 1997; Watanabe & Shimojo, 2001a, 2001b). However, the temporal characteristics of perceptual alternation, as well as a dominant interpretation, are also susceptible to external events, and these two aspects are dissociable. For example, visual transient events have been shown to modulate the temporal characteristics of perceptual alternation without any interpretation bias (Kanai et al., 2005). Therefore, it would be crucial to examine the effect of auditory modulation on both the temporal characteristics and interpretation bias in visual competition.

The goal of this study is to examine how auditory events affect visual competition. More specifically, we tested whether explicit and implicit auditory events influence when subjective perception changes (temporal characteristics) and what is seen (interpretation bias) for ambiguous visual patterns. For this purpose, we used a learning paradigm wherein observers were repeatedly exposed to synchronized audiovisual events. In Experiment 1, two auditory tones were alternately presented. Each tone was far above the threshold, but the switches between the tones were undetectable. In the learning session, two disambiguated apparent motions were alternately presented in synchronization with the undetectable switches in the auditory stimuli. In the pre- and post-learning sessions, visual stimuli consisted of ambiguous apparent motion. Observers traced the perceived direction of visual motion. We compared the timing of perceptual alternation and the dominant interpretation with regard to the auditory tones between the pre- and post-learning sessions. In Experiments 2 and 3, we tested the effects of auditory modulation using detectable switches in auditory tones in order to compare the effects of implicit and explicit auditory modulations in visual competition.

Experiment 1

Methods

Apparatus and stimuli

Visual stimuli were presented on a 21-inch CRT monitor (100 Hz; viewing distance, 57 cm). We used three types of visual motion patterns: ambiguous apparent motion (quartet dot), horizontal apparent motion, and vertical apparent motion (Figure 1a). Each visual pattern consisted of two blue dots (2.2 cd/m^2) on a gray background screen (2.5 cd/m^2). For the ambiguous motion pattern, the two dots alternately appeared on the top-left and bottom-right vertices or the bottom-left and top-right vertices of an imaginary square (2.7° × 2.7°). The visual stimulus was placed in the center of the screen. The duration of each display was 250 ms. For the horizontal (or vertical) motion pattern, two additional dots appeared in the middle of the top and bottom edges (or left and right edges) of the imaginary square between the displays on the vertices. The durations of the dots on the vertices and on the edges were 200 ms and 50 ms, respectively. The additional dots in the middle of the edges clearly disambiguated the motion direction.

Auditory stimuli were presented through headphones. In a single trial, two auditory tones were alternately presented (Figure 1b). One was a complex tone with seven components (CT7), and the other comprised of six components (CT6). The frequencies of the components of CT7 were equally spaced from 500 Hz to 900 Hz on a logarithmic scale, and the amplitude of the middle component was half of the amplitude of the others. The components of CT6 were the same as those of CT7, except that CT6 did not contain the middle component. It was virtually impossible to consciously discriminate between the two tones (Hoshiyama et al., 2007; Appendix A). The amplitudes of the tones were modulated by a 0.5-Hz sinusoidal wave. The auditory tone switched when the amplitude was zero, and the temporal intervals of the switches were sampled on a discrete (1-s bin) uniform distribution with a range from 4 to 12 s.

Procedure

Thirteen undergraduate students participated as paid volunteers. A trial began when observers pressed the enter key on the keyboard. The visual and auditory stimuli appeared after 1 s of the blank period. The observers reported the perceived direction of the apparent visual motion (horizontal or vertical) by pressing or releasing the space bar. The key assignment was counterbalanced across observers. Each trial lasted for 120 s. The observers performed three trials of the pre-learning session, followed by 6 trials of the learning session, and then 3 trials of the post-learning session, all in succession (Figure 1c). In the pre- and post-learning sessions, we presented the ambiguous motion concurrently with the alternating CT6 and CT7 tones. In the learning session, the disambiguated horizontal or vertical motion was alternately presented concurrent with CT6 or CT7. The pairing between visual and auditory stimuli (horizontal for CT6 and vertical for CT7, or vice versa) in the learning session was fixed for each observer and counterbalanced among observers. Before the experiment, we confirmed that the observers experienced perceptual alternation and could perceive both horizontal and vertical motions for ambiguous visual motion; then, the observers performed one trial with

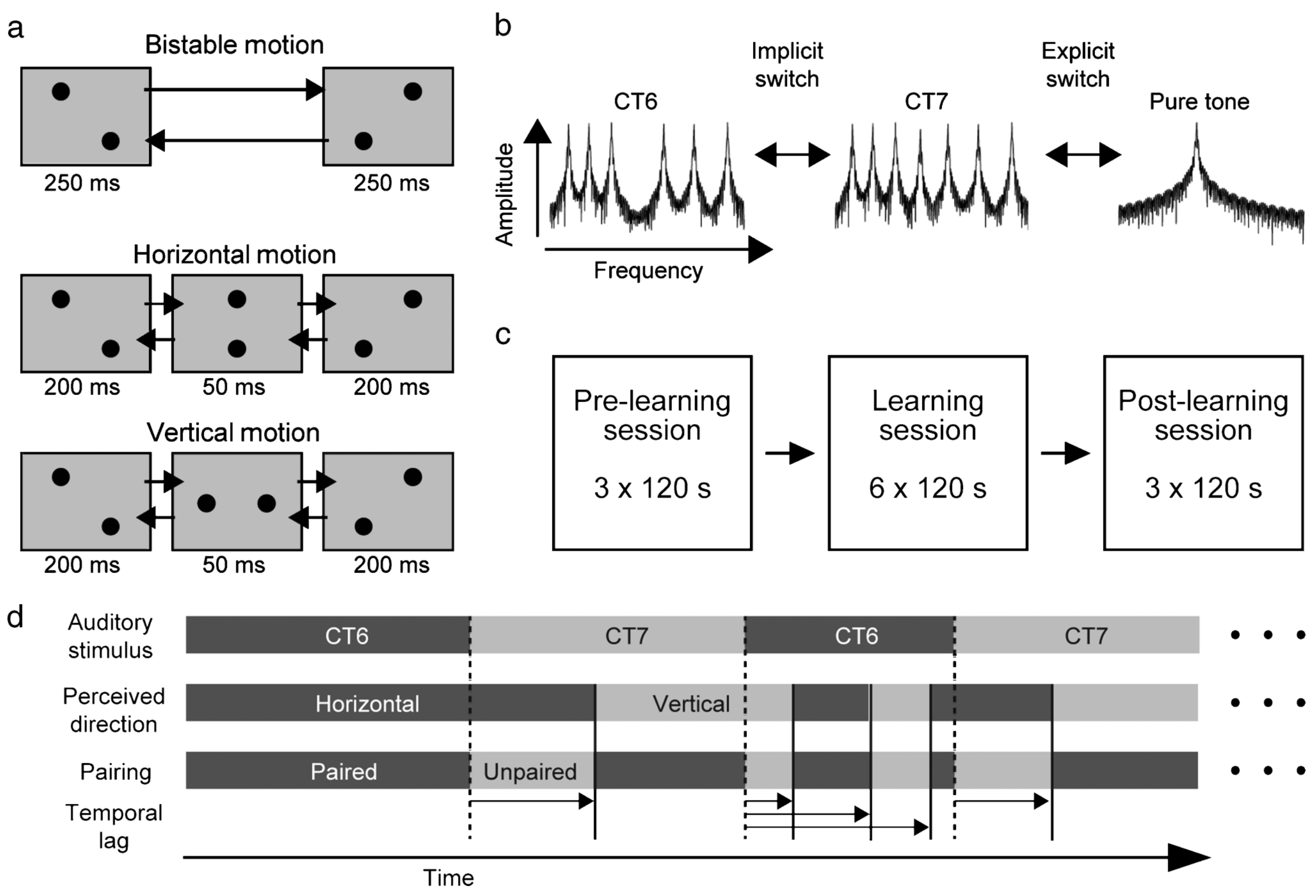

Figure 1. (a) Visual stimuli. For bistable ambiguous motion (quartet dot; top), two dots alternately appeared on the diagonal vertices of an imaginary square. For disambiguated motion (middle and bottom), additional dots briefly appeared in the middle of the edge lines of the imaginary square. (b) Frequency spectrum of auditory stimuli. For the implicit switch (Experiments 1 and C1), CT6 and CT7 were alternately presented. For the explicit switch (Experiments 2 and C2), CT7 and a pure tone were alternately presented. The amplitudes of the auditory stimuli were modulated by a 0.5-Hz sinusoidal wave. For transient stimulation (Experiment 3 and C3), a pure tone was presented for 100 ms. (c) Sequence of the experiment. (d) Schematic illustration of the method of analyses. The top bar shows the auditory stimulation, and the middle bar shows the perceived direction of bistable motion. The bottom bar shows whether the auditory stimulation and the perceived motion were paired. Assuming that horizontal (vertical) motion was coupled with CT6 (CT7) in the learning session, periods for which the perceived direction was horizontal (vertical) under CT6 (CT7) were defined as "pairing" (dark gray area). The arrows below the bars show the temporal lags of visual alternations (solid vertical lines) from the preceding auditory switches (dashed vertical lines).

ambiguous visual motion and one with disambiguated visual motion for practice.

After the experiment, the observers performed an auditory discrimination test. In the test, ambiguous visual motion was presented for 3 s, and the auditory tones (CT7 or CT6) were presented on the first and last seconds of the visual stimulus. The observers' task was to judge whether the two tones were identical. In total, 40 trials were conducted.

Data analysis

The methods of the data analysis are depicted in Figure 1d. We were mainly interested in (1) the temporal interval of perceptual alternations from auditory switches and (2) the interpretation bias induced by auditory tones.

To determine the temporal characteristics of perceptual alternation with regard to the auditory switches, we followed the index used in previous studies (Bruno et al., 2007; Kanai et al., 2005); we collected all perceptual alternations after the first auditory switch in each trial and calculated the temporal intervals of these visual alternations from the preceding auditory switches (the arrows in Figure 1d). We defined the mean of these temporal intervals as "mean temporal lag." The step-by-step calculation of the mean temporal lag is given as follows.

1. We collected auditory switch times (the dashed vertical lines in Figure 1d) in each trial (i.e., elapsed time from the beginning of a trial). Here, they are denoted as $AT(n)$ (s) (n = 1, 2, …, j; j is the total number of auditory switches in the trial).

2. We also collected perceptual alternation times (the solid vertical lines in Figure 1d), which are denoted as $PA(m)$ (s), ($m = 1, 2, \ldots k$; k is the total number of perceptual alternations in the trial). Note that we discarded perceptual alternations that took place before the first auditory switch.
3. For each $PA(m)$, we looked up $AT_{PA(m)}$ that was the maximum value of AT satisfying $AT(n) < PA(m)$ (i.e., $AT_{PA(m)}$ represents the timing of the last auditory switch that took place before the mth perceptual alternation).
4. We defined $PA(m) - AT_{PA(m)}$ as temporal lag. The arrows in Figure 1d indicate these temporal lags for each perceptual alternation.
5. These manipulations (steps 1 through 4) were repeated for all trials.
6. Finally, we averaged (arithmetic mean) all the temporal lags for each session (i.e., the pre-learning, learning, and post-learning sessions), which leads to the mean temporal lag.

Thus, we obtained the mean temporal lag of the pre-learning, learning, and post-learning sessions for each observer. The advantage of using the mean temporal lag was that it enabled us to define an expected mean lag assuming that the auditory switches had no effect on the timing of perceptual alternation. If perceptual alternation times are independent from auditory switch times, the distribution of temporal lag should be

$$P(lag) = \frac{1}{72} \sum_{\delta=4}^{12} H(\delta - lag), \tag{1}$$

where H is a Heaviside function. Note that the distribution of the temporal lag is determined based only on the distribution of the temporal intervals of auditory switches (i.e., uniform distribution with a range from 4 to 12 s), independent of the frequency of perceptual alternation. When the auditory switches did not modulate the timing of perceptual alternation, the expected value of the temporal lag was 4.41 s, independent of the frequency of perceptual alternation. If the mean temporal lag was smaller than this, it would indicate that perceptual alternation tended to take place closer in time to the preceding auditory switch.

Another index with regard to temporal modulation was the temporal lags of the first perceptual alternation after auditory switches. The method of calculation was the same as the mean temporal lag, with the exception that we submitted calculations only for the perceptual alternations that took place first for each auditory stimulus period and discarded the others. We defined the mean of these temporal lags as the "first temporal lag." The first temporal lag is useful to test how auditory switches selectively affect the timing of subsequent perceptual alternation. Note that the first temporal lag depends on the alternation frequency—a high alternation frequency accompanies a short first temporal lag.

In order to assess an interpretation bias, we calculated the duration of state in which the perceived direction of ambiguous visual motion was consistent with the direction paired with the auditory tone in the learning session (Figure 1d). The possible combinations of perceived motion and auditory tone were given as follows: horizontal–tone A (HA), horizontal–tone B (HB), vertical–tone A (VA), and vertical–tone B (VB). The pairing in the learning session was either HA with VB or HB with VA. In the pre- and the post-learning sessions, we defined two of them as the "paired state" and the others as the "non-paired state." For example, in the case that HA and VB were alternately presented in the learning session, they were regarded as the "paired state." We then summed the durations of HA and VB and divided them by the total duration of a single trial (i.e., 120 s), leading to the "pairing rate." A pairing rate higher (or lower) than the chance level (0.5) would indicate that auditory tones biased the dominant interpretation of ambiguous visual motion. Note that the interpretation bias may produce a distortion of the temporal lag (a temporal lag larger or smaller than chance); however, the auditory event can affect the timing of perceptual alternation without any interpretation bias (i.e., the distortion of the temporal lag can take place without distorting the pairing rate).

We also tested the interpretation bias using the first temporal lag. There were two types of perceptual alternations with regard to audiovisual pairing. One was the alternation from the paired state to the non-paired state, and the other was the alternation from the non-paired state to the paired state. For example, if the audiovisual pairing during the learning session was HA–VB, the perceptual alternation from vertical motion to horizontal motion during the presentation of tone A would be regarded as the alternation from the non-paired state to the paired state. We separately calculated the first temporal lag of these two types of perceptual alternations. If auditory switches could reverse perception from the "non-paired" to the "paired" state, then the first temporal lag of perceptual alternation to the paired state would be smaller than that of the alternation to the non-paired state.

The other indices were alternation frequency (the number of perceptual alternations for one trial) and the average duration of each percept period. When the alternation frequency was less than five, the observers were excluded from the analyses. We compared the mean lag and pairing rate in the pre- and post-learning sessions by performing a t-test with the Bonferroni correction.

Results and discussion

Two observers were excluded from the analyses owing to the frequency criterion. The mean correct rate for the

Experiment	Session	Mean lag (s)	First lag (s)	First lag [to paired (s)]	First lag [to unpaired (s)]	Pairing rate	Alternation frequency (times)	Average duration (s)
1	Pre	4.51	2.89	2.77	3.07	0.51	20.67	6.70
	Post	4.04	2.87	2.96	2.65	0.50	18.45	7.82
2	Pre	4.16	2.64	2.64	2.69	0.50	23.36	7.72
	Post	3.56	2.35	2.14	2.54	0.52	18.97	8.67
3	Pre	3.76	1.73				28.33	4.66
	Post	3.31	1.77				23.38	6.15
C1	Pre	4.46	3.21				18.07	7.00
	Post	4.27	3.31				17.97	8.19
C2	Pre	4.05	2.90				18.14	6.79
	Post	4.13	3.41				14.33	9.42
C3	Pre	3.66	2.39				19.52	6.97
	Post	3.68	2.53				16.05	8.19

Table 1. Summary of results.

auditory discrimination tests was 0.52, which was not different from the chance level, $t(9) = 0.35$, $p = 0.73$. In addition, no observers reported the auditory switch in the post-experiment interview. These results confirmed that the auditory switches between the two tones were an implicit event for the observers although the auditory tones themselves were far above the threshold (Appendix A).

The results are shown in Table 1 and Figure 2. With regard to the temporal modulation, the mean temporal lag was smaller in the post-learning session than in the pre-learning session (Figure 2a), $t(10) = 3.35$, $p < 0.01$; the mean temporal lag was smaller than the expected value (4.41 s) only in the post-learning session—pre-learning: $t(10) = 1.01$, $p = 0.34$; post-learning: $t(10) = 2.83$, $p < 0.05$. On the other hand, the first temporal lag did not differ between the two sessions (Figure 2b), $t(10) = 0.11$, $p = 0.91$.

With regard to the interpretation bias, the pairing rate did not differ between the two sessions (Figure 2d), $t(10) = 0.60$, $p = 0.56$. The pairing rates were also not higher than the chance level (0.5) in both the pre- and post-learning sessions—$t(10) = 0.71$, $p = 0.49$, and $t(10) = 0.05$, $p = 0.96$, respectively. We then conducted a two-way repeated measures ANOVA [session (pre- vs. post-learning) × alternation type (to paired vs. to non-paired)] on the first temporal lag (Figure 2c). There were no main effect of session, $F(1, 10) = 0.27$, $p = 0.62$, no alternation type, $F(1, 10) = 0.00$, $p = 0.99$, and no interaction, $F(1, 10) = 2.12$, $p = 0.18$. These results suggested that the effect of auditory modulation on the interpretation bias was absent in our experiment.

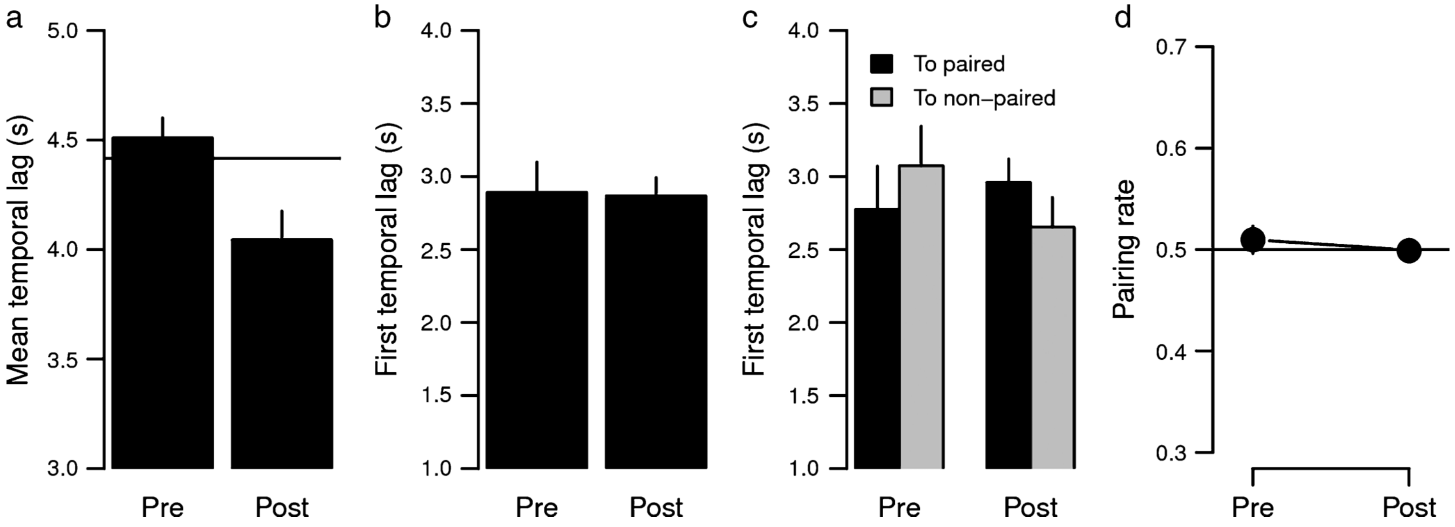

Figure 2. (a) Mean temporal lag, (b) first temporal lag, (c) first temporal lag from the non-paired to the paired state (black) and from the non-paired to the paired state (gray), and (d) pairing rates in the pre- and post-learning sessions in Experiment 1. The horizontal lines in (a) and (d) indicate the chance level. The error bars indicate the standard errors of the means.

Other indices, the alternation frequency, $t(10) = 0.98$, $p = 0.35$, and the average duration, $t(10) = 1.48$, $p = 0.16$, did not differ between the pre- and post-learning sessions (Table 1).

The results of Experiment 1 showed that the auditory events, even those that observers could not be aware of, affected the temporal characteristics of perceptual alternation after observers had experienced the synchronized audiovisual events. In particular, we observed the effect of auditory modulation for the mean temporal lag of perceptual alternation rather than for the first temporal lag, suggesting that the effect was not limited to the first perceptual alternation after the auditory event. To our knowledge, this is the first evidence of an implicit cross-modal interaction in visual competition. On the other hand, the auditory tones did not bias the interpretation of ambiguous visual patterns. Was the interpretation bias absent because the two tones were undiscriminable, and hence, the pairing between the directions of visual motion and the auditory tones was implicit? In the next experiment, to examine the potential difference between implicit and explicit auditory modulations, we modified the tones so that observers can easily detect auditory switches.

Experiment 2

Methods

Thirteen new observers were recruited. The apparatus, stimuli, and procedures were identical to those in Experiment 1 with one exception—one tone was CT7 and the other was a pure tone (Figure 1b). The frequency of the pure tone was the same as that of the middle component of CT7. These two tones were far above the threshold and, importantly, were easy to discriminate.

Results and discussion

The mean correct rate for the auditory discrimination test was 0.98, which was significantly higher than the chance level, $t(10) = 67.1$, $p < 0.001$. Thus, unlike in Experiment 1, the observers easily discriminated the two tones (Appendix A).

With regard to temporal modulation, the mean temporal lag was smaller in the post-learning session than in the pre-learning session (Figure 3a), $t(12) = 2.53$, $p < 0.05$; the mean temporal lag was smaller than the chance level (4.41 s) only in the post-learning session—pre-learning, $t(12) = 1.56$, $p = 0.14$; post-learning, $t(12) = 5.01$, $p < 0.0001$. On the other hand, the first temporal lag did not differ between the two sessions (Figure 3b), $t(10) = 1.32$, $p = 0.21$.

With regard to the interpretation bias, the pairing rate (Figure 3d) did not differ between the two sessions, $t(12) = 0.96$, $p = 0.36$. The pairing rates were also not higher than the chance level (0.5) in both the pre- and post-learning sessions: $t(12) = 0.08$, $p = 0.94$, and $t(12) = 1.68$, $p = 0.12$, respectively. A two-way repeated measures ANOVA of session (pre- vs. post-learning) and type of alternation (to paired vs. to non-paired) on the first temporal lag (Figure 3c) showed no effect of session ($F(1, 12) = 1.69$, $p = 0.22$), no effect of alternation type ($F(1, 12) = 1.34$, $p = 0.27$), and no interaction between them ($F(1, 12) = 0.48$, $p = 0.50$). These results, which were similar to Experiment 1, suggest that the effect of auditory modulation on the interpretation

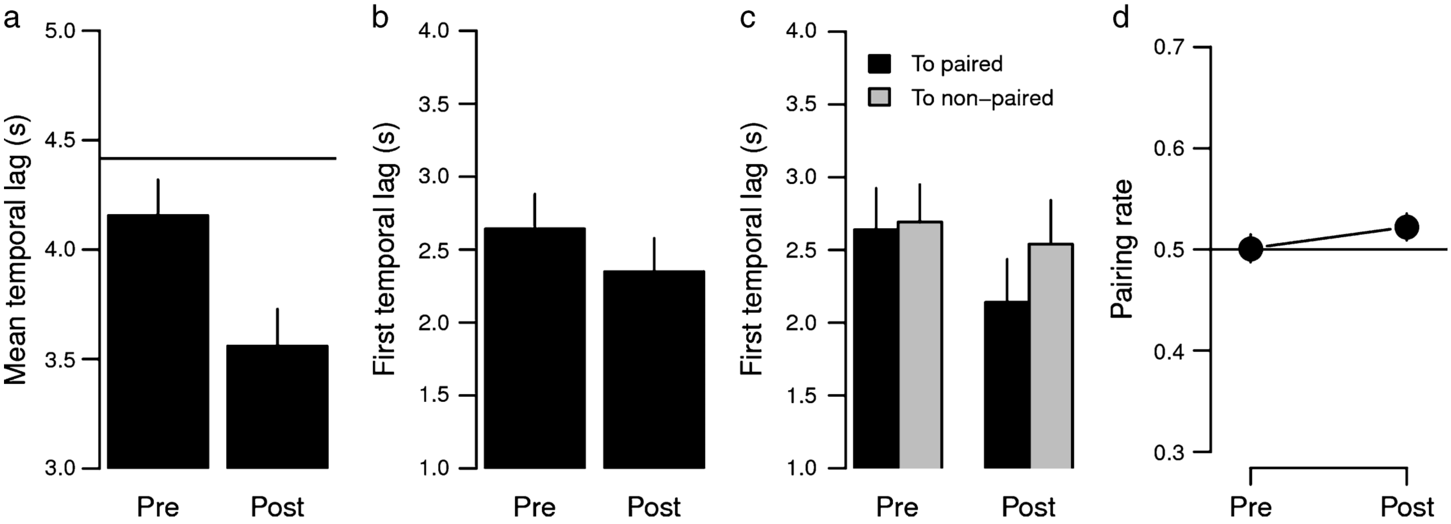

Figure 3. (a) Mean temporal lag, (b) first temporal lag, (c) first temporal lag from the non-paired to the paired state (black) and from the non-paired to the paired state (gray), and (d) pairing rates in the pre- and post-learning sessions in Experiment 2. The horizontal lines in (a) and (d) indicate the chance level. The error bars indicate the standard errors of the means.

bias was absent even when the two tones were easily discriminable, and hence, the audiovisual pairing was explicit. Other indices such as the alternation frequency, $t(12) = 2.07$, $p = 0.06$, and the average duration, $t(12) = 0.97$, $p = 0.35$, did not differ between the pre- and post-learning sessions (Table 1).

The results of Experiment 2 were similar to those of Experiment 1: after the observers had experienced the synchronized audiovisual events, the explicit auditory events distorted the timings of perceptual alternation. Although the tone and visual motion were explicitly coupled in the learning session, we did not observe any effect of auditory modulation on the dominant interpretation for the ambiguous visual motion.

Experiment 3

In the previous experiments, we associated the directions of visual motion with one of two tones; hence, the observers implicitly (Experiment 1) or explicitly (Experiment 2) encountered both temporal synchronization and audiovisual association. We observed only the temporal modulation; the interpretation bias in accordance with the audiovisual association was absent. Then, an experience of temporal synchronization may be sufficient for the subsequent effect of auditory modulation. For example, transient tones temporally synchronized with visual events could modulate the dominant percept in binocular rivalry (van Ee et al., 2009). In Experiment 3, we presented transient auditory beeps instead of the switches of continuous tones.

Methods

Seven new observers were recruited. The methods were identical to those used in the previous experiments except that the short beeps of the pure tone were presented for 100 ms at random intervals, instead of the alternate presentation of two tones. The pure tone was far above the threshold, and hence, the auditory event was explicit, as in Experiment 2. The distribution of the temporal interval of the beeps was identical to that of the auditory switches in Experiments 1 and 2 (4–12 s). In the learning session, the direction of apparent visual motion was switched in synchronization with the onset of the auditory beeps.

Results and discussion

In Experiment 3, since we did not associate the direction of visual motion with auditory tones, the interpretation bias could not be examined. With regard to temporal modulation, the mean temporal lag was smaller in the post-learning session than in the pre-learning session (Figure 4a), $t(6) = 3.31$, $p < 0.05$; the mean temporal lag was smaller than the chance level both in the pre- and post-learning sessions, $t(6) = 5.60$, $p < 0.01$ and $t(6) = 6.63$, $p < 0.001$, respectively. On the other hand, the first temporal lag did not differ between the two sessions (Figure 4b), $t(6) = 0.21$, $p = 0.84$. With regard to the other indices (Table 1), the alternation frequency was significantly smaller in the post-learning session than in the pre-learning session, $t(6) = 6.63$, $p < 0.01$, while the difference of the average duration was marginally significant, $t(6) = 1.97$, $p = 0.10$.

In Experiment 3, we found that the learning effect using the transient auditory event on the temporal characteristics of perceptual alternation was similar to those in Experiments 1 and 2; the learning of temporal synchronization between the switches of visual events and auditory transients also shortened the mean temporal lag. Thus, the association between the visual event and auditory tones were not necessary for the subsequent effect of auditory modulation; rather, these results suggest that temporal synchronization was the critical factor for audiovisual interaction in visual competition (van Ee et al., 2009). In addition, the mean temporal lag was smaller than chance, independent of learning. This observation indicates two types of auditory modulation: one, where the lag was smaller in the post-session than in the pre-session, emerges after experiencing audiovisual temporal synchronization, and the other, where the lag was smaller than chance level, is intrinsic to the auditory stimulation

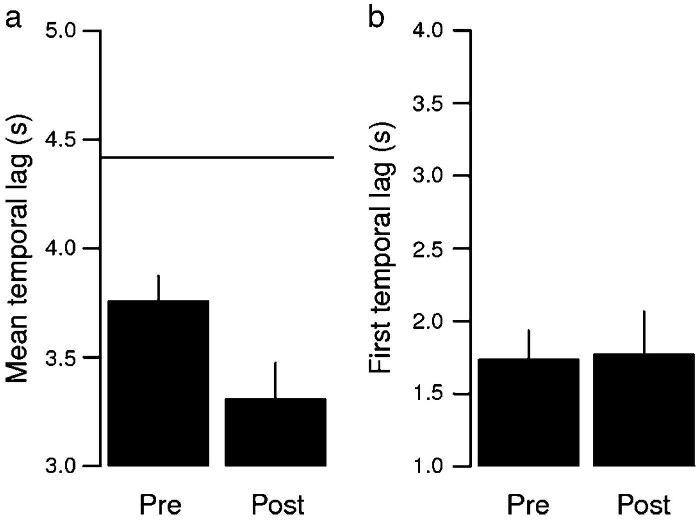

Figure 4. (a) Mean temporal lag and (b) first temporal lag in the pre- and post-learning sessions in Experiment 3. The horizontal line in (a) indicates the chance level. The error bars indicate the standard errors of the means.

itself. We further discussed these two types of auditory modulation in the section of between-experiment comparison and the General discussion section.

Control experiments

The results so far showed that the auditory events distorted the timing of perceptual alternation of ambiguous visual patterns after observers had experienced synchronized audiovisual events. To exclude the possibility that the learning effects were caused by repeatedly exposing the observers to auditory stimulation independent of audiovisual synchronization (e.g., increased sensitivity to the auditory stimulus or changes in auditory attention), we conducted control experiments wherein the auditory events were repeatedly presented, but there was no audiovisual coupling. The experimental procedures were similar to those of the main experiments (Experiments 1, 2, and 3). However, in the control experiments, we also presented ambiguous visual motion in the learning session. If the learning effect was simply due to repeated experience of auditory stimulation instead of audiovisual synchronization, the same effects would be observed in the control experiments. We conducted Experiments C1 (N = 10), C2 (N = 8), and C3 (N = 7), which corresponded to Experiments 1, 2, and 3, respectively. The observers in the control experiments were newly recruited.

Results and discussion

One observer in Experiment C2 was excluded from the analyses owing to the frequency criterion. The mean correct rate (0.47) for the auditory discrimination tests was not different from the chance level in Experiment C1, $t(9) = 0.91$, $p = 0.39$. Mean correct rate (0.87) was significantly higher than the chance level in Experiment C2, $t(4) = 3.73$, $p < 0.05$. These results replicated the results of Experiments 1 and 2.

In the control experiments, the pairing rate could not be calculated because we presented the ambiguous motion stimulus in the learning session, and hence, directions of visual motion were not associated with auditory tones. With regard to temporal modulation, we found no difference among them and the temporal lag between the pre- and post-learning sessions in all the control experiments (Figure 5): C1: $t(9) = 1.15$, $p = 0.28$; C2: $t(6) = 0.24$, $p = 0.82$; C3: $t(6) = 0.06$, $p = 0.96$. The mean temporal lag was not different from the chance level in Experiment C1—pre-learning: $t(9) = 0.27$, $p = 0.79$; post-learning: $t(9) = 1.53$, $p = 0.16$. The mean temporal lag in Experiment C2 was smaller than chance, although the differences did not reach significance—pre-learning: $t(6) = 1.88$, $p = 0.11$; post-learning: $t(6) = 1.68$, $p = 0.14$. In Experiment C3, the mean temporal lag was significantly smaller than the chance level in both the pre- and post-test sessions—pre-learning: $t(6) = 4.36$, $p < 0.01$; post-learning: $t(6) = 2.82$, $p < 0.05$. The first temporal lag did not differ between the two sessions: C1: $t(9) = 0.31$, $p = 0.76$; C2: $t(6) = 2.14$, $p = 0.08$; C3: $t(6) = 0.41$, $p = 0.70$.

These results showed that being repeatedly exposed to auditory stimulation without audiovisual temporal synchronization could not alter the effect of auditory modulation on the perceptual process of ambiguous visual figures. In other words, the learning effect observed in the main experiments, wherein the mean temporal lag was smaller in the post-learning session than in the pre-learning session,

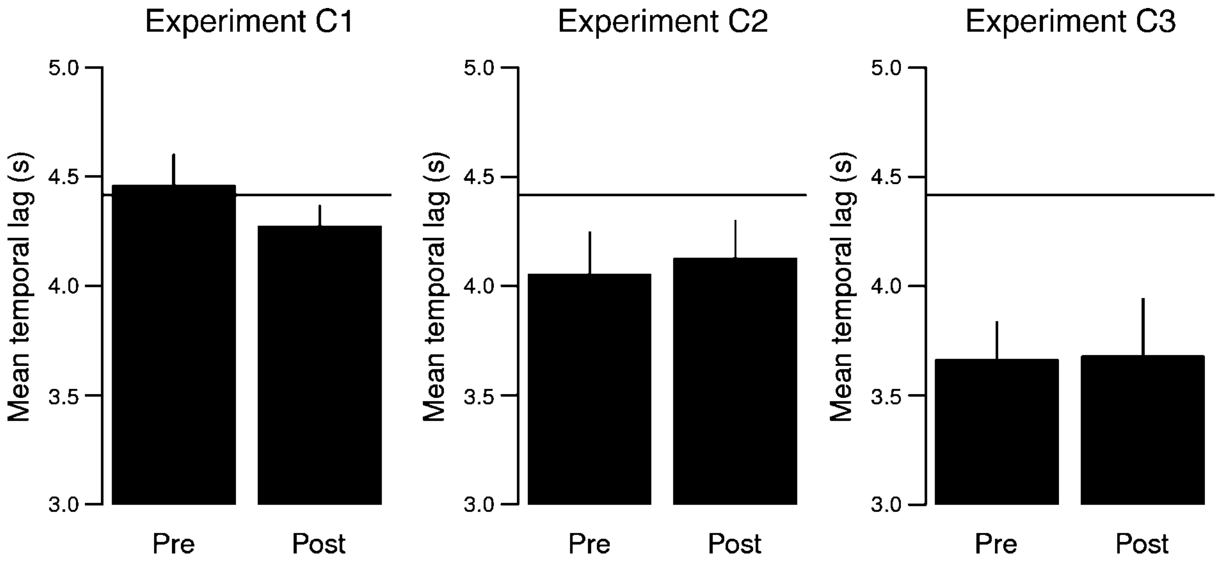

Figure 5. Mean temporal lag in the pre- and post-learning sessions in the control experiments. The horizontal line indicates the chance level. The error bars indicate the standard errors of the means.

emerged only after the observers experienced the audiovisual temporal synchronization.

Comparison among experiments

Through a series of experiments, we observed the effect of auditory modulation on the temporal characteristics of perceptual alternation, while the auditory bias for dominant perception was absent. Therefore, we sought to assess the factors influencing the mean temporal lag. There were three experimental factors: learning (main vs. control experiments; between-subject), auditory stimulus (implicit switch, explicit switch, and transient; between-subject), and session (pre- vs. post-learning; within-subject). A three-way mixed ANOVA on the mean temporal lag showed a significant main effect of session, $F(1, 49) = 12.4$, $p < 0.001$, and also a significant interaction between learning and session, $F(1, 49) = 6.55$, $p < 0.05$. Further analyses showed that the simple main effect of session was significant only in the main experiments, $F(1, 30) = 21.2$, $p < 0.001$, but not in the control experiments, $F(1, 23) = 0.14$, $p = 0.71$. Thus, the learning effect was observed only in the main experiments, where observers experienced temporally synchronized audiovisual events. Importantly, the learning effect in the main experiments was independent of auditory stimulus. We found no interaction involving session and auditory stimulus, and the difference in mean temporal lag between the pre- and post-learning sessions (about 0.5 s) was quantitatively similar, regardless of whether the auditory events were detectable.

In addition to the learning effect in the main experiments, we found a significant main effect of auditory stimulus, $F(2, 49) = 17.8$, $p < 0.001$. Importantly, the effect of auditory stimulus was independent of learning. In the transient conditions (Experiments 3 and 3C) and the explicit switch conditions (Experiments 2 and 2C), the mean temporal lag was smaller than in the implicit conditions (Experiments 1 and 1C). Although the effect of auditory stimulus, independent of learning, is of secondary importance in this research, it is noteworthy to discuss why the mean temporal lag after the explicit auditory event is shorter. The effect of explicit auditory modulation may be relevant to the induced perceptual alternation reported by Kanai et al. (2005). They found that the frequency of perceptual alternation increases immediately after the explicit visual transient event. In addition, in a different series of experiments in our laboratory that focused on the induced perceptual alternation by the explicit event, we found that both visual and auditory transient events were able to induce perceptual alternation, and the magnitudes of the visual and auditory effects were comparable and positively correlated on an individual basis (Takahashi & Watanabe, 2009). Thus, an explicit auditory event, as well as a visual transient, by itself induces perceptual alternation. This would be the reason why explicit auditory stimulation shortened the mean temporal lag independent of learning in this study. A transient event captures attention (Posner, 1980). In a cross-modal domain, the auditory transient event exogenously captures visual as well as auditory attention (cross-modal attentional link; Driver & Spence, 1998; Spence & Driver, 2004). Like the visual transient event, the explicit auditory events may induce perceptual alternation via the cross-modal attentional link. Our current conjecture is that the learning effect, which emerged only after exposure to audiovisual temporal synchronization and was caused by both implicit and explicit auditory inputs, and the auditory-induced alternation, which was independent of learning and was caused only by explicit input, are separate phenomena.

General discussion

In the present study, we investigated the effect of implicit auditory modulation on visual competition. We compared the temporal lag (i.e., the timing of perceptual alternation with regard to the auditory event) and the interpretation bias before and after the exposure to audiovisual synchronization. The main findings were given as follows: (1) After observers experienced the audiovisual synchronization, the perceptual alternation of ambiguous visual motion tended to occur closer to the auditory events in time. (2) The magnitude of the learning effects was similar for the implicit and explicit auditory events. (3) Neither implicit nor explicit auditory events produced an interpretation bias for ambiguous visual motion.

The main purpose of this study was to investigate whether implicit auditory events affect visual perception. The results suggest that implicit audiovisual interaction can influence the temporal characteristics of perceptual alternation for ambiguous visual patterns. The mean temporal lag of perceptual alternation from auditory events decreased after the observers had experienced audiovisual synchronization, despite the fact that the observers were unaware of the auditory events (Experiment 1). The finding that the magnitude of the effect was similar for implicit and explicit auditory events also supported the existence of implicit audiovisual influence, independent of auditory awareness.

How does learning of audiovisual synchronization cause the implicit auditory modulation effect on the timing of visual perceptual alternation? The reduction in temporal lag implies that visual alternations and auditory switches tend to occur closer in time. In the pre-learning session, the visual alternation occurred independent of the auditory events (i.e., the mean temporal lag did not differ from the chance level; Figure 2a) because the visual motion and the auditory tone had no inherent association. During the learning session, however, observers were repeatedly exposed to concurrent auditory switches and changes in physical motion direction, that is, synchronized (de) stabilization of auditory and visual events. Since arbitrary signals can be fused by concurrent learning (Ernst, 2007), switches of auditory tones and visual motions might be associated; this association may lead to fused stability, despite the fact that the auditory switches did not reach awareness.

The reduction in the mean temporal lag can be produced by increased alternation frequency immediately after the auditory events and later suppressed perceptual alternation, or both. To evaluate this, we compared the alternation frequencies in the pre- and post-learning sessions for <2 s (i.e., 0–2 s) and >2 s (i.e., 2–12 s) after auditory switches (Figure 6). More specifically, we divided each perceptual alternation into two bins on the basis of their temporal lag: <2 s and >2 s. Using the total presentation duration and the number of alternations of each bin, we calculated the alternation frequency per unit time (arbitrarily 1 s) for each bin. The criterion of 2 s was chosen because of the range of induced perceptual alternation (Kanai et al., 2005).

A three-way mixed ANOVA (auditory stimulus [implicit switch, explicit switch, vs. transient] × session [pre- vs. post-learning] × period [<2 s vs. >2 s]) conducted on the mean temporal lag showed a significant main effect of period, $F(1, 28) = 62.3$, $p < 0.001$, a significant two-way interaction of auditory event and period, $F(2, 28) = 23.2$, $p < 0.001$, and a significant two-way interaction of session and period, $F(1, 22) = 10.6$, $p < 0.01$. The other main effects and interactions did not reach significance. Further analyses of simple main effects revealed the following: (1) the alternation frequency was significantly larger in the period <2 s than in the period >2 s only when the auditory events were detectable (Figures 6a–6c); for Experiment 1, $F(1, 10) = 1.92$, $p = 0.19$; for Experiment 2, $F(1, 12) = 16.5$, $p < 0.01$; for Experiment 3, $F(1, 6) = 57.2$, $p < 0.001$ and (2) the alternation frequency was significantly smaller in the post-learning session compared to the pre-learning session only for the period >2 s (Figure 6d), independent of auditory event; for <2 s, $F(1, 30) = 0.28$, $p = 0.59$ and for > 2 s, $F(1, 30) = 13.9$, $p < 0.001$.

Earlier, we discussed that there were two types of auditory modulation, and they could be separate phenomena. The analyses here would support this conjecture. First, the alternation frequency for <2 s after the auditory event was higher than for >2 s after the auditory events, only if the auditory events were explicit (Experiments 2 and 3). In other words, the explicit auditory stimulation increased the alternation frequency immediately after (<2 s) the stimulation, independent of learning. This type of modulation was very similar to the induced perceptual alternation (Kanai et al., 2005; Takahashi & Watanabe, 2009). We therefore think that an explicit auditory event by itself induces perceptual alternation.

Second, and more importantly, experiencing audiovisual synchronization decreased the alternation frequency for >2 s after the auditory events, independent of whether the auditory events were detectable (Figure 6d). These results suggested that the learning effect was due to the temporally selective suppression of perceptual alternation. The suppression 2 s after the auditory events was partially consistent with the temporal constraints of the visual-haptic Necker cube (Bruno et al., 2007); in their research, observers touched and watched a Necker cube with their hand, moving or stationary. They showed that visual alternation from veridical to illusory was suppressed 2 s after the motor transition from stationary to moving, which implies that increasing the quality of haptic input that is consistent with visual input makes the visual perception more stable. In our study, the experience of audiovisual synchronization formed an association

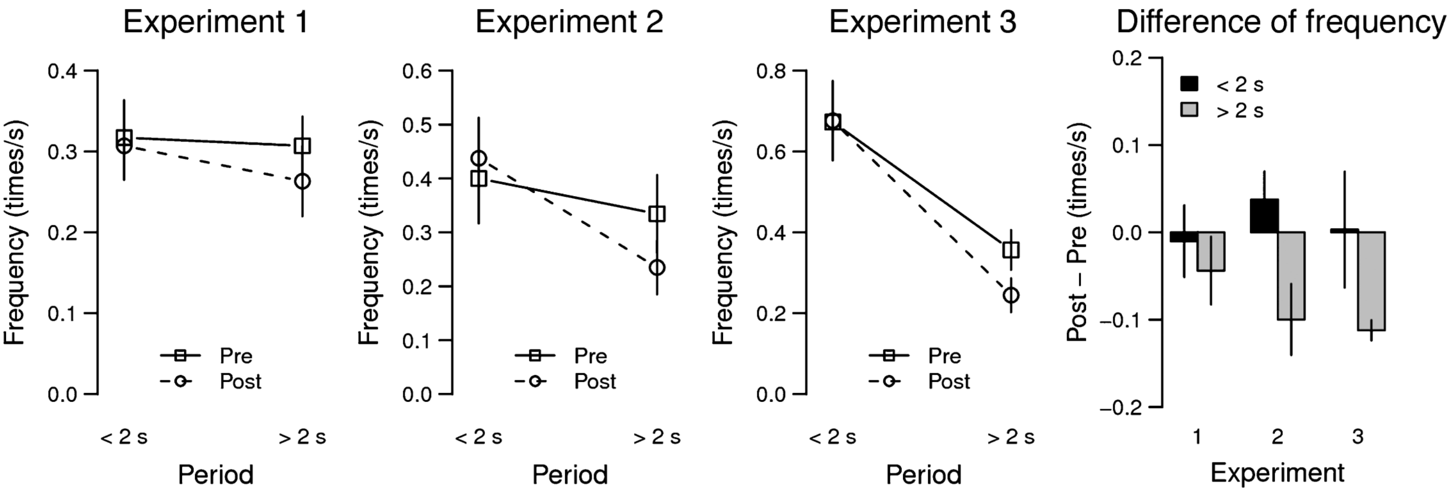

Figure 6. (a)–(c) Alternation frequencies of <2 s and >2 s after auditory switches, and (d) difference of the alternation frequencies between the pre- and post-learning sessions (i.e., frequency of post-learning − frequency of pre-learning) for each period. The error bars indicate the standard errors of the means.

between auditory events and visual stability, and the auditory quality in visual competition was relatively increased. Therefore, repeated exposure to cross-modal synchronization may allow constant auditory inputs (or absence of inputs in the case of auditory transient in Experiment 3) to function as a stabilizer for visual perception. In the visual domain, repetitive blanking (Leopold et al., 2002), moving stimuli (Blake et al., 2003), and short duration of presentation (Kanai et al., 2005) are reported to suppress perceptual alternation (or, in some cases, blanking increases alternation frequency; Orbach, Ehrlich, & Heath, 1963). The implication of these results is that the perceptual state starts fluctuating because of self-adaptation to the state, leading to alternation to other states (Blake et al., 2003; Kanai et al., 2005). The invariability of auditory information associated with visual stability may hinder self-adaptation in the visual process. It is worthwhile investigating whether visually induced suppression and cross-modal suppression are caused by a common mechanism.

Unlike previous research that reported a cross-modal interpretation bias (Ando & Ashida, 2003; Blake et al., 2004; Bruno et al., 2007; James & Blake, 2004; Sekuler et al., 1997; van Ee at al., 2009), we found that the auditory events did not affect the dominant interpretation of ambiguous visual motion. The reason for this variation could be the difference in semantic consistency between visual and other modalities' events. It is well known that when inputs to one modality are uncertain, the inputs to the other robust modalities capture the representation of the uncertain modality (e.g., ventriloquist effect, Ernst & Bülthoff, 2004; Howard & Templeton, 1966). Auditory and haptic stimuli used in the previous studies were semantically consistent with the visual stimuli (e.g., visual-haptic Necker cube), which resulted in the auditory and haptic capture of ambiguous visual events. We used vertical and horizontal visual motions and meaningless auditory tones, which were not semantically related, resulting in the absence of the interpretation bias. The semantic consistency across modalities, however, would be acquired through experience. Therefore, excessive learning of arbitrary audiovisual events might induce an interpretation bias.

The fact that the learning effect was specific to the stability of current percepts (i.e., temporal characteristics of perceptual alternation), rather than increased pairing for the conditioned association, may indicate how the temporal and semantic associations between modalities are acquired. In our experiments, temporal synchrony during learning made the implicit association between auditory and visual stabilities. However, what was heard in the auditory stability and what was seen in the visual stability (i.e., semantics of modalities' information) might not yet be associated. Thus, we speculate that the temporal synchrony of audiovisual events is implicitly learned first, after which semantic consistency among modalities is acquired. The results of Experiment 3, where the association of audiovisual inputs is unnecessary for the learning effect, are consistent with the notion that temporal synchronization plays a critical role in the audiovisual interaction in visual competition (van Ee et al., 2009).

The effect of implicit auditory modulation shown in the present study, together with previous findings indicating an interpretation bias (Ando & Ashida, 2003; Blake et al., 2004; Bruno et al., 2007; James & Blake, 2004; Sekuler et al., 1997; van Ee et al., 2009), points to the idea that cross-modal interaction contributes to the process of resolving visual ambiguity and sustaining stability in visual competition in an implicit as well as explicit manner. Anchored in the present results, there are multiple lines of further researches to clarify the underlying mechanism of implicit cross-modal interaction in visual competition.[1] For example, the time course of the learning effect should be assessed. Since we tested the learning effect immediately after the learning session, it is unclear how long the effect would last. Investigating the decay function of the learning effect would help understand the underlying mechanism, especially in clarifying the level of process relevant to the learning effect, e.g., sensory, perceptual, or cognitive level. The other line of investigation would be related to the endogenous (or intentional) perceptual decision-making process. Since studies have shown that observers intentionally bias perceptual alternation (Kornmeier et al., 2009; Meng & Tong, 2004; Toppino, 2003; Tsal & Kolbet, 1985), the learning effect may implicitly influence the intentional perceptual decision-making process. Alternatively, the effect of auditory modulation on perceptual alternation may take place at the earlier stage of the perceptual process, e.g., visual sensory area. Thus, investigating an endogenous aspect would help identify the pathway of the effect of implicit auditory modulation on visual perceptual alternation.

Appendix A

To confirm that the auditory switches between CT6 and CT7 were undetectable, three observers performed an auditory switch-detection task using two-interval, forced-choice design without any bias. In a trial, two stimuli were presented with a 0.5-s inter-stimulus interval. The stimuli consisted of two sequential tones (1 s) without temporal interval. The amplitudes of the tones were modulated by a 0.5-Hz sinusoidal wave. In the implicit condition, one of the two stimuli was CT6 followed by CT7 or CT7 followed by CT6 (with switch), and the other stimulus was two successive CT7s (without switch). In the explicit condition, one of the two stimuli was a pure tone followed by CT7 or CT7 followed by a pure tone (with switch), and the other was two successive CT7s. The observers were asked to determine the interval of the stimulus with an auditory switch. A total of 200 trials were conducted for each condition (400 trials in total). The order of the stimuli was randomized. The results showed that the auditory switches in the explicit condition, which corresponded to Experiment 2, were detected almost perfectly (the correct rates were 1.0 for the two observers and 0.92 for the other observer), and mean d' was greater than 2.5. On the other hand, the auditory switches in the implicit condition, which corresponded to Experiment 1, was virtually impossible to detect (correct rate = 0.53, d' = 0.11).

Acknowledgments

This work was supported by Shimojo Implicit Brain Function Project, ERATO, JST, and the Japan Society for the Promotion of Science. We thank Shinsuke Shimojo and Shohei Hidaka for their helpful comments and technical advice.

Commercial relationships: none.

Corresponding author: Kohske Takahashi.
Email: ktakahashi@fennel.rcast.u-tokyo.ac.jp.
Address: Research Center for Advanced Science and Technology, The University of Tokyo, 4-6-1, Komaba, Meguro-ku, 153-8904 Tokyo, Japan.

Footnote

[1] We thank the anonymous reviewers for these suggestions.

References

Ando, H., & Ashida, H. (2003). Touch can influence visual depth reversal of the Necker cube. *Perception, 32,* 97.

Blake, R., & Logothetis, N. K. (2002). Visual competition. *Nature Reviews. Neuroscience, 3,* 13–21. [PubMed]

Blake, R., Sobel, K. V., & Gilroy, L. A. (2003). Visual motion retards alternations between conflicting perceptual interpretations. *Neuron, 39,* 869–878. [PubMed]

Blake, R., Sobel, K. V., & James, T. W. (2004). Neural synergy between kinetic vision and touch. *Psychological Science, 15,* 397–402. [PubMed]

Bruno, N., Jacomuzzi, A., Bertamini, M., & Meyer, G. (2007). A visual-haptic Necker cube reveals temporal constraints on intersensory merging during perceptual exploration. *Neuropsychologia, 45,* 469–475. [PubMed]

Driver, J., & Spence, C. (1998). Crossmodal attention. *Current Opinion in Neurobiology, 8,* 245–253. [PubMed]

Ernst, M. O. (2007). Learning to integrate arbitrary signals from vision and touch. *Journal of Vision, 7*(5):7, 1–14, http://journalofvision.org/7/5/7/, doi:10.1167/7.5.7. [PubMed] [Article]

Ernst, M. O., & Bülthoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences, 8,* 162–169. [PubMed]

Hoshiyama, M., Okamoto, H., & Kakigi, R. (2007). Priority of repetitive adaptation to mismatch response following undiscriminable auditory stimulation: A magnetoencephalographic study. *The European Journal of Neuroscience, 25,* 854–862. [PubMed]

Howard, I. P., & Templeton, W. B. (1966). *Human spatial orientation*. New York: Wiley.

James, T. W., & Blake, R. (2004). Perceiving object motion using vision and touch. *Cognitive, Affective & Behavioral Neuroscience, 4,* 201–207. [PubMed] [Article]

Kanai, R., Moradi, F., Shimojo, S., & Verstraten, F. A. (2005). Perceptual alternation induced by visual transients. *Perception, 34,* 803–822. [PubMed]

Kim, C. Y., & Blake, R. (2005). Psychophysical magic: Rendering the visible "invisible". *Trends in Cognitive Sciences, 9,* 381–388. [PubMed]

Kornmeier, J., Hein, C. M., & Bach, M. (2009). Multistable perception: When bottom-up and top-down coincide. *Brain and Cognition, 69,* 138–147. [PubMed]

Leopold, D. A., & Logothetis, N. K. (1999). Multistable phenomena: Changing views in perception. *Trends in Cognitive Sciences, 3,* 254–264. [PubMed]

Leopold, D. A., Wilke, M., Maier, A., & Logothetis, N. K. (2002). Stable perception of visually ambiguous patterns. *Nature Neuroscience, 5,* 605–609. [PubMed]

Maruya, K., Yang, E., & Blake, R. (2007). Voluntary action influences visual competition. *Psychological Science, 18,* 1090–1098. [PubMed]

Meng, M., & Tong, F. (2004). Can attention selectively bias bistable perception? Differences between binocular rivalry and ambiguous figures. *Journal of Vision, 4*(7):2, 539–551, http://journalofvision.org/4/7/2/, doi:10.1167/4.7.2. [PubMed] [Article]

Orbach, J., Ehrlich, D., & Heath, H. A. (1963). Reversibility of the Necker cube. I. An examination of the concept of "satiation of orientation". *Perceptual and Motor Skills, 17,* 439–458. [PubMed]

Posner, M. I. (1980). Orienting of attention. *The Quarterly Journal of Experimental Psychology, 32,* 3–25. [PubMed]

Sasaki, H., Sakane, S., Ishida, T., Todorokihara, M., Kitamura, T., & Aoki, R. (2008). Subthreshold noise facilitates the detection and discrimination of visual signals. *Neuroscience Letters, 436,* 255–258. [PubMed]

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

Spence, C., & Driver, J. (2004). *Crossmodal space and crossmodal attention*. USA: Oxford University Press.

Sterzer, P., & Kleinschmidt, A. (2007). A neural basis for inference in perceptual ambiguity. *Proceedings of the National Academy of Sciences of the United States of America, 104,* 323–328. [PubMed] [Article]

Takahashi, K., & Watanabe, K. (2009). *Modality-independent modulation of unpredictable event on visual perceptual stability*. 13th Association for the Scientific Study of Consciousness (ASSC 13), Berlin, Germany.

Toppino, T. C. (2003). Reversible-figure perception: Mechanisms of intentional control. *Perception & Psychophysics, 65,* 1285–1295. [PubMed] [Article]

Tsal, Y., & Kolbet, L. (1985). Disambiguating ambiguous figures by selective attention. *The Quarterly Journal of Experimental Psychology Section A, 37,* 25–37.

Tsushima, Y., Sasaki, Y., Watanabe, T. (2006). Greater disruption due to failure of inhibitory control on an ambiguous distractor. *Science, 314,* 1786–1788. [PubMed]

van Ee, R., van Boxtel, J. J. A., Parker, A. L., & Alais, D. (2009). Multisensory congruency as a mechanism for attentional control over perceptual selection. *The Journal of Neuroscience, 29,* 11641–11649. [PubMed]

Watanabe, K., & Shimojo, S. (2001a). Postcoincidence trajectory duration affects motion event perception. *Perception & Psychophysics, 63,* 16–28. [PubMed] [Article]

Watanabe, K., & Shimojo, S. (2001b). When sound affects vision: Effects of auditory grouping on visual motion perception. *Psychological Science, 12,* 109–116. [PubMed]