

## Hearing Sound Alters Seeing Light

*Takahiro Kawabe*

### Abstract

This article introduces three audiovisual illusions in which hearing sound alters how light is seen: the audiovisual tau effect, sound-induced visual motion, and attenuation of the sense of agency by sounds. For each of them, we discuss a specific psychophysical mechanism called perceptual grouping that commonly underlies these illusions. Finally, we describe how scientific understanding of the audiovisual mechanism is contributing to future information technologies.

### 1. Introduction

The brain tries to comprehend the external world by interpreting sensory signals from each of the five senses—sight, hearing, touch, smell, and taste—which are also known by the scientific names: vision, audition, tactition, olfaction, and taste. These senses do not always work independently of each other. There have been many reports of psychophysical studies finding that an interpretation of sensory signals in one modality is strongly altered by sensory signals in the other modality when two senses are simultaneously stimulated. How are these senses able to interact with each other? We still have no clear answer to this question.

This article focuses on the interaction between audition and vision, and it introduces several illusions that reflect this interaction. We also discuss the idea that perceptual grouping is a fundamental psychophysical mechanism underlying a broad range of audiovisual interaction. Finally, we describe how an understanding of the psychophysical mechanism for sensory integration can contribute to future information technologies in a novel and user-friendly manner.

### 2. Vision and audition in space and time

Older studies assumed that vision is superior to audition in space perception and that audition is superior to vision in time perception (modality appropriateness hypothesis [1]). However, recent studies suggest that this assumption is not always accurate. Instead, the brain probably weights auditory and

visual signals according to the precision of each signal and judges an audiovisual event by relying on the weighted average of the sensory signals.

Recent studies have also reported an audiovisual illusion where vision and audition interact with each other beyond space and time. For example, when a sequence of flashes of light at constant space and time intervals (**Fig. 1(a)**) is accompanied by a sequence of tones at a non-constant time interval (**Fig. 1(b)**), the flash sequence is perceived to have non-constant space and time intervals (**Fig. 1(c)**). This illusion is called the audiovisual tau effect [2]. Auditory time can distort visual time judgments, and the distorted visual time judgments also alter the visual space judgments. Thus, the audiovisual tau effect is a kind of bypassed distortion of visual spatial judgments by auditory temporal information.

Interestingly, the effect is observed even in infancy. When infants are exposed to a novel stimulus, they continue to gaze upon it for a while. However, when they are exposed to a familiar stimulus, they get bored and exhibit less frequent gazes at it. By exploiting this characteristic of infants, a previous study [3] investigated whether the audiovisual tau effects occurs in infancy. Specifically, infants were exposed to an audiovisual stimulation that invokes the compelling audiovisual tau effect in adults. After a prolonged exposure to the stimulation, the infants were simultaneously presented with two spatial patterns: one involving the audiovisual tau effect and the other not involving it. The durations of gazes at each pattern were measured and it was found that the infants looked at the spatial pattern free of the audiovisual tau effect for less time than the spatial pattern with the

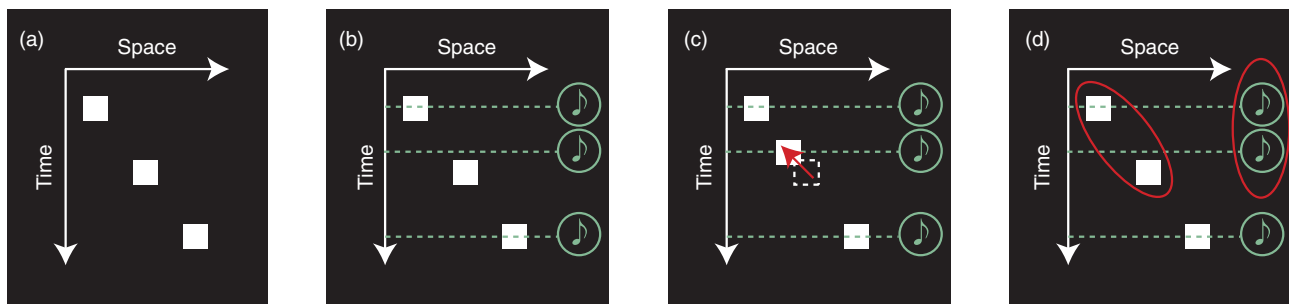


Fig. 1. Audiovisual tau effect.

effect, indicating that the audiovisual tau effect does occur even in infancy. These results indicate that audiovisual integration related to the audiovisual tau effect is located at an early level of processing in the brain, which is free from a knowledge-based interpretation of the relationship between audiovisual signals because it is implausible to assume that infants utilize some kind of knowledge.

### 3. Perceptual grouping across senses

The initial interpretation was that the audiovisual tau effect occurred because visual time representation in the brain was modulated by sounds. On the other hand, no study has ever demonstrated what kind of change in neural representation corresponds to the change in subjective experiences of time. Therefore, we should propose an explanation for the audiovisual tau effect without assuming modulation of the neural time presentation through audiovisual integration.

We think that perceptual grouping is a promising idea for sufficiently explaining the audiovisual tau effect. Perceptual grouping refers to one of the brain's strategies for treating elements with high similarity or high proximity as a group. As shown in Fig. 1(b), when two among three tones are temporally proximate, they are perceptually grouped together and the remaining one is perceptually assigned to another group.

We think that perceptual grouping in the temporal dimension affects perceptual grouping in the spatial dimension (Fig. 1(d)). Previous studies have demonstrated that elements within the same perceptual group were judged to be spatially proximate more than elements each belonging to different perceptual groups even though the physical spatial distance was identical [4]. Temporal grouping in audition might dictate spatiotemporal grouping in vision, and this

might cause the misjudgment of spatial intervals between flashes, leading to the audiovisual tau effect.

### 4. Sound alters visual motion perception

When viewing the alternate presentation of two flashes at the right and left sides of a display, we perceive the flashes to alternately move leftward and rightward (see Fig. 2(a)). This is a visual illusion called apparent motion. We see moving objects in the display of television; this is due to apparent motion. A recent study [5] discovered an interesting phenomenon whereby the apparent motion direction was subjectively fixed by presenting sounds with a non-uniform temporal interval (Figs. 2(b) and 2(c)). Specifically, a motion direction made of flashes that were temporally proximate to sounds tended to be dominant over the other. This phenomenon has been explained in terms of neural timing modulation. Specifically, sound timing might alter the timing of visual flashes, and the brain might determine the motion direction by relying on the altered visual timing.

On the other hand, we believe that this phenomenon can also be explained with perceptual grouping [6]. Our explanation is shown in Fig. 2(d). When temporally proximate sounds are grouped together (as represented by dotted ellipses), the grouping of sounds dictates the temporal grouping between flashes, and this consequently alters the spatiotemporal grouping between flashes, leading to a unidirectional apparent motion. The explanation based on perceptual grouping does not require modulation of the visual timing by sounds. Rather, the judgment for spatiotemporal *nearness* between flashes is affected by the grouping of sounds.

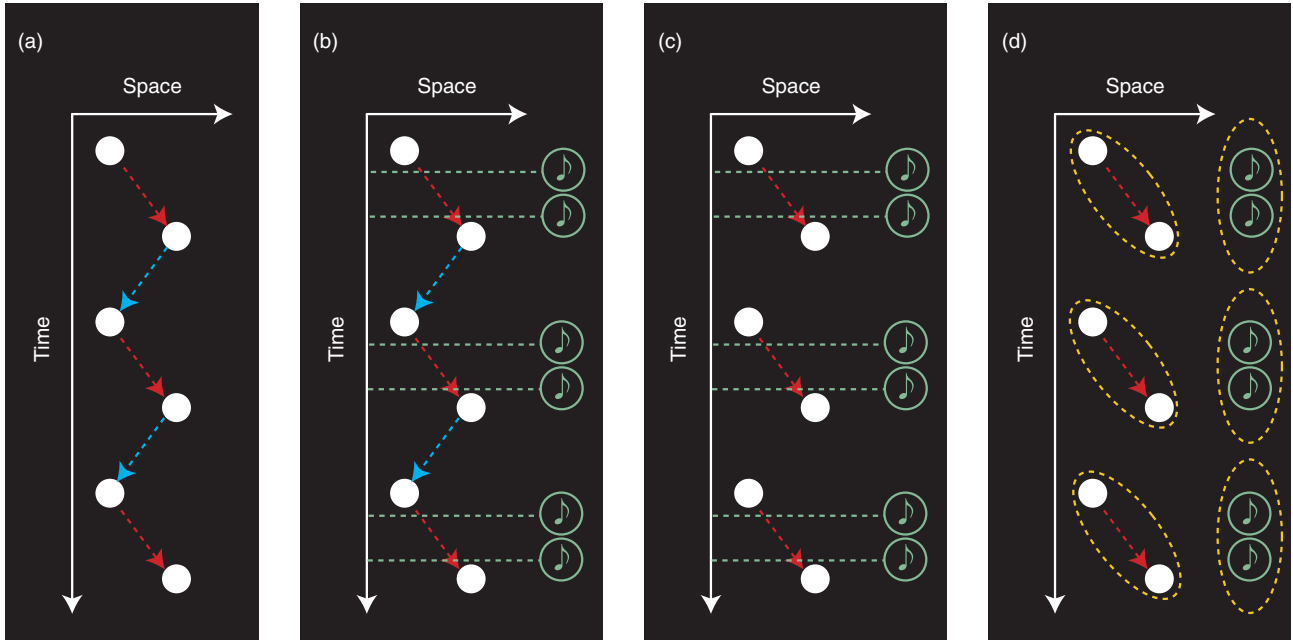


Fig. 2. Change in appearance of apparent motion direction caused by adding sounds.

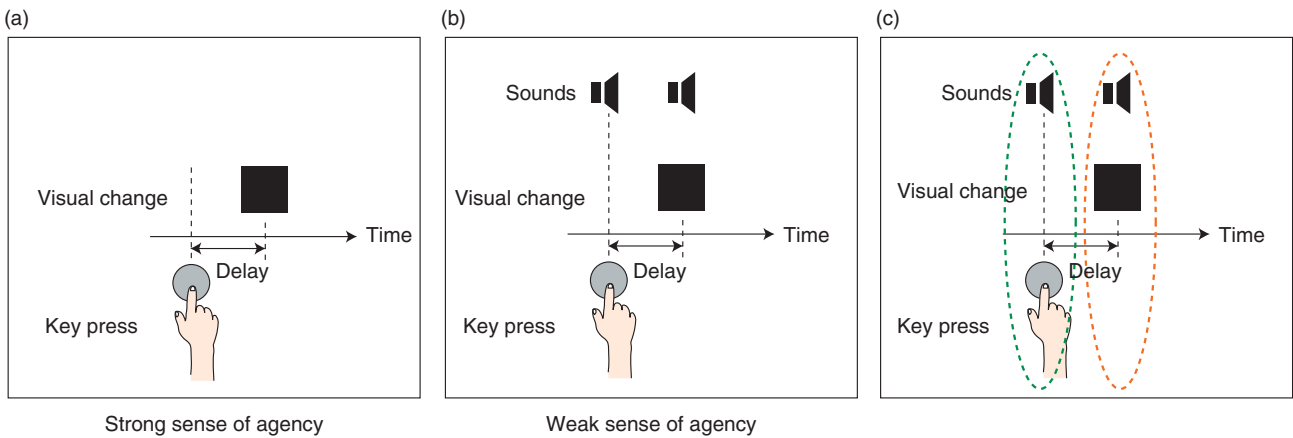


Fig. 3. Sound alters the sense of agency.

### 5. Sound alters the sense of agency

When you press a key on a computer keyboard, a visual change is often triggered on the display. In this situation, you feel as if there is a causal relationship between the key press and the visual change on the display. Counterintuitively, however, the sensation of causal relationship is also an illusion (or a product of mental processing). The sensation disappears when a

large temporal delay is inserted between the key press and the visual change on the display even though there is still a mechanical causal relationship between them. The sensation that an agent’s action triggers a change in the external world is called the sense of agency (**Fig. 3(a)**) [7].

We found that adding sounds reduced the sense of agency [8]. Specifically, when both the key press and the visual change on the display were accompanied

by pure tones, the sense of agency decreased (**Fig. 3(b)**). However, the sense of agency was not affected when only one action—either the key press or the visual change on the display—was accompanied by a pure tone.

We would like to interpret the modulation of the sense of agency by sounds in terms of perceptual grouping. When both the key press and visual change are accompanied by tones, the first tone is probably grouped together with the key press while the second tone is grouped together with the visual change (**Fig. 3(c)**). Thus, the key press belongs to a different perceptual group from the visual change, and this may result in the reduction in the sense of agency.

So far, one previous study has proposed a sensory-motor model for the sense of agency [9]. Specifically, the model assumes that the brain first predicts the outcome of an agent's action and then compares the prediction with actual sensory feedback. If the consistency between the prediction and actual feedback is low, the sense of agency is also decreased.

On the other hand, our results indicate that the sense of agency involves not only sensory-motor processing but also sensory processing. In particular, perceptual grouping is a key mechanism for the sense of agency. Investigating the interaction between sensory-motor and sensory processing will lead to a better understanding of the mechanism underlying the sense of agency.

## 6. Conclusion

My colleagues and I in our research group are trying to clarify psychophysical mechanisms for the integration of multimodal sensory signals. The clarified mechanism may be used to develop and/or improve various types of user interfaces. For example, the addition of sounds to silent video might enable us to control subtle visual representation. The temporal resolution of video in one-seg\* or augmented reality is often coarse, and the resultant subjective smoothness of the video is not high enough. Adding appropriate sounds to these videos may improve the smoothness of object motion in them. Moreover, we may be able to control the sense of agency by adding sounds in the interface. By applying the phenomenon that we have found, users can separate the information relevant to their action from irrelevant informa-

tion. Through an understanding of the principle of information processing in the brain, we would like to propose unique, interesting, and user-friendly information technology and devices in future.

## References

- [1] R. B. Welch and D. H. Warren, "Immediate perceptual response to intersensory discrepancy," *Psychological Bulletin*, Vol. 88, No. 3, pp. 638–667, 1980.
- [2] T. Kawabe, K. Miura, and Y. Yamada, "Audiovisual tau effect," *Acta Psychologica*, Vol. 128, No. 2, pp. 249–254, 2008.
- [3] T. Kawabe, N. Shirai, Y. Wada, K. Miura, S. Kanazawa, and M. K. Yamaguchi, "Audiovisual tau effect in infancy," *PLoS ONE*, Vol. 5, No. 3, e9503, 2010.
- [4] S. Coren and J. S. Girgus, "Principles of perceptual organization and spatial distortion: The gestalt illusions," *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 6, No. 3, pp. 404–412, 1980.
- [5] E. Freeman and J. Driver, "Direction of visual apparent-motion driven solely by timing of a static sound," *Current Biology*, Vol. 18, No. 16, pp. 1262–1266, 2008.
- [6] W. Roseboom, T. Kawabe, and S. Nishida, "Changes in visual apparent motion direction by cross-modal interaction are not dependent on temporal ventriloquism," Poster presented at the Vision Sciences Society Annual Meeting, Naples, FL, USA, May 2012.
- [7] P. Haggard and V. Chambon, "Sense of agency," *Current Biology*, Vol. 22, No. 10, pp. R390–R392, 2012.
- [8] T. Kawabe, W. Roseboom, and S. Nishida, "Crossmodal perceptual grouping modulates subjective causality between action and outcome," Poster presented at Asia-Pacific Conference on Vision, Incheon, Korea, July 2012.
- [9] C. D. Frith, S. -J. Blakemore, and D. M. Wolpert, "Abnormalities in the awareness and control of action," *Philosophical trans. of the Royal Society of London, Series B*, Vol. 355, No. 1404, pp. 1771–1788, 2000.
- [10] One-seg. NTT DOCOMO. <http://www.nttdocomo.co.jp/english/service/entertainment/1seg/index.html>
- [11] One-seg. Wikipedia. <http://en.wikipedia.org/wiki/1seg>



**Takahiro Kawabe**

Research Specialist, Sensory Representation Research Group, Human Information Science Laboratory, NTT Communication Science Laboratories.

He received the Ph.D. degree in psychology from Kyushu University, Fukuoka, in 2005. Since joining NTT Communication Science Laboratories in 2011, he has been studying cross-modal integration between vision and audition and multimodal integration to cause the sense of agency. He is also interested in material perception of fluids. He is a member of the Vision Science Society.

\* One-seg. A TV broadcasting service using one segment of the thirteen used for a high-definition television broadcast signal [10], [11].