Multisensory World: The Investigation of Factors Supporting Audiovisual Integration and the Establishment of Animal Model

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy in Engineering

by
Yuki Ito

Graduate School of Life and Medical Sciences
Doshisha University

November 2021

Doshisha University
Graduate School of Life and Medical Sciences

Abstract

Multisensory World: The Investigation of Factors Supporting Audiovisual Integration and the Establishment of Animal Model

by Yuki Ito
Supervisor: Professor Kohta I. Kobayasi

We are able to quickly and accurately perceive the situation of the external world by integrating multiple pieces of sensory information. However, when sensory modalities receive conflicting information, an illusion occurs in which perception can be altered. The integration of auditory and visual information cases two phenomena: perceptual and behavioral enhancement, and illusory phenomena. I investigated the factors that integrate auditory and visual information in humans. In addition, this study examined whether these two phenomena, that is, perceptual enhancement and illusion, are observed in gerbils as well, and assessed whether they are appropriate as good animal models for research on audiovisual integration.

First, I examined the effect of auditory frequency on a sound-induced flash illusion in humans. The sound-induced flash illusion is well known as illusion in which a single flash accompanied by two sounds is often perceived as two flashes. When two auditory stimuli of different frequencies were presented, the occurrence of illusion was reduced. In particular, as the frequency difference between the first and second sounds was larger, the illusory perception occurred infrequently. Furthermore, when a pure tone and noise were presented in combination, the stimuli in the combination containing the same frequency components were more likely to induce flash illusion than those not containing the same frequency components. These results suggest that sound-induced flash illusion is influenced by the frequency of the auditory stimuli.

In the second experiment, I manipulated subjective audibility and examined the relationship between the number of auditory stimuli and the visual stimuli that could be perceived. As the sound pressure level of the second auditory stimulus was low, the rate of perception of two sounds and the occurrence of illusory flash decreased. These data provide experimental evidence that susceptibility to the sound-induced flash illusion depends on subjective audibility.

Next, I investigated whether sound-induced flash illusion also occurred in gerbils. Rodents approach a novel object more frequently and spend more time exploring it than they approach an object to which they have previously been exposed. Using this property, gerbils were familiarized with a single flash, and subsequently exposed to double flashes. As a result, the contact duration increased.

This increase in exploration suggests that the gerbils recognized a change in the temporal pattern of flashing. Additionally, in the familiarization session, gerbils were allowed to explore, while single flash and double tone pips were asynchronously presented. Stimulus onset asynchrony (SOA) between sounds and a flash was the period that was considered to induce no illusion (SOA = 255 ms). When stimuli were presented synchronously after familiarization, the contact duration increased. On the other hand, when a flash and double tone pips were presented with a 100-ms SOA, the gerbil's exploration time decreased. These results suggest that sound-induced flash illusion also occurred in gerbils.

In the fifth experiment, I used flavoprotein fluorescence imaging to record the response of the visual cortex to audiovisual stimuli. The fluorescence activity to audiovisual stimuli in the visual cortex was greater than that to visual stimuli. This indicates that auditory stimulus influences the visual response.

In this dissertation, I found that frequency and subjective audibility play an important role in the sound-induced flash illusion in human studies. Furthermore, using behavioral measurements and optical imaging techniques, behavioral enhancement and illusion caused by audiovisual integration were also observed in Mongolian gerbils, and this demonstrates their effectiveness as an animal model. In summary, this thesis provides an important step towards understanding how and where auditory information and visual information are integrated.

TABLE OF CONTENTS

Chapter 1. Introduction

Chapter 2. How frequency processing affects the sound-induced flash illusion?

Chapter 3. Sound-induced flash illusion depends on frequency between first and second sounds

Chapter 4. Frequency congruency affects the sound-induced flash illusion

Chapter 5. Subjective audibility modulates the susceptibility of sound-induced flash illusion

Chapter 6. Audiovisual integration enhances behavioral performance in head-fixed Mongolian gerbil: three principles revisited

Chapter 7. Application of the novel object recognition paradigm to time-varying sensory stimulus

Chapter 8. Auditory-induced visual illusions in rodents measured by spontaneous behavioral response

Chapter 9. Audiovisual stimuli enhanced the neural activity in the gerbil’s visual cortex

Chapter 10. Conclusion

List of Figures

List of Tables

Acknowledgments

I would like to thank various individuals for their guidance and support in writing this dissertation. This thesis would not have been possible without them. I hereby convey my sincere thanks to them.

First, I would like to express my sincere gratitude to my supervisor, Professor Kohta I. Kobayasi, for his insightful advice and warm encouragement throughout the years. Without his knowledge and expertise, I would not have been able to carry out this research. I would also like to thank him for the many discussions we had, which improved my research, and for providing me with various opportunities. I am very grateful to Professor Shizuko Hiryu for teaching me important research-related methodology over the years. Owing to her support in my research, I was able to complete this thesis now. I would further like to thank my research mentor, Assistant Professor Takafumi Furuyama. He taught me all the basic skills required to perform rodent research. If it was not for his guidance, I would not have been able to perform the research, and I would not have reached this point where I am writing these acknowledgments. I am deeply grateful to Assistant Professor Koji Toda for giving technique of head-fixed operant conditioning. In particular, it was very difficult to adapt the experimental setup to the gerbil; however, I was able to complete it successfully with valuable assistance. I am profoundly thankful to Professor Tomoko Uekita for her excellent advice and valuable discussion on the novel object recognition procedure. I am very appreciative of Professor Katsuei Shibuki, Associate Professor Ryuichi Hishida, and Postdoctoral Research Associate Hiroaki Tsukano for their great technical support in flavoprotein fluorescence imaging.

Many thanks go to all of the members of the Neuroethology and Bioengineering Laboratory. I performed experiments and discussed human psychological experiments with Yu Masago and Hanaka Matsumoto. Thanks to them, we could come up with various ideas. I worked with Tomoki Osuka and Ryo Nishibori on the operant conditioning of gerbils. Their patience, excellent support, and advice were very helpful to me. I thank Yuta Tamai and Ryo Sato for their support in the novel object recognition. I was able to perform optical imaging with the help of Rikiya Uto and Junpei Ashida, and I would like to thank them for the same. I would like to thank the Japan Society for the Promotion of Science (JSPS) for the research grant.

Finally, I would like to express my gratitude to my parents, Toshiyuki and Masako, who allowed me to enter graduate school and warmly supported me. I would also like to express my heartfelt gratitude to my dear wife, Yuko, for her full support throughout my life.

Yuki Ito

Chapter 1

Introduction

We utilize multiple sensory information, such as seeing, hearing, and touching, to understand the external world. However, the information obtained from different sensory organs is not perceived separately but is integrated, and we reconstruct a coherent perception. For example, when I hit a desk, a sound is heard at the moment when my hand comes in contact with the desk. In that case, the event “my hand touched the desk” and “the sound was produced” are not perceived as separate events, but as a single event, such as “the sound was produced when my hand touched the desk”. Thus, in the physical world, it is extremely rare that only a single sensory signal is generated, and we are continuously exposed to varied sensory information, such as auditory and visual information in our daily life. In addition, the perception reconstructed by multisensory integration is often improved in most cases because the integration of the multisensory information can provide a more reliable estimate. However, when the brain receives conflicting information, the perception may be modified. This is called the “illusion” phenomenon. In this dissertation, I focus on auditory and visual integration and will separate and explain perceptual enhancement by audiovisual integration and illusion phenomena.

1.1 The characteristics of multisensory neurons

We perceive the situation of the external world by using multiple sensory information. The integration of multiple sensory information reduces the ambiguity of unimodal information and leads to a more stable perception of events. For example, in a place known to be a highly noisy environment, such as a station platform, it is difficult to hear the voice of friends compared to a quiet living room. However, we can carry on a conversation in such noisy environments since we subconsciously watch the other person's face, especially the movement of the lips, while listening to their voice. In other words, visual information helps us comprehend acoustically complex sound (speech). This effect can be observed in a laboratory setting. Several studies have demonstrated that the comprehension of speech stimuli is improved in noisy listening situations (Sumby and Pollack, 1954) and clear listening situations (Arnold and Hill, 2001) when speech stimuli are simultaneously presented with the lip of the speaker rather than when only the speech stimuli are presented. In a stimulus detection task, the accuracy of stimulus detection increases, and reaction time is faster when the multiple sensory stimuli are presented as compared to when only unisensory stimuli are presented. Thus, sensory integration improves the perception and behavior, such as stimulus detection (Frassinetti et al., 2002; Lovelace et al., 2003), localization (Harrington and Peck, 1988; Körding et al., 2007), and rapid response (Diederich and Colonius, 2004: Hershenson, 1962).

Historically, the mechanisms of sensory integration have been investigated in the cat (for a review, see Stein and Stanford, 2008). Meredith and Stein (1983) discovered multisensory neurons that respond to multiple sensory stimuli (visual, auditory, and tactile) in the deep layers of the superior colliculus (SC), located in the midbrain using single-unit recording (Meredith and Stein, 1983). Furthermore, multisensory neurons provide a significantly different firing rate in response to a unisensory stimulus than when multisensory stimuli were presented. SC is a subcortical region involved in the eye movement and is also a region converged in visual, auditory, and somatosensory information. Multisensory integration was defined by Meredith and Stein as follows. There is significant modulation of the number of impulses or the firing rate evoked by multisensory stimuli. Therefore, the impulses or firing rate of multisensory neurons was enhanced or depressed by multisensory stimuli. The higher neural response evoked by multisensory stimuli compared to unisensory stimuli is defined as enhancement, while the lower response is defined as depression (Stein and Stanford, 2008).

Meredith and Stein examined the response of these multisensory neurons in detail and found that their neural response was modulated according to three principles. Temporally proximate different sensory stimuli modulated the firing rate as compared to asynchronous stimuli (Meredith et al., 1987). When stimuli were presented from approximately the same location, the firing rate was stronger modulated than when stimuli sources were apart in space (Meredith and Stein, 1986a). These are called the temporal rule and spatial rule, respectively. It indicates that the sensory information is integrated when the stimulus source is spatiotemporally coincident. The third principle is known as the principle of inverse effectiveness. Inverse effectiveness states that multisensory gain is inversely related to the intensities of the stimulus being presented. Since a strong stimulus intensity can elicit a high neural impulse, the neural response evoked by the combination of these stimuli is only slightly stronger than the sum of the impulse evoked by the unisensory stimuli (Fig. 1.1A). In contrast, the combination stimuli in which low-intensity stimuli exhibit a relatively low firing rate produced clearly stronger neural activity than the sum of impulse when the individual stimuli were presented alone (Fig. 1.1B) (Meredith and Stein, 1986b).

Primary sensory cortices were classically, thought to process only sensory-specific information, which converged in multisensory areas; thus multisensory neurons were also thought to be present only in SC and association cortex, such as anterior ectosylvian sulcus (AES) (Stein and Wallace, 1996; Wallace et al., 1992). However, recent studies have also revealed the existence of multisensory neurons in the primary sensory cortices of the cerebral cortex. Wallace et al. (2004) found multisensory neurons in the primary auditory, visual, and somatosensory cortex; a lot of multisensory neurons were particularly observed in the intermediate cortical regions of the sensory areas (e.g., secondary visual cortex (V2), located between the primary auditory cortex and primary visual cortex; posterior parietal cortex (PPC), located between the primary visual cortex and primary somatosensory cortex) (Wallace et al., 2004). After the study by Wallace et al. study, many scientists have recorded the neural response in the primary sensory cortex in mice and rats, and multisensory neurons have been identified in each primary sensory cortex (Banks et al., 2011; Ibrahim et al., 2016; Iurilli et al., 2012; Morrill and Hasenstaub, 2018; Vasconcelos et al., 2011).

In a human imaging study, similar to rodent study, many high-order association areas (for instance, superior temporal sulcus (STS)) were involved in audiovisual integration (Calvert et al., 2000; Stevenson and James, 2009). Noesselt et al. (2007) showed that audiovisual stimuli having temporal correspondence affected not only the STS but also both the primary auditory and visual cortex using functional magnetic resonance imaging (fMRI) (Noesselt et al., 2007). It is important to note that the blood oxygenation level-dependent (BOLD) signal recorded by fMRI is a different signal, such as a spike obtained by a single-unit recording. In any case, auditory and visual information may be interacting throughout the cortex in both humans and rodents.

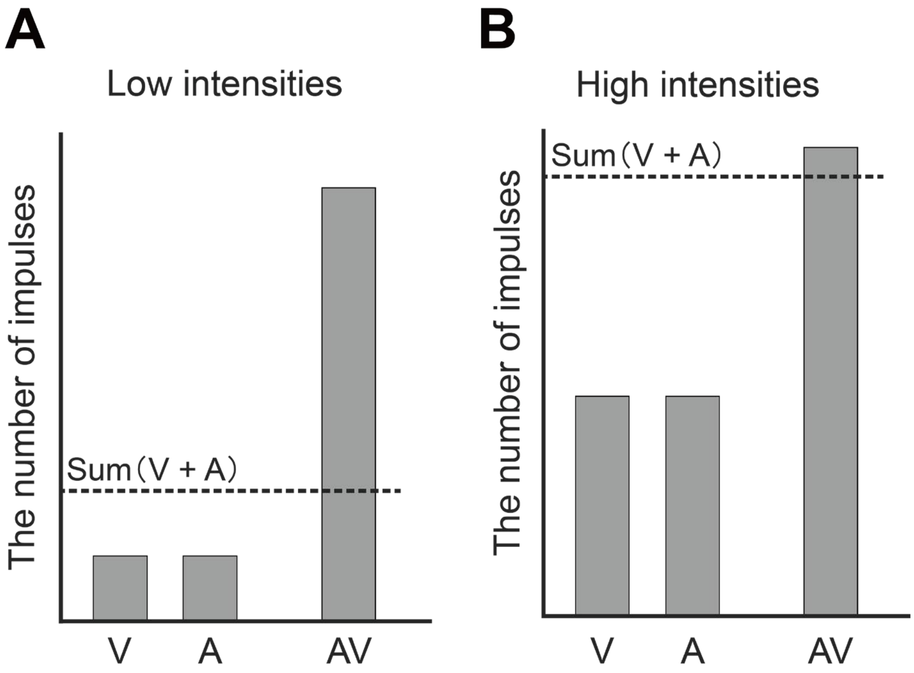

Figure 1.1. An image of the response of multisensory neurons.
(A) The response of multisensory neurons when the intensities of auditory and visual stimuli were low. (B) The response of multisensory neurons when the intensities of auditory and visual stimuli were high.

1.2 Behavioral benefits in audiovisual integration

These three principles of sensory integration are not rules that can be constrained to only a single neuron level, but also hold at the behavioral level. As explained in the introduction, audiovisual integration can improve the accuracy, localization of stimulus, and reaction time. With respect to humans, Frassinetti et al. (2002) demonstrated that the auditory stimuli presented from the same location as visual stimuli enhanced the detection performance of visual stimuli (following the spatial rule) (Frassinetti et al., 2002). Bolognini et al. (2005) showed that the sensitivity of stimulus detection was improved when simultaneous audiovisual stimuli were presented (following the temporal rule) (Bolognini et al., 2005). Furthermore, the accuracy in the localization of auditory stimuli was enhanced by visual stimuli having low intensity (following the principles of inverse effectiveness) (Bolognini et al., 2007). These studies showed that the presentation of auditory and visual stimuli in low-level features (simple sound and light) result in behavioral enhancement. It is possible to conduct and evaluate similar behavioral tasks to understand the mechanism of audiovisual integration in non-human animals.

Sakata et al. (2004) asked rats to answer the direction in which a stimulus was presented in a two-alternative forced choice (2AFC). They found that task-irrelevant sensory stimuli presented from the same direction at cue stimuli increased the success rate and reduced the reaction time compared to when the cue stimuli were presented alone (following spatial rule) (Sakata et al., 2004). The presentation of simultaneous audiovisual stimuli resulted in a rapid reaction time than the presentation of unisensory stimuli or asynchronous audiovisual stimuli (following temporal rule) (Hirokawa et al., 2008). In a study by Gleiss and Kayser, the intensity of the auditory stimuli was not changed, only the intensity of the visual stimuli. Rats showed higher correct responses when the audiovisual stimuli having weaker visual intensity were presented (following the principle of inverse effectiveness) (Gleiss and Kayser, 2012). Several studies demonstrated that behavioral gain was observed in also cats (Gingras et al, 2009) and mice (Meijer et al., 2018; Siemann et al., 2015). Therefore, audiovisual integration modulates individual neurons according to three principles and ultimately affects perception and behavior.

1.3 Illusion induced by audiovisual integration

In Chapter 1.2, I have shown that audiovisual integration can produce more reliable perceptions. However, when conflicting information between auditory and visual inputs to the brain, the perception could also be altered. Here, I will discuss three particularly famous illusions.

McGurk effect

The McGurk effect is a speech illusion in which the sound of syllable /ba/ is combined with video of lips uttering syllable /ga/ induces the perception of novel syllable /da/ (Fig. 1.2A) (McGurk and MacDonald, 1976). This means that the participants have perceived a third phoneme which is neither auditory nor visual stimulus. In particular, approximately 98 % of the adults reported that the McGurk effect occurred (McGurk and MacDonald, 1976). Furthermore, the McGurk effect occurs even when the gender of the auditory and visual stimuli is incongruent (Green et al., 1991) and when a simple light stimulus is used instead of a face stimulus (Rosenblum and Saldaña, 1996). In this illusion, there are large individual differences (Mallick et al., 2015) and also appear to be cultural differences (Sekiyama and Tohkura, 1991). In addition, spatial congruity does not seem to affect the Mcgurk effect (Jones and Munhall, 1977).

Auditory and visual stimuli must be presented within a certain time range in order to induce the McGurk effect, several previous studies reported about -60 ~ +200 ms (Jones and Callan, 2003; Munhall et al., 1996; Van Wassenhove et al., 2007). This time range is known as a multisensory time window. When auditory and visual stimuli are presented within the time window, the McGurk effect frequently occurs. However, as the temporal lag between the auditory and visual stimuli increases, the occurrence of illusion decreases. Audiovisual stimuli having a larger temporal lag are less likely to originate from a single event. This is a mechanism that is in line with the temporal rule in the three principles of multisensory integration, and this system makes a lot of sense. This is because there is a large difference in the neural conduction times in addition to the transmission speed between auditory and visual information. Therefore, even events that physically occur at the same time require different amounts of time to reach the cerebral cortex (King, 2005). It is believed that this problem is addressed by making the time window variable. In support of this idea, the temporal window is often asymmetric (the right side represents the visual stimulus precedes the auditory stimulus) and the effect of integration is often maximized when the visual stimulus slightly precedes the auditory stimulus (Zampini et al., 2003). In addition, the time required for neural processing differs depending on the feature of the stimulus; therefore, the width of the time window at simple stimuli such as pure tone is narrower than that at complex stimuli, such as speech (Conrey and Pisoni, 2006; Stevenson et al., 2012). In summary, multisensory stimuli presented within the time window are likely to be integrated to induce the improvement of percepts and illusions.

Ventriloquism effect

When a ventriloquist speaks without moving his mouth while moving the mouth of a doll held in his hand, we feel as if the doll is speaking instead of the ventriloquist. This phenomenon is called the ventriloquism effect; additionally, it is a phenomenon in which the position of the sound source is perceived by being dragged by the visual information (Jack and Thurlow, 1973). This illusion is an example of how vision dominates the final percept (Fig. 1.2B). The visual modality has higher spatial resolution than the auditory modality, showing greater reliability in spatial information relatively (Welch and Warren, 1980). Thus, in tasks involving spatial information, such as the estimation of auditory spatial source, vision is likely to take precedence. However, the dominance between these modalities is not fixed but varies from moment to moment based on the optimal Bayesian method. For instance, even when visual stimuli are blurred (i.e., unreliable), such as the situation of waking up from sleep, it is difficult to obtain an accurate perception if priority to visual stimuli is given. The reliability and accuracy of the sensory information are always weighted based on the situation, and the modality with the higher reliability dominates the final perception (Ernst and Banks, 2002). In fact, Alais and Burr (2004) revealed that a visual source is attracted to an auditory source when the reliability of visual stimuli was too weak (Alais and Burr, 2004).

These illusory phenomena arise as a result of trying to estimate perception. We may think that it would be better if an illusion did not occur because it forms a perception different from the physical world. However, illusion enriches our daily life in some aspects. For example, when we are watching TV at home, we feel as if the person on the display is talking without any discomfort. However, the sound is not presented from the person on the display, but from above, below, or behind the television, creating a spatial mismatch. It is through the influence of the ventriloquism effect that we are able to enjoy television and movies.

Sound-induced flash illusion

Shams et al. (2000) discovered sound-induced flash illusion, in which a single flash accompanied by two sounds is often perceived as two flashes (Fig. 1.2C) (Shams et al., 2000). Conversely, Andersen et al. (2004) demonstrated that the presentation of two flashes with a sound induced one flash (Fig. 1.2C) (Andersen et al., 2004). The two illusions are distinguished as fission illusion (one flash perceived as two flashes) and fusion illusion (two flashes perceived as one flash), respectively. In this dissertation, unless otherwise stated, the illusion refers to the fission illusion. Since this illusion is involved in temporal processing, the weight of the auditory modality is relatively higher than that of the visual modality and auditory information affects the final percept. Therefore, similar results have been obtained when auditory stimuli were changed to tactile stimuli, i.e., simultaneous presentation of one visual stimulus and two tactile stimuli induced the illusory flash (Touch-induced flash illusion; Violentyev et al., 2005). However, Andersen et al. (2004) showed that by reducing the amplitude of the auditory stimuli to 10 dB (lower reliability), the number of perceived auditory stimuli was altered by visual stimuli (Andersen et al., 2004). The Bayesian model devised by Shams determines the final perception based on prior expectation as well as the relative reliability of each modality (Shams et al., 2005a) and seems to be the best explanation of the sound-induced flash illusion.

Since the discovery of this illusion, several researchers have conducted various studies. As a result, various stimulus conditions affect the illusion, and among these the stimulus onset asynchrony (SOA) is a very important parameter. Shams et al. (2002) found that the susceptibility to illusion was reduced when the SOA exceeds 70 ms (Shams et al., 2002). Further, several previous studies have shown that the occurrence rate of fission illusion increased as the SOA was shorter (for a review, see Hirst et al., 2020). This temporal effect was consistent with the temporal rule, and the temporal window of illusion was reported to be approximately ±100 ms (Cecere et al., 2015). On the other hand, as the SOA becomes shorter, the interstimulus interval (ITI) between the two auditory stimuli becomes relatively shorter; eventually, the two auditory stimuli cannot be discriminated, resulting in a decrease in the relativity of auditory modality. Thus, the susceptibility to sound-induced flash illusion should depend on the ability to perceive the two stimuli. Supporting this idea, Gieseler et al. (2018) demonstrated that hearing aid users have a stronger susceptibility to illusion compared to age-matched, hearing-impaired non-users (Gieseler et al., 2018). Hirst et al. (2019) showed a relationship between self-reported hearing loss and decreased susceptibility to illusion with age (Hirst et al., 2019). However, there is no systematic evidence that the ability to detect two sounds is related to the sound-induced flash illusion.

The sound-induced flash illusion does not seem to be constrained by the spatial rule. Innes-Brown and Crewther (2009) found that an illusion occurred even when the source of visual and auditory stimuli was 20° apart (Innes-Brown and Crewther, 2009). Moreover, the flash illusion was observed at a spatial incongruence of 50° (Deloss and Andersen, 2015). This may be because the illusion paradigm is a task that does not require spatial information (Spence, 2013).

A commonly used visual stimulus in this paradigm is a high-contrast disk, but a number of studies manipulating visual stimulus have shown that the sound-induced flash illusion is robust to features of the visual stimulus (e.g., Shape: Takeshima, 2020; Gabor patch: Takeshima and Gyoba, 2015; Block pattern: Takeshima and Gyoba, 2013; Gaussian probs: Apthorp et al., 2013; Faces: Setti and Chan, 2011). For example, Setti and Chan (2011) showed that the image of buildings and faces could also induce the illusory flash (Setti and Chan, 2011). When buildings and faces as visual stimuli were presented, the occurrence rate of the illusion decreased than when the white circle was the visual stimulus (Setti and Chan, 2011). The stimuli that were familiar to the participants were less likely to induce the flash illusion than those that are unfamiliar, suggesting the involvement of familiarity (Setti and Chan, 2011). Additionally, increasing the contrast also increased the susceptibility to illusion (Pérez-Bellido et al., 2015). However, there are few studies on the effect of auditory stimulus features on this illusion even though the sound-induced flash illusion was induced by the auditory modality. As far as I know, Roseboom et al. (2013) revealed that sound-induced flash illusion did not occur when two auditory stimuli had a clear frequency difference were presented (Roseboom et al., 2013). Therefore, the frequency may be involved in the occurrence of illusory flash.

In a neurophysiological study on sound-induced flash illusion, event-related potential recordings (ERP) using electroencephalography (EEG) revealed similar patterns of activity when illusion occurred and when the visual stimulus was physically presented twice (Shams et al., 2001). Furthermore, early modulation of visual cortex activity has been found in several EEG studies (Arden et al., 2003; Mishra et al., 2007). Similar to the ERP study, the early interactions between the auditory and visual cortex were observed using magnetoencephalography (MEG) (Shams et al., 2005b). In support of this study, direct connections between auditory and visual cortical areas were found in humans (Eckert et al., 2008). In fMRI studies, Watkins et al. (2006) demonstrated that the activity in the primary visual cortex during the occurrence of fission illusion was higher than that during the presentation of one flash and was comparable to that during the physical presentation of two flashes (Watkins et al., 2006). Therefore, it is possible that the second auditory stimulus directly evokes the activity in the primary visual cortex. The BOLD signal in the primary visual cortex during the occurrence of fusion illusion was similar to the activity when one flash was physically presented (Watkins et al., 2007). Futhermore, it has been suggested that not only sound-induced flash illusion is involved in early audiovisual interaction, but also auditory and visual information was integrated at multiple stages (Mishra et al., 2007).

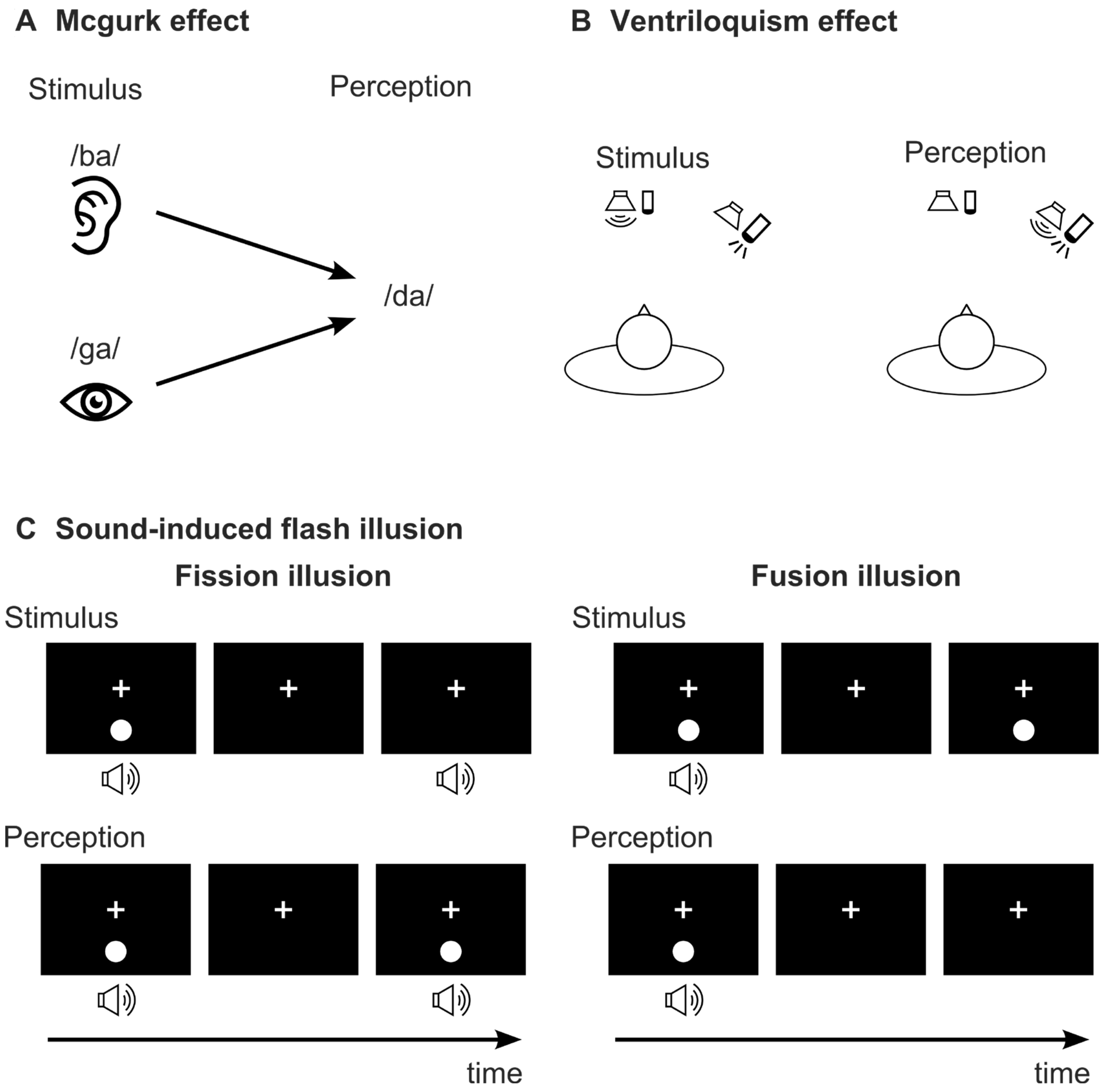

Figure 1.2. The illusions caused by audiovisual integration.

(A) McGurk effect. The sound of syllable /ba/ is combined with video of lips uttering syllable /ga/ induces the perception of novel syllable /da/. (B) Ventriloquism effect. The position of the sound source is perceived by being dragged by the visual information. (C) Sound-induced flash illusion. Fission illusion indicated that single flash accompanied by two sounds is often perceived as two flashes. Fusion illusion indicated that double flashes accompanied by one sound are often perceived as one flash.

1.4 Acceptability in the animal model: Mongolian gerbil

The Mongolian gerbil (*Meriones unguiculatus*) is a well-known common animal model for auditory research (Budinger and Scheich, 2009). Gerbils are ascribed to the family of Muridae, subfamily of Gerbillinae (Fabre et al., 2012) and live in the Mongolia desert. Compared to mice and rats, which are widely treated in rodent studies, the hearing range of the gerbil was wider from 0.1 to 60 kHz, and the sensitivity of low frequencies between 1 and 4 kHz is similar to that to humans (Ryan, 1976). Particularly, due to the audibility to the frequencies of human speech including harmonics (250 - 3000 Hz), gerbils have been used as subjects to elucidate the human speech perception in behavioral and neurophysiological studies (Ohl and Scheich, 1997; Sinnott et al., 1997). Budinger et al., (2006; 2008) assessed the direct connection of the primary auditory cortex with non-auditory cortices of gerbils by anatomical methods (Budinger et al., 2006; Budinger et al., 2008) and found that the primary auditory cortex has a connection to non-auditory sensory areas (visual and somatosensory and olfactory) and multisensory regions. In addition, Henschke et al. (2015) provided detailed anatomical pathways along the primary auditory, visual, and somatosensory cortex (Henschke et al., 2015). The multisensory neurons, which responded to not only auditory stimulus but also visual stimulus, were discovered in the primary auditory cortex (Kobayasi et al., 2013).

Furthermore, their visual system is also unique. While the majority of rodents such as mice or rats are nocturnal, gerbils are mainly diurnal. Gerbils have the acuity of about 1.5 - 2 cycles per degree, and it is suggested that the visual system of gerbils is well adapted to a diurnal lifestyle (Baker and Emerson, 1983). The circadian rhythms of their activities under natural light conditions also indicate that gerbils are not fully nocturnal (Pietrewicz et al., 1982), and they show a greater diurnal tendency than domestic mice or laboratory rats (Refinetti, 2006). The gerbil's retina also has a well-developed cone system. The retina of gerbils comprised many cones than that of mice and rats (Govardovskii et al., 1992). Additionally, the peak of cones was approximately 360 nm and 490 nm (Jacobs and Deegan II, 1994). The i-wave and d-wave found in human and other diurnal species were observed in electroretinogram (ERG) recordings of gerbils (Yang et al., 2015). The fusion frequency of flicker in the gerbil was higher than that in mice, and the response at 55 Hz flicker was observed (Yang et al., 2015). Gerbils are able to perform decision making based on the brightness, color, and contrast (Garbers et al., 2015). Furthermore, Nishiyama et al. (2011) found that gerbils modulate the frequency of vocalizations depending on the distance of another gerbil (Nishiyama et al., 2011). These behavioral and physiological features in auditory and visual senses suggest that gerbils have the ability to integrate auditory and visual information and are a good animal model for audiovisual integration research in rodents.

1.5 Outline this dissertation

This thesis aimed to clarify the factors that integrate auditory and visual information and to establish and evaluate a new animal model in order to elucidate the neural mechanisms which support the audiovisual integration. The various improvements of behavior and perception and different illusions caused by audiovisual integration have been discovered. The illusion phenomenon, in particular, is a good research model because it distinctly alters perception and serves as a key to reveal how and where the information from different sensory modalities is integrated in the brain and subsequently generates our perception. At the neural level, however, the mechanism of how the integration takes place and how the final perception is formed is not yet elucidated. One of the reasons may be that it is unclear whether the same illusion occurs in animals such as rodents established various neurophysiological methods. If an animal model which experiences the illusion can be established, the elucidation of neural mechanism in audiovisual integration is expected using a well-developed neurophysiological method. It is also important to investigate the factors that cause integration in a paradigm that is possible in rodents. Therefore, this dissertation consists of psychophysical experiments in human and behavioral and physiological experiments in Mongolian gerbils as the subject.

In Chapters 2-5, I examined the factors that constrain the occurrence of audiovisual integration in humans using the sound-induced flash illusion as a model. First, in Chapters 2-4, I examined how the frequency factors are involved in the illusion. In Chapter 5, I examined the subjective audibility and susceptibility to the illusion by focusing on the sound pressure information. Subsequently, in Chapters 6~9, I evaluated whether the gerbil is an appropriate animal model for the study of audiovisual integration. Subsequently, I investigated the neural mechanism of audiovisual integration. In Chapter 6, I examined whether audiovisual stimuli improve gerbil's behavioral performance through audiovisual integration. In Chapter 7-8, I developed and examined a novel experimental paradigm to examine whether sound-induced flash illusion occurred in gerbils. Chapter 9 examined the neural response to enhance behavioral performance by audiovisual integration. Finally, in Chapter 10, the results of this paper will be summarized, and future issues and directions to be taken will be discussed.

Chapter 2

How frequency processing affects the sound-induced flash illusion?

2.1 Introduction

We perceive the environment via multiple sensory modalities and integrate information from different modalities to reconstruct a coherent multisensory world. The integration of visual and auditory modalities plays an especially important role in our everyday lives and has received much attention (for a review, see Shams and Seitz, 2008). Several illusions are reported to occur when two sensory modalities receive conflicting information simultaneously (the 'ventriloquism effect'; Jack and Thurlow, 1973, the 'McGurk effect'; Mcgurk and Macdonald, 1976). One of the most well-known examples of this phenomenon is the sound-induced flash illusion (SIFI) (Shams et al., 2000). If two brief tones are accompanied by a single flash, the single flash is often perceived as two flashes. This illusion was discovered by Shams et al. (2000), who speculated that the auditory information affects and alters visual perception (Shams et al., 2002). Recent research into SIFI has demonstrated that when two tones of distinctly different frequencies are used as the auditory stimuli, the illusory perception occurs significantly less often, suggesting that SIFI depends on the sound frequency (Roseboom et al., 2013). However, no study has systematically investigated the effect of the stimulus frequency on SIFI. In this study, I examined whether the occurrence of SIFI changes when auditory stimuli of different frequencies are presented. I discuss how auditory frequency processing is related to auditory-visual integration.

2.2 Materials and methods

2.2.1 Participants

Thirteen adults (3 women and 10 men, 22–27 years old, all right-handed) with normal hearing and normal (or corrected normal) vision participated in the experiments. All the experiments were conducted in accordance with the guidelines for human experiments approved by the Ethics Committee of Doshisha University.

2.2.2 Stimuli

The visual stimulus was a uniform white disk (55mm diameter) displayed on a black background using a liquid crystal display (Foris FG2421, Eizo). The refresh rate of the monitor was set at 120 Hz. A fixation point (6mm white cross) was displayed in the center of the screen throughout the entire session, and the white disk was presented at 21mm below the fixation point for 8ms (= 1 frame). The visual stimulus was presented in two ways: flashing once (single flash) or twice (double flash). In the double flash, the interstimulus interval between the flashes was set at 58 ms. I confirmed the duration of the stimulus and interstimulus interval with a high-speed camera (EX-F1, Casio) with a temporal resolution of 0.83 ms (1,200 fps).

The auditory stimulus was a brief tone burst presented via headphones (SR-507, STAX). The sound pressure level of the tone burst was measured with a microphone (ER-7C Series B, Etymotic Research) and was adjusted to 80 dB SPL. The duration was set at 20 ms (the rise and fall times were 5 ms), which allowed stable pitch perception (Krumbholz et al., 2003). The auditory stimulus was always presented twice, and the interval between the tones was set at 48 ms. The first tone burst preceded the first flash by 40 ms. These temporal parameters were selected based on previous research (Shams et al., 2000) and are shown in Fig. 2.1. Three sets of stimuli were prepared, and each set had a different base frequency: 500 Hz (low frequency: LOW), 3,000 Hz (mid frequency: MID), or 5,000 Hz (high frequency: HIG). Within each stimulus set, at least the first or second sound had the base frequency, and the non-base-frequency sound had a frequency ranging from 1/2 an octave to 2 octaves above the base frequency in increments of 1/3 octave (a total of seven frequency configurations). Depending on when the base-frequency tone was presented (i.e., first or second), each stimulus set was further divided into two (denoted by ‘‘f’’ and ‘‘s,’’ representing the first and second, respectively); hence, six different stimulus configurations were used: LOWf, LOWs, MIDf, MIDs, HIGf, and HIGs (see Table 2.1 for details). In all, a total of 40 types of auditory stimuli (the 39 combinations and the no-sound condition) were paired with two types of visual stimuli (single or double flashes) to produce 80 types of stimuli. Each stimulus type was presented 11 times in a random order, resulting in a total of 880 trials for each participant.

2.2.3 Procedure

The participants sat in front of a monitor placed 60 cm from their faces and were asked to state the number of flashes they perceived (one or two) by pressing a button with their right hand. After they had pressed the button, the next trial began automatically after a 1,000 ms interval. All experiments were controlled with the Presentation software (Neurobehavioral Systems).

Table 2.1. Frequency combinations of the first and second tone stimuli.
In LOWf, MIDf, and HIGf, the frequency of the first sound was 500, 3,000, and 5,000 Hz (base frequency), respectively. In the LOWs, MIDs, and HIGs, the frequency of the second sound was base frequency.

LOWf	MIDf	HIGf	LOWs	MIDs	HIGs
2nd sound [Hz]	2nd sound [Hz]	2nd sound [Hz]	1st sound [Hz]	1st sound [Hz]	1st sound [Hz]
250	1500	2500	250	1500	2500
315	1890	3150	315	1890	3150
397	2381	3969	397	2381	3969
500	3000	5000	500	3000	5000
630	3780	6300	630	3780	6300
794	4762	7939	794	4762	7939
1000	6000	10000	1000	6000	10000

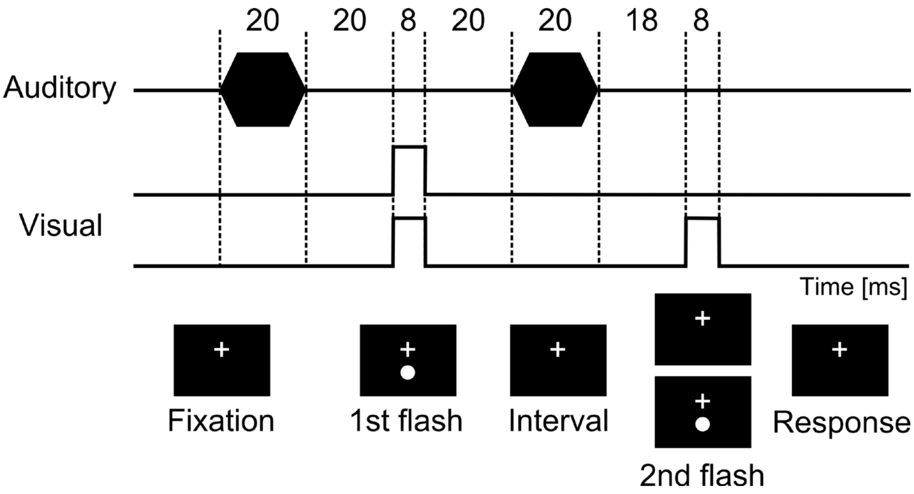

Figure 2.1. Schematic representation of the experimental procedure.
The durations of the auditory and visual stimuli were 20 and 8 ms, respectively. The first tone burst was presented 40 ms before the first flash. A white cross was always presented at the center of the screen as the fixation point. Participants responded with their right hand, using a keyboard.

2.3 Results

The results are summarized in Fig. 2.2. When tone bursts of the same frequency (= base frequency) were presented twice, the average percentage of double-flashes perception was higher than when different frequency combinations at any base frequency were presented. Analysis of the variance of the same frequency trial identified a significant effect of the base frequency ($F_{(2,36)} = 3.63$, $p < 0.05$). A multiple comparison showed a significantly higher SIFI rate when the MID stimuli or HIG stimuli were presented than when the LOW stimuli were presented ($p < 0.05$ with Bonferroni's correction), but no significant difference was observed between MID and HIG. The SIFI rate also increased as the base frequency increased (average ± SEM: LOW, 59.4 ± 8.5 %; MID, 80.4 ± 5.5 %; HIG, 82.5 ± 5.6 %). As the frequency difference between the first and second sounds increased, the illusion occurred less frequently at all base frequencies. I compared these trends between different base frequencies. While the difference is not statistically significant, the kurtosis of the response curve increased as the base frequency increased (LOWf, 1.81 ± 0.83; LOWs, 1.60 ± 0.75; MIDf, 2.29 ± 0.74; MIDs, 1.97 ± 0.60; HIGf, 3.73 ± 0.77; HIGs, 2.67 ± 0.71).

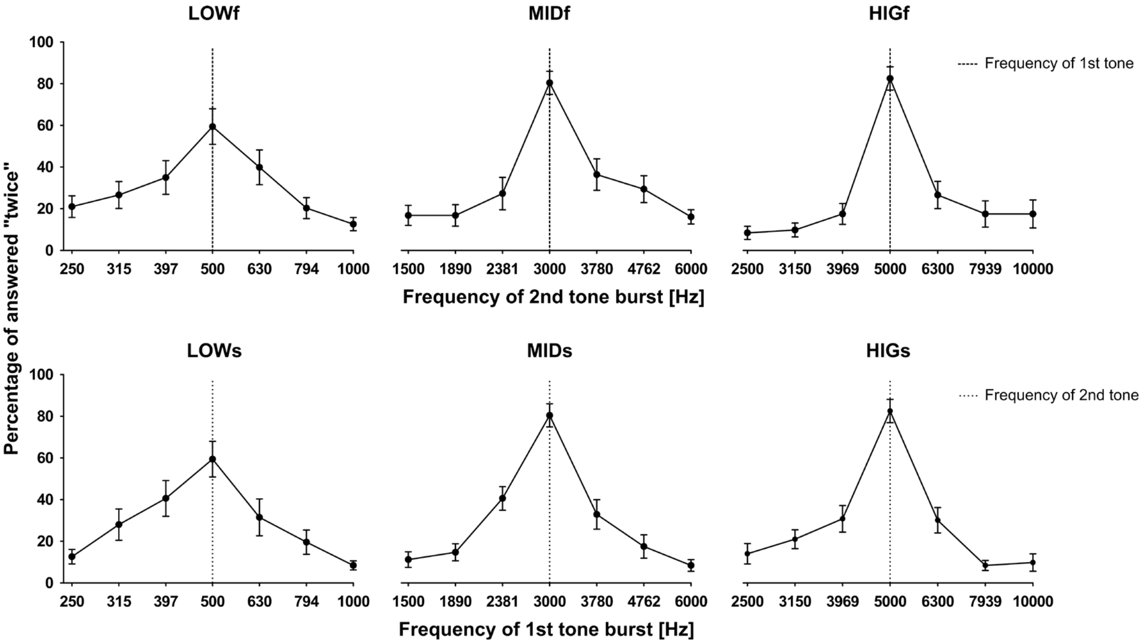

Figure 2.2. Effect of frequency on the sound-induced flash illusion.

The horizontal axes show the frequency of the auditory stimulus of the non-base-frequency sound and the vertical axes show the rate at which the participants perceived two visual stimuli. Error bars indicate the standard error of the mean (n = 13).

2.4 Discussion

When the frequencies of the first and second tone bursts were the same, the occurrence rate of SIFI was high. However, as the frequency difference between the first and second sounds increased, the illusion occurred less frequently (Fig. 2.2). A previous study by Roseboom et al. (2013) demonstrated that SIFI did not occur when two auditory stimuli had a clear frequency difference (300/3,500 Hz) (Roseboom et al., 2013), and they argued that the perceptual grouping of the two sounds was related to SIFI. Rose and Moore (2000) investigated the effect of sound frequency on the auditory stream segregation between two tone sequences and showed that the frequency difference must be small for the two sequences to be perceptually grouped together in a high-frequency range (> ca. 500 Hz) than in a low frequency range (Rose and Moore, 2000). A similar trend was observed in the kurtosis of my results, in that SIFI became more sensitive to frequency difference at a higher base frequency (Fig. 2.2). This correspondence supports the idea that the perceptual grouping of the sound stimulus is a critical component underlying SIFI (Roseboom et al., 2013).

My data also show that the SIFI rate increased as the base frequency increased (from 500 to 5,000 Hz). Several researchers have demonstrated that the auditory temporal resolution increases as the stimulus frequency decreases (Tyler et al., 1982; Viemeister, 1976). Because lower temporal resolution leads to a temporally smoother auditory image (Moore et al., 1988), the temporal dynamics of the 500 Hz tone pairs may be less distinct than those of higher-frequency tone pairs. Therefore, the lower SIFI rate at the low base frequency could be attributable to the difference in the perceptual distinctiveness of the temporal properties, or the ‘‘two-ness,’’ of the stimulus. In total, my results demonstrate that SIFI is affected by both the absolute frequency and the frequency difference of the auditory stimuli, suggesting that auditory-visual integration is not independent of auditory frequency processing.

Chapter 3

Sound-induced flash illusion depends on frequency between the first and second sounds

3.1 Introduction

Audiovisual integration plays a critical role in the accurate perception of the external world. For this reason, the process of audiovisual integration has attracted a lot of scientific interest. The illusory phenomenon in which multisensory integration modulates perception has been focused on in order to elucidate this mechanism. For example, when a brief two sound is presented with a brief single flash, we often perceive two flashes (Shams et al., 2000). This illusion was well known as a sound-induced flash illusion. Since this illusion is easily induced, the effects of various stimulus parameters on the illusion have been examined (for a review, see Hirst et al., 2020). In contrast, there are few studies focused on auditory stimuli despite the fact that auditory stimuli modulate visual perception. As far as I know, only frequency has been examined among auditory parameters. Roseboom et al. (2013) demonstrated that the flash illusion less occurred when two tones of distinctly different frequency are presented (Roseboom et al., 2013). In Chapter 2, I showed that the occurrence of illusion was reduced as the frequency difference between the first and second auditory stimuli increased. These studies suggest that sound-induced flash illusion depends on the sound frequency.

The stimulus onset asynchrony (SOA) between auditory and visual stimuli is an important parameter for the flash illusion. When the SOA exceeds about 70 ms, the illusory flash was not perceived (Shams et al., 2002). In addition, it was reported that participants having a longer temporal binding window frequently perceived the illusory flash (Stevenson et al., 2012), and this illusion was shown to be highly influenced by individual difference (Cecere et al., 2015; Huang et al., 2021; Mishra et al., 2007). Therefore, the susceptibility of illusion might also depend on the sensitivity of frequency in the task in which the sound frequency is systematically manipulated. Humans having a higher ability of frequency discrimination should less perceive illusory flash when auditory stimuli with different frequencies between the first and second sound were presented. In this study, a combination of auditory stimuli that differed in frequency between the first and second auditory stimuli was presented, and I examined the occurrence of sound-induced flash illusion and sensitivity to frequency differences.

3.2 Materials and methods

3.2.1 Participants

Twenty-one adults (11 men and 10 women, 21-29 years old), in all, participated in this experiment. All participants have normal hearing and normal (or corrected normal) vision. All experiments were conducted in accordance with the guidelines for human experiments approved by the Ethics Committee of Doshisha University.

3.2.2 Stimuli

The basic stimulus conditions were the same as the stimulus condition used in Chapter 2. The visual stimulus was a uniform white disk (55 mm diameter) displayed on a black background using a liquid crystal display (Foris FG2421, Eizo). A fixation point (6 mm white cross) was displayed in the center of the screen throughout the entire session, and the white disk was presented at 21 mm below the fixation point for 8 ms. The inter-stimulus interval (ISI) between flashes was set at 58 ms in the trial in which double flashes were presented. The auditory stimulus was a brief tone burst. The duration of the auditory stimulus was 20 ms (rise and fall time were 5 ms) and the sound pressure level was 80 dB SPL. The auditory stimulus was always presented twice via headphones (SR-507, STAX), and ISI between the tones was set at 48 ms. 3000 Hz was defined as base frequency, and the non-base-frequency was ranged from 1500 ~ 6000 Hz, and the detail was as follows; 1500, 1750, 2000, 2250, 2500, 2750, 3500, 4000, 4500, 5000, 5500, 6000 Hz. The combination of two tone bursts consisted of the same condition and different condition. In the same condition, the frequencies of two tone bursts were both base frequency, while the frequencies of two tone bursts were base frequency and non-base-frequency in the different condition. The first tone burst always preceded the first flash by 40 ms. The refresh rate of the monitor was set at 120 Hz, and the duration of visual stimulus and time profile of the stimuli were confirmed with a high-speed camera (EX-F1, Casio) and oscilloscope. The sound pressure level was measured and adjusted with a microphone (ER-7C Series b, Etymotic research).

3.2.3 Task and Procedure

The experimental procedure was the same in Chapter 2. The participants sat in front of a monitor placed 60 cm from their faces, and the height of the monitor was adjusted such that the position of their eye was the same height as the fixation point. Participants performed illusion task, and followed by frequency discrimination task. In the illusion task, the participants were instructed to answer the number of flashes they perceived by pressing a keyboard. The visual stimulus was presented once (single flash) or twice (double flashes). In the frequency discrimination task, the participants were instructed to report whether the frequencies of the presented two tone bursts were the 'same' or 'different'. In this task, the visual stimulus was not presented. After they had answered, the next trial began automatically after an interval of 1 s. All experiments were controlled with the Presentation software (Neurobehavioral Systems) and were conducted in the sound-proof room (170 [W] × 150 [L] × 230 cm [H]). The noise level at the sound-proof room was below 38.3 dB SPL.

3.2.4 Data Analysis

Psychometric functions of the percentage of two flashes participants perceived were fitted to a Gaussian curve for each individual in the illusory task. In the frequency discrimination task, psychometric functions of the percentage of 'same' that the frequencies of two auditory stimuli were fitted to a Gaussian curve. The fitting was performed by the formula using *fit* function in MATLAB (MathWorks) as follow:

$$F(x) = a \times \exp\,(-\frac{x-b}{c})^2$$

where, a is the amplitude of the fitting curve, and b indicated the peak of frequencies. c indicated the sigma in which was the width of the fitting curve.

3.3 Results

In the illusion task, when the frequencies of the first and second tone bursts were the same, the occurrence rate of the sound-induced flash illusion was significantly higher than when they were different (p < 0.001 in all frequencies conditions, Tukey's test for post-hoc comparison). Consistent with the previous study (Chapter 2), as the frequency difference between the first and second tone burst increased, the illusion occurred less frequently. There was no difference in the distribution of the susceptibility of illusion between the conditions in which the tone burst of base frequency was presented first or second time (kurtosis: p = 0.32; skewness: p = 0.23, paired t-test). In the frequency discrimination task, as the frequency difference increased, the percentage of 'same' was low (p < 0.001 in all frequencies conditions, Tukey's test for post-hoc comparison). I estimated the sigma as the width of frequency in both susceptibilities of the illusion and frequency discrimination ability. No significant correlations were observed between the width in susceptibility of the illusion and frequency discrimination ability (r = -0.12, p = 0.73).

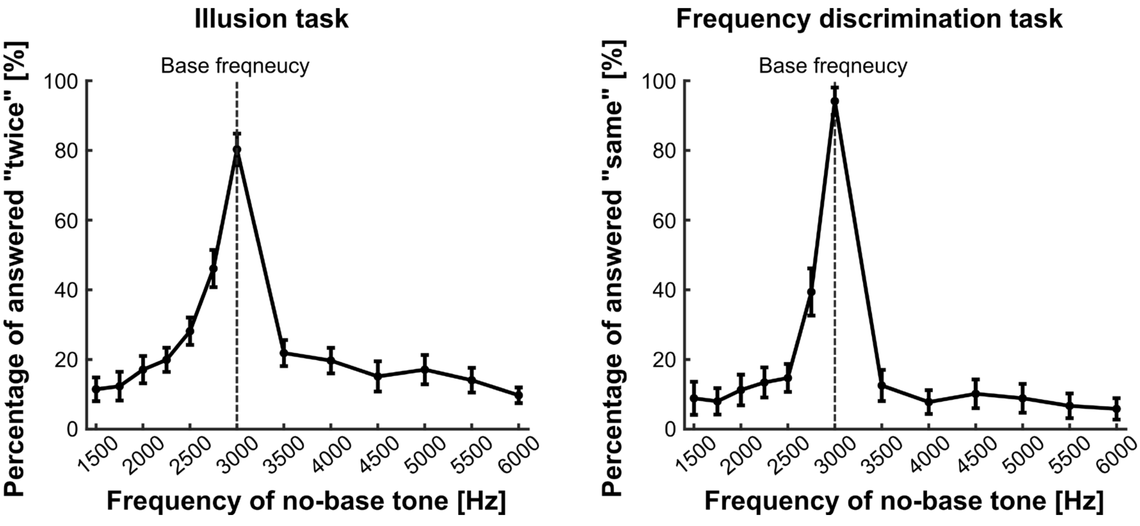

Figure 3.1. Mean percentage of occurrence in illusory flashes (left panel) and mean percentage of the same perception in frequency between the first and second tone burst (right panel).

The base frequency was 3000 Hz in both experimental tasks. When the frequencies of the first and second tone bursts were the same, the occurrence rate of the sound-induced flash illusion and same-perception was the high. Error bars indicate the standard error of the mean.

3.4 Discussion

In the present study, I examined the relationship between the susceptibility of sound-induced flash illusion and the discrimination of frequency. The auditory stimuli varying the frequency difference between the first and second sound were presented. As the result, as the frequency difference between the first and second sound increased, participants noticed the difference in frequency and did not perceive illusory flash. However, there was no correlation between the frequency discrimination ability and the susceptibility of the illusion.

Consistent with previous research in Chapter 2, the present study showed that the larger the frequency difference between the first and second tone, the less illusory flash was perceived. Great interindividual differences exist in the magnitude of sound-induced flash illusion (Cecere et al., 2015; Huang et al., 2021; Mishra et al., 2007), and interindividual variability ranging from 13.6 ~ 100 % at the same frequency condition was also observed in the present study. Sound-induced flash illusion most frequently occurred for all participants when auditory stimuli having base frequency were presented twice, compared that two auditory stimuli having a frequency difference were presented. Therefore, this result suggested that frequency congruency was an important parameter in sound-induced flash illusion. I hypothesized that the ability to recognize the frequency difference at auditory stimuli presented in a short period may affect the sound-induced flash illusion. The results in the frequency discrimination task showed the percentage at which participants perceived that the frequencies of the auditory stimuli presented twice were the same decreased as the frequency difference was large. The psychometric function of frequency discrimination ability and the occurrence of the illusion was fitted, and I estimated the sigma as indices and compared the sigma between frequency discrimination ability and the occurrence of the illusion. Contrary to my expectation, no correlation between the two. One interpretation is the frequency used by this study. In the present study, the tone burst of the base frequency (3000 Hz) and pure tone ranging from 1/2 octave to 2 octaves based on base frequency were presented, and non-base-frequency was determined by the physical quantity in increments of 1/6. Since most of the participants discriminated the difference between the base frequency and 2750 Hz and 3500 Hz (average ± SEM: 2750 Hz, 39.4 ± 6.8 %; 3500 Hz, 12.6% ± 4.45 %), the interindividual difference in frequency discrimination might have been not observed. The sensitivity of frequency discrimination to 2750 Hz was lower than that to 3500 Hz ($t_{(20)} = 3.79$, $p < 0.01$, paired t-test). Human sensory quantity is proportional to the logarithm of the intensity of the stimulus, according to

Weber-Fechner's law (Moore, 2012). In addition, the sensitivity to frequency is also proportional to logarithm of the frequency. According to Weber-Fechner, sensitivity to 2750 Hz should be lower than that to 3500 Hz based on 3000 Hz. Thus, the performance to low frequency was worse than high frequency. Future studies should examine the discriminability of frequency and the susceptibility of the flash illusion, as individual differences might be clarified by using stimulus with more difficulty to discriminate frequencies (e. g. 1/12 octave).

A direction for future research is to compare a musician having high temporal resolution and a non-musician. Bidelman (2016) revealed that musicians less perceived illusory flash, compared to non-musicians (Bidelman, 2016). In addition, musicians had a narrower temporal binding window and particularly, weak susceptibility to flash illusion at long SOAs (Bidelman, 2016). Bidelman (2016) explained that long-term musical training led to a refined temporal resolution in audiovisual integration, and reduced the sensitivity to the illusion. Similarly, musical training might lead to enhancement of frequency discrimination. In this study, most of the participants had no musical training, and the difference in the musical experience could not be evaluated. It will be necessary to compare musicians and non-musician to examine the sensitivity of frequency discrimination and the susceptibility to the illusion about frequency factor.

A neurophysiological study on sound-induced flash illusion showed the early interactions between auditory and visual cortical regions by an event-related potential recording (Mishra et al., 2007). In support of this study, direct connections between auditory and visual cortical areas were found in humans (Eckert et al., 2008). Watkins et al. (2006) reported that the blood oxygenation level-dependent (BOLD) signal in the primary visual cortex when sound-induced flash illusion occurred was similar to that when physical two flashes were presented in the functional magnetic resonance imaging study (Watkins et al., 2006). Therefore, direct projections from the primary auditory cortex to the primary visual cortex may modulate visual perception. In light of the present study, twice activation of the same neuron population might induce the illusory flash. The frequency information is encoded on the basilar membrane and retains until the primary auditory cortex. Therefore, individual neurons in the primary auditory cortex have a specific frequency tuning (Merzenich et al., 1975). When the auditory stimulus of the same frequency was presented twice, the neurons corresponding to the frequency of auditory stimulus were activated twice, and this activity could directly modulate the response in the primary visual cortex. Nevertheless, in future studies, for example, neurophysiological studies focused on the frequency are needed by using the animal model.

Chapter 4

Frequency congruency affects the sound-induced flash illusion

4.1 Introduction

The integration of multiple sensory information provides behavioral benefits, but receiving conflicting sensory stimuli produces perceptions that differ from the physical world. Sound-induced flash illusion, in which two brief flash accompanied by two sounds is often perceived as two flashes, is well known (Shams et al., 2000; Shams et al., 2002). In this illusion, auditory information alters visual perception. A previous study focused on auditory information demonstrated that the illusory flash was diminished when clearly different auditory stimuli like a combination of pure tone and noise were presented (Roseboom et al., 2013). Additionally, in Chapters 2 and 3, the occurrence of illusion decreased as the frequency difference between the first and second auditory stimulus increased. I investigated the frequency in detail since sound frequency obviously affects the sound-induced flash illusion. In this study, I examined the effect of whether or not the first and second auditory stimuli contained the same frequency on the sound-induced flash illusion. I would expect that the occurrence of illusion was higher when the noise, including the frequency of tone burst, and tone burst were presented than when the noise, without the frequency of tone burst, and tone burst were presented.

4.2 Materials and methods

4.2.1 Participants

In all, 15 adults (11 women and 4 men) ranging from 22-25 years old, participated in this experiment. All participants showed normal hearing and normal (or corrected normal) vision. All experiments were conducted in accordance with the guidelines for human experiments approved by the Ethics Committee of Doshisha University.

4.2.2 Stimuli

Almost all stimuli feature was the same as in my previous research (Chapter 2 and 3). The visual stimulus was a uniform white disk (55 mm diameter) displayed on a black background using a liquid crystal display (Foris FG2421, Eizo). A fixation point (6 mm white cross) was displayed in the center of the screen throughout the entire session, and the white disk was presented at 21 mm below the fixation point for 8 ms. The visual stimulus was presented once (single flash) or twice (double flashes). In the double flashes trial, the inter-stimulus interval (ISI) between flashes was set at 58 ms. A total of three auditory stimuli were used as follow: the brief tone burst (T; frequency: 3 kHz); the broadband noise (BN) ranging 1.5 ~ 6 kHz (covering ±1 octave from centered at the tone burst frequency); the notched noise (NN) in which ±1 octave of frequency centered at the tone burst was removed (frequency range: 0.75 ~ 1.5, 6 ~ 12 kHz). The combination for the auditory stimulus included 5 combinations as follows: tone-tone (TT), tone-broadband noise (T-BN), tone-notched noise (T-NN), and broadband noise-broadband noise (BN-BN). All auditory stimuli consisted of 20 ms duration with 5 ms rise/fall time and 80 dB SPL. The sound pressure level was measured and adjusted with a microphone (ER-7C Series b, Etymotic research). The auditory stimulus was always presented twice via headphones (SR-507, STAX). ISI between auditory stimuli was set at 48 ms. In all stimulus combinations, the first auditory stimulus always preceded the first flash by 40 ms. The refresh rate of the monitor was set at 120 Hz, and the duration of the visual stimulus and time profile of the stimuli were confirmed with a high-speed camera (EX-F1, Casio) and oscilloscope.

4.2.3 Procedure

The experimental procedure was the same as my previous research (Chapter 2 and 3). The participants sat in front of a monitor placed 60 cm from their faces, and the height of the monitor was adjusted with that the position of their eye was the same height as the fixation point. The task of participants was to count the number of perceiving flashes by pressing a button. The next trial automatically began after they responded. All experiments were controlled with the Presentation software (Neurobehavioral Systems) and were conducted in a sound-proof room (170 [W] × 150 [L] × 230 cm [H]). The noise level in sound-proof room was below 38.3 dB SPL.

4.3 Results

In the condition in which the first and second auditory stimuli were different, the flash illusion less occurred than in the same condition (Fig. 4.1). A one-way analysis of variance (ANOVA) with stimulus conditions was conducted. The results revealed a significant main effect ($F_{(3,56)} = 42.58$, $p < 0.001$). A multiple comparison showed that the occurrence of illusion for T-T and BN-BN as the same condition was significantly higher than that for T-BN and T-NN ($p < 0.001$, Tukey's test for post-hoc comparison) as the different condition. Compared with in the same condition, no significant difference was observed between T-T and BN-BN ($p = 0.99$, Tukey's test for post-hoc comparison). In addition, there was also no significant difference between T-BN and T-NN ($p = 0.76$, Tukey's test for post-hoc comparison), but the percentage of perceived two flashes for T-BN was higher than that for T-NN (average ± SEM: 16.7 ± 3.0 % vs 9.4 ± 3.0 %).

In the condition where the visual stimulus was presented twice, the same tendency as when the visual stimulus was presented once was observed (Fig. 4.1). A one-way ANOVA identified a significant effect of stimulus condition ($F_{(3,56)} = 27.49$, $p < 0.001$). Multiple comparison indicated that the percentage of two flashes for T-T and BN-BN was higher than that for T-BN and T-NN ($p < 0.001$, Tukey's test for post-hoc comparison). The difference between T-T and BN-BN conditions was not significant ($p = 0.99$, Tukey's test for post-hoc comparison), and the difference between T-BN and T-NN condition was also not observed ($p = 0.23$, Tukey's test for post-hoc comparison).

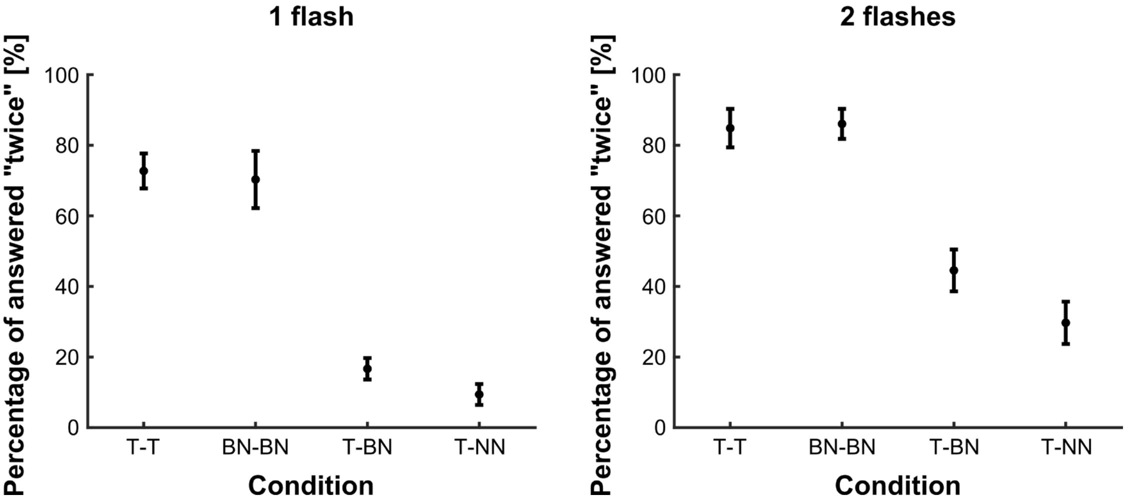

Figure 4.1. Mean proportion of two flashes perception in 1 flash (left panel) and 2 flashes condition (right panel).

T: tone burst, BN: broadband noise, NN: notched noise. The same auditory stimuli evoke the high two-flashes perception. Error bars represent the standard error of the mean.

4.4 Discussion

This study examined the effect of whether or not the first and second auditory stimuli contained the same frequency on the sound-induced flash illusion. The results showed that the illusion frequently occurred when the same auditory stimulus (e.g. BN-BN condition) was presented. This result replicated previous studies (Shams et al., 2000). Many previous studies used beeps around 3.5 kHz (Hirst et al., 2020; Keil, 2020), but auditory stimuli for inducing the illusion were not limited to the pure tone because my result revealed that even broadband noise induced the flash illusion and the susceptibility between pure tone and the noise was not different. In contrast, when the first and second auditory stimuli were not the same, perception of illusion less occur. This result was consistent with Roseboom et al. (2013) and my previous studies (Roseboom et al., 2013; Chapters 2 and 3). In Chapters 2 and 3, the occurrence of illusion reduced as the frequency difference between the first and second auditory stimuli increased. It is possible that frequency congruence affects the flash illusion.

In a neurophysiological study, Mishra et al. (2007) found that early interactions between auditory and visual cortical areas by event-related potential recording (Mishra et al., 2007). Furthermore, studies using electroencephalography (EEG) (Arden et al., 2003) and magnetoencephalography (MEG) (Shams et al., 2005b) also suggest the illusion is related to early interactions. Therefore, direct connections from the auditory cortex to the visual cortex were associated with the perception of sound-induced flash illusion. The sound frequency is encored on the basilar membrane, and the neurons in the primary auditory cortex have selectivity to certain sound frequencies (Merzenich et al., 1975). Therefore, two consecutive activations in the same neurons might trigger the perception of illusory flash. To investigate this hypothesis, the auditory stimuli, which were a combination of tone burst (T) and broadband noise (BN) containing the frequency of tone burst (3 kHz) or notched noise (NN) without frequency of tone burst were presented. The result showed that auditory stimuli of the combination including the same frequency frequently induced illusory flash than the combination not including the frequency of tone burst. these data supported the finding that the presentation of the same frequency was an important factor in the flash illusion. However, the occurrence of the illusion in even the auditory stimuli of the combination containing the same frequency was significantly lower than that in the auditory stimuli of the combination exactly the same sound. This result may be explained by the power of frequency in noise. The broadband noise used in this study was created such that the power of all frequencies was the same, and thus the power of frequency (3000 Hz) of broadband noise was much weaker than that of tone burst. It was possible that neurons activated by tone burst as the first auditory stimulus were not activated by broadband noise as the second auditory stimulus. Future studies need to investigate the frequency in detail using noise with varying power at specific frequencies.

In conclusion, my data suggests that sound-induced flash illusion is robust to characteristics of the auditory stimulus. Alternatively, the first and second sounds need to be the same frequency. Lower-order regions like the primary cortices may play a critical role in the illusion, but this remains to be investigated.

Chapter 5

Subjective audibility modulates the susceptibility of sound-induced flash illusion

5.1 Introduction

We utilize the information from multiple senses when we perceive the world. In most cases, the integration from the auditory and visual senses reduces the ambiguity of unimodal information and reconstructs the reliable perceptual world. It is found that sensory integration improves behavioral tasks such as stimulus detection (Bolognini et al., 2005; Frassinetti et al., 2002) and speech perception (Ross et al., 2007; Sumby and Pollack, 1954). However, in some cases, the perception of one sensory modality is affected by the other modality when we receive conflicting information from different modalities. This phenomenon is shown as an "illusion", and various illusions have been reported such as Ventriloquism effect (Jack and Thurlow, 1973), McGurk effect (Mcgurk and Macdonald, 1976), Stream-bounce effect (Sekuler et al., 1997). Recently, sound-induced flash illusion that the number of flashes is often perceived as two flashes when a brief flash accompanied by two brief sounds is presented is attracted the attention of many scientists because this illusion is induced by a simple task and is highly repeatable (Shams et al., 2000). It is proposed that sound-induced flash illusion follows the optimal rule of multisensory integration. In other words, the weight of sensory information depends on its reliability and precision, and the more reliable modality dominates the final perception depending on the situation. In the task involving temporal processing such as sound-induced flash illusion, the auditory modality becomes greater reliable because the auditory modality has a better temporal resolution than the visual modality (Welch and Warren, 1980). Therefore, the audition is relatively strongly weighted and tends to affect the final perception. Sound-induced flash illusion can be explained by a causal inference model that takes into account the relativity, prior expectation, and prior knowledge to determine the perception (Odegaard et al., 2016; Shams et al., 2005a). In addition, Shams et al. (2002) argue that discontinuous stimulus easily tends to alter the percept of continuous stimulus (Shams et al., 2002). Therefore, it is considered that sound-induced flash illusion is affected by the sensitivity to perceive the two auditory stimuli and two visual stimuli. Since hearing loss can lead to a higher threshold of gap detection, and the reliability of auditory senses reduces, the susceptibility of this illusion may decrease. However, because the effect of hearing loss on gap detection is mixed (similar: Harris et al., 2010; Moore and Glasberg, 1988; Moore et al., 1992, worse: Feng et al., 2010; Irwin and McAuley, 1987), no conclusion can be drawn, but at least hearing aid users have an improvement in the temporal resolution after habituating with the hearing aid (Pinheiro et al., 2012). Thus, the susceptibility to sound-induced flash illusion in hearing aid users is stronger than that in hearing-impaired non-user (Gieseler et al., 2018).

One of the important parameters well known to modulate the occurrence of sound-induced flash illusion is Stimulus Onset Asynchrony (SOA). In several studies focusing on SOA, SOA is defined as the lag between a flash-sound pair and another sound. Shams et al. (2002) found that SOA exceeding 70 ms reduced the susceptibility of the illusion (Shams et al., 2002), and many previous studies demonstrated that the flash illusion frequently occurred as the SOA was shorter (e.g. Hirst et al., 2020). This temporal effect was consistent with the temporal rule that integrates temporally close events into the same event included by the principles of multisensory integration (Meredith et al., 1987; Stein and Stanford, 2008). On the other hand, a shorter SOA indicates that the inter-stimulus interval between two auditory stimuli relatively becomes shorter, and since we cannot eventually discriminate two auditory stimuli, the relativity of the auditory modality decreases. As a result, perception of illusory flash is thought to reduce. However, it is hard to investigate this hypothesis by manipulating the SOA since the gap detection threshold of auditory stimulus was about 5 ms (Shailer and Moore, 1987). From this background, there is no systematical evidence of whether the audibility to detect the two sounds is associated with the sound-induced flash illusion. Therefore, I proposed the hypothesis that the audibility could be modulated by manipulating the sound pressure level.

The sound is masked by the presence of another sound in daily life, making the signal less audible. For example, when a weaker sound is presented right after a louder sound, the second sound cannot be perceived. This phenomenon is called forward masking, which means that the preceding sound raises the threshold for detecting the following sound (Moore, 2012). Specifically, as the sound pressure level of the second sound is low than that of the first sound, the detectability of the second sound is decreased, and then the susceptibility of sound-induced flash illusion reduces. In this study, I estimate whether the masking affects the perception of illusion in experiment I. In experiment II, I assessed the perception of two sounds and investigated whether the audibility was related to the susceptibility of the sound-induced flash illusion.

5.2 Materials and methods

5.2.1 Participants

Twenty adults (11 women and 9 men, 21-29 years old), in all, participated in this experiment, and each participant took part in only one of the two experiments. Ten adults (4 women and 6 men) participated in experiment I, and ten individuals (7 women and 3 men) participated in experiment II. All participants have normal hearing and normal (or corrected normal) vision. All experiments were conducted in accordance with the guidelines for human experiments approved by the Ethics Committee of Doshisha University.

5.2.2 Stimuli

The visual stimulus was a uniform white disk (55 mm diameter) displayed on a black background using a liquid crystal display (Foris FG2421, Eizo). A fixation point (6 mm white cross) was displayed in the center of the screen throughout the entire session, and the white disk was presented at 21 mm below the fixation point for 8 ms. The visual stimulus was presented in two ways: flashing once (single flash) or twice (double flashes). The inter-stimulus interval (ISI) between flashes was set at 58 ms in the trial in which double flashes were presented. The auditory stimulus was a brief tone burst at the frequency of 3 kHz, and the duration of auditory stimulus was 20 ms (the rise and fall time were 5 ms). The auditory stimulus was always presented twice via headphones (SR-507, STAX), and ISI between the tones was set at 48 ms. The sound pressure level of tone burst was different from experiments I to II. In experiment I, the sound pressure level within a range from 70 to 90 dB SPL was used. There were two conditions for the combination of sounds. In the same condition that the sound pressure level of the first and second sounds was the same, the intensity of sounds was 70, 80, and 90 dB SPL. In the different condition that the sound pressure level of the first auditory stimulus was different from that of the second sounds, the auditory stimulus of 80 dB SPL was always presented as a first tone burst, and the intensities of second sounds were 70, 73, 75, 78, 82, 85, 87, and 90 dB SPL. In experiment II, the sound pressure level was the range from 55 to 95 dB SPL. The participants were asked to perform a visual task in which they answered the number of visual stimuli and an auditory task in which they answered the number of auditory stimuli. Auditory stimulus having 75 dB SPL was always presented as a first tone burst, and the sound pressure level of second sounds was as follows: Visual task; 55, 59, 63, 67, 71, 75, 79, 83, 87, 91, and 95 dB SPL; Auditory task; 55, 63, 75, 87, and 95 dB SPL. In all stimulus combinations, the first tone burst always preceded the first flash by 40 ms. The refresh rate of the monitor was set at 120 Hz, and the duration of visual stimulus and time profile of the stimuli were confirmed with a high-speed camera (EX-F1, Casio) and oscilloscope. The sound pressure level was measured and adjusted with a microphone (ER-7C Series b, Etymotic research).

5.2.3 Task and procedure

The experimental procedure was the same as my previous research (Chapters 2-4). The participants sat in front of a monitor placed 60 cm from their faces, and the height of the monitor was adjusted such that the position of their eye was the same height as the fixation point. In experiment I, participants were instructed to state the number of flashes they perceived by pressing a button. After they had answered, the next trial began automatically after a 1 s interval. In experiment II, the participants were asked to state the number of flashes (Visual task) or sounds (Auditory task) they perceived. They could know the task in the response window for the first time. In other words, they were given the task after the stimuli had been presented. However, I instructed the participants to pay attention to the visual stimuli in order to examine the number of auditory stimuli perceived by participants while their attention was on the flashes. Visual and Auditory tasks were conducted in about 8:2. All experiments were controlled with the Presentation software (Neurobehavioral Systems) and were conducted in the sound-proof room (170 [W] × 150 [L] × 230 cm [H]). The noise level at the sound-proof room was below 38.3 dB SPL.

5.3 Results

5.3.1 Experiment I

When the sound pressure level of the first and second sound was the same, the average percentage of double-flashes perception in the trial of 90 dB SPL was higher than that in the trial of 70 dB SPL (42.7 ± 9.1 % vs 49.1 ± 9.0 % in 1 flash trial; 70.9 ± 8.3 % vs 80.9 ± 4.6 % in 2 flashes trial). However, An analysis of the variance (one-way ANOVA) did not identify a significant effect of the sound pressure level (1 flash trial, $F_{(2,27)} = 0.19$, $p = 0.83$; 2 flashes trial, $F_{(2,27)} = 0.70$, $p = 0.50$). In the different condition in which the sound pressure level of the first auditory stimulus was different from that of the second auditory stimulus, the illusion occurred less as the intensity of the second tone burst was lower than that of the first tone burst, and the illusion occurred more as the intensity of second tone burst was higher than that of first tone burst. A one-way

ANOVA on the number of perceived flashes revealed no significant effect of the sound pressure level ($F_{(8,89)} = 1.08$, $p = 0.38$). In the trial in which physical double flashes were presented, the average percentage of perceived double flashes was reduced as the intensity of the second sound was decreased, and there was a near significant effect of the intensity ($F_{(8,89)} = 1.93$, $p = 0.07$). In addition, participants often perceived double flashes in double flashes trials than single flash trials.

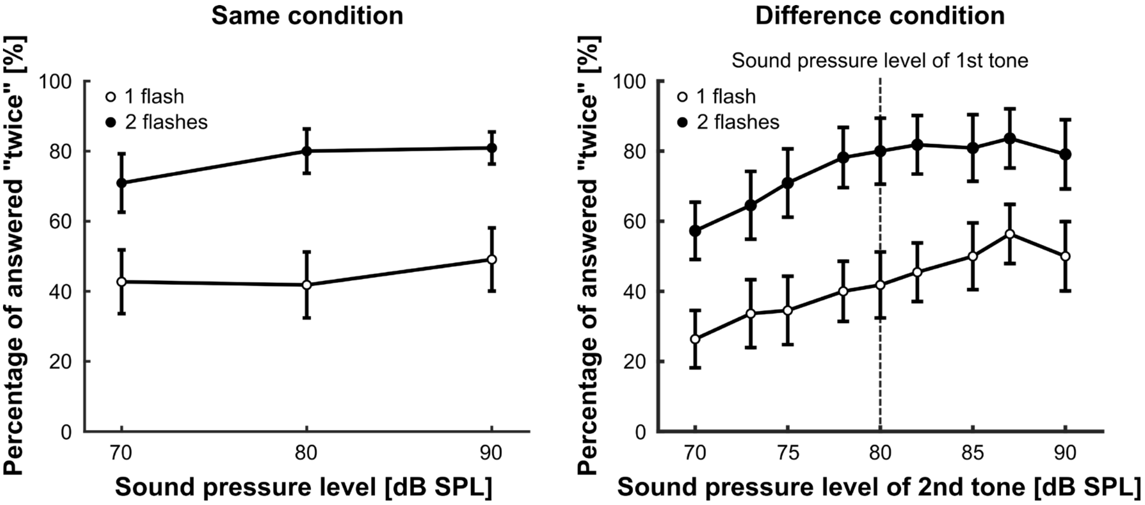

Figure 5.1. Mean proportion of two flashes perception in the same and different condition.

In same condition, the frequency of the first and second auditory stimulus was the same. Open circle indicates the results in 1 flash trial, and closed circle indicates the results in 2 flashes trial. In different condition, the intensity of first and second auditory stimulus was different, the sound pressure level of first tone burst was always 80 dB SPL. Error bars indicate ±S.E.M.

5.3.2 Experiment II

The rate of perception of two flashes in both trials in which the brief visual stimulus was presented once and trials in which the flash was presented twice was also reduced as the sound pressure level of the second auditory stimulus was low in the visual task. There was a clear significant effect of sound pressure level (1 flash trial: $F_{(10,99)} = 5.24$, $p < 0.001$; 2 flashes trial: $F_{(10,99)} = 1.92$, $p = 0.05$). In the auditory task, the rate of perception of two sounds decreased as the intensity of the second tone burst was low ($F_{(4,45)} = 8.03$, $p < 0.001$), and this result was consistent with the result in the visual task and Experiment I. In addition, I investigated the relationship between the occurrence of the illusion and the sensitivity to auditory stimuli. The percentage of the perception of two sounds significantly correlated with the occurrence of sound-induced flash illusion ($r = 0.66$, $p < 0.001$). Similar to the result of the trial in which a single flash was presented, a correlation was found between the rate of perception of two sounds and the rate of perception of two flashes ($r = 0.57$, $p < 0.001$).

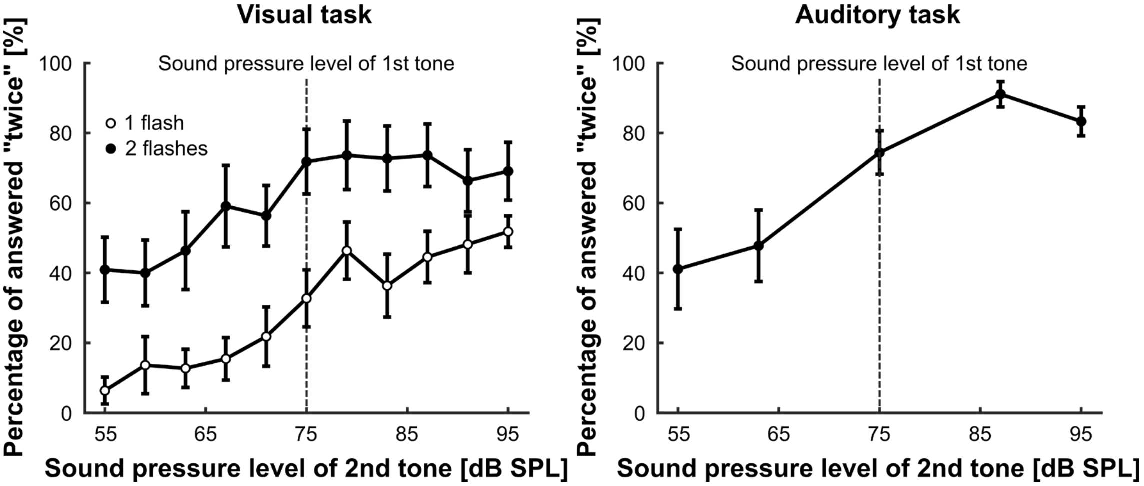

Figure 5.2. Mean proportion of two stimuli response in visual and auditory task. Two auditory stimuli in which have different sound pressure level were presented, the intensity of the first tone burst was always 75 dB SPL. As the sound pressure level of the second sound was low, the audibility perceiving two sounds decreased, and fusion frequently occurred but fission less occur. Error bars indicate ± S.E.M.

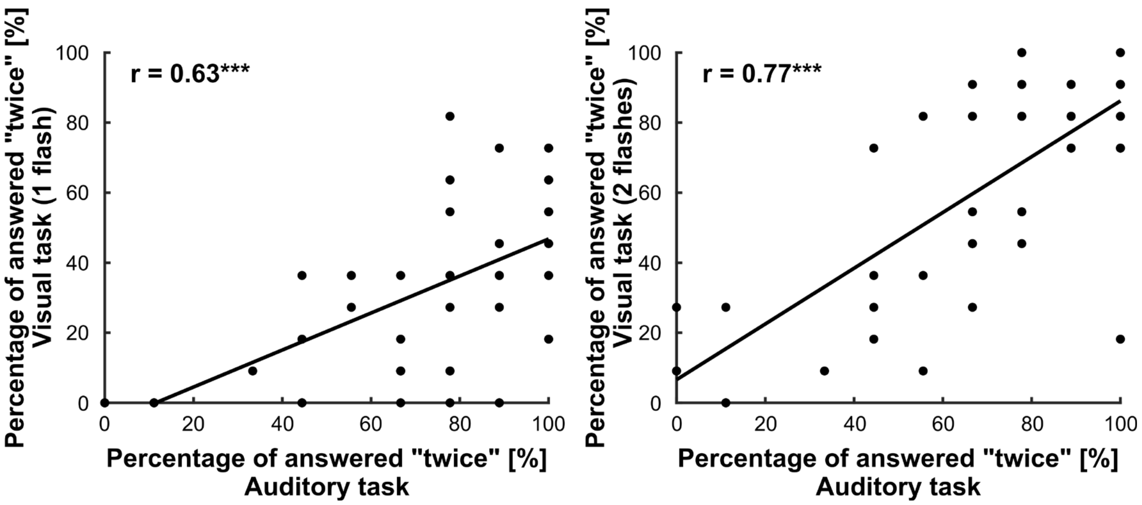

Figure 5.3. Correlations between the audibility and the susceptibility of illusions for 1 flash (left panel) and 2 flashes condition (right panel).

As the perception perceiving two auditory stimuli was higher, the two-flashes perception increased, and there was a correlation between them. ***: $p < 0.001$

5.4 Discussion

In this study, I examined whether the subjective perceptibility of the two sounds affects the occurrence of sound-induced flash illusion by varying the sound pressure level of the auditory stimuli. Experiment I revealed that it was difficult to induce the double flash illusion when the sound pressure level of the second auditory stimuli was lower than that of the first sound. By experiment II, it was found that this effect was due to forward masking, and the audibility of two sounds depended on the occurrence of the illusion.

When tone burst of high intensity was presented, the percentage of perceived two flashes was higher than when low-intensity stimuli were presented in the same condition. Andersen et al. (2004) showed that the illusion often occurred when auditory stimuli having sufficiently audible intensity were presented, but the illusion did not occur when the intensity of auditory stimuli was near the threshold (Andersen et al., 2004). My data was coincident with Andersen's results, but a significant effect was not found because the sound pressure level used in this study was of sufficient enough intensity that participants perceived two sounds. In future studies, I should examine this effect by systematically manipulating the sound pressure level.

In the different condition in experiment I, the sound pressure level of the second tone burst was not the same as that of the first tone burst. My data showed that the occurrence of illusory flash decreased as the intensity of the second tone burst was low. The shorter inter-stimulus interval (ISI) between consecutively presented two sounds, the less likely we are to notice the presence of the auditory stimulus presented by a second time (Moore and Glasberg, 1983). The amount of masking increases as the intensity of the first stimulus (masker) increases (Moore and Glasberg, 1983). Therefore, the second auditory stimuli were not subjectively perceived due to forward masking, and the occurrence of sound-induced flash illusion might be reduced. Forward masking occurs when ISI between auditory stimuli is within 200 ms (Durrant and Lovrinic, 1995), while masking is little seen when ISI exceeds 25 ms in backward masking (Durrant and Lovrinic, 1995). Since ISI between auditory stimuli used in this study were 48 ms, these results were probably affected by forward masking. However, it was unclear in experiment I whether the subjective audibility was altered. Accordingly, I conducted experiment II in order to confirm that the second sound was harder to perceive in my experimental procedure. Specifically, the auditory task in which participants answered the number of auditory stimuli was inserted into the visual task in which they answered the number of visual stimuli as similar to experiment I. I could evaluate the perceptual accuracy for auditory stimuli in the same task. Furthermore, I expected that the effect of masking would be more pronounced by increasing the difference of the sound pressure level between the first and second sounds.

In coincidence with the result of experiment I, as the sound pressure level of the second tone burst was weak, the susceptibility of illusion was low in the visual task. The auditory task was interleaved at a low probability to probe that this effect was due to the masking. As expected, the rate at which participants could perceive the two sounds decreased as the sound pressure level of the second sound decreased. In addition, as the perception of perceiving two auditory stimuli was high, the illusion more frequently occurred, and there was a correlation between them. Thus, these results indicated that forward masking reduced the audibility of the second auditory stimulus. In consequence, the illusion less occurred because the participants perceive only one sound even though the auditory stimuli were presented twice. These results suggest that the two sounds need to be perceived for seeing the illusory second flash.

Notably, in the double flashes trial in which visual stimulus was physically presented twice, the rate of two flashes perception also decreased as the intensity of the second auditory stimulus was low in both experiment I and II. Thus far, I focused on the fission illusion in which presenting one flash accompanied with two sounds induces the perception of two flashes. On the other hand, the previous study found fusion illusion in which the number of flashes was perceived as once when two flashes accompanied with one sound (Andersen et al., 2004; Shams et al., 2005a). It was possible that the decrease of two flashes perception was caused by fusion illusion because participants tended to perceive two sounds as one sound by forward masking. My data showed that the occurrence of both fission and fusion illusion was modulated by manipulating the sound pressure level of the second auditory stimulus.

It was found that sound-induced flash illusion was robust in the feature of visual stimulus from manipulating visual stimulus (e.g. Shape: Takeshima, 2020; Gabor patch: Takeshima and Gyoba, 2015, Block pattern: Takeshima and Gyoba, 2013; Gaussian probs: Apthorp et al., 2013; Faces: Setti and Chan, 2011). However, there are few studies investigating the effect of auditory stimulus features on this illusion although sound-induced flash illusion was induced by auditory modality. Roseboom et al. (2013) revealed that sound-induced flash illusion did not occur when two auditory stimuli had a clearly frequency difference were presented (Roseboom et al., 2013). In addition, I demonstrated that the susceptibility decreased as the frequency difference between two auditory stimuli increased (Chapters 2 and 3). These studies suggested that the similarity between two auditory stimuli influences the occurrence of illusory flash. Thus, as the similarity of the first and second sounds is low, the susceptibility to the illusion may be low. However, my results did not seem to partially support this conception. The changes in sound pressure level might diminish the sound characteristics. My data showed that fission illusion frequently occurred even though the similarity of two auditory stimuli decreased along with increasing the sound pressure level of the second tone burst and suggested that the subjective audibility of the sounds is a more important factor in the sound-induced flash illusion than the similarity of the sound.

In conclusion, the present study demonstrated that the subjective audibility was altered by manipulating the intensity of the sounds, and the sound-induced flash illusion (both fission and fusion) was affected by subjective audibility. Specifically, as the sound pressure level of the second sound was low, the audibility perceiving two sounds decreased, and fusion frequently occurred but fission less occur. These results were consistent with the results of the clinical studies of age-related hearing loss (Hirst et al., 2019) and hearing aid users (Gieseler et al., 2018), which showed that the susceptibility of fission was reduced as the audibility decreased, and expanded to young adults with healthy sensory abilities.

Chapter 6

Audiovisual integration enhances behavioral performance in head-fixed Mongolian gerbil: three principles revisited

6.1 Introduction

We are continuously exposed to various sensory information, such as auditory and visual information. We detect the stimuli of the external world which are filled with the information of these modalities and determine our behavior based on our perception of the outer environment. Sensory information is appropriately separated and integrated in the brain to reconstruct a coherent perceptual world, and provides an advantage in terms of perception and behavior.

For example, multisensory stimuli, such as auditory and visual stimuli, improve the accuracy and response speed (Frassinetti et al., 2002). In addition, the lip movements of the speaker enhanced comprehend speech perception in noisy environments (Sumby and Pollack, 1954). From these results, audiovisual processing plays an important role in our life.

Historically, the neural mechanism of multisensory integration has been investigated in the cat superior colliculus (SC) (Stein and Stanford, 2008). Meredith and Stein performed detailed single-unit recording in SC and discovered multisensory neuron which responded to stimuli from multiple senses (Meredith and Stein, 1983). Multisensory neurons provide a significantly different firing rate than in response to a unisensory stimulus when multisensory stimuli were presented. The response of the neuron to multisensory stimuli which is higher than the response to unisensory stimuli is defined as enhancement, while which is lower than the response to unisensory stimuli is defined as depression (Stein and Stanford, 2008). This multisensory modulation was found to be based on three principles. The multisensory stimuli must be presented in a spatiotemporally congruent manner for multisensory modulation. In other words, the modulation of the firing rate was maximum when the stimuli were presented from the same spatial location (Meredith and Stein, 1986a). Similarly, temporally proximate stimuli induced the modulation of neural response (Meredith and Stein, 1987). These are known as the spatial rule and temporal rule, respectively. Finally, the modulation of the firing rate was strong when the intensities of unisensory stimuli are weak (Meredith and Stein, 1986b), and this phenomenon is called by the principle of inverse effectiveness.

It is found that these three principles of sensory integration apply not only to single neurons but also to psychophysical paradigms. Frassinetti et al. (2002) demonstrated that the auditory stimuli with the presentation from the same location as visual stimuli enhanced detection performance of visual stimuli (Frassinetti et al., 2002). Bolognini et al. (2005) showed that the sensitivity of stimulus detection was improved when auditory and visual stimuli were simultaneously presented (Bolognini et al., 2005). Furthermore, the accuracy in the localization of auditory stimuli was enhanced by visual stimuli having low intensity (Bolognini et al., 2007). The behavioral performance in rodents was constrained by three principles, similar to humans (Hirokawa et al., 2008; Meijer et al., 2018; Sakata et al., 2004).

Recent rapid developments in neurophysiology and the use of molecular genetics and photogenetic manipulation have made it possible to understand detailed neural mechanisms at the level of neurons and networks in rodents. However, the mice and rats are nocturnal and the sensitivity of hearing is very different from that of humans. Here, I used Mongolian gerbil (*Meriones unguiculatus*) as an animal model for the audiovisual integration study. Gerbils have a wider audible range (0.1 ~ 60 kHz), and particularly the sensitivity of low frequencies between 1 and 4 kHz comparable to humans (Ryan, 1976). In addition, the gerbil's primary auditory cortex has a connection to non-auditory sensory areas (e.g. primary visual cortex and primary somatosensory cortex) (Budinger and Scheich, 2009; Henschke et al., 2015). Kobayasi et al. (2013) found that the multisensory neurons in the gerbil's primary auditory cortex (Kobayasi et al., 2013). The visual system in gerbils was adapted to a diurnal system since gerbils have the acuity of about 1.5 cycles per degree (Baker and Emerson, 1983). The gerbil's retina also has a well-developed cone system (Govardovskii et al., 1992; Jacobs and Deegan II, 1994), and i-wave and d-wave, found in human and other diurnal species, was observed in gerbil's electroretinogram (ERG) recordings (Yang et al., 2015). Furthermore, Nishiyama et al. (2011) found that gerbils modulate the occurrence frequency of vocalizations depending on the distance of another gerbil (Nishiyama et al., 2011). According to these behavioral and physiological features in auditory and visual senses, it is considered that gerbils were a suitable animal model in audiovisual research.

Despite these facts, the sensory integration of gerbils has not been assessed. The purpose of this study was to establish a setup in which three principles of sensory integration could be evaluated using head-fixed operant conditioning and to assess behavioral performance to audiovisual stimuli. I found that the gerbil's stimulus detection performance was constrained and enhanced by inverse effectiveness and temporal rule. A whole gerbil genome has been sequenced (Zorio et al., 2019) and optogenetic tools to gerbils have been applied recently (Brunk et al., 2019; Keplinger et al., 2018; Wrobel et al., 2018). Thus, the field of neuroscience in gerbils will be developed in future studies. My findings provide an important step to reveal the mechanism of audiovisual integration.

6.2 Materials and methods

6.2.1 Subjects

All experimental procedures were performed in accordance with guidelines established by the Ethics Review Committee of Doshisha University, and the experimental protocols were approved by the Animal Experimental Committee of the university. A total of 3 Mongolian gerbils, *Meriones unguiculatus*, (2 males and 1 female) aged 8 weeks at the beginning of the training were used in this study. Gerbils were bred and reared in our laboratory, and were maintained on a 12-h light/dark cycle, and were housed at 22-23 °C with approximately 50 % relative humidity. For behavioral experiments, gerbils were placed on a water restriction with free access to food. The body weight of gerbils was daily monitored and maintained within 80 % of their weight before restriction.

6.2.2 Surgery

Mongolian gerbils were anesthetized with 2 % isoflurane. I maintained anesthesia and body temperature throughout the surgical procedure. Their hair was shaved off, and the skin on the head of the gerbils was removed. The skull was exposed, cleaned, and dried. In order to fix animals' heads, a custom-designed titanium head plate was attached to the skull using acrylic dental resin (C&B Super-Bond, Sun Medical, Shiga, Japan) and dental cement (ADFA, Shofu, Kyoto, Japan). Gerbils were allowed to recover for at least a week after the surgical procedure.

6.2.3 Behavioral apparatus

All behavioral training and test were conducted under a dark condition in a soundproof box (60 [W] × 60 [L] × 60 cm [H]). Gerbils were inserted into a custom-made platform and the head plate was clamped by two stabilized clamps (PC2, Thorlabs, Japan). A spout (diameter: 0.90 mm) in order to present liquid reward was positioned in front of the mouth of animals, and the location of the spout was calibrated before the behavioral task started. A lick was detected by a contact lick meter (DCTS-10, Sankei Kizai, Tokyo, Japan) which was connected between the spout and copper sheet of the stage. 10 % sucrose solution was delivered through the tube under the control of a solenoid valve. Approximately 4μl was given as a reward when gerbils licked. The amount of reward was adjusted periodically. I delivered air-puff (15 psi) via the tips (diameter: 1.10 mm) located 3 cm from the gerbil’s left face as a punishment. Air-puff was also under the control of a solenoid valve. A white-light-emitting diode (LED) attached to a diffuser was placed about 13 cm from the right eye of gerbils (vertical axis is 0°, horizontal axis is about 60°).

Spatial congruency of auditory and visual stimuli is an important principle of audio-visual integration (for review, see Spence, 2013; Stein and Stanford, 2008). Therefore, a loudspeaker (FT28D, Fostex, Tokyo, Japan) was set next to the LED. In addition, another loudspeaker was set at the left side of gerbils to investigate “spatial rule”.

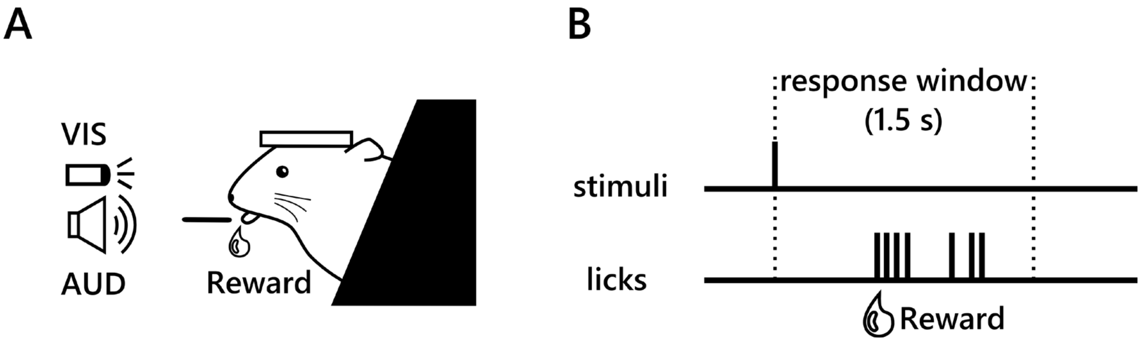

Figure 6.1. Experimental apparatus and behavioral task.
(A) Behavioral setup for head-fixed gerbils. VIS: visual stimulus, AUD: auditory stimulus.
(B) Task design. The response window is set to 1.5 s after stimulus offset, and reaction time is defined to the period between stimulus onset and the first lick.

6.2.4 Stimuli

The duration of auditory and visual stimuli was 500 ms at initial training, and finally was shortened to 10 ms. The visual stimulus was a brief flash and was presented by the white LED. The brightness of visual stimuli was measured with a lux meter (GL-03, be-s Co., Ltd., Tokyo, Japan), and was adjusted by modulating the voltage. A tone burst of 4 kHz was used as auditory stimuli. The sound pressure level was calibrated with a microphone (Type 1, ACO Pacific Inc., Aichi, Japan) placed at the location of the animal’s head. In a training session, the intensities of auditory and visual stimuli were 254 lx and 80 dB SPL, respectively. Several different stimulus parameters were used in each session (See 6.2.4 for details).

6.2.5 Behavioral training paradigm

In all tasks, Gerbils were instructed to detect the auditory and/or visual stimulus. Gerbils were water deprived for the day before the beginning of behavioral training. I give additional water to bring the total to 1ml when they received a liquid reward below 0.4 ml. The training stages were as follows:

Stage1: Habituation (2~3 days)

Gerbils were given 4 µl water when gerbils lick a spout. In this stage, auditory and visual stimuli were not presented. When gerbils voluntarily took a licking behavior and were given rewards more than 200 trials, they were advanced to the next stage.

Stage2: Classical conditioning (2~4 days)

In the second stage, gerbils learned the relationship between stimulus and reward by classical conditioning. Specifically, gerbils were trained to lick when they perceived auditory or visual stimuli. The duration of auditory and visual stimuli was 500 ms and presented a probability of 50% each, and the same modality stimulus was never presented more than three consecutive times to avoid bias from a specific stimulus. Regardless of whether or not gerbil's licking behavior to the stimulus was observed, they always received a water reward after stimulus presentation. The inter-trial interval was 5 ~ 9 s. The trial in which gerbils responded within 1.5 s after stimulus offset (i.e. response window) was defined with "Hit". Whereas, the trial was defined with "Miss" when gerbils failed to lick during the response window. In this session, the sound pressure level of the auditory stimulus was 80 dB SPL, and the intensity of the visual stimulus was 254 lx. Once the probability of Hit exceeded 80% in a session, gerbils were advanced to the next stage.

Stage3: Operant conditioning (7 days)

In the third stage, gerbils were trained to detect the stimulus to get the reward. Gerbils could only receive a water reward when they responded to the stimuli within the response window (Fig. 6.1). A no-stimulus trial (NoGo trial) in which the auditory or visual stimulus was not presented was interleaved. Air puff and timeout were presented as a punishment when gerbils licked the spout within the response window in no-stimulus trial (False Alarm: FA), while neither reward nor punishment was presented when gerbils did not lick in no-stimulus trial (Correct rejection: CR). Go and NoGo stimuli were presented at a ratio of 70 % and 30 % of a total trial, respectively. Gerbils were advanced to the next stage after 7 sessions of this training were completed. I calculated Hit and FA rates, and the discriminability (d’) to investigate the learning of animals.

Hit rate = number of Hit / (number of Hit + number of Miss)

FA rate = number of FA / (number of CR + number of FA)

d’ = Z’(Hit rate) – Z’(FA rate)

Higher d’ indicates a better detection performance. Z’ is the inverse of the cumulative normal distribution. I counted the sessions in which their performance was above criterion (Hit rate of auditory and visual stimuli > 80 % and FA rate < 30 %).

Stage4: Operant conditioning in short duration of the stimulus (~7 days)

In the final stage, the duration of the stimulus was shortened step by step from 500 ms to 10 ms. I gradually shortened the duration if Hit rate > 80 % and FA rate < 30 % in every 100 trials. Gerbils were advanced to the test session when they performed above this criterion to both auditory and visual stimuli for 2 consecutive sessions. Usually, gerbils completed this phase within 1 week. In all training phases, the day’s training was over when gerbils missed 10 consecutive to GO stimuli.

6.2.6 Test session

The test session consisted of 4 parts.

6.2.6.1 Measuring stimulus detection threshold

In this session, I examined the response rate of auditory and visual stimuli of various intensities to measure the perceptual sensitivity of each gerbil to the stimuli. The stimulus intensities were as follows:

Auditory stimulus: 20, 25, 30, 35, 40, 45, 50, 55, 60, 70, 80 dB SPL.

Visual stimulus: 3, 8, 10, 17, 31, 49, 103, 164, 180, 210, 254 lx.

In shortly, 11 different intensities of auditory and visual stimuli were presented, respectively. Except for the intensity of the stimulus, the behavioral paradigm was the same as in the final training session. Gerbils performed this test 3 ~ 5 times to examine the individual gerbils' threshold of stimulus more precisely.

6.2.6.2 Principle of inverse effectiveness

Auditory-only (AUD), visual-only (VIS), and audiovisual (simultaneous stimuli; AUDVIS) stimuli were presented in this phase, and classified into 4 bins based on the perceptual level of each gerbil: Sub-threshold (under 40 % response rate), around-threshold (40 ~ 65 %), Supra-threshold (65 ~ 90 %) and max intensity (the intensities used by training sessions). In summary, all trials consisted as follows: auditory-only trial (20 %), visual-only trial (20 %), audiovisual trial (20 %), no-stim trial (30 %), and max intensity trial (10 %). This test session was conducted 3 ~ 5 times.

6.2.6.3 Temporal rule

Audiovisual stimuli in which had a variable stimulus onset asynchrony (SOA) between the auditory and visual stimulus were presented in this phase. The SOAs were 0, ±40, ±80, ±120, ±160, ±200. Plus indicated that the auditory stimulus preceded the visual stimulus. The stimulus intensities were used supra-threshold in which detection rate was about 65 ~ 90 % based on 6.2.6.1 (see method) in fear of a loss of the motivation of gerbils. all trials consisted as follows: auditory-only trial (33 %), visual-only trial (33 %), audiovisual trial (33 %), and no-stim trial (30 %). This test session was also conducted 3 ~ 5 times.

6.2.6.4 Spatial rule

In this phase, auditory stimuli were presented from ipsilateral or contralateral to the location of visual stimuli. The stimuli were 5 patterns as follows in this phase: Visual-only (VIS), ipsilateral auditory-only (AUDi), contralateral auditory-only (AUDc), ipsilateral audiovisual (AUDVISi), contralateral audiovisual (AUDVISc). “i” indicated the ipsilateral, while “c” indicated the contralateral. As with the temporal rule session, the stimulus intensities were used supra-threshold. Ipsilateral stimuli and contralateral stimuli were presented at a ratio of 50 %. This test session was also conducted 3 ~ 5 times.

In all test sessions, the session in which the number of trials was less than 200 or FA rate was greater than 30 % were excluded from the analysis.

6.2.7 Psychometric curve and perceptual thresholds

To determine the threshold to stimulus, logistic regression was performed. The logistic function was fitted to maximize β0 and β1 using MATLAB *glmfit* function. The function was as follows:

$$f(x) = \frac{1}{1 + e^{-(\beta_0 + x \times \beta_1)}}$$

The variable x was assigned the stimulus intensity.

6.3 Results

Training session

Gerbils were trained in 4 steps before they could perform a task in the test phase. Table 1 showed the period taken of each stage. It took approximately two weeks to reach the final training stage. The behavioral performance at stage 3 was shown in Fig. 6.2 and Fig. 6.3. All gerbils showed stable performance within a few sessions and reached the criterion of stage 3 in 5-6 sessions. In stage 4, gerbils accurately detected the brief stimuli in which the duration was 10 ms in a few training sessions, and performed many trials in each behavioral session (765 ± 92 trials per session). The reaction times to auditory and visual stimuli for each gerbil were analyzed. The reaction time to the auditory stimulus was faster than that to visual stimulus in all gerbils (Fig. 6.4). The percentage of responses below 200 ms was 1.4 ± 0.3 % of all trials, and I excluded responses from the analysis because it was too fast for a response to a stimulus. Since a large individual difference in reaction time was found, z-scored reaction time was employed in future results.

Table 6.1. The sessions in which gerbils spent achieving a criterion in training phase.

Gerbils reached the test session within 3 weeks of training. 7 sessions of stage 3 were conducted in all gerbils, but gerbils reached the criterion within 7 sessions.

Gerbil Number	Stage Number				
	1	2	3	criterion	4
Gerbil 1	2	2	7	5	7
Gerbil 2	2	4	7	6	6
Gerbil 3	3	3	7	6	5

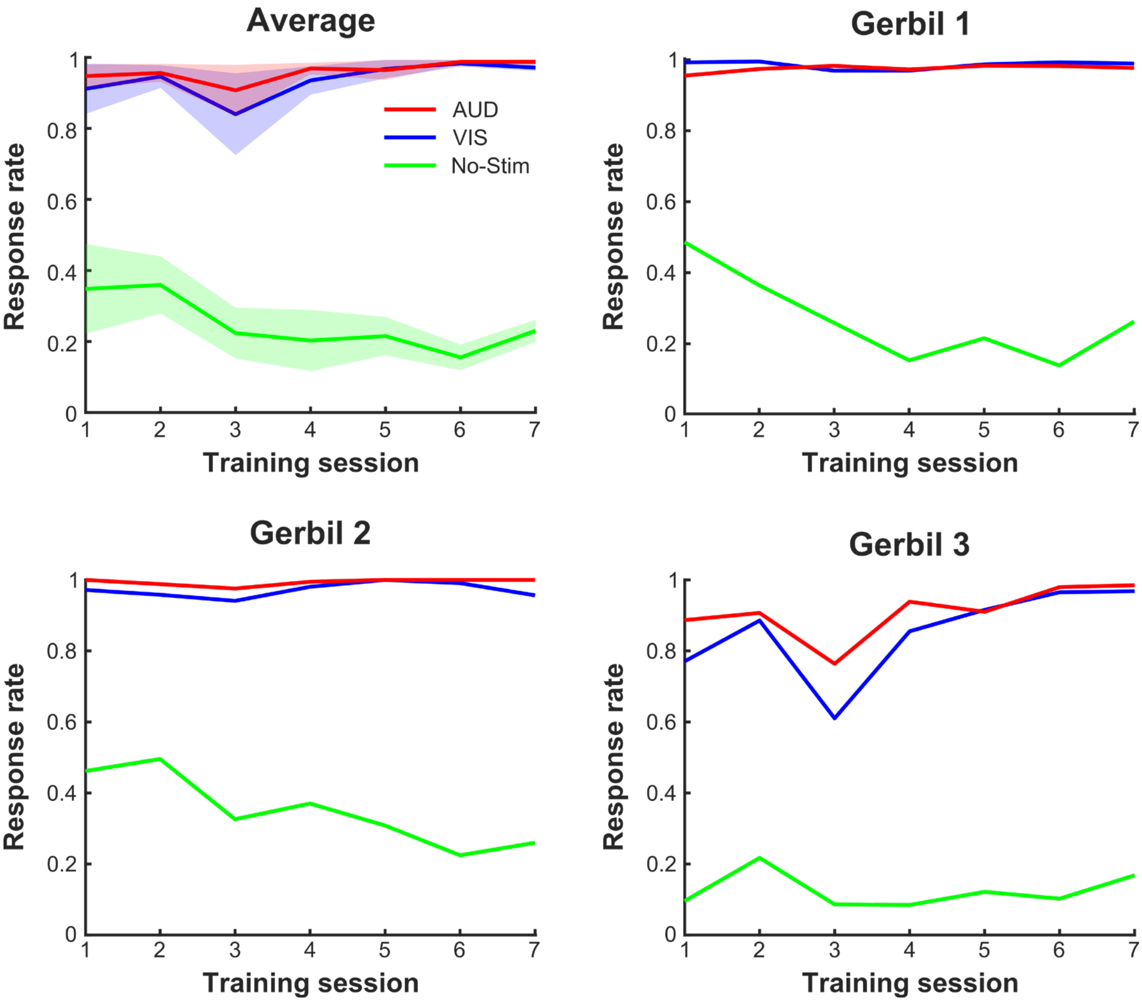

Fig. 6.2. Learning curve of behavioral performance for each gerbil in stage 3. AUD: auditory stimulus, VIS: visual stimulus, No-Stim: no-presentation of stimulus. Gerbils responded Go-stimulus such as AUD (red) and VIS (bule), while response rate of No-Stim condition (green) decreased as gerbils passed training session. Shading lines indicate SEM (n = 3).

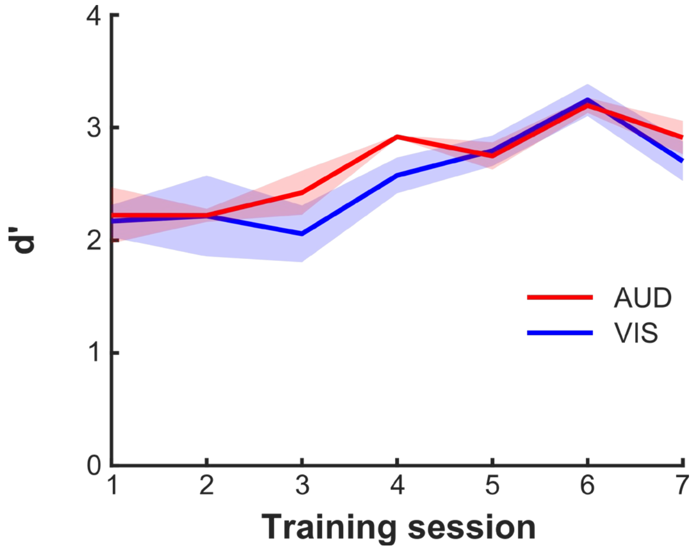

Fig. 6.3. Learning curve of d' in stage 3 of training session.

d' of both AUD (auditory stimulus: red line) and VIS (visual stimulus: blue line) increased as gerbils passed training session. Shading lines indicate SEM across subjects (n = 3).

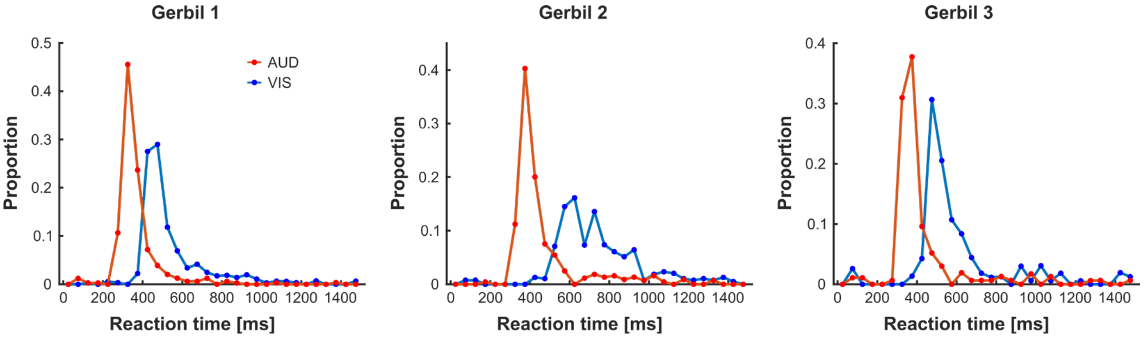

Fig. 6.4. Reaction time distribution in training session for each gerbil.

The reaction time to AUD (auditory stimulus: red) was faster than that to VIS (visual stimulus: blue line).

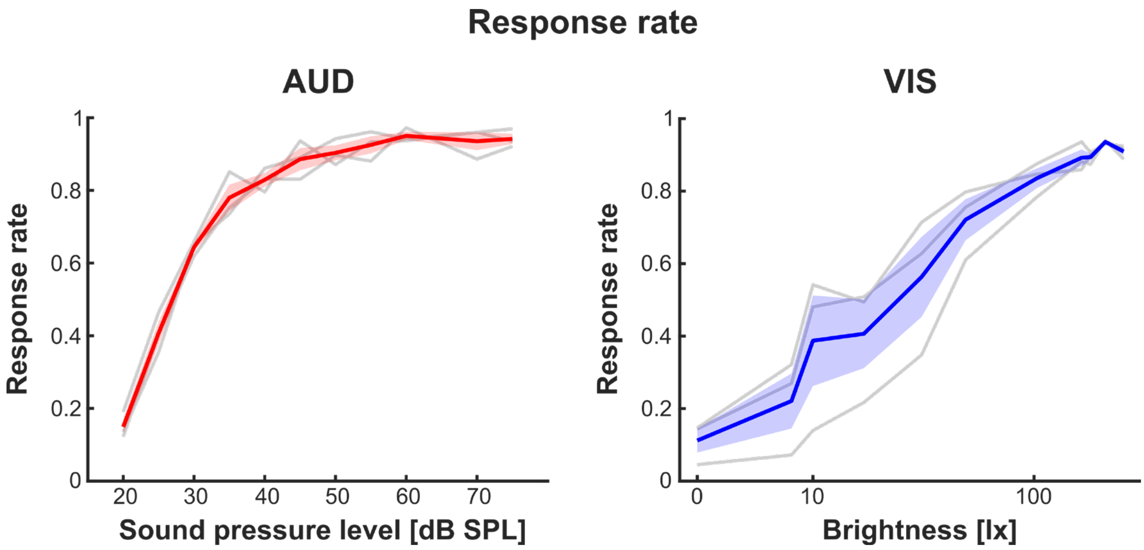

Fig. 6.5. Response rate in AUD (red) and VIS (blue) at different stimulus intensities. As the intensities of stimulus increased, response rate of both AUD and VIS increased. Shading line denote the SEM across subjects (n = 3).

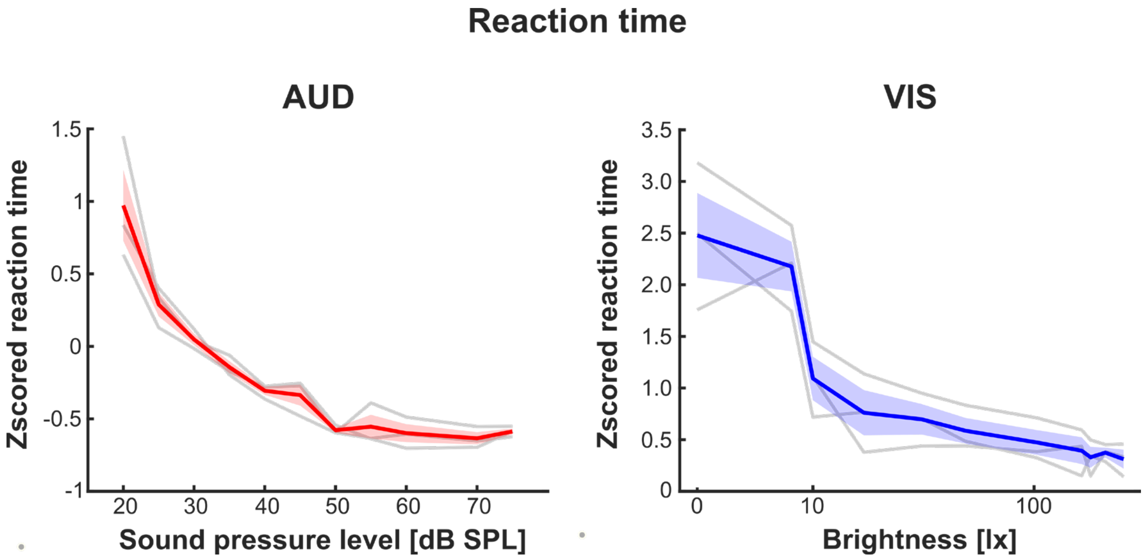

Fig. 6.6. Reaction time to AUD (red) and VIS (blue) at different stimulus intensities. As the intensities of stimulus increased, reaction time to both AUD and VIS shortened. Shading line denote the SEM across subjects (n = 3).

Measuring stimulus detection threshold

The response rate of both auditory and visual stimuli increased as stimulus intensity increased ($p < 0.001$ in all gerbils). In the fitted psychometric function, the threshold for the auditory stimulus was 27.5 ± 0.7 dB SPL, and the visual perceptual threshold was 49.1 ± 14.0 lx (n = 3). The reaction time decreased to both auditory and visual stimuli decreased as stimulus intensity increased ($p < 0.01$ in gerbil2 and gerbil3). The reaction time to auditory stimuli was faster than that to the visual stimuli, and even the reaction time at the maximum intensity of the visual stimuli was comparable to the reaction time when the auditory stimulus of 25 dB SPL was presented (average ± SEM: 0.31± 0.09 vs 0.29 ± 0.08 in Zscored reaction time, n = 3).

Inverse effectiveness

The stimulus intensities which were appropriate for each gerbil obtained in the measuring detection stimulus session were presented. Similar to the results in the measuring detection threshold session, all gerbils showed a higher response rate and shorter reaction time as stimulus intensity increased ($p < 0.01$ in gerbil1 and gerbil2) regardless of the stimulus type. In all stimulus intensities, the response rate to audiovisual stimulus was higher than that to auditory or visual stimulus in all gerbils. Specifically, gerbil1 significantly responded to audiovisual stimulus in the intensity of sub-threshold ($p < 0.05$), around-threshold ($p < 0.01$), and supra-threshold ($p < 0.05$), compared to best unisensory stimuli. Gerbil2 showed a higher response rate to audiovisual stimulus when the intensities were around-threshold ($p < 0.01$). In gerbil3, audiovisual stimulus of the intensity of only sub-threshold induced a higher response rate ($p < 0.01$). All three gerbils consistently slower responded to visual stimulus than that to auditory and audiovisual stimulus at all intensities (*p* < 0.01 except sub-threshold). In gerbil1 and gerbil2, the reaction time to audiovisual stimulus was faster than that to auditory stimulus. On the other hand, reaction time to audiovisual stimulus in gerbil3 was that to auditory stimulus at only max intensity. The false alarm rate was less than 10 % on average (9.4 ± 1.6 %), and gerbils performed the task reliably.

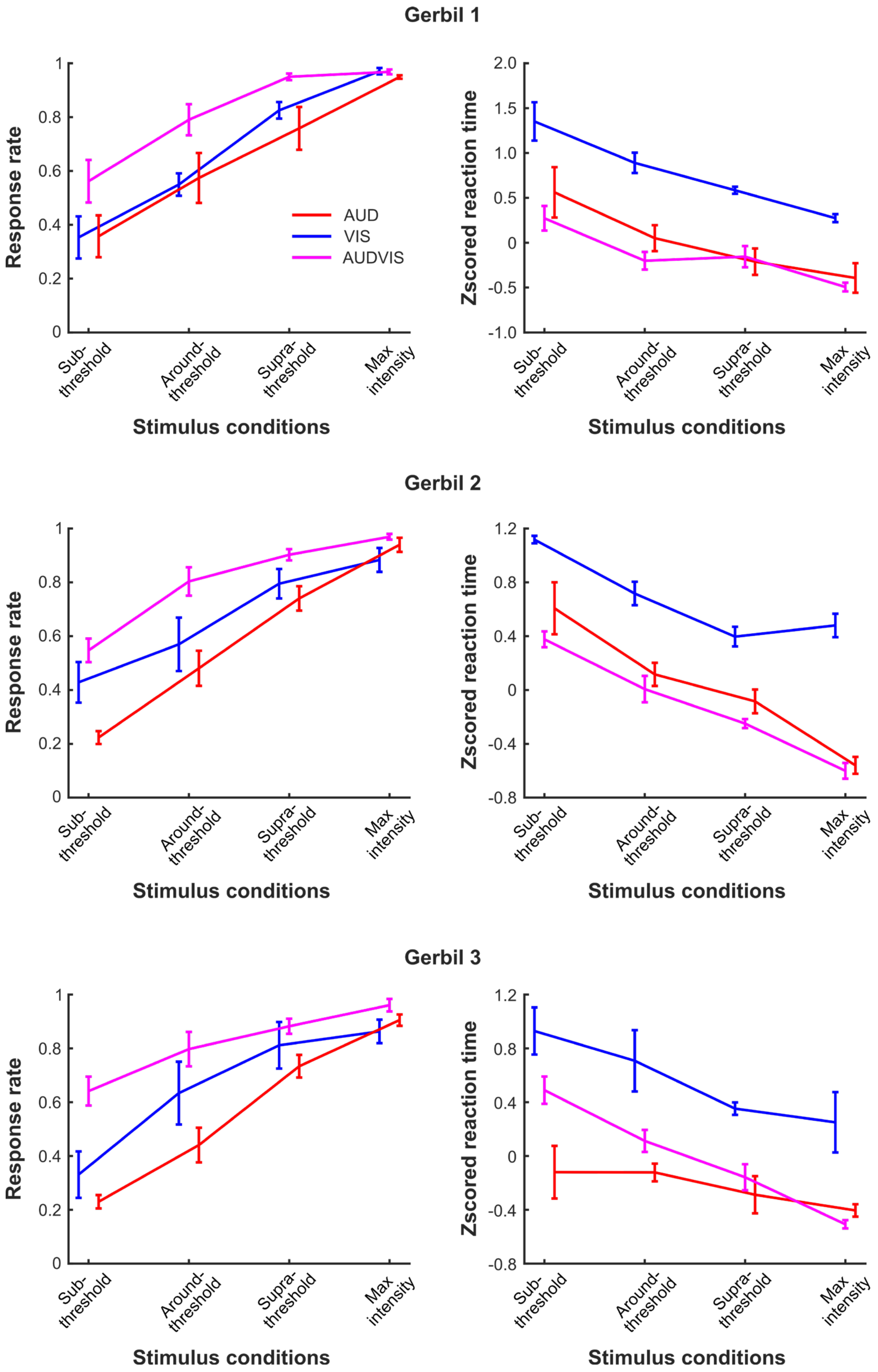

Fig. 6.7. Behavioral performance in the principle of inverse effectiveness session. Response rate of each gerbil is shown in left panel. Reaction time of each gerbil is shown in right panel. Error bars denote within-subject SEM.

Temporal rule

All gerbils showed a higher response rate to audiovisual stimulus compared to that to unimodal stimulus, but there was no significant difference along SOAs at gerbil1 and gerbil3. The reaction time in the condition in which auditory and visual stimuli were synchronized (SOA = 0 ms) was the fastest in gerbil1 (-0.08 vs -0.32) and gerbil3 (-0.24 vs -0.28). In contrast, the reaction time of gerbil2 to auditory stimulus was faster than that to SOA = 0 (-0.34 vs -0.28), but a significant difference was not observed (p = 1.00). As the SOA increased, the reaction time tends to become slower. The false alarm rate was less than 10 % on average (7.2 ± 0.7 %).

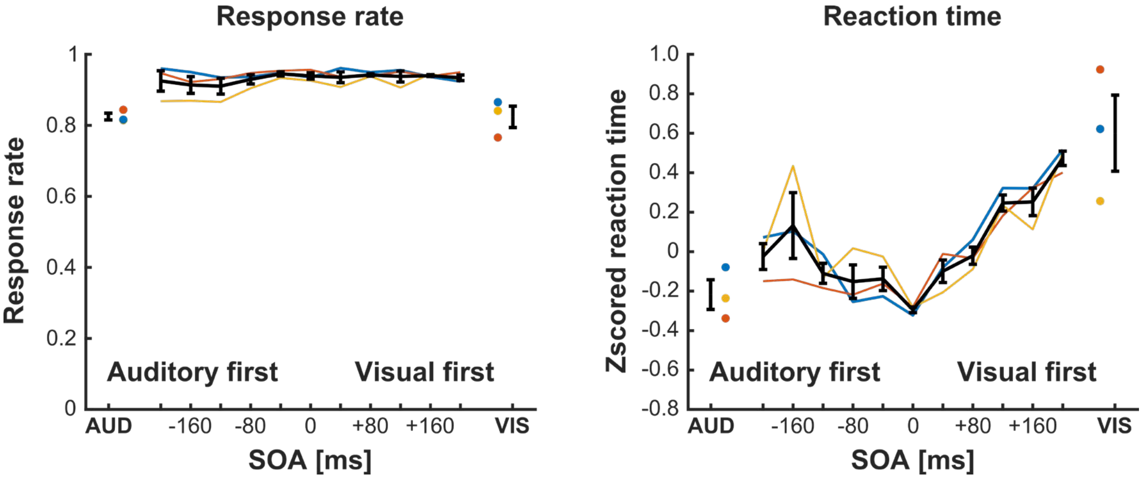

Fig. 6.8. Behavioral performance in the temporal rule session.

Response rate is shown in left panel, and reaction time is shown in right panel. Black line indicates average data, and colored line (red, bule, yellow) represents the data of each gerbil. Plus indicated that auditory stimulus preceded visual stimulus. AUD: auditory stimulus, VIS: visual stimulus. Error bars denote across animal SEM (n = 3).

Spatial rule

The response rate to audiovisual stimulus was higher than that to auditory-only or visual-only stimulus, but there was no significant difference on detection performance between unisensory stimulus and multisensory stimulus at gerbil1 and gerbil3 ($p = 0.99$). In gerbil2, a significant difference between unisensory and multisensory stimuli ($p < 0.05$), except between AUDc and AUDVISc ($p = 0.96$). The difference between ipsilateral and contralateral audiovisual condition was not observed in all gerbils ($p > 0.05$ in all gerbils). Reaction time to visual stimulus was slower than that to auditory and audiovisual stimulus ($p < 0.01$ in gerbil1 and gerbil2). It was not found spatial influence on reaction time in all gerbils ($p > 0.05$). The false alarm rate was less than 10 % on average (3.6 ± 1.1 %).

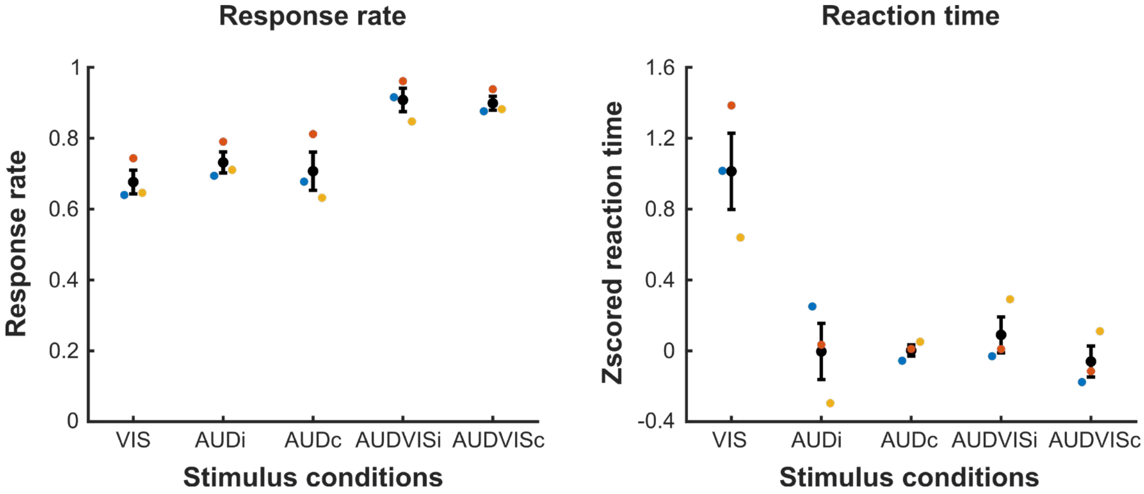

Fig. 6.9. Behavioral performance in the spatial rule session.
Response rate is shown in left panel, and reaction time is shown in right panel. Black line indicates average data, and colored plot (red, bule, yellow) represents the data of each gerbil. AUD: auditory stimulus, VIS: visual stimulus, AUDVIS: audiovisual stimulus. 'i' indicates the ipsilateral stimuli. 'c' indicates that contralateral stimuli. Error bars denote across animal SEM (n = 3).

6.4 Discussion

For Mongolian gerbils, I developed an experimental setup for head-fixed operant conditioning with water as a reward. The water reward allows the carrying out of a large number of trials (765 ± 92 trials per session) than the pellet reward (150 ~ 200 trials per day) (Kobayasi et al., 2012; Carney et al., 2011; Yao et al., 2020), and this system makes it possible to obtain sufficient data for psychophysical evaluation in more stimulus conditions. In addition, the setup I developed was able to evaluate three principles of sensory integration in the same experimental environment. Spatiotemporal congruency and intensities of auditory and visual stimuli were important factors in sensory integration (Stein and Stanford, 2008). The intensities and temporal information of stimuli can be easily controlled, but the control of spatial information is hard because the receptive position must be fixed. In the present setup, since the gerbil's head was fixed, the auditory and visual organs always receive relevant stimuli in the same spatial arrangement. Therefore, the fixation of the receptive position of the stimuli could strictly control the audiovisual integration. The present study is the first study to evaluate behavioral performance of gerbils in audiovisual integration.

Inverse effectiveness

I found that the stimulus detection performance was improved by the presentation of audiovisual stimulus. Although individual differences were observed, the effect of audiovisual integration was obtained when auditory and visual stimulus intensities were between subjective sub-threshold and supra-threshold. In particular, these results were consistent with the principle of inverse effectiveness where a behavioral gain was obtained as the intensity of unimodal stimulus decreased. Therefore, my data was similar to those of mice and rats evaluated with the simple stimulus detection paradigm (Meijer et al., 2018) and the discrimination paradigm (Gleiss and Kayser, 2012; Siemann et al., 2015). However, my results differed in part from mice studies in the simple stimulus detection task. Meijer et al. (2018) reported that a behavioral gain was diminished when auditory and visual stimulus intensities were below the perceptual threshold (i.e. sub-threshold) (Meijer et al., 2018). This result may be explained by the difference of experimental conditions. Because I set the stimulus intensity based on the psychological curve of each gerbil measured beforehand in measuring detection threshold session, the perceptual threshold might fluctuate depending on the condition of the day (e.g., physical condition and waking up from sleep). On the other hand, as Meijer et al. (2018) adopted the Bayesian staircase procedure, the stimulus intensities could be calibrated according to the individual perception rate during 1 session, and the perceptual threshold according to the condition of the day was evaluated. In addition, in mice studies, it was found that the maximum behavioral benefits were obtained when the intensities of each stimulus were around the perceptual threshold (Meijer et al., 2018; Siemann et al., 2015). In my data, two gerbils showed the greatest behavioral gain when audiovisual stimulus having the intensity of individual perceptual threshold was presented, suggesting that gerbils' behavior enhanced in accordance with principles of inverse effectiveness. In the future, I also need to study the sub-threshold performance by using the adaptive staircase method.

The gerbil's reaction time to visual stimulus was consistently slower than that to auditory and audiovisual stimulus. The velocity at which stimulus reaches the receptors is faster for light than for sound, while auditory conduction velocity is processed faster than visual, and the response in the primary sensory cortex is also faster in auditory stimulus than visual stimulus (Inui and Kakigi, 2006; Inui et al., 2006). Additionally, the reaction time to auditory stimulus was faster than that to visual stimulus in simple stimulus detection task in human studies (Diederich and Colonius, 2004; Hansonn et al., 2009). Therefore, the results of my study are considered to be appropriate. Gerbil1 and gerbil2 showed faster responses to audiovisual stimulus, compared to auditory stimulus. In contrast, reaction time to audiovisual stimulus in gerbil3 was slower than that to auditory stimulus except for max intensity condition. Previous human studies in audiovisual integration have shown that reaction time was shortened when the multisensory stimulus was presented (Diederich and Colonius, 2004; Hershenson, 1962). In rodents study, head-fixed mice did not show audiovisual facilitation of the reaction time in simple stimulus detection task (Meijer et al., 2018), while rats showed audiovisual facilitation of reaction time in stimulus location detection task using a two-alternative forced choice (2AFC) (Gleiss and Kayser, 2012; Hirokawa et al., 2008; Sakata et al., 2004). The present study was used a simple stimulus detection task under the head fixation condition, and two gerbils showed faster reaction time, but not significantly, to audiovisual stimulus. These mixed results may be due to differences in the subjects and experimental methods. Future studies are needed to investigate this difference in detail.

Temporal rule

The temporal rule, one of the principles of sensory integration, enhanced the behavioral response to temporally congruent audiovisual stimuli. Thus, I presented audiovisual stimuli with a variety of SOAs. Two out of three gerbils responded most quickly to audiovisual stimulus that the stimulus onset was synchrony. In human studies, the reaction time decreased as SOA was shortened (Bazilinskyy and Winter, 2018). In rodent study, Hirokawa et al. (2008) showed that the rats' reaction time was the fastest when the simultaneous audiovisual stimulus was presented (Hirokawa et al., 2008). My results were consistent with these previous studies since reaction time was shortened as SOA became shorter in all gerbils, although there was no significant difference between reaction times to audiovisual and auditory stimuli. Therefore, it is possible that behavioral performance of gerbils was facilitated based on temporal rule.

Spatial rule

Spatial coincidence (spatial rule) is one of the basic principles of sensory integration. According to recording single neuron, the modulation of the firing rate in multisensory neuron was maximal when each unisensory stimulus was presented from the same location in space (Stein and Stanford, 2008). The behavioral study also showed that spatially congruent audiovisual stimuli enhanced behavioral performance in humans (Bolognini et al., 2005) and rats (Sakata et al., 2004). My data showed no difference between spatially congruent and incongruent audiovisual stimuli in either response rate or reaction time. In particular, the effects of spatial coincidence on behavioral performance remain mixed in human studies. Spence (2013) explained that audiovisual integration was facilitated for the task to which space was related, but not for the task which does not require spatial attention (Spence, 2013). In other words, the spatial rule is a task-depended phenomenon. Since the paradigm of the present study did not require spatial information, the spatial rule might be not applied. Human studies in a simple stimulus detection task like my paradigms demonstrated that spatially coincidence did not influence the behavioral performance (Teder-Sälejärvi et al., 2005), and was consistent with my data. The spatial rule will be examined in detail by conducting the 2AFC for stimulus spatial location detection task that requires attention to the spatial location of the stimulus (Sakata et al., 2004).

The present study is the first to evaluate the behavioral performance of audiovisual integration in Mongolian gerbils. My data demonstrated that gerbils improved their behavioral performance by audiovisual stimuli. In the simple stimulus detection task, their behavior was enhanced according to the principle of inverse effectiveness and temporal rule, while their behavior was not influenced by spatial rule. Head-fixed operant conditioning is a suitable setup for such multisensory studies of spatiotemporal coincidence because the stimulus source is unified, and it can also be combined with neurophysiological methods. Studies combining behavioral and neurophysiological methods have shown that the secondary visual cortex is involved in audiovisual integration (Hirokawa et al., 2008; Meijer et al., 2020). Since I established a setup to evaluate the three principles of multisensory integration in gerbils, future studies combining behavioral and neurophysiological methods are expected to elucidate the neural mechanisms of audiovisual integration in each principle of audiovisual integration.

Chapter 7

Application of the novel object recognition paradigm to time-varying sensory stimulus

7.1 Introduction

Rats approach a novel object more frequently and spend more time exploring it than they do an object to which they have been previously exposed (Berlyne, 1950). Such novelty-seeking behavior has been widely used in the novel object recognition (NOR) task to investigate perception and recognition of many rodents species. Because the NOR is based on animals' innate preference for novelty, this method does not require laborious prior training such as operant conditioning and does not reinforce an ability associated with the training. Therefore, NOR paradigm is capable of studying animals' relatively natural or innate perception and memory. Previous NOR task tested rather static stimulus such as characteristics of object (shape, size, color, etc.) or position of the object (Antunes and Biala, 2012; Winters et al., 2008). Thus, it is not known whether change in temporal dynamics motivate novel seeking behavior to the stimulus.

The purpose of this study was to develop a version of the NOR task that is applicable to the time-varying stimulus. Mongolian gerbil (*Meriones unguiculatus*), one of the standard laboratory rodent species, was chosen for my demonstration of the novel recognition paradigm. The gerbil has relatively good eyesight because the species is not entirely nocturnal (Pietrewicz et al., 1982). Here I demonstrated that NOR paradigm using a light-emitting diode as the visual stimulus was effective to evaluate animal's perception of time-varying sensory stimulus.

7.2 Materials and methods

7.2.1 Subjects

Mongolian gerbils (n = 11), which were bred and reared in our laboratory, were used in this study. Each animal was housed in a home cage with two to five other gerbils. Water and food pellets were available ad libitum.

All experimental procedures were performed in accordance with guidelines established by the Ethics Review Committee of Doshisha University, and the experimental protocols were approved by the Animal Experimental Committee of Doshisha University.

7.2.2 Experimental apparatus

A behavioral test was conducted in a square arena (45 cm (W) × 45 cm (L) × 55 cm (H)) located in a sound-proof room. Three kinds of the object representing different shapes (rectangular, four-sided pyramid, and rectangular with a dumbbell shape) were constructed using black stainless mesh sheets. Two different objects were placed in assigned positions

7.2.3 Stimuli

Visual stimulus (10 ms flashing light) was presented from the white light-emitting diode (LED) embedded inside of an object. Single flashes were presented repeatedly in familiarization sessions with an inter-onset interval (IOI) of 1010 ms from each object.

The stimulus-onset asynchrony (SOA) between objects was 505 ms. In test session, paired flashes with 160 ms IOI (double flash) were presented from one of the objects. The temporal center of the double flash and single flash had 505 ms SOA (Fig. 7.1).

7.2.4 Experimental procedure

Each animal was allowed to explore the arena, which was free of objects and stimuli for 30 minutes per day for 4 days as a habituation session before behavioral testing. Then animals underwent a familiarization session consisting of 5 trials, where the animal was familiarized to the visual stimulus presented from LED. After the familiarization, I changed the flash patterns of one of the objects (test trial, Fig. 7.1). Each trial lasted 5 minutes, and the inter-trial interval was 2 minutes. The contact duration was defined as the total time during which the subject was touching the object with its snout or forepaw within the first 90 seconds of each trial, and was measured by counting the number of flames showing the exploratory behavior on recorded movies.

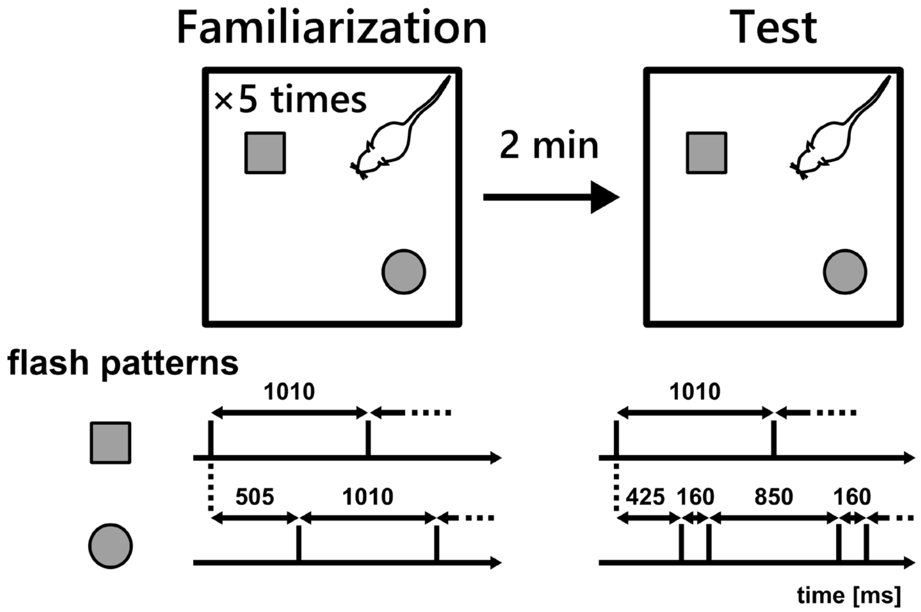

Figure 7.1. Experimental schedule and stimulus.

Object shape (rectangle, dumbbell, and pyramid) and location were counterbalanced among animals. In familiarization session, single flash was presented from both objects. The stimulus inter-onset interval was 1010 ms, and stimulus onset asynchrony between the objects was 505 ms. In test session, double flashes were presented from one of the objects instead of single flash.

7.3 Results

In familiarization session, animals were exposed to the repetition of single flashes from two objects, and contact duration decreased as the trial proceeded (1st to 5th session of Fig. 7.2). After the number of flashes with one of the object changed from single to double (in test session), exploration of that object increased. Contact duration of novel-flash object is significantly longer than that of familiarized flash object (n = 11, $p < 0.05$ with one tailed t test).

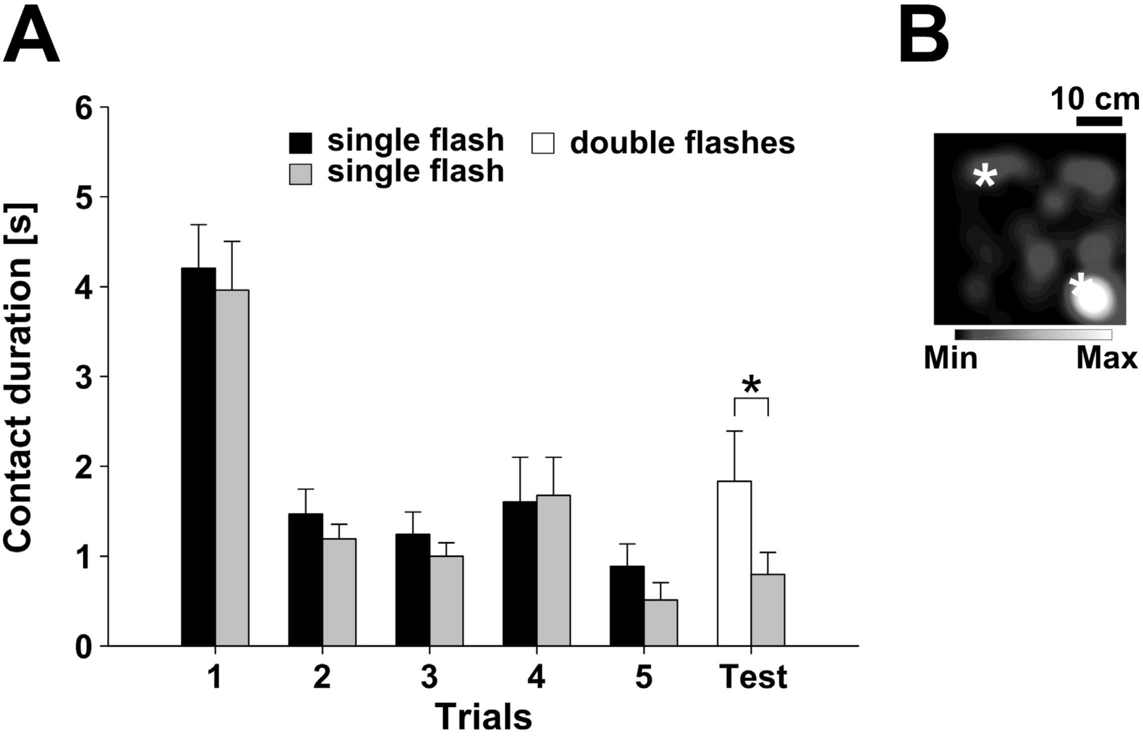

Figure 7.2. Increase in exploration of the time-varying stimulus.

(A) Series of bar graphs show the relationship between stimuli and behavioral responses (n = 11). Paired bars indicated that stimuli were presented simultaneously. Error bars indicate standard error. *p < 0.05. (B) Example of an occupation map during a test session. Positions of the objects are denoted by asterisks; novel and familiarized flash patterns were produced from the lower right and upper left objects, respectively. Scale bar shows 10 cm. Gray scale bar indicates staying time.

7.4 Discussion

Gerbils showed a decrease in contact time for both objects as a result of the familiarization process (Fig. 7.2). That was commonly observed in a typical object recognition paradigm. When the temporal pattern of a flashing object was altered, the animals explored the temporally novel object more than the other familiarized object. These results indicated that gerbils perceived the difference of the temporal pattern in visual stimuli. The difference of contact duration, however, was marginally significant. Further data (i.e., more subjects) are needed to draw more definite conclusions.

My data shows that time-varying visual stimulus could attract the interest of gerbils, and therefore promote the exploration of the object. This study suggests that the NOR paradigm could be used to study animal's perception of not only the static characteristic but also time-varying features of stimulus. The result also demonstrated that the NOR paradigm may have greater potential than previously assumed. One of the useful expansions is applying the NOR paradigm to auditory modality because the auditory stimulus is inherently temporally dynamic. In all, this study will provide a useful opportunity to investigate the animal's perception of temporally-dynamic stimulus, as conventional NOR paradigm have provided (Cohen and Stackman Jr., 2015).

Chapter 8

Auditory-induced visual illusions in rodents measured by spontaneous behavioral response

8.1 Introduction

When two sensory modalities receive conflicting information simultaneously, the perception in one modality is sometimes modified to align with the information in the other modality to construct a coherent multi-modal percept (Jack and Thurlow, 1973). A prominent example, the “McGurk effect” (or “McGurk–MacDonald illusion”), demonstrates that listening to the sound /ba/ with a video clip showing a person’s lip uttering /ga/ often results in a combined auditory perception, such as “da” (Mcgurk and Macdonald, 1976). That audio-visual integration shows how significantly visual information (i.e., articulatory movement) contributes to auditory speech perception. More recently, Shams et al. (2000; 2002) reported the “sound-induced flash illusion”, which demonstrates that the opposite interaction (i.e., an auditory modality altering a visual perception) can also occur. In the illusion, a brief flash accompanied by two brief sounds is often perceived as two flashes (Shams et al., 2000; Shams et al., 2002). Because the sound-induced flash illusion is not related to human-specific perception (i.e., speech perception), unlike the McGurk effect, it is reasonable to assume that this type of multimodal integration is fairly common in many animal species. However, as far as I know, there is no experimental evidence that animals other than humans are capable of experiencing this illusion. The lack of an appropriate animal model has hindered our understanding of this integration at the cellular and network levels.

Here, I used the Mongolian gerbil, *Meriones unguiculatus*, as a subject because it has sensitive low-frequency hearing comparable to humans (Ryan, 1976) and is considered a standard laboratory rodent, particularly in auditory neuroscience (Maier and Klump, 2006; Ohl and Scheich, 1997; Sakai and Suga, 2002;). The gerbil's retina also has a well-developed cone system (Govardovskii et al., 1992; Jacobs and Deegan II, 1994), and behavioral measurements of the animal's grating acuity have suggested that the visual system of the gerbil is well adapted to a diurnal lifestyle (Baker and Emerson, 1983). The circadian rhythms of their activities under natural light conditions also indicate that gerbils are not fully nocturnal (Pietrewicz, 1982), and they show a greater diurnal tendency than domestic mice or laboratory rats (Refinetti, 2006). Therefore, these behavioral and physiological features make the gerbil a practical and valuable animal model of auditory-visual integration in non-human animals.

In this study, I used the novel object recognition (NOR) paradigm to measure the gerbil's perception. Many rodents, including rats (Ennaceur and Delacour, 1988), mice (Dodart et al., 1997; Messier, 1997), and degus (Uekita and Okanoya, 2011), approach a novel object more frequently and spend more time exploring it than they do an object to which they have previously been exposed (Berlyne, 1950). The NOR paradigm relies on the animal's innate preference for novelty. Therefore, no prior training, such as operant conditioning, is required, which means potentially high throughput. More importantly, the paradigm is suitable for evaluating an animal's natural (untrained) cognitive potential acquired through normal development. The analogue of the NOR paradigm in human infants and non-human primates, the preferential looking time paradigm, reveals subjects’ natural cognitive functions (e.g., syntax learning (Marcus et al., 1999), voice–face pairing (Sliwa et al., 2010), and cross-modal numerical matching (Jordan et al., 2005; Moore et al., 1987). In most of rodent studies investigating multimodal integration, the animals were trained to evaluate their perception (Siemann et al., 2015; Wada et al., 2016). Thus, their untrained sensory potential for a multisensory environment is relatively unknown. I used the NOR paradigm to investigate whether the Mongolian gerbil has the capacity to perceive the sound-induced flash illusion.

8.2 Materials and methods

8.2.1 Subjects

In total, 43 Mongolian gerbils (27 males and 16 females), ranging in age from 1 to 4 months, were used in this study. All gerbils were bred and reared in our laboratory, and all were experimentally naïve. Each animal was housed with two to five other gerbils in a 20-cm width [W] × 40-cm length [L] × 17-cm height [H] cage with free access to food and water. The animal room was maintained on a 12-h light-dark schedule, and the temperature in the room was maintained at 22–23 °C with approximately 50 % relative humidity. The gerbils were handled for at least 5 days before testing to reduce handling stress. All experimental procedures were performed in accordance with guidelines established by the Ethics Review Committee of Doshisha University, and the experimental protocols were approved by the Animal Experimental Committee of the university.

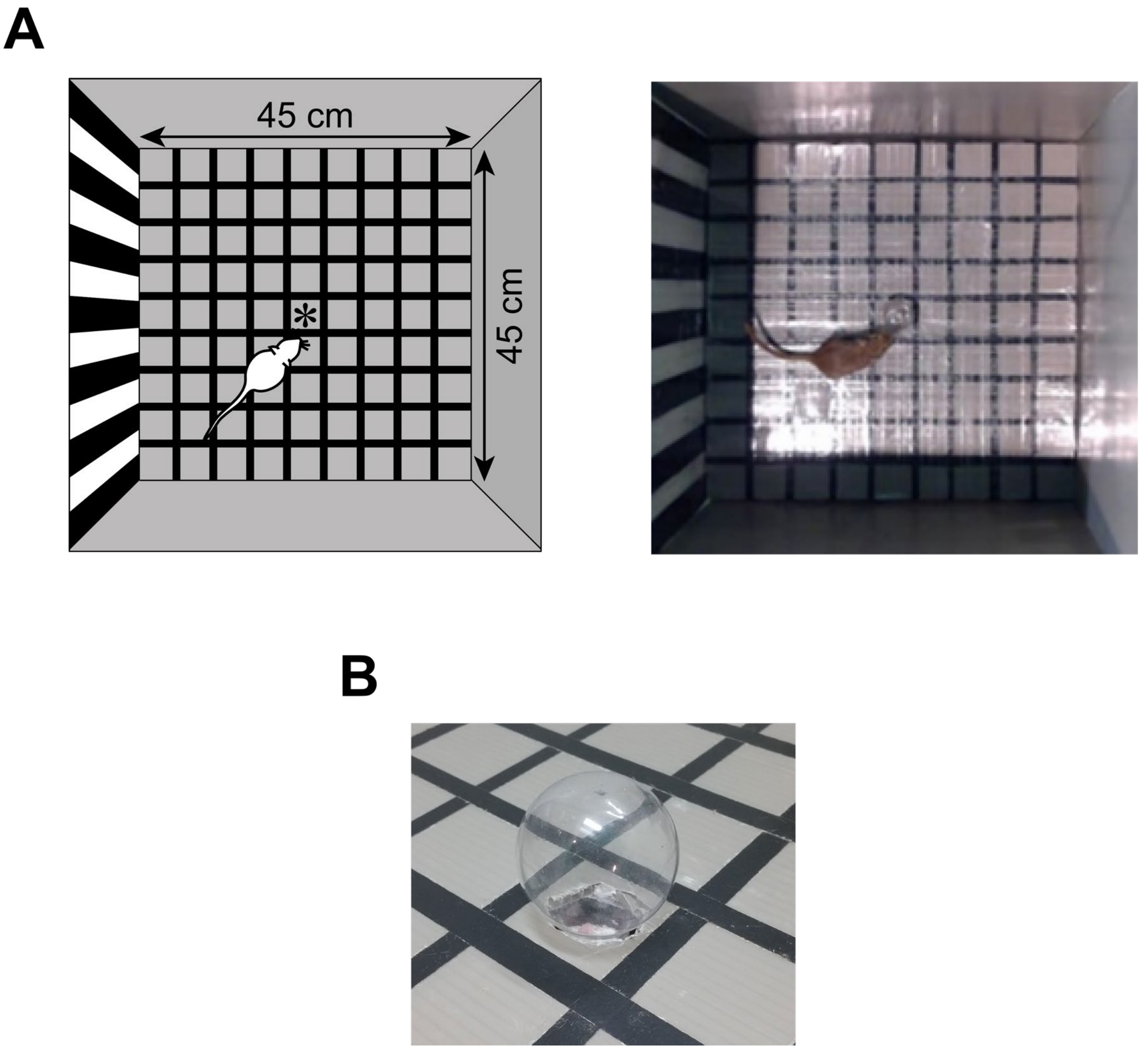

Figure 8.1. Experimental setting.
(A) Schematized figure (left) and a picture (right) of the behavioral arena. (B) A picture of the object to be explored by the subject. The object was a glass bulb of 2.2-cm radius, and a white-light-emitting diode was embedded in the object.

8.2.2 Experimental conditions

All behavioral tests were conducted in a square behavioral arena (Fig. 8.1A; 45 [W] × 45 [L] × 55 cm [H]) located in a soundproof room. The brightness at floor level in the arena was 120 lx (measured with a lux meter; GL-03, be-s Co., Ltd., Tokyo, Japan). One wall of the arena was painted with vertical black and white stripes (3.5 cm [W]) as landmarks by which the gerbils could orient themselves. The floor was colored with a square grid pattern (each square $5 \times 5\ \mathrm{cm}^2$) marked with black tape (0.8 cm [W]). A bulb-shaped object (2.2-cm radius [R]) made of glass was placed at the center of the arena to be explored (Fig. 8.1B). A white-light-emitting diode (LED) attached to a diffuser embedded within the object provided the visual stimulus. A loudspeaker (FT28D, Fostex, Tokyo, Japan) for presenting the auditory stimulus was set 100 cm above the arena. The amplitude of the sound stimulus was calibrated with a microphone (Type 1, ACO Pacific Inc., Aichi, Japan) placed at the center of the arena at the height of the animal's head (3 cm above the floor). The animal's position during the experiment was monitored and recorded at 30 frames per second with a video camera (LifeCam HD-5000, Microsoft Inc., WA, USA) set next to the loudspeaker.

8.2.3 Stimuli

The visual stimulus was a brief light (duration, 10 ms) presented by the LED inside the object. The intensity of the flash, measured 1 cm from the object, was 25 lx. The auditory stimulus was a tone pip (duration, 7 ms; frequency, 4 kHz; amplitude, 75 dB SPL). The visual and auditory stimuli were repeatedly and continuously presented during the familiarization and test sessions. Seven different stimulus configurations were used (Fig. 8.2). In the visual modality (VM) experiment, only the visual stimulus (i.e., no auditory stimulus) was presented. A single flash was repeated (VMs) with an inter-onset interval (IOI) of 510 ms during the familiarization sessions for the experiment. A double flash (VMd) was repeated at the same IOI during the test session. Each double flash consisted of paired single flashes with an IOI of 160 ms (Fig. 8.2A).

In the multimodal (MM) experiment, both the visual and auditory stimuli were presented, and the stimulus onset asynchrony (SOA) between them was changed for the test session. The IOI of the visual stimulus (flash) was 510 ms. Five stimulus configurations were used: MMd255, MMd0, MMd100, MMs255, and MMs0, where “d” and “s” represent double and single tone pips, respectively, and the number represents the SOA in ms. In the familiarization session with double tone pips (MMd), a single flash and a double tone pip consisting of two tones separated by a 64-ms IOI were repeated with an SOA of 255 ms (MMd255; Fig. 8.2B). In the test trials, an SOA of 0 ms (MMd0) or 100 ms (MMd100) was tested. In the familiarization session with a single tone pip (MMs), a single flash and a single tone pip were repeated with a 255-ms SOA (MMs255; Fig. 8.2C). In the test session, the SOA was 0 ms (MMs0). All visual and auditory stimuli were generated with customized programs (MATLAB; MathWorks, MA, USA).

8.2.4 Behavioral procedure

My experimental paradigm was a standard NOR procedure (Antunes and Biala, 2012; Bevins and Besheer, 2006). A habituation session was conducted first, followed by a familiarization session, and then a test session. In the habituation session, each animal was allowed to explore the behavioral arena freely, with no flash-emitting object, for 30 min on 4 days. No visual or auditory stimulus was presented during these sessions. In the familiarization session, the animals were habituated to the arena, and a flash-emitting object was set at the center of the arena. Either the visual and auditory stimuli were presented (MM experiment) or only the visual stimulus (VM experiment). Each animal was released at the same position and allowed to explore the arena. Each trial lasted 5 min and was repeated five times, with a 2-min inter-trial interval. During the interval, the floor and object were cleaned with 80% ethanol to remove any odors, and the animal was isolated from its cage-mates in a cylinder-shaped box (12.4 [R] × 37.8 cm [H]). The same stimuli were presented during the familiarization session. After the fifth familiarization trial, the test session began following an identical 2-min interval. The test session lasted 5 min and was conducted only once.

8.2.5 Analysis

Each animal's exploration of the object was quantified as the duration of contact with the object. 'Contact duration' was defined as the total period during which the subject was touching the object with its snout or forepaw within the first 90 s of each trial; it was measured by counting the number of frames showing exploratory behavior. Two experts, blinded to the experimental conditions, manually evaluated the behavior, and their scores correlated significantly ($r = 0.91$, $p < 0.01$). Changes in contact duration (test minus fifth familiarization trial) were tested if different from zero using t-test with a significance level of $p < 0.05$.

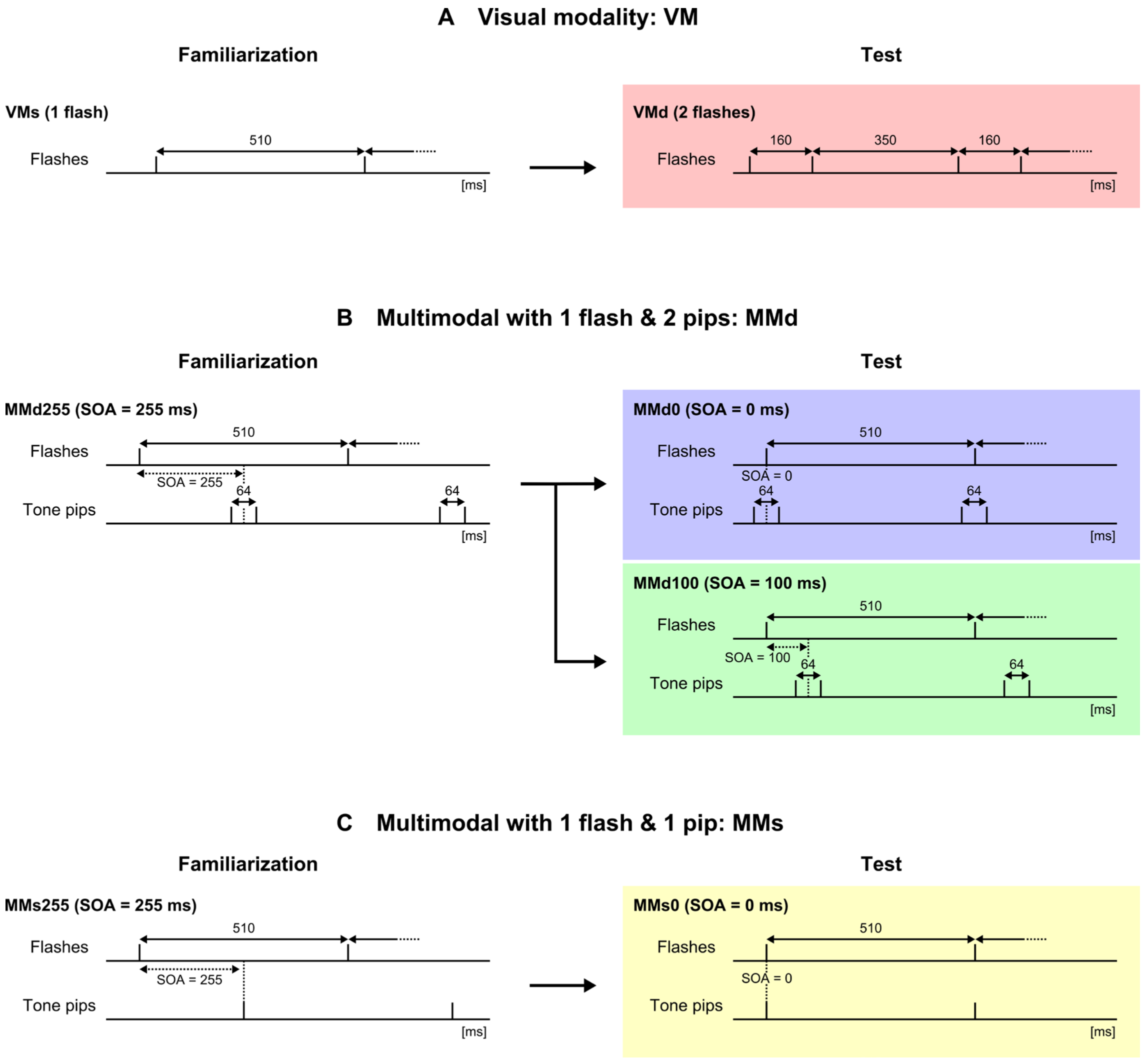

Figure 8.2. Temporal profile of each stimulus.

The flash and tone pip were repeatedly presented during the trial. (A) Stimulus in the visual modality (VM) experiment. Only the visual stimulus was presented. A single flash (VMs) was presented during the familiarization session, and a double flash (VMd) was presented during the test session. (B) The stimulus in the multimodal experiment using double tone pips (MMd). The visual stimulus was a single flash, and the auditory stimulus was double tone pips. Three types of stimulus onset asynchrony (SOA) were employed: 0 ms (MMd0), 100 ms (MMd100), and 255 ms (MMd255). (C) The stimulus in the multimodal experiment using a single tone pip (MMs). The visual stimulus was a single flash, and the auditory stimulus was a single tone pip. Two SOAs were employed: 0 ms (MMs0) and 255 ms (MMs255). The SOA was changed in the test session in both multimodal experiments (B, C).

8.3 Results

Two gerbils experienced seizures during the habituation or experimental period and were excluded from the analysis. In the VM experiment, the animals were repeatedly exposed to single flashes (VMs) during the familiarization trial, and they explored the object less as the trial proceeded (Fig. 8.3A VM). After the number of flashes changed (VMd) in the test trial, their exploration of the object increased significantly compared with that in the fifth familiarization trial (1.90 ± 0.35 s vs. 3.01 ± 0.54 s, mean ± standard error of the mean; $t_{(8)} = 2.29$, $p < 0.05$; Fig. 8.3B VM). These results confirmed that the NOR paradigm is suitable for assessing the perception of a temporally changing visual stimulus.

In the MM experiments, the gerbils' exploration of the object decreased as the trial progressed during the familiarization sessions, as was observed in the VM experiments (MMd255: 6.24 ± 0.6 s vs. 1.46 ± 0.25 s; MMs255: 7.81 ± 1.01 s vs. 1.77 ± 0.51 s). The contact durations in the first and fifth familiarization trials and the test trial under all conditions are shown in Fig. 8.3A. With test stimulus MMd0, in which a flash and double tone pips were presented with 0-ms SOA, the animals explored the object significantly longer than they had in the fifth familiarization trial (1.28 ± 0.24 s vs. 1.78 ± 0.37 s, $t_{(10)} = 1.83$, $p < 0.05$; Fig. 8.3B MMd0). With test stimulus MMd100, in which a flash and double tone pips were presented with 100-ms SOA, no significant increase in the exploratory period was observed (1.68 ± 0.49 s vs. 1.16 ± 0.25 s, $t_{(8)} = 1.29$, $p = 0.12$; Fig. 8.3B MMd100). In the MMs experiments, in which a single flash and a single tone pip were presented, reducing the SOA from 255 ms (familiarization trial) to 0 ms (test trial), the exploratory period did not increase (1.77 ± 0.51 s vs. 1.42 ± 0.36 s, $t_{(11)} = 0.93$, $p = 0.19$; Fig. 8.3B MMs0)

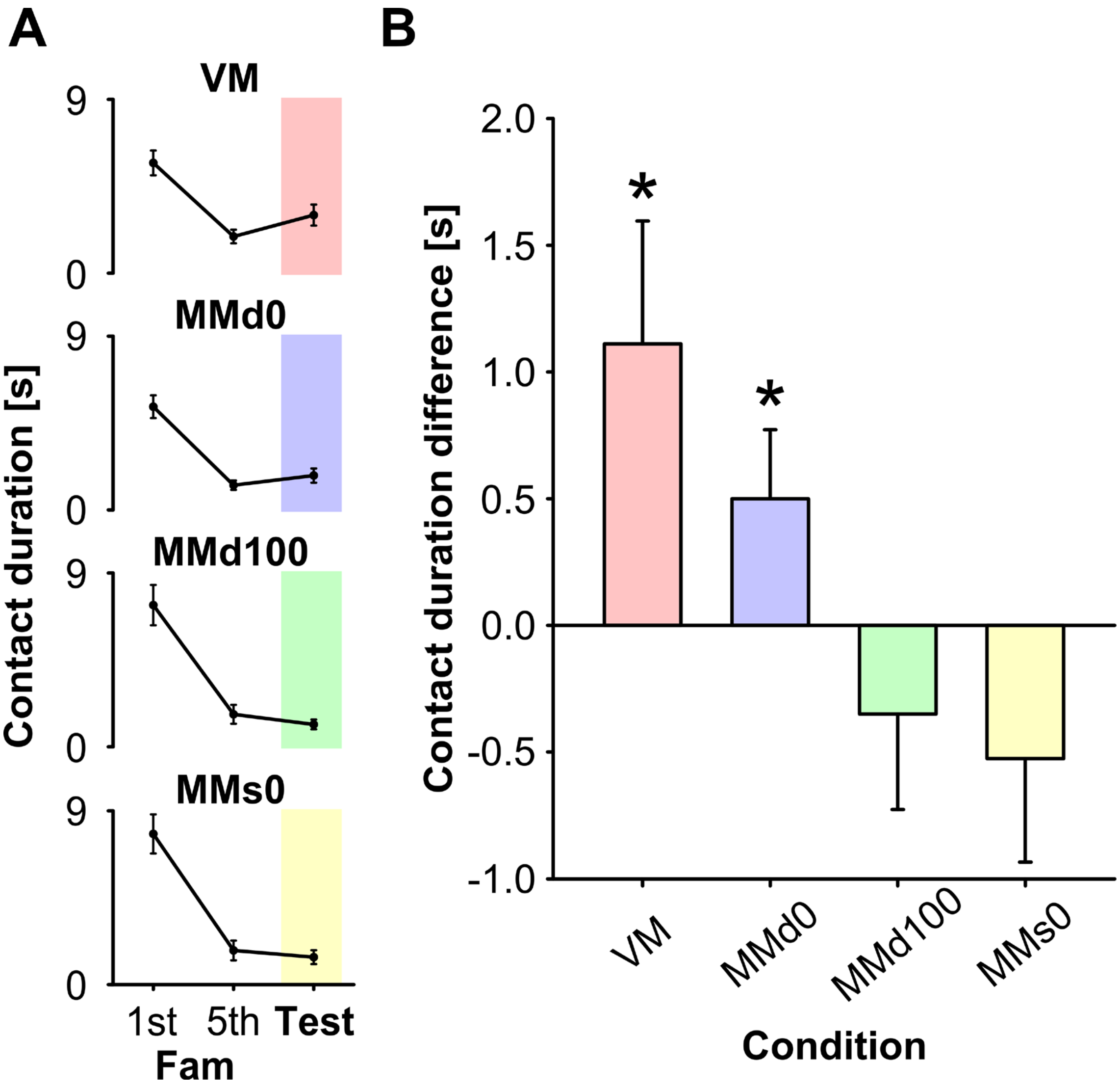

Figure 8.3. Effect of stimulus onset asynchrony (SOA) on the recognition of a flashing pattern.

Error bars indicate standard errors. (A) The contact durations in the first and fifth (the last) familiarization trial (Fam) and the test trial under all stimulus conditions. (B) Change in contact duration (test minus fifth familiarization trial) with different SOA. Contact duration increased significantly in the test trial compared with that in the fifth familiarization trial (255-ms SOA) when the animals were exposed to the 0-ms SOA (MMd0), but not when they were exposed to the 100-ms SOA (MMd100). In addition, a noncontradictory audio–visual stimulus pair (MMs0) did not produce a significant increase in the test trial compared with the fifth familiarization trial. *$p < 0.05$.

8.4 Discussion

The results of the VM tests showed that changing the number of flashes triggered a significant increase in the gerbils' exploratory behavior. Previous studies using the NOR paradigm have demonstrated that animals are more interested in an object and explore it more thoroughly when the object's properties (shape, size, and color) and/or position are altered (for a review, see Ennaceur, 2010). Recent studies that combined a Y-shaped apparatus with a NOR task demonstrated that rats visually recognize an object based on their perception of its features with tactile senses (Jacklin et al., 2016; Winters and Reid, 2010), demonstrating that this task can be used to test multimodal recognition. Because I wanted to use the NOR paradigm to evaluate the perceptual content of a temporally dynamic visual signal, particularly its multimodal effect, the pattern of flashes emitted by an object was altered as the experimental variable. As observed in the typical NOR paradigm, the gerbils showed decreasing contact with the object during familiarization, but when the number of flashes was altered from one to two, the gerbils explored the object for a longer period (Fig. 8.3A VM). This increase in exploration suggests that the gerbils recognized the change in the temporal pattern of flashing, and the difference was sufficient to solicit exploratory behavior toward the object. This result demonstrated that the NOR paradigm can be used to assess this animal's visual temporal perception and recognition.

In humans, when two beeps are presented together with a single flash, the flash is often perceived twice (Shams et al., 2000). Whereas, as the amount of SOA between the auditory and visual stimuli increases, the illusory perception occurs less frequently (Shams et al., 2002). A study by Bidelman (2016) systematically analyzed the effect of SOA in musicians and non-musicians and demonstrated that the temporal window of sensory integration was <100 ms in musicians and ~200 ms in non-musician subjects (Bidelman, 2016). My results demonstrate that exploration of the stimulus object increased significantly under the MMd0 condition, but not under the MMd100 condition (Fig. 8.3B). I interpret these data as indicating that the animals perceived the MMd255 stimulus as a single flash during the familiarization session, but experienced the MMd0 stimulus as a double flash (an illusory flash), as humans do. However, they did not experience the MMd100 stimulus as a double flash (or did so to a much smaller extent). These results support the idea that animals perceive the sound-induced flash illusion and that the temporal window for audio-visual integration is similar to that in humans.

The contact duration in MMd0 (Fig. 8.3B) is almost half that in VM (while the difference was not statistically significant ($t_{(18)} = 1.15$, $p = 0.26$)). Many human studies have shown that the illusory perception (double flash illusion) does not always occur. For example, Shams et al. (2002) reported that the illusion occurred in ca. 70 % of tests, even under the best SOA conditions (Shams et al., 2002). Other research by Andersen et al. (2004) reported that subjects perceived two flashes in 55 % of all trials (Andersen et al., 2004). Recent research confirmed that the illusory flash was perceived only in ca. 60 % of events (Roseboom et al., 2013). Therefore, even if gerbils experience the illusory flashes as often as humans, they still perceive MMd0 as double flashes in 55-70 % of all presentations, and the probabilistic nature of the illusion could explain the response difference between VM and MMd0. The greater increase from a familiarized stimulus (i.e., single flash) in VM could promote stronger exploratory behavior than in MMd0.

One might argue that the data only suggest that the animals perceived flashes in

MMd0 as different from those in MMd255, not necessarily as a double flash. For example, a simultaneous auditory stimulus might somehow enhance the saliency of the visual stimulus enough to trigger exploration (Stein et al., 2014). Because those multimodal enhancements are reported to be the most prominent at the 0-ms SOA, I introduced the MMs0 (0-ms SOA) as a control. The results under the MMs condition demonstrate that changing the SOA from 255 to 0 ms was insufficient to produce a significant behavioral response (Fig. 8.3B MMs0), suggesting the showing that stimulus simultaneity alone was insufficient to change exploratory behavior. The incongruence of the two sensory cues (i.e., single or double) and the stimulus timing are important to affect the behavior, and the requirements are comparable to those observed during a sound-induced flash illusion in human research (Shams et al., 2005a). One might also argue that a double tone pip alone is sufficient to induce exploration, and a flash presented with a certain delay somehow cancels the effect. My additional analysis suggests that this is rather unlikely. Specifically, the contact duration in the first familiarization trial in VMs (single flash without double tone pips: Fig. 8.3A VM first) was not statistically significantly different from the first familiarization in MMd255 (single flash with double tone pips: Fig. 8.3A MMd0 first and MMd100 first; $t_{(27)} = 0.53$, $p = 0.60$). Whereas this analysis on MMd cannot be a complete substitute for testing on additional control stimulus (i.e., unimodal double-tone-pip), the statistical result suggests that the double tone pips did not significantly increase exploration. Overall, my data should not be taken as direct evidence that animals perceive the auditory-induced illusory flash, and my results require further behavioral research, such as establishing whether the animals can discriminate the physically presented double flashes from the illusory one (Rosenthal et al., 2009).

The neural mechanism of sensory integration has been the focus of many studies. Rodents integrate multisensory input both behaviorally and physiologically (Hirokawa et al., 2008; Schormans et al., 2017; Song et al., 2017). However, as far as I know, there is no behavioral evidence showing that rodents experience multisensory illusions, except the study by Wada et al. (2016) on a visual-tactile illusion (Wada et al., 2016). My results present the possibility that the visual recognition of untrained gerbils was altered by the auditory stimulus, suggesting that the species have the capacity to integrate temporal content of perception in a sophisticated manner as do humans, and also that the species is a good animal model for investigating the auditory-induced flash illusion. Several human studies have already investigated the brain regions involved in this illusion. Using functional magnetic resonance imaging, Zhang and Chen (2006) showed that elevated activity in the visual cortex is associated with this illusory perception (Zhang and Chen, 2006). Mishra et al. (2007) provided evidence that rapid input from the auditory and multisensory areas modifies the activity of the visual cortex and promotes the illusion (Mishra et al., 2007). Because neither of these regions nor the proposed mechanism is specific to humans, these systems could serve as the neural basis for the integration in rodent and as well. Future research in rodents may provide cellular-level insight into the sound-induced flash illusion.

Chapter 9

Audiovisual stimuli enhanced the neural activity in the gerbil’s visual cortex

9.1 Introduction

Audiovisual integration reduces the ambiguity of unisensory information and enhances behavioral performance. When a visual stimulus was presented with an auditory stimulus, the perceptual intensity of the visual stimulus was higher than when the visual stimulus was presented alone (Stein et al., 1996). In addition, task-irrelevant visual stimulus enhanced the detectability to the auditory stimulus (Lovelace et al., 2003). From these results, the information obtained by different sensory organs appropriately separates or integrates, and ultimately, greater benefits in terms of behavior were provided. Thus, understanding the mechanism of audiovisual integration has become increasingly important in recent years. Historically, multisensory neurons that respond to multiple sensory stimuli have been found in the superior colliculus and cortical association area in cat studies (Meredith and Stein, 1983; Wallace et al., 1992). Additionally, in also primary sensory cortices, it was shown that multisensory neurons were obtained in rodents (Wallace et al., 2004). Such multisensory neurons showed the modulation of the firing rate when multisensory stimuli were presented, compared to when unisensory stimuli were presented. Upon comparison of the firing rate to unisensory stimulus, a high firing rate to multisensory stimulus is defined as enhancement, while a low firing rate is defined as depression (Stein and Stanford, 2008).

Since multisensory interactions in both the primary sensory cortices and the association cortices are revealed, sensory integration may occur in multiple stages. Therefore, simultaneous recording of multiple brain regions may reveal the question in which how auditory and visual information is integrated.

I used flavoprotein fluorescence imaging for functional brain imaging *in vivo*. Flavoproteins are a family of proteins in the mitochondrial electron transport system. As the neuron is activated, the density of calcium ion [Ca2+] increases in a cell. Subsequently, energy metabolism is activated and the flavoprotein is oxidized. Oxidized flavoprotein emits endogenous green fluorescence in blue light. Notably, Shibuki et al. (2003) showed that brain activity can be visualized by recording the endogenous green fluorescence in blue light (Shibuki et al., 2003). This imaging technique requires no exogenous fluorophore and the intensity of endogenous fluorescence in flavoprotein fluorescence imaging is 10 times stronger than that in intrinsic signal imaging (Tohmi et al., 2006). Flavoprotein fluorescence imaging does not require surgical operations, such as craniotomy, and can record the neural responses in the mouse visual (Tohmi et al., 2006; Tohmi et al., 2014; Yoshitake et al., 2013), auditory (Takahashi et al., 2006; Tsukano et al., 2015; Yanagawa et al., 2016), and somatosensory cortex (Kitaura et al., 2010; Komagata et al., 2011; Shibuki et al., 2006). In addition, the neural activities in wider brain regions were simultaneously obtained (Hishida et al., 2014; Hishida et al., 2019). Because this imaging technique can observe the neural response of a wide cortical surface, the activities elicited by the sensory stimuli are simultaneously recorded in different sensory cortices. As a first step, I identified the visual, auditory, and somatosensory cortex by flavoprotein fluorescence imaging, and recorded a retinotopic map in the gerbil's visual cortex. Next, I examined how the auditory stimulus affects the visual response in the visual cortex.

9.2 Materials and methods

9.2.1 Subjects

All experimental procedures involving animal were conducted in accordance with the guide lines of the animal care and use committee of Doshisha University. Seven 6-10-week-old Mongolian gerbils, *Meriones unguiculatus*, were used for all experiments. Gerbils were bred and reared in our laboratory, and were maintained on a 12-h light/dark cycle, and were housed at 22-23 °C with approximately 50 % relative humidity. Each animal was housed with 2-5 other gerbils in a cage of 20 cm (W) × 40 cm (L) × 17 cm (H), with free access to food and water.

All experiments were conducted in an acoustically and electrically shielded box. Gerbils were anesthetized with urethane (1.5 g/kg, i.p.). Through surgery, the body temperature was maintained by placing the subject on a heating pad. Their hair was shaved off, and the skin on the head of the gerbils was removed. The skull was exposed and cleanly maintained. A metal plate was attached using dental cement (Shofu, Kyoto, JPN) to the skull, and the gerbil’s head was fixed by custom-designed platform. The surface of the skull was covered with liquid paraffin (Wako, Osaka, JPN) to keep the skull transparent. After the recording, gerbils were euthanized with an overdose of pentobarbital (i.p.).

9.2.2 Flavoprotein fluorescence imaging

A cooled charge-coupled device (CCD) camera system (BU-61, Bitran, JPN) was mounted on a stereoscopic microscope (SZX16, Olympus, JPN) with a 1× objective (numerical aperture, 0.15; Olympus, Japan). Cortical images (240 × 135 pixel after binning) of endogenous green fluorescence (λ= 495-540 nm) in blue light (λ= 460-480 nm) using a 130 W mercury lamp for light source (SHI-130 OL, Tokyo, JPN) through fiber optic cables were recorded from the area of left cortex at a rate of 10 frames per second using the CCD camera system. The radiant exposure of blue light was set at 5.7 W/cm^2 using a digital power meter (PM100D, Thorlabs, Newton, NJ, US) with a thermal power sensor (S302C, Thorlabs). The cortical images were recorded for 8 s and averaged for 30-40 trials. The recording of fluorescence images was repeated at 25 s intervals. A spatial filter (averaging in 5 × 5 pixels) was performed to improve image quality. The baseline intensity (F0) in each pixel was determined by averaging the 5 images before stimulus onset. The fluorescence change (ΔF) was obtained by as difference between the fluorescence signal (F) and the baseline intensity (F0) in each pixel. The images were normalized as $\Delta F/F0$ and shown in a pseudocolor scale. The fluorescence response was evaluated as values of relative fluorescence change ($\Delta F/F0$) in a circular window with a diameter of 10 pixels. The circular window was placed on the brain region including the highest response peak in the images after 2 s from stimulus onset.

9.2.3 Stimuli

A red light-emitting diode (LED) (λ= 613 nm; diameter, 5 mm) emitted blinking light on the animal's right eye. The LED was placed 5 cm from the eye in the horizontal plane. The auditory stimulus was broadband noise (4-60 kHz) and presented from a loudspeaker (ES1, Tucker-Davis Technologies, USA) which was located 5 cm from the right ear of the animal. The sound pressure level of auditory stimuli was measured by the microphone (Type1, ACO Pacific, Japan) and was adjusted to 85 dB SPL. For whisker stimulation, I formed a custom-made stimulation system. The whisker stimuli were delivered to the right whiskers by a motor. The duration of each stimulus was set to 200 ms in all conditions. The trial and each stimulator were controlled by a microcontroller (Arduino Uno, Ivrea, Italy).

9.2.4 Analysis

A paired t-test was used to assess the change in fluorescence responses between visual stimulus and audiovisual stimulus, with a significance level of $p < 0.05$. Analysis was performed using Rstudio (Rstudio Team, 2018).

9.3 Results

The flavoprotein fluorescence responses when the whisker stimulus was presented were shown in Fig. 9.1 (0 s indicated the stimulus onset). Fluorescence responses were activated until about 2 s after stimulus onset, but not observed 4 s after stimulus onset. The responses within regions of interest (ROI) rapidly increased after the stimulus was presented, and the response peak was about 1 s later (Fig. 9.1B). Response amplitude in $\Delta F/F0$ was 0.84 % (1.1 s). After the response peak, the level of fluorescence gradually decreased and returned to the baseline level. Auditory-only, tactile-only, visual-only, and multimodal (simultaneous stimulus of 3 modalities) stimuli were presented, respectively, to identify the auditory, somatosensory, and visual cortex. The fluorescence responses in different cortical regions were obtained when each unisensory stimulus (auditory, tactile, and visual) was presented. The fluorescence responses appeared in all regions activated by each unisensory stimulation when multisensory stimuli were presented (Fig. 9.2). From these results, the following experiment was conducted, assuming that the cortical region responded by visual stimulation was the visual cortex.

Next, I attempted to obtain retinotopic maps for the visual cortical areas. Visual stimuli were presented from LED which was placed 0° (front), 45°, and 90° (right) in the horizontal plane (Fig. 9.3A). The fluorescence responses to the stimuli from each direction were observed (Fig. 9.3B). The location of the response field was shifted toward the medial side as the spatial location of visual stimulation was shifted from 0° to 90°. The areas with responses > 50 % of the maximum amplitude were defined as the regions activated by the horizontal axis of stimulation.

In addition, we investigated how the auditory stimulus influenced visual response in the visual cortex (Fig. 9.4). Fig. 9.4B showed the flavoprotein fluorescence response in 1.2 s after each stimulus onset. Fluorescence responses were clearly observed only when the visual stimulus was presented (visual-only or auditory-visual stimulation), whereas almost no response was observed at no visual stimulus (auditory-only stimulation). I compared the response amplitude between audio-visual and visual conditions. The flavoprotein fluorescence response to audio-visual stimuli was larger than that to visual stimuli (0.20 ± 0.05 % vs 0.18 ± 0.03 %, mean ± standard error of the mean; $t_{(4)} = 0.65$, $p = 0.05$; Fig. 9.4D).

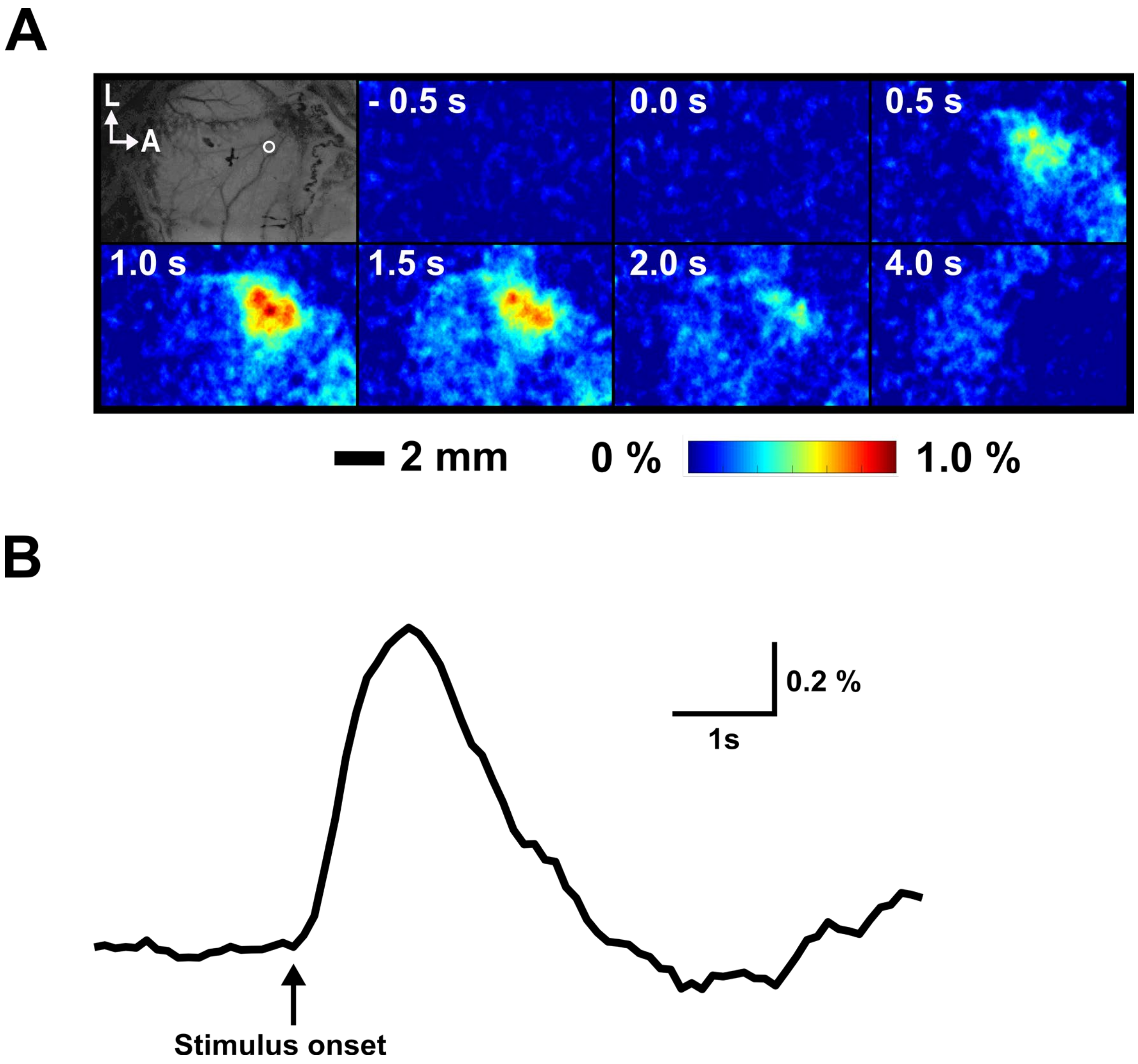

Figure 9.1. Time course of flavoprotein fluorescence responses induced by whisker stimulation.

(A) Original image and each pseudocolor images in ΔF/F0 at several times. 0.0 s indicated the stimulus onset. Regions of interest (ROI) was placed at the highest peak response. (B) Time course of ΔF/F0 in a region of interest (ROI). The horizontal bar shows the time, and the vertical bar shows the ΔF/F0. Stimulus onset was 0.0 s. The peak of response amplitude was 1.1 s after stimulus onset. Data were obtained by averaging 36 trials.

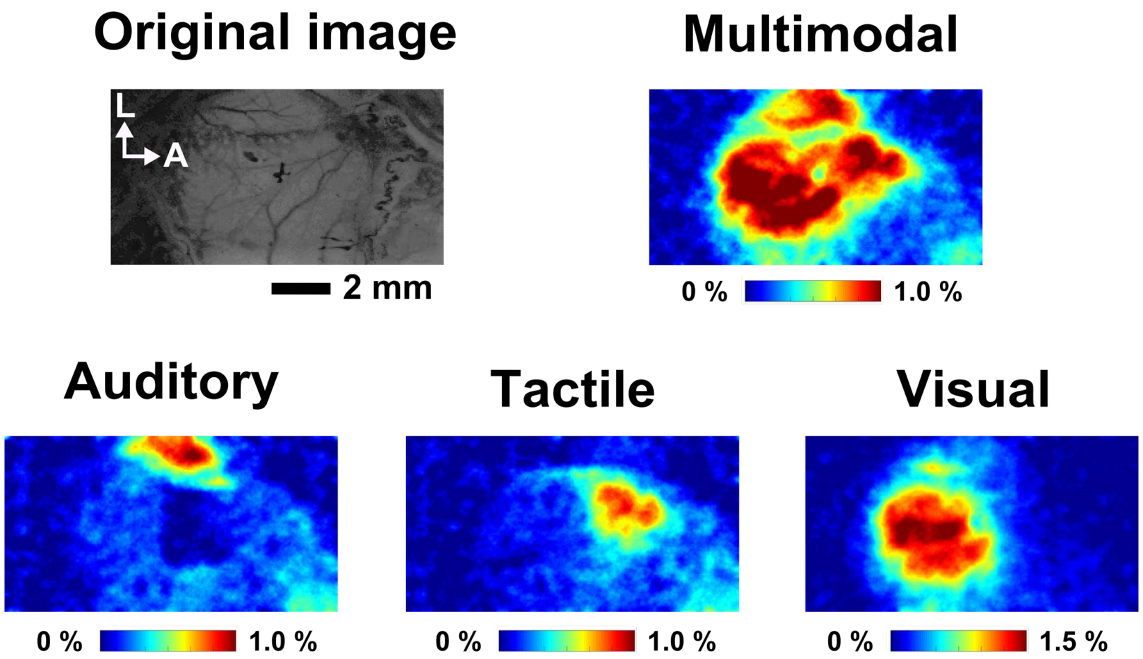

Figure 9.2. Activated brain region in left hemisphere induced by auditory, tactile, visual and multimodal stimulation.

The fluorescence responses in different cortical region were activated when each stimulus presented. Each pseudocolor image were the image in highest peak response.

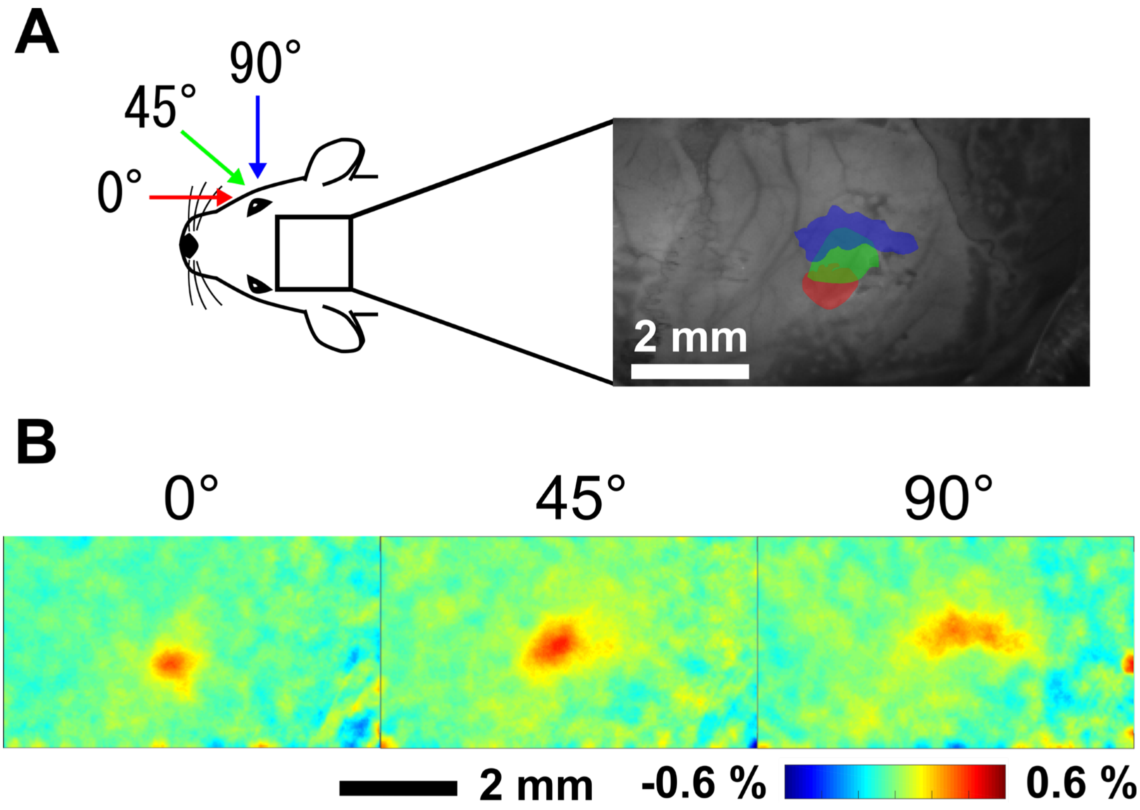

Figure 9.3. Retinotopic maps.

(A) Retinotopic map superimposed on an original fluorescence image. Fluorescence responses were elicited via the right eye using LED stimuli placed at 0° (red), 45° (green) and 90° (blue). Cortical areas in which ΔF/F0 was stronger than 50 % of peak amplitudes in each location of LED were superimposed on the original image in each color. Fluorescence responses was shifted from lateral to medial corresponding to the locations on the LED from front to right. (B) Original fluorescence responses evoked by the stimulation from the three LED location.

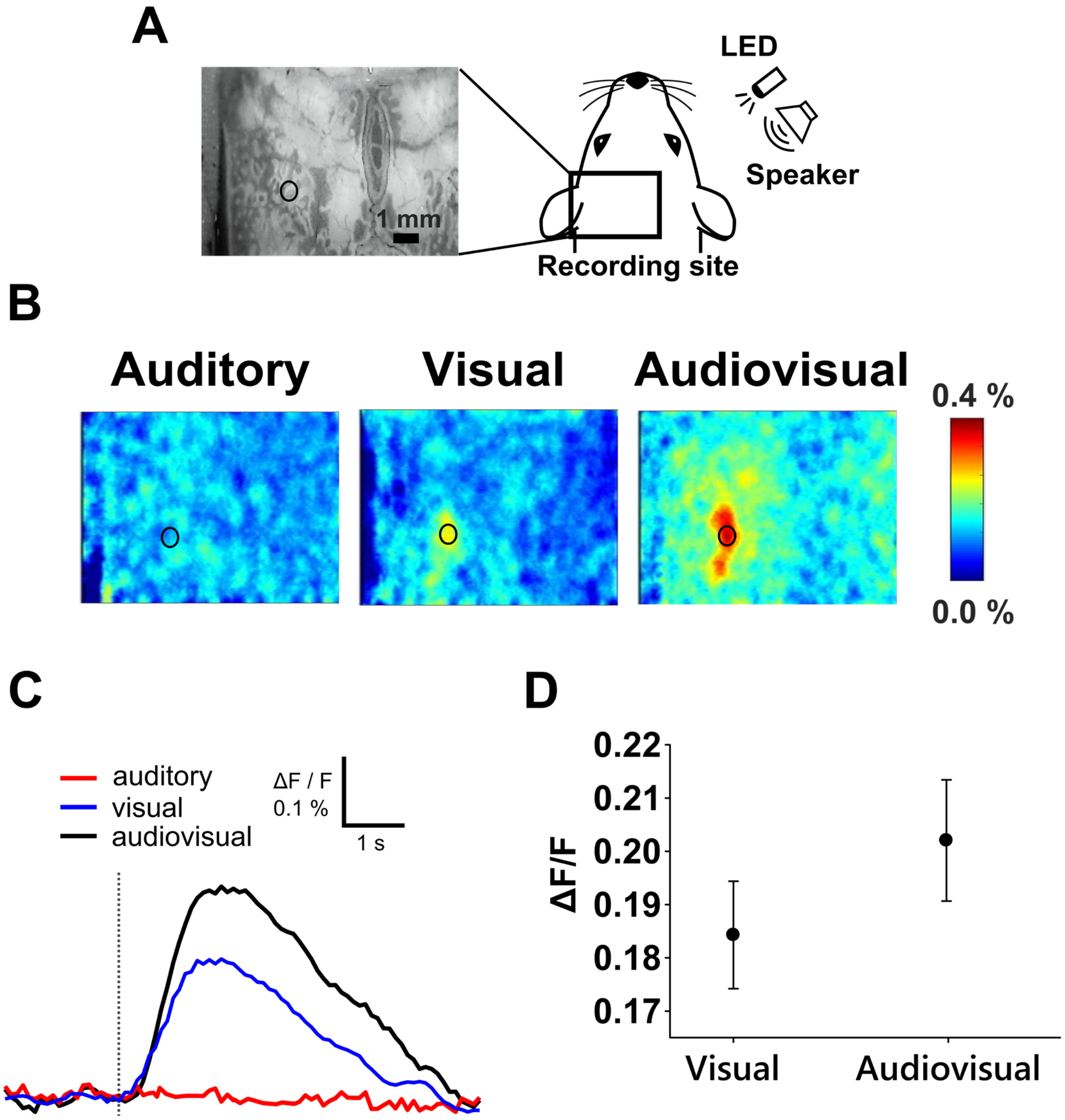

Figure 9.4. Fluorescence responses to auditory, visual, and audiovisual stimuli. (A) Original cortical image and experimental setting. Stimuli were presented from right side on the animal. Region of interest (ROI) was placed at the highest peak response. (B) pseudocolor images in ΔF/F0 when auditory, visual, and audiovisual stimuli were presented, respectively. Visual cortex was activated only when visual stimulus presented. Data were obtained by averaging 30 trials. (C) Time courses of ΔF/F0 in a ROI to auditory, visual and audiovisual stimuli. The horizontal bar shows the time, and the vertical bar shows the ΔF/F0. Dashed line was the time of stimulus onset. (D) Comparison of max fluorescence responses between visual and audiovisual. The amplitude of fluorescence responses measured in ROI. The activation in visual areas was enhanced when a visual stimulus accompanied by auditory stimulus was presented.

9.4 Discussion

In our previous study, flavoprotein fluorescence responses were obtained in also Mongolian gerbil, and flavoprotein fluorescence imaging was found to be effective in gerbils (Tamai et al., 2020). The functional localization of the cerebrum exists in rodents. It is revealed by optical imaging that the different brain regions respond by presenting different sensory stimuli (Kuroki et al., 2018). In the method using flavoprotein fluorescence imaging, the cortical sensory response to somatosensory, visual, and auditory stimuli has been investigated, and it has been reported that the different brain regions were activated (Shibuki et al., 2006; Takahashi et al., 2006; Tohmi et al., 2006). My results showed that the respective different regions responded when whisker, auditory and visual stimuli were presented (Fig. 9.2). The relative location of each sensory cortex as estimated by my flavoprotein fluorescence imaging was consistent with the location on recently reported gerbil atlas (Radtke-Schuller et al., 2016). These suggest that the region responded by whisker stimulation was the somatosensory cortex; the region activated by auditory stimulation was the auditory cortex, and the region activated by visual stimulation was the visual cortex.

I measured the retinotopic map in order to investigate whether the visual cortex considered by the results of imaging was the true visual cortex. In visual areas, the neurons are assigned to each location on the cortical surface a point in visual space. This regular arrangement is called a ‘Retinotopic map’ (Dräger, 1975). My results showed that the location of the response field was shifted toward the medial side as the spatial location of visual stimuli was shifted from 0° to 90° and were consistent with the characteristics of the retinotopic map reported by previous studies (Dräger, 1975; Schuett et al., 2002).

Moreover, I focused on the visual cortex and examined the effect of audiovisual integration in this area. Fluorescence responses in the visual cortex increased when visual stimuli were accompanied by auditory stimuli compared to visual stimuli alone (Fig. 9.4). A functional magnetic resonance imaging (fMRI) study in humans showed that simultaneous presentation of auditory and visual stimuli evoked stronger neural activity in the visual cortex, compared to visual-only stimulation (Shams et al., 2010; Watkins et al., 2006). My results using flavoprotein fluorescence imaging also are consistent with previous observations.

Neuroanatomical studies have shown that the primary visual cortex of the gerbils is directly connected to the primary auditory cortex (Budinger and Scheich, 2009; Henschke et al., 2015). Therefore, it is possible that the neural response in the visual cortex was modulated by the audiovisual stimulus. In mice studies, the response of neurons in the visual cortex was enhanced (Ibrahim et al., 2016) or inhibited (Iurilli et al., 2012) by the presentation of simultaneous audiovisual stimuli using an electrophysiological recording. In this study, the population recording using optical imaging was conducted, and the involvement of neuronal levels such as excitatory and inhibitory neurons was unknown.

Mongolian gerbils may have a mechanism of sensory integration similar to that of humans. Since this study only examined the neural response in the visual cortex, future studies with simultaneous recording of multiple regions (i.e., auditory and visual cortices) are needed to examine the activities of each region.

Chapter 10

Conclusion

In this dissertation, I first investigated the important factors of audiovisual integration in psychophysical experiments in humans. In particular, I focused on sound-induced flash illusion, in which visual perception is altered by auditory stimuli, and examined the influence of various factors of auditory information on the illusion. Next, I investigated whether Mongolian gerbils (*Meriones unguiculatus*), a common animal model for auditory physiological studies, could also be suitable as an animal model for the study of audiovisual integration, using behavioral and optical imaging methods. Finally, I briefly summarize the results obtained in this thesis and discuss future issues and developments.

10.1 Summary of main results

10.1.1 Effects of frequency on sound-induced flash illusion (Chapters 2-4)

In Chapters 2-4, I presented various characteristic auditory stimuli. In Chapter 2, the combination of various frequency differences between the first and second auditory stimuli was used at three base frequencies (500, 3000, and 5000 Hz). I found that the occurrence of illusion reduced as frequency difference increased. This effect was similar to all base frequency conditions. Therefore, it was suggested that flash illusion was affected by the frequency difference of presenting two auditory stimuli.

In Chapter 3, I assessed the relationship between the ability to discriminate frequencies and the susceptibility to illusion. Two auditory stimuli containing various frequencies difference were presented as the base frequency was 3000 Hz. Consistent with the results in Chapter 2, as the frequency difference between the first and second auditory stimuli was large, the illusion occurred less frequently. In the frequency discrimination task, as the frequency difference increased, participants accurately notice the difference in frequency between the first and second sounds. However, there was no correlation between the ability of frequency discrimination the susceptibility to illusion.

In Chapter 4, noise with various frequency components and pure tone were presented. When different auditor stimuli were presented the first time and second time, the illusion was diminished compared to when the identical auditory stimuli were presented twice. I also found that when two auditory stimuli with a combination of pure tone and noise were presented, stimuli of a combination containing the same frequency component were more likely to induce the illusion than stimuli of the combination having not the same frequency component. These results indicate that sound-induced flash illusion depends on the coincidence in the frequency of auditory stimulus presented twice.

10.1.2 Effects of subjective audibility on sound-induced flash illusion (Chapter 5).

I investigated whether manipulation of subjective audibility altered the occurrence rate of fission illusion and fusion illusion. Specifically, the sound pressure level of the second auditory stimulus was manipulated. The rate of perceiving two sound decreased as the sound pressure level of the second auditory stimulus was low, although auditory stimuli were physically presented twice. This result suggested that the subjective audibility could be controlled by manipulating the auditory intensities. In addition, I found that as the sound pressure level of the second auditory stimulus was weak, the proportion of two flashes perception also decreased in both one flash and two flashes trials. These results indicated that decreasing the intensities of the second auditory stimuli reduced the fission illusion and induced the fusion illusion frequently. I also found a correlation between subjective audibility and the occurrence rate of illusion. It was suggested that sound-induced flash illusion was affected by the number of auditory stimuli participants could be subjectively perceived.

10.1.3 Audiovisual stimuli enhanced behavioral performance in Mongolian gerbil (Chapter 6).

I investigated whether the behavioral performance in gerbils is constrained by the three principles of multisensory integration in a simple stimulus detection task. When audiovisual stimuli were presented, the detection rate was improved and the reaction time was shortened compared to when auditory-only and visual-only stimuli were presented. This data suggests that audiovisual integration enhanced behavioral performance. As stimulus intensities decreased, strong behavioral gain for response rate was obtained, and especially, it was often maximal at stimuli of around threshold. Subsequently, I examined the temporal rule. Various audiovisual stimuli with different onset time between auditory and visual stimuli were presented. The reaction time to synchronized audiovisual stimuli was faster than that to asynchronous audiovisual stimuli. In contrast, the detection rate and reaction time were not different in spatial congruency, when the auditory stimulus was presented from the ipsilateral or contralateral position as the visual stimulus. These results indicate that behavioral performance in gerbils improves based on inverse effectiveness and temporal rule, but not spatial rule. However, the spatial rule may depend on the behavioral paradigm (see Chapter 6 for details), and caution is needed. At the least, gerbils may have the ability to integrate auditory and visual stimuli, and their behavioral performance is enhanced through audiovisual integration.

10.1.4 Sound-induced flash illusion occurred in Mongolian gerbil (Chapters 7 and 8).

In Chapter 7-8, I established a novel behavioral paradigm for examining time-varying sensory stimuli and investigated whether gerbils perceived illusory flash. In Chapter 7, I examined whether gerbils could discriminate between temporally varying visual stimuli. After gerbils were exposed to a single flash at the familiarization session, the contact duration for the object in which stimuli were changed increased when the pattern of blinking was changed from single flash to double flashes. This result indicates that gerbils perceived the change of blinking pattern as a novel event. In shortly, the gerbils may have been able to recognize the difference between the single flash and double flashes. In Chapter 8, I developed the behavioral paradigm established in Chapter 7 and investigated whether sound-induced flash illusion also occurred in gerbils. In coincidence with the result of Chapter 7, changing the number of flashes triggered an increase in the gerbils' exploratory behavior. When a flash and double tone pips were altered from 255-ms SOA to 0-ms SOA, gerbils explored the object longer. In contrast, when the SOA was changed from 255 to 100 ms, the increase of contact duration was not observed. These results suggest that sound-induced flash illusion also occurred in gerbils, and gerbils may have a mechanism of audiovisual integration similar to that in humans.

10.1.5 Neural mechanisms involved in audiovisual integration (Chapter 9)

Using an optical imaging technique, I investigated the effects of audiovisual integration in the visual cortex. First, the auditory, visual, and tactile stimuli were presented, respectively, to identify the rough location of the gerbil's visual cortex. Subsequently, I obtained a retinotopy map specific to the visual cortex. Finally, the response to auditory, visual, and audiovisual stimuli in the visual cortex was recorded. I found that when the audiovisual stimuli were presented, the activity was enhanced more than when the visual stimuli were presented. This result can be attributed to the effect of audiovisual integration.

10.2 The value of the gerbil as an animal model for audiovisual integration

In this dissertation, I examined whether gerbils are an appropriate animal model for studying audiovisual integration. Specifically, I investigated whether behavioral performance was enhanced and the illusion also occurred in gerbils. These two phenomena are common examples of multisensory integration. Gerbils have behavioral and physiological features in auditory (Ryan, 1976; Budinger and Scheich, 2009) and visual senses (Baker and Emerson, 1983; Refinetti, 2006; Yang et al., 2015), and may also have the capacity to integrate auditory and visual information. I found that audiovisual stimuli enhanced behavioral performance in stimulus detection task (Chapter 6) and neural activity in the primary visual cortex in flavoprotein fluorescence imaging (Chapter 9). Furthermore, I demonstrated that sound-induced flash illusion occurred in gerbils (Chapter 8). My data showed that the two phenomena of audiovisual integration observed in human studies also occurred in gerbils, and indicated that gerbils were a good animal model for audiovisual integration. I plan to research two future directions using the animal model of gerbils.

10.3 How similar are the mechanisms of audiovisual integration in humans and other animals?

In humans, a variety of psychophysical experiments have been developed and illusions involved in audiovisual integration have been discovered. Several studies using EEG, fMRI, and other techniques have revealed important regions for audiovisual integration, but the mechanisms at the neuronal level have not been clearly described. In rodent research, on the other hand, the development of behavioral paradigms similar to those used in human studies in recent years has enabled us to investigate the effects of audiovisual integration at the behavioral level. Furthermore, with the development of neuroscience, it becomes possible to record and manipulate single neurons, and to understand them at the circuit level. The results obtained from common tasks to humans and rodents such as the stimulus detection task are quite similar in humans and rodents (human: Bolognini et al., 2005; Bolognini et al., 2007; Frassinetti et al., 2002, rodent (rat): Gleiss and Kayser, 2012; Hirokawa et al., 2008; Sakata et al., 2004). In the future, audiovisual integration for behavioral enhancement will be elucidated at the neuronal level by studies in rodents. In fact, the neural mechanisms involved in the principle of inverse effectiveness have been investigated. Meijer et al. (2020) found that the neural populations in the anterolateral area (AL) which is part of the lateral secondary visual cortex (V2L) of mice showed stronger neural correlates of detection performance to audiovisual stimuli than the primary visual cortex (V1) (Meijer et al., 2020). This evidence indicates that the area AL is part of important regions for audiovisual integration. In addition, several rat studies showed V2L corresponding to the AL region in mice plays an important role in temporal processing in audiovisual integration (Hirokawa et al., 2008; Schormans et al., 2017). Future studies should focus on V2L and consider the role of V2L in the three principles of audiovisual integration.

On the other hand, even with the double-flash illusion induced by relatively simple stimuli (flashes and beeps), the detailed neural mechanisms have not been clarified due to the absence of an animal model. I have demonstrated that the double-flash illusion may also occur in rodents, gerbils. However, my research has a limitation. The novel object recognition procedure used in Chapter 8 cannot be investigated quantitatively. The novel object recognition is based on the curiosity of the animals and recording quantitative neural activity is difficult because one neural response to one stimulation cannot be obtained. In Chapter 6, I also, however, developed head-fixed operant conditioning in gerbils. This behavioral setting is compatible with neurophysiological methods, and one stimulus - one response (both behavioral and neural response) can be obtained. Human studies have accumulated a lot of evidence for the neural mechanisms involved in the sound-induced flash illusion (for review, see Hirst et al., 2020; Keil, 2020), and these hypotheses can be investigated at the neuronal level using gerbils as an animal model. Furthermore, I revealed the effect of frequency congruency on the sound-induced flash illusion in humans in Chapters 2-4. Since the greater frequency difference between the first and second auditory stimulus reduced the illusory flash, the illusory flash may be elicited by direct projection from A1 to V1. I would like to investigate this hypothesis to record the perceptual response and the neural response of the primary visual cortex in gerbils. In the future, to understand sensory integration, behavioral neuroscience addressing the two aspects of behavioral performance enhancement and illusions (sound-induced flash illusion) in both humans and rodents contribute to the question of how integration takes place and forms the final perception.

10.4 Applications in the clinical field

Multisensory integration is an essential mechanism for enriching our life. Understanding the mechanisms of multisensory integration is not simply multisensory perception, and also allows us to enter into the differences in sensory processing between several clinical groups. In a noisy environment, visual information, such as the movement of the lips, improves the comprehend speech perception (Sumby and Pollack, 1954). In contrast, autism spectrum disorder (ASD) mainly represents the impairment in social communication (APA, 2013). Individuals with ASD are unable to hear sounds in even weak background noise (Alcántara et al., 2004). Furthermore, Smith and Bennetto (2007) showed that the audiovisual gain of speech perception in individuals with ASD was weaker than that in individuals with typical development (TD) (Smith and Bennetto, 2007). This may be due to impaired multisensory integration causing them difficulty to communicate with others. In other words, audiovisual integration may contribute to more complex functions such as smooth communications. There is now evidence that all sensory modalities are impacted in ASD (Robertson and Baron-Cohen, 2017). Based on this background, multisensory integration itself may have been affected. Indeed, behavioral benefits from audiovisual integration in ASD have been reported to be weak (Collignon et al., 2013; Stevenson et al., 2014a). Studies of the illusion have shown that the McGurk effect tends to occur less frequently in individuals with ASD compared to individuals with TD (Mongillo et al., 2008; Stevenson et al., 2014b). In ASD, sound-induced flash illusion has also been examined. Some studies have shown high susceptibility to illusion (Foss-Feig et al., 2010), while others have shown a low susceptibility (Stevenson et al., 2014c), and no unified conclusion has yet been reached. One possibility may reflect the strong individual differences in ASD (Yaguchi and Hidaka, 2018).

Recently, there has been evidence investigating the effects of audiovisual enhancement using ASD model mice. Siemann et al. (2017) demonstrated that a weaker enhancement of behavioral performance was observed in serotonin transporter Ala56 knock-in mice, a type of ASD model, compared to wild type (WT) (Siemann et al., 2017). This result is consistent with human studies, and future studies are needed to investigate the neural mechanisms by which audiovisual integration is impaired.

Furthermore, it is also a good idea to examine the effect of illusion in WT and ASD model animals. In visuotactile integration, mice experienced the rubber hand illusion, in which body ownership was altered (Botvinick and Cohen, 1988), in the tail (Rubber tail illusion: Wada et al., 2016), but rubber tail illusion was less likely to occur in Ca2+-dependent activator protein for secretion 2-knockout (Caps2-KO) mice, a type of animal model of ASD (Wada et al., 2019). Individuals with ASD tended to less experience the rubber hand illusion (Paton et al., 2012), providing similar results in humans and mice. The mechanism of sensory integration impairment with ASD has been investigated using the illusion caused by visuotactile integration. However, the illusion of audiovisual integration has not been reported in rodents. I showed that behavioral enhancement and sound-induced flash illusion caused by audiovisual integration were observed in gerbils. A whole gerbil genome has been sequenced (Zorio et al., 2019) and optogenetic tools to gerbils have been applied recently (Brunk et al., 2019; Keplinger et al., 2018; Wrobel et al., 2018). Eventually, we may be able to apply the technology for genetic modification. It seems that ASD can be triggered by the administration of valproate or thalidomide to the mother in early childhood (Narita et al., 2002). By applying this technology, it will be possible to study ASD in gerbils. In other words, the comparison of ASD and WT gerbil to two phenomena, i.e., behavioral performance enhancement and illusions including the sound-induced flash illusion, respectively, provides clues to better understand the impairment of audiovisual integration of individuals of ASD have. The results I have shown in this dissertation are the first step to understand the audiovisual integration not only in ASD but also in various psychiatric disorders such as schizophrenia.

References

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology*, 14(3), 257-262.

Alcántara, J. I., Weisblatt, E. J., Moore, B. C., & Bolton, P. F. (2004). Speech-in-noise perception in high-functioning individuals with autism or Asperger's syndrome. *Journal of Child Psychology and Psychiatry*, 45(6), 1107-1114.

American Psychiatric Association, A. (2013). *Diagnostic and statistical manual of mental disorders* (Vol. 5). Washington, DC: American Psychiatric Association.

Andersen, T. S., Tiippana, K., & Sams, M. (2004). Factors influencing audiovisual fission and fusion illusions. *Cognitive Brain Research*, 21(3), 301-308.

Antunes, M., & Biala, G. (2012). The novel object recognition memory: neurobiology, test procedure, and its modifications. *Cognitive Processing*, 13(2), 93-110.

Apthorp, D., Alais, D., & Boenke, L. T. (2013). Flash illusions induced by visual, auditory, and audiovisual stimuli. *Journal of Vision*, 13(5), 3-3.

Arden, G. B., Wolf, J. E., & Messiter, C. (2003). Electrical activity in visual cortex associated with combined auditory and visual stimulation in temporal sequences known to be associated with a visual illusion. *Vision Research*, 43(23), 2469-2478.

Arnold, P., & Hill, F. (2001). Bisensory augmentation: A speechreading advantage when speech is clearly audible and intact. *British Journal of Psychology*, 92(2), 339-355.

Baker, A. G., & Emerson, V. F. (1983). Grating acuity of the Mongolian gerbil

(*Meriones unguiculatus*). *Behavioural Brain Research*, 8(2), 195-209.

Banks, M. I., Uhlrich, D. J., Smith, P. H., Krause, B. M., & Manning, K. A. (2011). Descending projections from extrastriate visual cortex modulate responses of cells in primary auditory cortex. *Cerebral Cortex*, 21(11), 2620-2638.

Bazilinskyy, P., & Winter, J. D. (2018). Crowdsourced measurement of reaction times to audiovisual stimuli with various degrees of asynchrony. *Human Factors*, 60(8), 1192-1206.

Berlyne, D. E. (1950). Novelty and curiosity as determinants of exploratory behaviour 1. *British Journal of Psychology. General Section*, 41(1-2), 68-80.

Bevins, R. A., & Besheer, J. (2006). Object recognition in rats and mice: a one-trial non-matching-to-sample learning task to study'recognition memory'. *Nature Protocols*, 1(3), 1306-1311.

Bidelman, G. M. (2016). Musicians have enhanced audiovisual multisensory binding: experience-dependent effects in the double-flash illusion. *Experimental Brain Research*, 234(10), 3037-3047.

Bolognini, N., Frassinetti, F., Serino, A., & Làdavas, E. (2005). “Acoustical vision” of below threshold stimuli: interaction among spatially converging audiovisual inputs. *Experimental Brain Research*, 160(3), 273-282.

Bolognini, N., Leor, F., Passamonti, C., Stein, B. E., & Làdavas, E. (2007). Multisensory-mediated auditory localization. *Perception*, 36(10), 1477-1485.

Botvinick, M., & Cohen, J. (1998). Rubber hands ‘feel’touch that eyes see. *Nature*, 391(6669), 756-756.

Brunk, M. G., Deane, K. E., Kisse, M., Deliano, M., Vieweg, S., Ohl, F. W., ... & Happel, M. F. (2019). Optogenetic stimulation of the VTA modulates a frequency-specific gain of thalamocortical inputs in infragranular layers of the auditory cortex. *Scientific Reports*, 9(1), 1-15.

Budinger, E., Heil, P., Hess, A., & Scheich, H. (2006). Multisensory processing via early cortical stages: connections of the primary auditory cortical field with other sensory systems. *Neuroscience*, 143(4), 1065-1083.

Budinger, E., Laszcz, A., Lison, H., Scheich, H., & Ohl, F. W. (2008). Non-sensory cortical and subcortical connections of the primary auditory cortex in Mongolian gerbils: bottom-up and top-down processing of neuronal information via field AI. *Brain Research*, 1220, 2-32.

Budinger, E., & Scheich, H. (2009). Anatomical connections suitable for the direct processing of neuronal information of different modalities via the rodent primary auditory cortex. *Hearing Research*, 258(1-2), 16-27.

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(11), 649-657.

Carney, L. H., Sarkar, S., Abrams, K. S., & Idrobo, F. (2011). Sound-localization ability of the Mongolian gerbil (*Meriones unguiculatus*) in a task with a simplified response map. *Hearing research*, 275(1-2), 89-95.

Cecere, R., Rees, G., & Romei, V. (2015). Individual differences in alpha frequency drive crossmodal illusory perception. *Current Biology*, 25(2), 231-235.

Cohen, S. J., & Stackman Jr, R. W. (2015). Assessing rodent hippocampal involvement in the novel object recognition task. A review. *Behavioural Brain Research*, 285, 105-117.

Collignon, O., Charbonneau, G., Peters, F., Nassim, M., Lassonde, M., Lepore, F., ... &

Bertone, A. (2013). Reduced multisensory facilitation in persons with autism. *Cortex*, 49(6), 1704-1710.

Conrey, B., & Pisoni, D. B. (2006). Auditory-visual speech perception and synchrony detection for speech and nonspeech signals. *The Journal of the Acoustical Society of America*, 119(6), 4065-4073.

DeLoss, D. J., & Andersen, G. J. (2015). Aging, spatial disparity, and the sound-induced flash illusion. *PLoS One*, 10(11), e0143773.

Diederich, A., & Colonius, H. (2004). Bimodal and trimodal multisensory enhancement: effects of stimulus onset and intensity on reaction time. *Perception & Psychophysics*, 66(8), 1388-1404.

Dodart, J. C., Mathis, C., & Ungerer, A. (1997). Scopolamine-induced deficits in a two-trial object recognition task in mice. *Neuroreport*, 8(5), 1173-1178.

Dräger, U. C. (1975). Receptive fields of single cells and topography in mouse visual cortex. *Journal of Comparative Neurology*, 160(3), 269-289.

Durrant, J. D., & Lovrinic, J. H. (1995). *Bases of hearing science*. Williams & Wilkins.

Eckert, M. A., Kamdar, N. V., Chang, C. E., Beckmann, C. F., Greicius, M. D., & Menon, V. (2008). A cross-modal system linking primary auditory and visual cortices: Evidence from intrinsic fMRI connectivity analysis. *Human Brain Mapping*, 29(7), 848-857.

Ennaceur, A. (2010). One-trial object recognition in rats and mice: methodological and theoretical issues. *Behavioural Brain Research*, 215(2), 244-254.

Ennaceur, A., & Delacour, J. (1988). A new one-trial test for neurobiological studies of memory in rats. 1: Behavioral data. *Behavioural Brain Research*, 31(1), 47-59.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415(6870), 429-433.

Fabre, P. H., Hautier, L., Dimitrov, D., & Douzery, E. J. (2012). A glimpse on the pattern of rodent diversification: a phylogenetic approach. *BMC Evolutionary Biology*, 12(1), 1-19.

Feng, Y., Yin, S., Kiefte, M., & Wang, J. (2010). Temporal resolution in regions of normal hearing and speech perception in noise for adults with sloping high-frequency hearing loss. *Ear and hearing*, 31(1), 115-125.

Foss-Feig, J. H., Kwakye, L. D., Cascio, C. J., Burnette, C. P., Kadivar, H., Stone, W. L., & Wallace, M. T. (2010). An extended multisensory temporal binding window in autism spectrum disorders. *Experimental Brain Research*, 203(2), 381-389.

Frassinetti, F., Bolognini, N., & Làdavas, E. (2002). Enhancement of visual perception by crossmodal visuo-auditory interaction. *Experimental Brain Research*, 147(3), 332-343.

Garbers, C., Henke, J., Leibold, C., Wachtler, T., & Thurley, K. (2015). Contextual processing of brightness and color in Mongolian gerbils. *Journal of Vision*, 15(1), 13-13.

Gieseler, A., Tahden, M. A., Thiel, C. M., & Colonius, H. (2018). Does hearing aid use affect audiovisual integration in mild hearing impairment?. *Experimental Brain Research*, 236(4), 1161-1179.

Gingras, G., Rowland, B. A., & Stein, B. E. (2009). The differing impact of multisensory and unisensory integration on behavior. *Journal of Neuroscience*, 29(15), 4897-4902.

Gleiss, S., & Kayser, C. (2012). Audio-visual detection benefits in the rat. *PLoS One*, 7:e45677.

Govardovskii, V. I., Ro, P., Sze, A., & Khokhlova, T. V. (1992). Cones in the retina of the Mongolian gerbil, *Meriones unguiculatus*: an immunocytochemical and electrophysiological study. *Vision Research*, 32(1), 19-27.

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 voices in the McGurk effect. *Perception & Psychophysics*, 50(6), 524-536.

Harrington, L. K., & Peck, C. K. (1998). Spatial disparity affects visual-auditory interactions in human sensorimotor processing. *Experimental Brain Research*, 122(2), 247-252.

Harris, K. C., Eckert, M. A., Ahlstrom, J. B., & Dubno, J. R. (2010). Age-related differences in gap detection: Effects of task difficulty and cognitive ability. *Hearing Research*, 264(1-2), 21-29.

Henschke, J. U., Noesselt, T., Scheich, H., & Budinger, E. (2015). Possible anatomical pathways for short-latency multisensory integration processes in primary sensory cortices. *Brain Structure and Function*, 220(2), 955-977.

Hershenson, M. (1962). Reaction time as a measure of intersensory facilitation. *Journal of Experimental Psychology*, 63(3), 289.

Hishida, R., Kudoh, M., & Shibuki, K. (2014). Multimodal cortical sensory pathways revealed by sequential transcranial electrical stimulation in mice. *Neuroscience Research*, 87, 49-55.

Hishida, R., Horie, M., Tsukano, H., Tohmi, M., Yoshitake, K., Meguro, R., ... & Shibuki, K. (2019). Feedback inhibition derived from the posterior parietal cortex regulates the neural properties of the mouse visual cortex. *European Journal of Neuroscience*, 50(6), 2970-2987.

Hirokawa, J., Bosch, M., Sakata, S., Sakurai, Y., & Yamamori, T. (2008). Functional role of the secondary visual cortex in multisensory facilitation in rats. *Neuroscience*, 153(4), 1402-1417.

Hirst, R. J., McGovern, D. P., Setti, A., Shams, L., & Newell, F. N. (2020). What you see is what you hear: Twenty years of research using the Sound-Induced Flash Illusion. *Neuroscience & Biobehavioral Reviews*, 118, 759-774.

Hirst, R. J., Setti, A., Kenny, R. A., & Newell, F. N. (2019). Age-related sensory decline mediates the Sound-Induced Flash Illusion: Evidence for reliability weighting models of multisensory perception. *Scientific Reports*, 9(1), 1-12.

Huang, J., Wang, E., Lu, K., Wang, A., & Zhang, M. (2021). Long-term training reduces the responses to the sound-induced flash illusion. *Attention, Perception, & Psychophysics*, 1-11.

Ibrahim, L. A., Mesik, L., Ji, X. Y., Fang, Q., Li, H. F., Li, Y. T., ... & Tao, H. W. (2016). Cross-modality sharpening of visual cortical processing through layer-1-mediated inhibition and disinhibition. *Neuron*, 89(5), 1031-1045.

Innes-Brown, H., & Crewther, D. (2009). The impact of spatial incongruence on an auditory-visual illusion. *PloS One*, 4(7), e6450.

Inui, K., & Kakigi, R. (2006). Temporal analysis of the flow from V1 to the extrastriate cortex in humans. *Journal of neurophysiology*, 96(2), 775-784.

Inui, K., Okamoto, H., Miki, K., Gunji, A., & Kakigi, R. (2006). Serial and parallel processing in the human auditory cortex: a magnetoencephalographic study. *Cerebral Cortex*, 16(1), 18-30.

Irwin, R. J., & McAuley, S. F. (1987). Relations among temporal acuity, hearing loss, and the perception of speech distorted by noise and reverberation. *The Journal* *of the Acoustical Society of America*, 81(5), 1557-1565.

Iurilli, G., Ghezzi, D., Olcese, U., Lassi, G., Nazzaro, C., Tonini, R., ... & Medini, P. (2012). Sound-driven synaptic inhibition in primary visual cortex. *Neuron*, 73(4), 814-828.

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(3), 967-979.

Jacklin, D. L., Cloke, J. M., Potvin, A., Garrett, I., & Winters, B. D. (2016). The dynamic multisensory engram: neural circuitry underlying crossmodal object recognition in rats changes with the nature of object experience. *Journal of Neuroscience*, 36(4), 1273-1289.

Jacobs, G. H., & Deegan II, J. F. (1994). Sensitivity to ultraviolet light in the gerbil (*Meriones unguiculatus*): characteristics and mechanisms. *Vision Research*, 34(11), 1433-1441.

Jones, J. A., & Callan, D. E. (2003). Brain activity during audiovisual speech perception: an fMRI study of the McGurk effect. *Neuroreport*, 14(8), 1129-1133.

Jones, J. A., & Munhall, K. G. (1997). Effects of separating auditory and visual sources on audiovisual integration of speech. *Canadian Acoustics*, 25(4), 13-19.

Jordan, K. E., Brannon, E. M., Logothetis, N. K., & Ghazanfar, A. A. (2005). Monkeys match the number of voices they hear to the number of faces they see. *Current Biology*, 15(11), 1034-1038.

Keil, J. (2020). Double flash illusions: current findings and future directions. *Frontiers in Neuroscience*, 14, 298.

Keplinger, S., Beiderbeck, B., Michalakis, S., Biel, M., Grothe, B., & Kunz, L. (2018).

Optogenetic control of neural circuits in the Mongolian gerbil. *Frontiers in Cellular Neuroscience*, 12, 111.

King, A. J. (2005). Multisensory integration: strategies for synchronization. *Current Biology*, 15(9), R339-R341.

Kitaura, H., Hishida, R., & Shibuki, K. (2010). Transcranial imaging of somatotopic map plasticity after tail cut in mice. *Brain Research*, 1319, 54-59.

Kobayasi, K. I., & Riquimaroux, H. (2013). Audiovisual integration in the primary auditory cortex of an awake rodent. *Neuroscience Letters*, 534, 24-29.

Kobayasi, K. I., Usami, A., & Riquimaroux, H. (2012). Behavioral evidence for auditory induction in a species of rodent: Mongolian gerbil (*Meriones unguiculatus*). *The Journal of the Acoustical Society of America*, 132(6), 4063-4068.

Komagata, S., Chen, S., Suzuki, A., Yamashita, H., Hishida, R., Maeda, T., ... & Shibuki, K. (2011). Initial phase of neuropathic pain within a few hours after nerve injury in mice. *Journal of Neuroscience*, 31(13), 4896-4905.

Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., & Shams, L. (2007). Causal inference in multisensory perception. *PLoS One*, 2(9), e943.

Krumbholz, K., Patterson, R. D., Seither-Preisler, A., Lammertmann, C., & Lütkenhöner, B. (2003). Neuromagnetic evidence for a pitch processing center in Heschl's gyrus. *Cerebral Cortex*, 13(7), 765-772.

Kuroki, S., Yoshida, T., Tsutsui, H., Iwama, M., Ando, R., Michikawa, T., ... & Itohara, S. (2018). Excitatory neuronal hubs configure multisensory integration of slow waves in association cortex. *Cell Reports*, 22(11), 2873-2885.

Lovelace, C. T., Stein, B. E., & Wallace, M. T. (2003). An irrelevant light enhances auditory detection in humans: a psychophysical analysis of multisensory integration in stimulus detection. *Cognitive Brain Research*, 17(2), 447-453.

Maier, J. K., & Klump, G. M. (2006). Resolution in azimuth sound localization in the Mongolian gerbil (*Meriones unguiculatus*). *The Journal of the Acoustical Society of America*, 119(2), 1029-1036.

Mallick, D. B., Magnotti, J. F., & Beauchamp, M. S. (2015). Variability and stability in the McGurk effect: contributions of participants, stimuli, time, and response type. *Psychonomic Bulletin & Review*, 22(5), 1299-1307.

Marcus, G. F., Vijayan, S., Rao, S. B., & Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science*, 283(5398), 77-80.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264(5588), 746-748.

Meijer, G. T., Marchesi, P., Mejias, J. F., Montijn, J. S., Lansink, C. S., & Pennartz, C. M. (2020). Neural correlates of multisensory detection behavior: comparison of primary and higher-order visual cortex. *Cell Reports*, 31(6), 107636.

Meijer, G. T., Pie, J. L., Dolman, T. L., Pennartz, C., & Lansink, C. S. (2018). Audiovisual integration enhances stimulus detection performance in mice. *Frontiers in Behavioral Neuroscience*, 12, 231.

Meredith, M. A., Nemitz, J. W., & Stein, B. E. (1987). Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. *Journal of Neuroscience*, 7(10), 3215-3229.

Meredith, M. A., & Stein, B. E. (1983). Interactions among converging sensory inputs in the superior colliculus. *Science*, 221(4608), 389-391.

Meredith, M. A., & Stein, B. E. (1986a). Spatial factors determine the activity of multisensory neurons in cat superior colliculus. *Brain Research*, 365(2), 350-354.

Meredith, M. A., & Stein, B. E. (1986b). Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *Journal of Neurophysiology*, 56(3), 640-662.

Merzenich, M. M., Knight, P. L., & Roth, G. L. (1975). Representation of cochlea within primary auditory cortex in the cat. *Journal of neurophysiology*, 38(2), 231-249.

Messier, C. (1997). Object recognition in mice: improvement of memory by glucose. *Neurobiology of Learning and Memory*, 67(2), 172-175.

Mishra, J., Martinez, A., Sejnowski, T. J., & Hillyard, S. A. (2007). Early cross-modal interactions in auditory and visual cortex underlie a sound-induced visual illusion. *Journal of Neuroscience*, 27(15), 4120-4131.

Mongillo, E. A., Irwin, J. R., Whalen, D. H., Klaiman, C., Carter, A. S., & Schultz, R. T. (2008). Audiovisual processing in children with and without autism spectrum disorders. *Journal of autism and developmental disorders*, 38(7), 1349-1358.

Moore, B. C. (2012). *An introduction to the psychology of hearing*. Brill.

Moore, D., Benenson, J., Reznick, J. S., Peterson, M., & Kagan, J. (1987). Effect of auditory numerical information on infants' looking behavior: Contradictory evidence. *Developmental Psychology*, 23(5), 665.

Moore, B. C., & Glasberg, B. R. (1983). Growth of forward masking for sinusoidal and noise maskers as a function of signal delay; implications for suppression in noise. *The Journal of the Acoustical Society of America*, 73(4), 1249-1259.

Moore, B. C., & Glasberg, B. R. (1988). Gap detection with sinusoids and noise in normal, impaired, and electrically stimulated ears. *The Journal of the Acoustical Society of America*, 83(3), 1093-1101.

Moore, B. C., Glasberg, B. R., Plack, C. J., & Biswas, A. K. (1988). The shape of the ear's temporal window. *The Journal of the Acoustical Society of America*, 83(3), 1102-1116.

Moore, B. C., Peters, R. W., & Glasberg, B. R. (1992). Detection of temporal gaps in sinusoids by elderly subjects with and without hearing loss. *The Journal of the Acoustical Society of America*, 92(4), 1923-1932.

Morrill, R. J., & Hasenstaub, A. R. (2018). Visual information present in infragranular layers of mouse auditory cortex. *Journal of Neuroscience*, 38(11), 2854-2862.

Munhall, K. G., Gribble, P., Sacco, L., & Ward, M. (1996). Temporal constraints on the McGurk effect. *Perception & Psychophysics*, 58(3), 351-362.

Narita, N., Kato, M., Tazoe, M., Miyazaki, K., Narita, M., & Okado, N. (2002). Increased monoamine concentration in the brain and blood of fetal thalidomide-and valproic acid–exposed rat: putative animal models for autism. *Pediatric Research*, 52(4), 576-579.

Nishiyama, K., Kobayasi, K. I., & Riquimaroux, H. (2011). Vocalization control in Mongolian gerbils (*Meriones unguiculatus*) during locomotion behavior. *The Journal of the Acoustical Society of America*, 130(6), 4148-4157.

Noesselt, T., Rieger, J. W., Schoenfeld, M. A., Kanowski, M., Hinrichs, H., Heinze, H. J., & Driver, J. (2007). Audiovisual temporal correspondence modulates human multisensory superior temporal sulcus plus primary sensory cortices. *Journal of Neuroscience*, 27(42), 11431-11441.

Odegaard, B., Wozny, D. R., & Shams, L. (2016). The effects of selective and divided attention on sensory precision and integration. *Neuroscience Letters*, 614, 24-28.

Ohl, F. W., & Scheich, H. (1997). Orderly cortical representation of vowels based on formant interaction. *Proceedings of the National Academy of Sciences*, 94(17), 9440-9444.

Paton, B., Hohwy, J., & Enticott, P. G. (2012). The rubber hand illusion reveals proprioceptive and sensorimotor differences in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 42(9), 1870-1883.

Pérez-Bellido, A., Ernst, M. O., Soto-Faraco, S., & López-Moliner, J. (2015). Visual limitations shape audio-visual integration. *Journal of Vision*, 15(14), 5-5.

Pietrewicz, A. T., Hoff, M. P., & Higgins, S. A. (1982). Activity rhythms in the Mongolian gerbil under natural light conditions. *Physiology & Behavior*, 29(2), 377-380.

Pinheiro, M. M. C., Dias, K. Z., & Pereira, L. D. (2012). Acoustic stimulation effect on temporal processing skills in elderly subjects before and after hearing aid fitting. *Brazilian Journal of Otorhinolaryngology*, 78, 09-16.

Radtke-Schuller, S., Schuller, G., Angenstein, F., Grosser, O. S., Goldschmidt, J., & Budinger, E. (2016). Brain atlas of the Mongolian gerbil (*Meriones unguiculatus*) in CT/MRI-aided stereotaxic coordinates. *Brain Structure and Function*, 221(1), 1-272.

Refinetti, R. (2006). Variability of diurnality in laboratory rodents. *Journal of Comparative Physiology A*, 192(7), 701-714.

Robertson, C. E., & Baron-Cohen, S. (2017). Sensory perception in autism. *Nature Reviews Neuroscience*, 18(11), 671-684.

Rose, M. M., & Moore, B. C. (2000). Effects of frequency and level on auditory stream segregation. *The Journal of the Acoustical Society of America*, 108(3), 1209-1214.

Roseboom, W., Kawabe, T., & Nishida, S. Y. (2013). The cross-modal double flash illusion depends on featural similarity between cross-modal inducers. *Scientific Reports*, 3(1), 1-5.

Rosenblum, L. D., & Saldaña, H. M. (1996). An audiovisual test of kinematic primitives for visual speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, 22(2), 318.

Rosenthal, O., Shimojo, S., & Shams, L. (2009). Sound-induced flash illusion is resistant to feedback training. *Brain Topography*, 21(3), 185-192.

Ross, L. A., Saint-Amour, D., Leavitt, V. M., Javitt, D. C., & Foxe, J. J. (2007). Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. *Cerebral Cortex*, 17(5), 1147-1153.

Ryan, A. (1976). Hearing sensitivity of the mongolian gerbil, *Meriones unguiculatis*. *The Journal of the Acoustical Society of America*, 59(5), 1222-1226.

Sakai, M., & Suga, N. (2002). Centripetal and centrifugal reorganizations of frequency map of auditory cortex in gerbils. *Proceedings of the National Academy of Sciences*, 99(10), 7108-7112.

Sakata, S., Yamamori, T., & Sakurai, Y. (2004). Behavioral studies of auditory-visual spatial recognition and integration in rats. *Experimental Brain Research*, 159(4), 409-417.

Schormans, A. L., Scott, K. E., Vo, A. M., Tyker, A., Typlt, M., Stolzberg, D., & Allman, B. L. (2017). Audiovisual temporal processing and synchrony perception in the rat. *Frontiers in Behavioral Neuroscience*, 10, 246.

Schuett, S., Bonhoeffer, T., & Hübener, M. (2002). Mapping retinotopic structure in mouse visual cortex with optical imaging. *Journal of Neuroscience*, 22(15), 6549-6559.

Sekiyama, K., & Tohkura, Y. I. (1991). McGurk effect in non - English listeners: Few visual effects for Japanese subjects hearing Japanese syllables of high auditory intelligibility. *The Journal of the Acoustical Society of America*, 90(4), 1797-1805.

Sekuler, R., Sekuler, A. B., & Lau, R. (1997). Sound alters visual motion perception. *Nature*, 385(6614), 308.

Setti, A., & Chan, J. S. (2011). Familiarity of objects affects susceptibility to the sound-induced flash illusion. *Neuroscience Letters*, 492(1), 19-22.

Shailer, M. J., & Moore, B. C. (1987). Gap detection and the auditory filter: Phase effects using sinusoidal stimuli. *The Journal of the Acoustical Society of America*, 81(4), 1110-1117.

Shams, L., Iwaki, S., Chawla, A., & Bhattacharya, J. (2005b). Early modulation of visual cortex by sound: an MEG study. *Neuroscience letters*, 378(2), 76-81.

Shams, L., Kamitani, Y., & Shimojo, S. (2000). What you see is what you hear. *Nature*, 408(6814), 788-788.

Shams, L., Kamitani, Y., & Shimojo, S. (2002). Visual illusion induced by sound. *Cognitive Brain Research*, 14(1), 147-152.

Shams, L., Kamitani, Y., Thompson, S., & Shimojo, S. (2001). Sound alters visual evoked potentials in humans. *Neuroreport*, 12(17), 3849-3852.

Shams, L., Ma, W. J., & Beierholm, U. (2005a). Sound-induced flash illusion as an optimal percept. *Neuroreport*, 16(17), 1923-1927.

Shams, L., & Seitz, A. R. (2008). Benefits of multisensory learning. *Trends in Cognitive Sciences*, 12(11), 411-417.

Shibuki, K., Hishida, R., Murakami, H., Kudoh, M., Kawaguchi, T., Watanabe, M., ... & Tanaka, R. (2003). Dynamic imaging of somatosensory cortical activity in the rat visualized by flavoprotein autofluorescence. *The Journal of Physiology*, 549(3), 919-927.

Shibuki, K., Ono, K., Hishida, R., & Kudoh, M. (2006). Endogenous fluorescence imaging of somatosensory cortical activities after discrimination learning in rats. *Neuroimage*, 30(3), 735-744.

Siemann, J. K., Muller, C. L., Bamberger, G., Allison, J. D., Veenstra-VanderWeele, J., & Wallace, M. T. (2015). A novel behavioral paradigm to assess multisensory processing in mice. *Frontiers in Behavioral Neuroscience*, 8, 456.

Siemann, J. K., Muller, C. L., Forsberg, C. G., Blakely, R. D., Veenstra-VanderWeele, J., & Wallace, M. T. (2017). An autism-associated serotonin transporter variant disrupts multisensory processing. *Translational Psychiatry*, 7(3), e1067-e1067.

Sinnott, J. M., Street, S. L., Mosteller, K. W., & Williamson, T. L. (1997). Behavioral measures of vowel sensitivity in Mongolian gerbils (Meriones unguiculatus): effects of age and genetic origin. *Hearing Research*, 112(1-2), 235-246.

Sliwa, J., Duhamel, J. R., Pascalis, O., & Wirth, S. (2011). Spontaneous voice–face identity matching by rhesus monkeys for familiar conspecifics and humans. *Proceedings of the National Academy of Sciences*, 108(4), 1735-1740.

Smith, E. G., & Bennetto, L. (2007). Audiovisual speech integration and lipreading in autism. *Journal of Child Psychology and Psychiatry*, 48(8), 813-821.

Song, Y. H., Kim, J. H., Jeong, H. W., Choi, I., Jeong, D., Kim, K., & Lee, S. H. (2017).

A neural circuit for auditory dominance over visual perception. *Neuron*, 93(4), 940-954.

Spence, C. (2013). Just how important is spatial coincidence to multisensory integration? Evaluating the spatial rule. *Annals of the New York Academy of Sciences*, 1296(1), 31-49.

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(6), 497-506.

Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: current issues from the perspective of the single neuron. *Nature Reviews Neuroscience*, 9(4), 255-266.

Stein, B. E., Stanford, T. R., & Rowland, B. A. (2014). Development of multisensory integration from the perspective of the individual neuron. *Nature Reviews Neuroscience*, 15(8), 520-535.

Stein, B. E., & Wallace, M. T. (1996). Comparisons of cross-modality integration in midbrain and cortex. *Progress in Brain Research*, 112, 289-299.

Stevenson, R. A., & James, T. W. (2009). Audiovisual integration in human superior temporal sulcus: Inverse effectiveness and the neural processing of speech and object recognition. *Neuroimage*, 44(3), 1210-1223.

Stevenson, R. A., Siemann, J. K., Schneider, B. C., Eberly, H. E., Woynaroski, T. G., Camarata, S. M., & Wallace, M. T. (2014b). Multisensory temporal integration in autism spectrum disorders. *Journal of Neuroscience*, 34(3), 691-697.

Stevenson, R. A., Siemann, J. K., Woynaroski, T. G., Schneider, B. C., Eberly, H. E., Camarata, S. M., & Wallace, M. T. (2014a). Brief report: Arrested development of audiovisual speech perception in autism spectrum disorders. *Journal of*

Autism and Developmental Disorders, 44(6), 1470-1477.

Stevenson, R. A., Siemann, J. K., Woynaroski, T. G., Schneider, B. C., Eberly, H. E., Camarata, S. M., & Wallace, M. T. (2014c). Evidence for diminished multisensory integration in autism spectrum disorders. *Journal of autism and developmental disorders*, 44(12), 3161-3167.

Stevenson, R. A., Zemtsov, R. K., & Wallace, M. T. (2012). Individual differences in the multisensory temporal binding window predict susceptibility to audiovisual illusions. *Journal of Experimental Psychology: Human Perception and Performance*, 38(6), 1517.

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *The Journal of the Acoustical Society of America*, 26(2), 212-215.

Takahashi, K., Hishida, R., Kubota, Y., Kudoh, M., Takahashi, S., & Shibuki, K. (2006). Transcranial fluorescence imaging of auditory cortical plasticity regulated by acoustic environments in mice. *European Journal of Neuroscience*, 23(5), 1365-1376.

Takeshima, Y. (2020). Emotional information affects fission illusion induced by audio-visual interactions. *Scientific Reports*, 10(1), 1-10.

Takeshima, Y., & Gyoba, J. (2013). Complexity of visual stimuli affects visual illusion induced by sound. *Vision Research*, 91, 1-7.

Takeshima, Y., & Gyoba, J. (2015). Spatial frequency modulates the degree of illusory second flash perception. *Multisensory Research*, 28(1-2), 1-10.

Tamai, Y., Ito, Y., Furuyama, T., Horinouchi, K., Murashima, N., Michimoto, I., ... & Kobayasi, K. I. (2020). Auditory cortical activity elicited by infrared laser irradiation from the outer ear in Mongolian gerbils. *PloS One*, 15(10), e0240227.

Teder-Sälejärvi, W. A., Di Russo, F., McDonald, J. J., & Hillyard, S. A. (2005). Effects of spatial congruity on audio-visual multimodal integration. *Journal of Cognitive Neuroscience*, 17(9), 1396-1409.

Tohmi, M., Kitaura, H., Komagata, S., Kudoh, M., & Shibuki, K. (2006). Enduring critical period plasticity visualized by transcranial flavoprotein imaging in mouse primary visual cortex. *Journal of Neuroscience*, 26(45), 11775-11785.

Tohmi, M., Meguro, R., Tsukano, H., Hishida, R., & Shibuki, K. (2014). The extrageniculate visual pathway generates distinct response properties in the higher visual areas of mice. *Current Biology*, 24(6), 587-597.

Tsukano, H., Horie, M., Bo, T., Uchimura, A., Hishida, R., Kudoh, M., ... & Shibuki, K. (2015). Delineation of a frequency-organized region isolated from the mouse primary auditory cortex. *Journal of neurophysiology*, 113(7), 2900-2920.

Tyler, R. S., Summerfield, Q., Wood, E. J., & Fernandes, M. A. (1982). Psychoacoustic and phonetic temporal processing in normal and hearing-impaired listeners. *The*

Uekita, T., & Okanoya, K. (2011). Hippocampus lesions induced deficits in social and spatial recognition in Octodon degus. *Behavioural Brain Besearch*, 219(2),

Van Wassenhove, V., Grant, K. W., & Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia*, 45(3), 598-607.

Vasconcelos, N., Pantoja, J., Belchior, H., Caixeta, F. V., Faber, J., Freire, M. A. M., ... & Ribeiro, S. (2011). Cross-modal responses in the primary visual cortex encode complex objects and correlate with tactile discrimination. *Proceedings of the National Academy of Sciences*, 108(37), 15408-15413.

Viemeister, N. F. (1979). Temporal modulation transfer functions based upon modulation thresholds. *The Journal of the Acoustical Society of America*, 66(5), 1364-1380.

Violentyev, A., Shimojo, S., & Shams, L. (2005). Touch-induced visual illusion. *Neuroreport*, 16(10), 1107-1110.

Wada, M., Ide, M., Atsumi, T., Sano, Y., Shinoda, Y., Furuichi, T., & Kansaku, K. (2019). Rubber tail illusion is weakened in Ca 2+-dependent activator protein for secretion 2 (Caps 2)-knockout mice. *Scientific Reports*, 9(1), 1-9.

Wada, M., Takano, K., Ora, H., Ide, M., & Kansaku, K. (2016). The rubber tail illusion as evidence of body ownership in mice. *Journal of Neuroscience*, 36(43), 11133-11137.

Wallace, M. T., Meredith, M. A., & Stein, B. E. (1992). Integration of multiple sensory modalities in cat cortex. *Experimental Brain Research*, 91(3), 484-488.

Wallace, M. T., Ramachandran, R., & Stein, B. E. (2004). A revised view of sensory cortical parcellation. *Proceedings of the National Academy of Sciences*, 101(7), 2167-2172.

Watkins, S., Shams, L., Josephs, O., & Rees, G. (2007). Activity in human V1 follows multisensory perception. *Neuroimage*, 37(2), 572-578.

Watkins, S., Shams, L., Tanaka, S., Haynes, J. D., & Rees, G. (2006). Sound alters activity in human V1 in association with illusory visual perception. *Neuroimage*, 31(3), 1247-1256.

Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological bulletin*, 88(3), 638.

Winters, B. D., & Reid, J. M. (2010). A distributed cortical representation underlies crossmodal object recognition in rats. *Journal of Neuroscience*, 30(18), 6253-6261.

Winters, B. D., Saksida, L. M., & Bussey, T. J. (2008). Object recognition memory: neurobiological mechanisms of encoding, consolidation and retrieval. *Neuroscience & Biobehavioral Reviews*, 32(5), 1055-1070.

Wrobel, C., Dieter, A., Huet, A., Keppeler, D., Duque-Afonso, C. J., Vogl, C., ... & Moser, T. (2018). Optogenetic stimulation of cochlear neurons activates the auditory pathway and restores auditory-driven behavior in deaf adult gerbils. *Science Translational Medicine*, 10(449).

Yaguchi, A., & Hidaka, S. (2018). Distinct autistic traits are differentially associated with the width of the multisensory temporal binding window. *Multisensory Research*, 31(6), 523-536.

Yang, S., Luo, X., Xiong, G., So, K. F., Yang, H., & Xu, Y. (2015). The electroretinogram of Mongolian gerbil (*Meriones unguiculatus*): comparison to mouse. *Neuroscience Letters*, 589, 7-12.

Yanagawa, Y., Osanai, H., & Tateno, T. (2016). Transcranial flavoprotein-autofluorescence imaging of sound-evoked responses in the mouse auditory cortex under three types of anesthesia. *Neuroscience Letters*, 633, 189-195.

Yao, J. D., Gimoto, J., Constantinople, C. M., & Sanes, D. H. (2020). Parietal cortex is required for the integration of acoustic evidence. *Current Biology*, 30(17), 3293-3303.

Yoshitake, K., Tsukano, H., Tohmi, M., Komagata, S., Hishida, R., Yagi, T., & Shibuki, K. (2013). Visual map shifts based on whisker-guided cues in the young mouse visual cortex. *Cell Reports*, 5(5), 1365-1374.

Zampini, M., Shore, D. I., & Spence, C. (2003). Audiovisual temporal order judgments. *Experimental Brain Research*, 152(2), 198-210.

Zhang, N., & Chen, W. (2006). A dynamic fMRI study of illusory double-flash effect on human visual cortex. *Experimental Brain Research*, 172(1), 57-66.

Zorio, D. A., Monsma, S., Sanes, D. H., Golding, N. L., Rubel, E. W., & Wang, Y. (2019). De novo sequencing and initial annotation of the Mongolian gerbil (*Meriones unguiculatus*) genome. *Genomics*, 111(3), 441-449.

Curriculum Vitae

Yuki Ito

Data of Birth: July 18, 1994
Place of Birth: Fukushima, Japan

Education

2017	Doshisha University, Kyotanabe B.S. in Engineering
2019	Doshisha University, Kyotanabe M.S. in Engineering

Publications

1. **Ito, Y**., Sato, R., Tamai, Y., Hiryu, S., Uekita, T., & Kobayasi, K. I. (2019). Auditory-induced visual illusions in rodents measured by spontaneous behavioural response. *Scientific Reports*, 9(1), 1-7.
2. **Ito, Y**., Masago, Y., Furuyama, T., & Kobayasi, K. I. (2020). How frequency processing affects the sound-induced flash illusion?. *Acoustical Science and Technology*, 41(1), 334-336.
3. **Ito, Y**., Sato, R., & Kobayasi, K. I. (2020). Application of the novel object recognition paradigm to time-varying sensory stimulus. *Acoustical Science and Technology*, 41(1), 388-389.
4. Tamai, Y., **Ito, Y**., Furuyama, T., Horinouchi, K., Murashima, N., Michimoto, I., ... & Kobayasi, K. I. (2020). Auditory cortical activity elicited by infrared laser irradiation from the outer ear in Mongolian gerbils. *PloS One*, 15(10), e0240227.

Professional memberships

The Japanese Society for Animal Psychology
The Japanese Society for Comparative Physiology and Biochemistry
The Japan Neuroscience Society
Society for Neuroscience

November 2021