

## Effect of visual stimuli on temporal order judgments of a sequence of pure tones

Shoya Tanabe

Graduate School of Science and Technology, Niigata University, 8050, Ikarashi 2-Nocho, Nishi-ku, Niigata 950-2181, Japan;  
e-mail: [f11c103b@alumni.niigata-u.ac.jp](mailto:f11c103b@alumni.niigata-u.ac.jp)

Mamoru Iwaki

Graduate School of Science and Technology, Niigata University, 8050, Ikarashi 2-Nocho, Nishi-ku, Niigata 950-2181, Japan;  
e-mail: [iwaki@eng.niigata-u.ac.jp](mailto:iwaki@eng.niigata-u.ac.jp)

Received 11 March 2012; in revised form 29 March 2013; published online 3 June 2013.

**Abstract.** We investigated what effect visual spatial information had on auditory temporal order judgments (TOJs) and examined what effect visual stimuli had on the TOJs of sequences of pure tones in experiment 1. The auditory stimuli were sequences of four distinct pure tones. The visual stimuli consisted of two vertically aligned flashes: one flashed before the first tone and the other flashed after the last tone. Participants judged whether the temporal order of the second and third tones in auditory stimuli occurred with the higher tone being first or the lower tone being first. As a result, the proportion of responses for higher-tone-first increased when the flash of the upper LED preceded that of the lower LED, independent of the actual temporal order. Participants in experiment 2 were asked to make simultaneity judgments instead, which were also affected by visual stimuli. The auditory stimuli in experiment 3 were the same as those in experiment 1, whereas the visual stimuli consisted of two horizontally aligned flashes. Furthermore, the participants made TOJs, which were not affected by the horizontally aligned visual stimuli. We concluded that vertically aligned visual stimuli had an effect on auditory TOJs with some response bias.

**Keywords:** audio-visual interaction, temporal order judgment, pure tone sequence, visual flash.

### 1 Introduction

The perception of the temporal order of external successive events does not always follow its physical temporal order. The perceived temporal order of external successive events of a sensory modality could be affected by the stimuli of another sensory modality. In regard to vision and audition, for example, Morein-Zamir, Soto-Faraco, and Kingstone (2003) found that visual temporal order judgments (TOJs) were improved by brief tones. Also, Van der Burg, Olivers, Bronkhorst, and Theeuwes (2008) demonstrated that visual TOJs were affected when a preceding color change in a distractor was accompanied by a spatially non-specific auditory signal. Moreover, Harrar and Harris (2008) discovered that the audio-visual TOJs were modulated after exposure to time-staggered sound–light stimuli.

Teramoto, Watanabe, Umemura, and Kita (2008) recently reported that self-motion induced by a large field is able to affect the perception of the temporal order of tones. They used a rotating virtual cylinder with random dots as the visual stimuli and two successive sounds as the auditory stimuli; the first auditory stimulus was presented to the right ear and the second was presented to the left ear. They observed the perceptual bias of auditory TOJs when the participants experienced visually induced self-motion; thus, the perceived temporal order of auditory events was modulated by the direction of visual motion (or self-motion). The sound presented to the ear in the direction opposite to visual motion was specifically perceived prior to the sound presented to the ear in the same direction. They concluded that visual input modifies the perception of the temporal characteristics of sounds. In other words, it can be interpreted as the influence of visual spatial information on auditory spatial perception because their auditory stimuli were spatially informative.

Our aim in this paper is to present an example where visual spatial information can affect the TOJ tasks of auditory stimuli that are spatially uninformative unlike those in Teramoto et al. (2008). The auditory stimuli in our experiments were a sequence of pure tones that were spatially uninformative, since the same auditory stimuli were presented to both ears. The visual stimuli consisted of two flashes, which were vertically or horizontally aligned.

If the TOJ tasks of spatially uninformative auditory stimuli are affected by visual spatial information, this is an interesting finding that is distinct from the modality appropriateness hypothesis (Welch & Warren, 1980) that states that auditory modality is expected to be more dominant than visual modality when a temporal task is solved (e.g., the temporal ventriloquism effect; Morein-Zamir et al., 2003, and for a review, Shimojo et al., 2001) and visual modality is more dominant than auditory modality when a spatial task is solved (e.g., the ventriloquism effect; Jack & Thurlow, 1973).

There have also been several studies showing that auditory timing affects visual spatial perception. For example, Freeman and Driver (2008) found that the timing of static auditory events that is spatially uninformative influenced the perceived direction of visual apparent motion and induced visual aftereffects. Kawabe, Miura, and Yamada (2008) demonstrated that the audio-visual tau effect, i.e. shorter temporal intervals between pure tones, caused the perception of shorter spatial intervals between visual stimuli. Roseboom, Kawabe, and Nishida (2013) discovered that modulation of the direction of apparent visual motion with auditory signals could be accounted for by the grouping of apparently related auditory and visual sequences, and not by temporal capture explanations. However, the direction of the effect we investigated in this paper was opposite to that in these studies. Consequently, our results should contribute to the field of audio-visual interaction studies, especially that for visual spatial and auditory temporal patterns.

We then examined what effect vertically aligned visual stimuli had on TOJs for the two auditory targets of the sequence of pure tones in experiment 1, the effect on SJs in experiment 2 to exclude an alternative explanation for response bias, and the effect of horizontally aligned visual stimuli on TOJs in experiment 3 to verify whether the effect observed in experiment 1 ever occurred due to horizontally aligned visual stimuli.

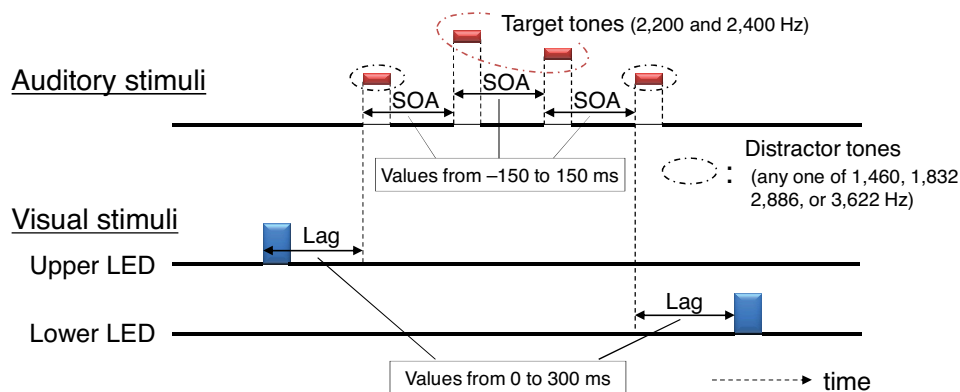
## 2 Experiment 1: Auditory temporal order judgments—Effect of vertically aligned visual stimuli

### 2.1 Participants

Seven males (21–24-years-old) with normal hearing and normal or corrected-to-normal vision participated in this experiment.

### 2.2 Stimuli and apparatus

The auditory stimuli were presented via headphones (Sony, MDR-Z900). They were in a sequence of four distinct pure tones with different frequencies (Figure 1). Each tone lasted for 30 ms (60 dB SPL) including 3-ms linear onset and offset ramps. The frequencies of the second and third tones (target tones) were mutually exclusive at 2,200 or 2,400 Hz, and those of the first and fourth tones (distractor tones) were one out of 1,460, 1,832, 2,886, and 3,622 Hz frequencies. The frequencies of both distractor tones were the same. The stimulus onset asynchronies (SOAs) between one tone and the next were  $-150$ ,  $-110$ ,  $-80$ ,  $-60$ ,  $-30$ ,  $30$ ,  $60$ ,  $80$ ,  $110$ , and  $150$  ms. Here, negative SOAs indicate that the lower tone (2,200 Hz) of target tones was presented first, whereas positive SOAs indicate that the higher tone (2,400 Hz) of target tones was presented first. The auditory system is very sensitive to



**Figure 1.** Schematic of trials in experiment 1. This figure is one example of positive SOA and the upper-LED-preceding condition.

temporal order when the pair of two target tones is solely presented. Therefore, we made the TOJ tasks more difficult by presenting the distractor tones before and after the target tones.

The visual stimuli were presented with two blue LEDs. They were placed 10 cm above and below a red LED. The red LED served as a fixation target and was approximately positioned at eye level. The blue LEDs flashed for 30 ms in two different orders: the upper LED preceded the lower LED (upper-LED-preceding condition) and vice versa (lower-LED-preceding condition). The flashes were presented with one out of three different lags, i.e. 0, 100, and 300 ms before the first tone and after the fourth tone (Figure 1). This lag value was the interval between the first flash and the first tone and also that between the second flash and the fourth tone. This is because there was a possibility that the participants could not perceive the spatial order of LED flashes when the interval between the first and second flashes was very short (thus, the SOA value was very small), if the visual stimuli were presented with the target tones.

### 2.3 Procedure

This experiment was performed in a dimly lit sound-proofed room. The participants were seated about 57 cm from the fixation LED that was approximately positioned at eye level. They pressed a keyboard to indicate which tone was presented first with regard to the second and third tones (i.e., whether the higher tone or the lower tone was first). Each TOJ trial began with the fixation LED being illuminated. The audio-visual stimulus was presented after a period that varied randomly between 1,000 and 1,500 ms for each trial. The fixation LED remained lit until the participants responded. The next trial began 1500 ms after the responses. The participants were informed that they had enough time to make judgments and that they had to watch the fixation LED during trials. This experiment consisted of 840 trials (10 auditory stimuli, 7 visual stimuli, and 12 repetitions), and was divided into 15 blocks of 56 trials. The participants took part in three blocks per day.

### 2.4 Results

The upper row in Figure 2 plots the mean and standard error for the proportion of responses for the higher tone occurring first as a function of SOA, and the sigmoid function  $F(\text{SOA}) = 100/(1 + \exp(-b(\text{SOA} - x_0)))$  fits the experimental data. The inflection point ( $x_0$ ) was taken as the point of subjective simultaneity (PSS). The PSS means that the relative timing of the second and third tones was simultaneous. Half the difference between the SOAs at 75% and 25% of the responses for the higher tone occurring first was taken as the just noticeable difference (JND). The JND means the minimal SOA at which the auditory TOJ could be reliably determined. The JNDs are in the middle row and the PSSs are in the lower row in Figure 2.

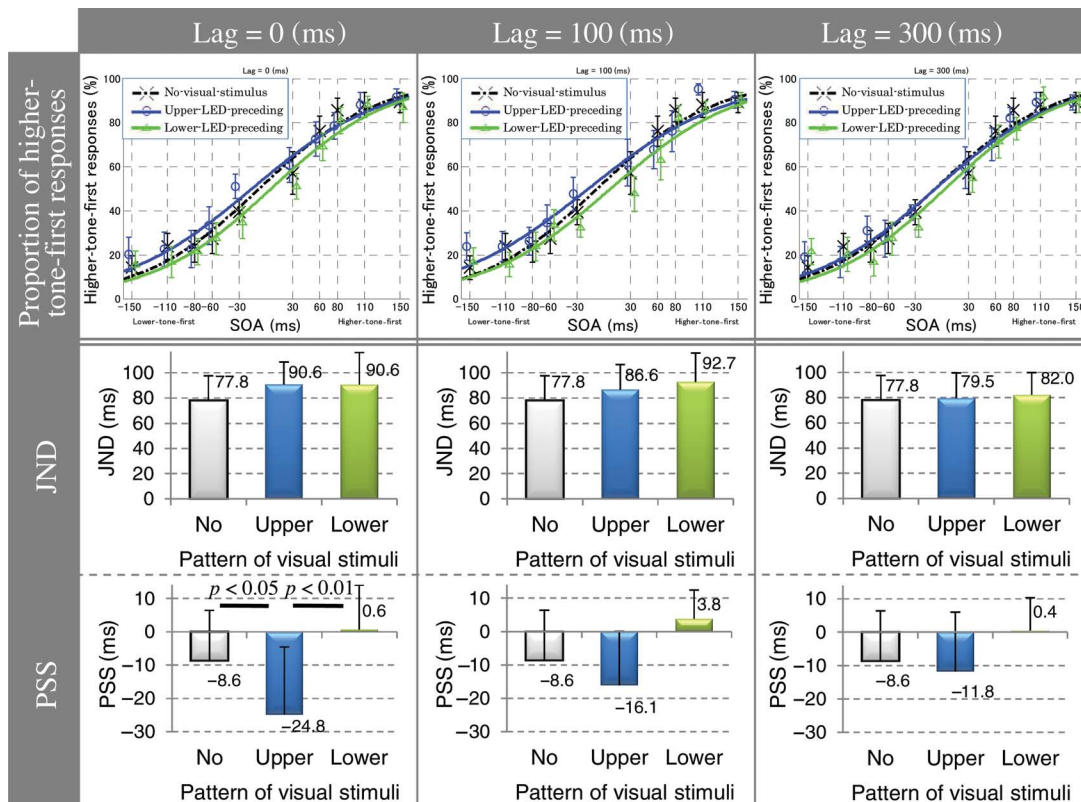
The JNDs and PSSs were submitted to repeated-measures analysis of variance (ANOVA) with the pattern of visual stimuli. The results indicate that there were no significant effects on JNDs at any lag values or for PSS at lags = 100 and 300 ms. More importantly, there were significant effects on PSSs at lag = 0 ms,  $F(2, 12) = 6.18$ , and  $p = 0.014$ . Least significant difference (LSD) post-hoc analyses revealed that there were significant differences between upper-LED-preceding versus no-visual-stimulus conditions ( $p < 0.05$ ) and between upper-LED-preceding versus lower-LED-preceding conditions ( $p < 0.01$ ). There were no significant differences between no-visual-stimulus versus lower-LED-preceding conditions.

### 2.5 Discussion

We examined the effect of vertically aligned visual stimuli on the TOJs of the sequence of pure tones in experiment 1. The auditory stimuli were sequences of four distinct pure tones. The visual stimuli consisted of two vertically aligned flashes: one flashed before the first tone and the other flashed after the last tone; therefore, the flash of the upper LED preceded that of the lower LED and vice versa. The participants judged whether the temporal order of the second and third tones in the auditory stimuli was for the higher tone to occur first (higher-tone-first) or the lower tone to occur first (lower-tone-first).

As a result, there were significant differences on PSSs at lag = 0 ms, whereas non-significant differences on JNDs are plotted in Figure 2. Thus, the proportion of higher-tone-first responses for the upper-LED-preceding condition was significantly larger than that for the no-visual-stimulus and lower-LED-preceding conditions at lag = 0 ms. This means that the vertically aligned visual stimuli affected the auditory TOJ tasks.

This visual effect that represented the differences in the JNDs and PSSs of the visual conditions was observed in the small lag values between visual and auditory stimuli; thus, a maximum effect was



**Figure 2.** Results from experiment 1. The upper row indicates the mean and standard error of the proportion of higher-tone-first responses as a function of SOA (data points have been slightly offset to improve legibility) and the sigmoid function fits to data. JNDs obtained from the curve are in the middle row and PSSs are in the lower row. Three columns are arranged according to lag values.

gained when the visual and auditory stimuli were presented closely in time. This matches the idea that the temporal co-occurrence of stimuli is important for audio-visual interaction (e.g., Calvert, Spence, & Stein, 2004) and suggests that the effect observed in experiment 1 occurred at the perceptual level.

However, TOJs can be biased on the basis of the observer's prior expectancy or the pattern of attentional deployment (e.g., Shore, Spence, & Klein, 2001). It is possible that the participants simply had a bias to respond to the spatial order of flashes rather than the temporal order of pitches when the temporal aspects of auditory signals were ambiguous. This would mean that the results were due to a response bias rather than due to an altered perception of temporal order. Therefore, we conducted simultaneity judgment (SJ) tasks for the sequence of pure tones in experiment 2 instead of TOJs to clarify this issue.

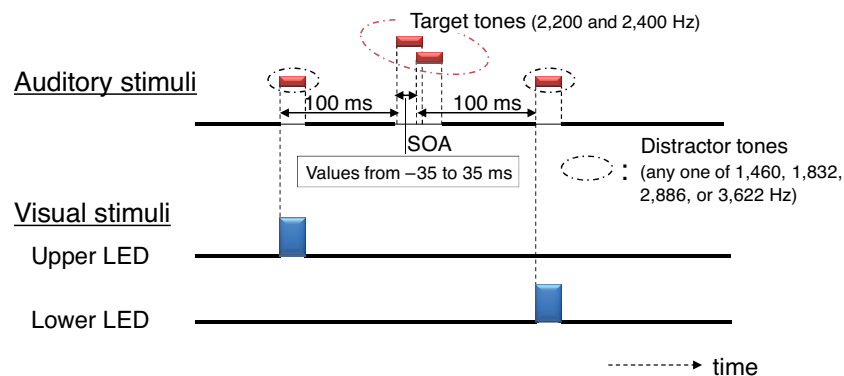
### 3 Experiment 2: Auditory simultaneity judgments—Effect of vertically aligned visual stimuli

#### 3.1 Participants

Five males (22–24-years-old) with normal hearing and normal or corrected-to-normal vision participated in this experiment. All of them took part in experiment 1.

#### 3.2 Stimuli and apparatus

The auditory stimuli were a sequence of four distinct pure tones with different frequencies (Figure 3). Each tone lasted for 30 ms (60 dB SPL) including 3-ms linear onset and offset ramps. The frequencies of the second and third tones (target tones) were mutually exclusive at 2,200 or 2,400 Hz, and those of the first and fourth tones (distractor tones) were one out of 1,460, 1,832, 2,886, and 3,622 Hz frequencies. The frequencies of both distractor tones were the same. The second tone was always presented 100 ms after the first tone and the fourth tone was always presented 100 ms after the third tone. The SOAs between the second and the third tones were  $-35$ ,  $-30$ ,  $-25$ ,  $-20$ ,  $0$ ,  $20$ ,  $25$ ,  $30$ , and  $35$  ms.



**Figure 3.** Schematic of trials in experiment 2. This figure is one example of positive SOA and the upper-LED-preceding condition.

Here, the negative SOAs indicate that the lower tone (2,200 Hz) of target tones was presented first, whereas the positive SOAs indicate that the higher tone (2,400 Hz) of target tones was presented first, and zero SOA indicates that the second and third tones were presented simultaneously.

The visual stimuli were the same as those in experiment 1. However, the lag value (in Figure 1), which is the interval between the first flash and the first tone and also that between the second flash and the fourth tone, was always 0 ms; thus, the first flash was presented simultaneously with the first tone and the second flash was presented simultaneously with the fourth tone (Figure 3). The reason the lag value was only limited to 0 ms is that the effect in experiment 1 was only obtained for the lag = 0 ms condition.

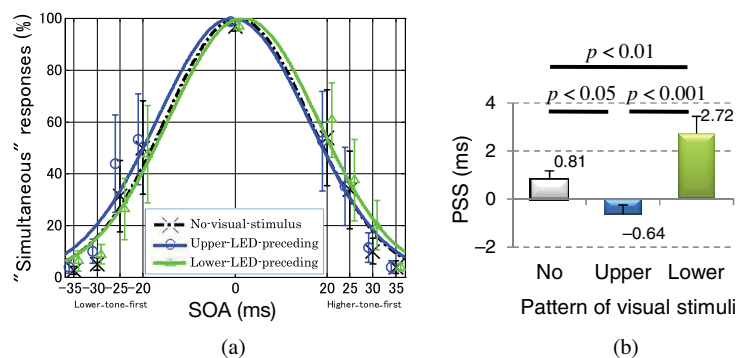
### 3.3 Procedure

The participants judged whether the onsets of the second and third tones were presented simultaneously or not (simultaneous or successive). This experiment consisted of 432 trials (9 auditory stimuli, 3 visual stimuli, and 16 repetitions), and was divided into 8 blocks of 54 trials. The participants took part in two blocks per day. The other procedure was the same as that in experiment 1.

### 3.4 Results

Figure 4(a) plots the mean and standard error for the proportion of simultaneous responses as a function of SOA, and the Gaussian function  $G(\text{SOA}) = 100 \times \exp(-\text{SOA} - \mu)^2 / 2\sigma^2$  fits the experimental data. The SOA at maximum of simultaneous responses,  $\mu$ , was taken as the PSS. The PSSs are shown in Figure 4(b).

The PSSs were submitted to repeated-measures ANOVA with the pattern of visual stimuli. The results indicate significant effects on the PSSs,  $F(2, 8) = 19.51$  and  $p = 0.001$ . LSD post-hoc analyses revealed that there were significant differences in all combinations:  $p < 0.05$  between no-visual-stimulus and upper-LED-preceding conditions,  $p < 0.01$  between no-visual-stimulus and lower-LED-preceding conditions, and  $p < 0.001$  between upper-LED-preceding and lower-LED-preceding conditions.



**Figure 4.** Results from experiment 2. (a) Mean and standard error of the proportion of simultaneous responses as a function of SOA (data points have been slightly offset to improve legibility) and the Gaussian function fitted to data. (b) PSSs obtained from panel (a).



### 3.5 Discussion

We examined what effect vertically aligned visual stimuli had on the SJs of sequences of pure tones in experiment 2. The auditory stimuli were sequences of four distinct pure tones. The visual stimuli consisted of two vertically aligned flashes: one flashed simultaneously with the first tone and the other flashed simultaneously with the fourth tone; therefore, the flash of the upper LED preceded that of the lower LED and vice versa. The participants judged whether the onsets of the second and third tones were presented simultaneously or not (simultaneous or successive).

As a result, the differences in PSSs were significant, as seen in [Figure 4\(b\)](#). Therefore, the proportion of simultaneous responses for the upper-LED-preceding condition was significantly shifted horizontally to the left compared with the no-visual-stimulus condition, and that for the lower-LED-preceding condition shifted horizontally to the right compared with the no-visual-stimulus condition.

This means that the vertically aligned visual stimuli also affected the auditory SJ tasks. This result ruled out an alternative explanation for response bias and the audio-visual interaction we report in this paper occurred at the perceptual level.

We used vertically aligned visual stimuli in experiments 1 and 2. However, as it is natural to conjecture if this effect is specific to vertically aligned visual stimuli, we used horizontally aligned visual stimuli in experiment 3.

## 4 Experiment 3: Auditory temporal order judgments—Effect of horizontally aligned visual stimuli

### 4.1 Participants

Five males (22–24-years-old) with normal hearing and normal or corrected-to-normal vision participated in this experiment. All of them took part in experiments 1 and 2.

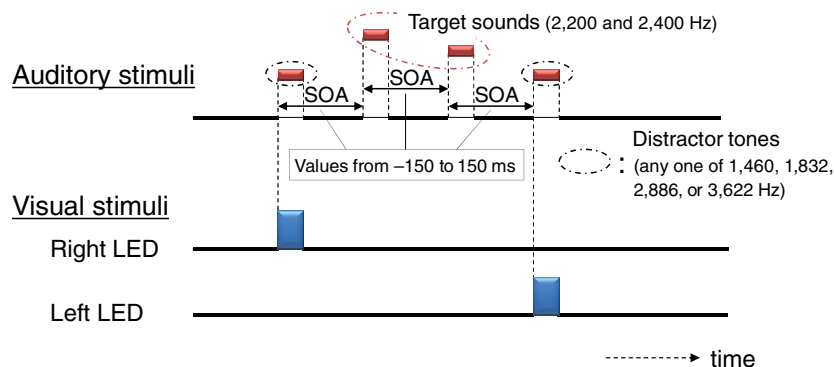
### 4.2 Stimuli and apparatus

The auditory stimuli were the same as those in experiment 1, except that the SOAs were  $-150$ ,  $-90$ ,  $-60$ ,  $-45$ ,  $-30$ ,  $30$ ,  $45$ ,  $60$ ,  $90$ , and  $150$  ms ([Figure 5](#)). Here, negative SOAs indicate that the lower tone (2,200 Hz) of target tones was presented first, whereas positive SOAs indicate that the higher tone (2,400 Hz) of target tones was presented first.

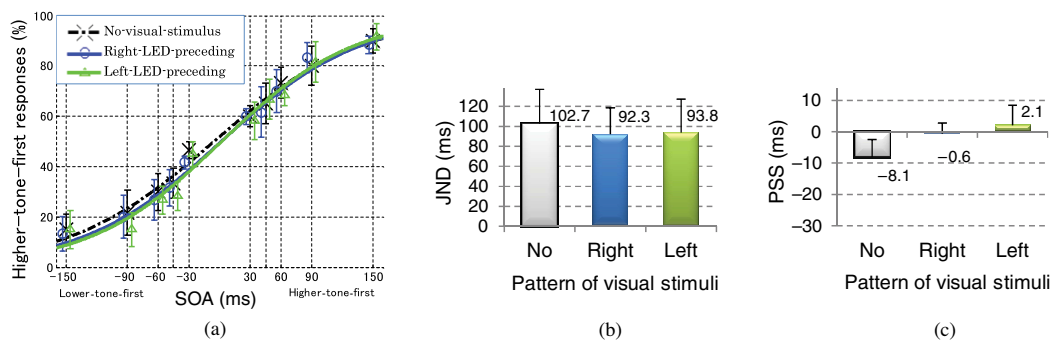
The visual stimuli were presented with two blue LEDs. They were placed 10 cm to the right and left of the red LED, which served as a fixation target and was positioned approximately at eye level. The blue LEDs flashed for 30 ms on two different orders: the right LED preceded the left LED (right-LED-preceding condition) and vice versa (left-LED-preceding condition). The first flash was presented simultaneously with the first tone and the second flash was presented simultaneously with the fourth tone ([Figure 5](#)).

### 4.3 Procedure

The participants judged which of the second and third tones were presented first (higher-tone-first or lower-tone-first). This experiment consisted of 360 trials (10 auditory stimuli, 3 visual stimuli, and



**Figure 5.** Schematic of trials in experiment 3. This figure is one example of positive SOA and the right-LED-preceding condition.



**Figure 6.** Results from experiment 3. (a) Mean and standard error of the proportion of higher-tone-first responses as a function of SOA (data points have been slightly offset to improve legibility) and the sigmoid function fitted to data. (b) JNDs obtained from panel (a). (c) PSSs obtained from panel (a).

12 repetitions), and was divided into six blocks of 60 trials. The participants took part in two blocks per day. The other procedure was the same as in experiment 1.

#### 4.4 Results

Figure 6(a) plots the mean and standard error for the proportion of higher-tone-first responses as a function of SOA, and the sigmoid function  $F(\text{SOA}) = 100/(1 + \exp(-b(\text{SOA} - x_0)))$  fits the experimental data. The JNDs and PSSs were calculated in the same way as that in experiment 1 and are in Figures 6(b) and (c).

The JNDs and PSSs were submitted to repeated-measures ANOVA with the pattern of visual stimuli. The results indicate there were no significant effects of visual stimuli on either JNDs or PSSs, i.e.  $F(2, 8) = 0.758$  and  $p = 0.500$  for JNDs, and  $F(2, 8) = 2.77$  and  $p = 0.122$  for PSSs.

#### 4.5 Discussion

We examined what effect horizontally aligned visual stimuli had on the TOJs of the sequence of pure tones in experiment 3. The auditory stimuli were the same as those in experiment 1, except for the SOAs. The visual stimuli consisted of two horizontally aligned flashes: one flashed simultaneously with the first tone and the other flashed simultaneously with the fourth tone; therefore, the flash of the right LED preceded that of the left LED and vice versa. The participants judged whether the temporal order of the second and third tones in the auditory stimuli was higher-tone-first or lower-tone-first.

As a result, The JNDs and PSSs also did not differ under the visual conditions, as can be seen from Figures 6(b) and (c). Thus, the proportion of responses for higher-tone-first did not differ under the visual conditions, as shown Figure 6(a).

This means that the horizontally aligned visual stimuli did not affect the auditory TOJs, despite the fact that the vertically aligned visual stimuli affected the auditory TOJs in experiment 1.

### 5 General discussion

The main goal of this paper is to explain how visual spatial information is able to affect auditory TOJs for a tone sequence with different frequencies. We examined the effect of vertically aligned visual stimuli on the TOJs of the sequence of pure tones in experiment 1. The proportion of responses for the higher-tone-first for the upper-LED-preceding condition was significantly larger than that for the no-visual-stimulus and lower-LED-preceding conditions as a result of experiment 1. Next, we examined what effect vertically aligned visual stimuli had on the SJs of the sequence of pure tones in experiment 2 to exclude an alternative explanation for response bias. The proportion of simultaneous responses for the upper-LED-preceding condition was significantly shifted horizontally to the left compared with the no-visual-stimulus condition as a result of experiment 2, and that for the lower-LED-preceding condition shifted horizontally to the right compared with the no-visual-stimulus condition. Finally, we examined what effect horizontally aligned visual stimuli had on the TOJs of the sequence of pure tones in experiment 3 to verify whether the effect observed in experiments 1 and 2 occurred even with horizontally aligned visual stimuli. The proportion of responses for the higher-tone-first did not differ across the pattern of visual stimuli as a result of experiment 3. Experiments 1–3 indicated that the

---

spatial information of not horizontally but vertically aligned visual stimuli is able to affect the TOJs of the sequence of pure tones and that this effect does not occur at the response or decision level but at the perceptual level.

This study demonstrated a cross-modal effect of vertically aligned visual stimuli on the TOJs of the sequence of pure tones at the perceptual level. The cross-modal effect is one example that is not explained by the modality appropriateness hypothesis (Welch & Warren, 1980) because the auditory TOJ task should not be affected by visual stimuli if the hypothesis holds. Also, the results are possibly related to the assertion made by Wada, Kitagawa, and Noguchi (2003), where they stated that the dominance of auditory processing over the visual one in time perception is based on stimulus uncertainty, not on the difference in sensory modality to which stimuli are presented. The perceived temporal order of auditory stimuli in our experiments could have higher uncertainty than the order of visual stimuli. Moreover, Alais and Burr (2004) reported that visual localization was captured by audition for severely blurred visual stimuli (that were poorly localized), which could not be accounted for by the modality appropriateness hypothesis. They discussed the idea that multimodal information is optimally combined by summing the independent stimulus estimates from each modality according to an appropriate weighting scheme. The weight of each modality should be smaller for larger uncertainties. The temporal order of auditory stimuli in our experiments was probably ambiguous due to distractor tones being presented whereas the order of visual stimuli could be easily perceived. The uncertainty of auditory stimuli should lead to results where the perceived temporal order of auditory TOJ tasks was affected by visual stimuli. The results of our study may support the idea that Alais and Burr (2004) discussed. However, we only found that auditory timing affects visual spatial perception with regard to the audio-visual interaction of visual spatial and auditory temporal patterns. Freeman and Driver (2008) found that the timing of static auditory events, which is spatially uninformative, influenced the perceived direction of visual apparent motion and induced visual aftereffects. Kawabe et al. (2008) demonstrated that the audio-visual tau effect with shorter temporal intervals between pure tones caused shorter spatial intervals to be perceived between visual stimuli. Roseboom et al. (2013) discovered that modulation of the direction of apparent visual motion with auditory signals could be accounted for the grouping of apparently related auditory and visual sequences, and not by temporal capture explanations. The direction of the effect in our study was opposite to that in the literature, indicating that auditory timing affects visual spatial perception. Considering the reports by Wada et al. (2003) or Alais and Burr (2004), it is natural that visual spatial and auditory temporal patterns affect each other. However, such examples had never been reported previously.

Several studies have shown an intimate cross-modal association between visual spatial height and auditory pitch frequency. For example, Su, Kim, and Iwamiya (2011) demonstrated that the vertical correspondence of the direction between the movement of images and the pitch shift of sound had a strong effect on perceived congruence; the combination of a rising image and an ascending pitch, and that of a falling image and a descending pitch, created greater congruence than alternative combinations. Also, Maeda, Kanai, and Shimojo (2004) reported that changing pitch systematically affected the perception of visual motion. Furthermore, Evans and Treisman (2010) discovered cross-modal mappings of auditory pitch to the visual position. They found that their participants responded faster in speeded classification tasks for bimodal congruent pairs and that audio-visual interaction occurred at the perceptual level. The combination of upper-LED-preceding versus higher-tone-first conditions in our experiments probably created higher cross-modal congruence. This congruence would have helped the flash of the upper LED shorten the response time for the higher tones of target tones, causing the proportion of responses for the higher tone occurring first to increase. Similar responses to the lower tone of target tones would have occurred for the flash of the lower LED.

The effect obtained in experiment 1 was only observed when the visual and auditory stimuli were presented closely in time (lag = 0 ms). This matches the idea that the temporal co-occurrence of stimuli is important for audio-visual interaction (e.g., Calvert et al., 2004) and highlights the fact that the effect we reported in this paper should occur at the perceptual level together with the results in experiment 2. However, the maximum difference in experiment 1 in the visual effect (the difference in PSSs in the pattern of visual stimuli) was 25.4 ms (upper-LED-preceding versus lower-LED-preceding conditions) and in experiment 2, this was 3.35 ms (upper-LED-preceding versus lower-LED-preceding conditions). The magnitude of the effect was attenuated by employing SJ tasks in experiment 2, instead of TOJ tasks in experiment 1. A large part possibly stemmed from the response bias on the basis of the visual spatial pattern, and residual perceptual effects were relatively small. However, the participants



were likely to respond to SJ tasks as being “successive” when there was even slight SOA in the target tones. The duration of the pure tone was 30 ms. It would be useful for participants to perceive the offset timing of target tones as well as that of the onset to simplify SJ tasks. Thus, the visual effect in SJ tasks is likely to be smaller than that in TOJ tasks. Moreover, the auditory system is typically very sensitive to temporal order. For example, it is known that the temporal order of brief click (20  $\mu$ s) pairs could be discriminated with 0.25 ms for click pair durations (Henning & Gaskell, 1981), and that single clicks could be distinguished from click pairs when the gap between the two clicks in a pair was only a few tens of microseconds (Leshowitz, 1971). Thus, the difference of 3.35 ms is sufficiently large to modulate auditory temporal perception.

Our experiments revealed that the temporal perception of auditory events affected visual spatial information even if it was not a sufficiently large-field visual stimulus to be perceived as self-motion, whereas Teramoto et al. (2008) reported that auditory TOJs were not affected by visual stimulus when their participants did not perceive self-motion. Thus, large-field visual stimuli are not always necessary to affect the temporal perception of auditory stimuli.

## 6 Conclusion

This study demonstrated a cross-modal effect of the spatial information of vertically aligned visual stimuli on the TOJs of the sequence of spatially uninformative pure tones at the perceptual level. The effect on TOJ tasks was diminished for horizontally aligned visual stimuli. Although response bias was found in the TOJ tasks for the vertically aligned visual stimuli, it is difficult to disregard the effect on SJ tasks. We still need to find what mechanism induced the audio-visual interaction we reported in this paper.

## References

- Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology*, *14*, 257–262. doi:10.1016/j.cub.2004.01.029 ◀
- Calvert, G., Spence, C., & Stein, B. E. (2004). *The handbook of multisensory processes*. London: MIT Press. ◀
- Evans, K. K., & Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features. *Journal of Vision*, *10*, 1–12. doi:10.1167/10.1.6 ◀
- Freeman, E., & Driver, J. (2008). Direction of visual apparent motion driven solely by timing of a static sound. *Current Biology*, *18*, 1262–1266. doi:10.1016/j.cub.2008.07.066 ◀
- Harrar, V., & Harris, L. R. (2008). The effect of exposure to asynchronous audio, visual, and tactile stimulus combinations on the perception of simultaneity. *Experimental Brain Research*, *186*, 517–524. doi:10.1007/s00221-007-1253-0 ◀
- Henning, G. B., & Gaskell, H. (1981). Monaural phase sensitivity measured with Ronken’s paradigm. *Journal of the Acoustical Society of America*, *70*, 1673. doi:10.1121/1.387231 ◀
- Jack, C. E., & Thurlow, W. R. (1973). Effects of degree of visual association and angle of displacement on the “ventriloquism” effect. *Perceptual and Motor Skills*, *37*, 967–979. doi:10.2466/pms.1973.37.3.967 ◀
- Kawabe, T., Miura, K., & Yamada, Y. (2008). Audiovisual tau effect. *Acta Psychologica*, *128*, 249–254. doi:10.1016/j.actpsy.2008.01.004 ◀
- Leshowitz, B. (1971). Measurement of the two-click threshold. *Journal of the Acoustical Society of America*, *49*, 515–527. doi:10.1121/1.1912374 ◀
- Maeda, F., Kanai, R., & Shimojo, S. (2004). Changing pitch induced visual motion illusion. *Current Biology*, *14*, R990–R991. doi:10.1016/j.cub.2004.11.018 ◀
- Morein-Zamir, S., Soto-Faraco, S., & Kingstone, A. (2003). Auditory capture of vision: examining temporal ventriloquism. *Cognitive Brain Research*, *17*, 154–163. doi:10.1016/S0926-6410(03)00089-2 ◀
- Roseboom, W., Kawabe, T., & Nishida, Y. (2013). Direction of visual apparent motion driven by perceptual organization of cross-modal signals. *Journal of Vision*, *13*, 1–13. doi:10.1167/13.1.6 ◀
- Shimojo, S., Scheier, C., Nijhawan, R., Shams, L., Kamitani, Y., & Watanabe, K. (2001). Beyond perceptual modality: Auditory effects on visual perception. *Acoustical Science and Technology*, *22*, 61–67. doi:10.1250/ast.22.61 ◀
- Shore, D. I., Spence, C., & Klein, R. M. (2001). Visual prior entry. *Psychological Science*, *12*, 205–212. doi:10.1111/1467-9280.00337 ◀
- Su, X., Kim, K., & Iwamiya, S. (2011). Perceived congruence between the changing patterns of visual image and pitch of a sound. *i-Perception*, *2*, 763. ◀
- Teramoto, W., Watanabe, H., Umemura, H., & Kita, S. (2008). Change of temporal-order judgment of sounds during long-lasting exposure to large-field visual motion. *Perception*, *37*, 1649–1666. doi:10.1068/p5692 ◀
- Van der Burg, E., Olivers, C. N. L., Bronkhorst, A. W., & Theeuwes, J. (2008). Audiovisual events capture attention: Evidence from temporal order judgments. *Journal of Vision*, *8*, 1–10. doi:10.1167/8.5.2 ◀

- 
- Wada, Y., Kitagawa, N., & Noguchi, K. (2003). Audio-visual integration in temporal perception. *International Journal of Psychophysiology*, 50, 117–124. doi:10.1016/S0167-8760(03)00128-4 ◀
- Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin*, 88, 638–667. doi:10.1037//0033-2909.88.3.638 ◀



**Shoya Tanabe** studied bio-cybernetics, human sciences, and assistive technology at Niigata University. He received his master's degree in engineering from Niigata University in 2013. His research interests are in audio-visual interaction and its engineering application. He has studied multimodal processing in visual spatial and auditory temporal information.



**Mamoru Iwaki** received the Dr. Eng. degrees in electronic and information engineering from the University of Tsukuba, Japan, in 1994. He is an associate professor in the Graduate School of Science and Technology, Niigata University, Japan. His research interests include speech signal processing, human auditory systems, and bio-signal interpretation.