Low-Level Visual Processing Speed Modulates Judgment of Audio-Visual Simultaneity

Yasuhiro TAKESHIAM and Jiro GYOBA

Department of Psychology, Graduate School of Arts and Letters, Tohoku University, Sendai 980-8576, Japan

Received December 5, 2014; final version accepted May 26, 2015

Temporal consistency between visual and auditory presentations is necessary for integration of visual and auditory information. Subjective simultaneity perception is more important than the synchrony of physical inputs for temporal consistency. Our previous studies have shown that audio-visual integration is difficult even if the visual and auditory inputs are physically synchronous when visual processing is slow. In the present study, we examined the effects of visual processing speed on audio-visual integration using a simultaneity judgment task. Visual processing speed was manipulated by varying the spatial frequency of visual stimuli. High spatial frequency stimuli require a longer processing time because visual responses to high spatial frequencies are slow. The results indicated that the difference between subjective and physical synchrony was larger in high spatial frequency than in low spatial frequency. Thus, the spatial frequency of the visual stimulus affected the judgments of simultaneity for visual and auditory stimuli. The effects of visual processing speed on audio-visual integration are believed to occur at a lower-order stage of sensory processing.

KEYWORDS: processing speed, simultaneous judgment task, gabor patch

1. Introduction

When perception in one sensory modality is ambiguous, information from other sensory modalities reduces this ambiguity [1]. Therefore, humans perceive a stable outer environment via multisensory inputs. Multisensory integration has been studied extensively. In particular, many studies have examined the process involved in integrating visual and auditory information [e.g., 2–4].

However, audio-visual integration does not always occur. For example, temporal consistency between visual and auditory information is necessary. In single-cell recording studies, the neuronal responses in the superior colliculus (SC) are enhanced by temporal consistency between visual and auditory inputs [5]. Moreover, human behavioral data show that visual target sensitivity can be facilitated by temporally consistent auditory input [6]. In human neural networks, SC activity is related to audio-visual integration [7]. Therefore, temporal consistency is important for audio-visual integration. However, a range of subjective synchrony perception between the temporal onset of visual and auditory stimuli is tolerated [8], from −130 ms (sound preceding) to +250 ms (sound following) [9]. Several cross-modal illusions induced by sound (i.e., audio-visual interactions) have temporal windows within this range [e.g., 10–14]. Therefore, subjective synchrony seems to be more important for audio-visual integration than physical synchrony.

In a previous study, we hypothesized that the timing of audio-visual onset would be judged as more asynchronous when visual processing was slower [15]. Takeshima and Gyoba [15] investigated the effects of processing speed on the fission illusion, which is induced in audio-visual integration by manipulating the complexity of visual stimuli. In the fission illusion, when a single flash is presented simultaneously with two beeps, two flashes are perceived [11, 12]. The results indicated that it was difficult to induce the fission illusion with complex visual stimuli, suggesting that audio-visual integration was occurs with more difficulty with complex visual stimuli, for which processing speed is slow. Our hypothesis, that slow processing of sensory stimuli affects the simultaneous perception of auditory and visual information, was supported in a study investigating the effects of object quantity [16] in addition to complexity [15].

In the present study, we re-examined the effects of processing speed on audio-visual integration with respect to the following. First, it has been suggested that audio-visual integration occurs at multiple stages of sensory processing. In various anatomical and physiological studies, multisensory information converges not only in higher-order association cortices, but also in early sensory areas (i.e., primal visual, auditory, or somatosensory areas) [17–19 for a review]. Moreover, Mishra et al. [20] have measured the event-related potentials (ERPs) in the fission illusion. The results indicated that some ERP components were observed at various latencies. Sanabria et al. [21] have examined the effects...
of audio-visual integration on motion perception using signal detection theory. In signal detection theory, two indices are calculated: the d-prime and criterion [22]. The d-prime indicates the effects at the perceptual level, and the criterion indicates the effects at the cognitive level. The results indicated that both d-prime and criterion scores were changed by audio-visual integration in motion perception. Therefore, effects of processing speed should be observed at both higher-order (i.e., complexity of visual stimuli) and lower-order levels of processing. We manipulated the spatial frequency of visual stimuli in this study. It is well known that spatial frequency information is processed in the early visual cortex (V1) and responses to high spatial frequency visual stimuli are slow [23]. Second, we used a simultaneity judgment (SJ) task in the present study, while in our previous studies the effects of processing speed were measured by the strength of the fission illusion or performance on a same-different task [15, 16]. If visual processing speed influences subjective temporal synchrony perception, simultaneity judgments for visual and auditory stimuli should also be altered.

2. Method

2.1 Participants

Nine observers (four women and five men, mean age = 23.9 ± 1.36 years) participated in this experiment. All participants reported normal or corrected-to-normal vision and normal audition. None were informed of the purpose of the experiment except for one participant (the first author).

2.2 Apparatus

Stimuli were generated and controlled by a custom-made program written with MATLAB (MathWorks, Inc.), Cogent Graphics and 2000 toolbox (www.vislab.ucl.ac.uk/cogent.php), and a PC (XPS720, Dell; OS: Windows Vista, Microsoft). The visual stimuli were displayed on a CRT-display (Trinitron GDM-F520, Sony; resolution: 1024 × 768 pixels; refresh rate: 60 Hz). The auditory stimuli were conveyed through an audio interface (Edirol FA-66, Roland) and headphones (HDA200, Sennheiser). Simultaneity of the visual and auditory stimuli was confirmed using a digital oscilloscope (TS-80600, Iwatsu). The experiment was conducted in a dark room with 43.6 dB (A) of background noise. Participants viewed the monitor binocularly at a distance of 60 cm with their heads stabilized on a chin rest.

2.3 Stimuli

Each trial consisted of the presentation of a red (15.3 cd/m²) double circle (1.1 deg in diameter; for fixation) and visual stimuli. All stimuli were presented on a gray (17.9 cd/m²) background. Visual stimuli were Gabor patches (Figure 1a) of two spatial frequencies, 1.0 and 5.0 cycles per degree (c/deg). The visual stimuli were 2.0 deg in diameter, and were presented for 17 ms. The auditory stimulus was a pure tone at 2000 Hz. The duration of the auditory stimulus was 17 ms (including ramp times of 1.7 ms at the beginning and end of the sound wave envelope), and the sound pressure level was 75 dB (A). There were nine stimulus onset asynchronies (SOAs) between visual and auditory stimuli: −350, −233, −150, −67, 0, +67, +150, +233, +350 ms (negative SOAs indicate that the auditory stimulus was presented before the visual stimulus and vice versa).

2.4 Procedure

A trial schematic is shown in Figure 1b. Trials were initiated by pressing the “0” key. Each trial consisted of a 500 ms fixation followed by blank and target displays. The duration of the blank display was randomized (500–1000 ms). Each participant completed 640 trials; there were more synchronous (N = 320) than asynchronous (N = 40 for each SOA) trials to approximately equal the a priori number of synchronous and asynchronous trials (for a similar methodology, see [24–26]). In the SJ task, participants were instructed to press “1” for simultaneity and “3” for asynchrony. After the SJ task, each participant performed a response time (RT) task for each spatial frequency Gabor patch. The trial sequence is shown in Fig. 1c. A fixation cross was presented at the center of the screen (500 ms), followed by a blank display (randomized duration, 500–1000 ms). Then, the visual stimulus was presented until a response was made. Participants were instructed to press the “5” key as soon as the visual stimulus was presented. There were 60 RT task trials (N = 30 for each spatial frequency).

3. Results

To compute the point of subjective simultaneity (PSS), the data from each participant were estimated by fitting the following four parameters Gaussian function to each individual’s data (see also [25, 26]) by minimizing the root-mean-square-error (RMSE) in Microsoft Excel Solver:

\[ P(\text{response}|\text{SOA}) = \text{blink rate} + a \cdot e^{-\frac{(\text{SOA}-\text{PSS})^2}{2\sigma^2}} \]

The SOA parameter was equal to the experimental condition (−350 to +350 ms). The parameters \(a\) and \(b\) and blink rate required estimation. The blink rate parameter was included to consider a small proportion of noisy trials in which participants were not attending to the actual stimuli (e.g., due to eye blinks or other artifacts), which would otherwise lead to an overestimation of the other parameters [27]. The same analysis methodology, including this parameter, was
The blink rate was restricted to a minimum of 0% and maximum of 2.5% and was estimated to be 1.8% (1 c/deg: 1.5%; 5 c/deg: 2.1%) in this experiment. The other parameters were restricted to a minimum of 0.

The results are shown in Figure 2a, which presents the mean percentage of simultaneity responses, as a function of spatial frequency and SOA, together with fitted psychometric functions (1 c/deg: Mean RMSE = 0.10, SD = 0.06; 5 c/deg: Mean RMSE = 0.10, SD = 0.05). A two-way analysis of variance (ANOVA) with spatial frequency and SOA as within-participants factors was conducted. The results revealed a significant main effect of SOA ($F(8,64) = 122.87, p < .001, \eta_p^2 = .93$), indicating that participants complied with the SJ task. Furthermore, the interaction between spatial frequency and SOA was also significant ($F(8,64) = 3.84, p < .001, \eta_p^2 = .33$). The simple main effects of spatial frequency were significant in the +233 and +350 ms SOA conditions (+233 ms SOA: $F(1,72) = 7.39, p < .05, \eta_p^2 = .09$; +350 ms SOA: $F(1.72) = 5.94, p < .05, \eta_p^2 = .08$), indicating that the percentage of simultaneity responses was higher for 5 c/deg than 1 c/deg Gabor patches. However, a main effect of spatial frequency was not significant ($F(1.8) = 0.35, p = .57, \eta_p^2 = .04$). Moreover, PSS scores were computed as SOA score corresponding to the peak position of simultaneous judgment in each individual’s psychometric function. A two-tailed $t$-test was conducted to compare the PSS values between spatial frequency conditions. Mean PSS is shown in Figure 2b. PSS was larger in the 5 c/deg than in the 1 c/deg condition ($t (8) = 2.82, p < .05, d = .23$). The 95% confidence interval was from 74.53 ms to 112.12 ms in the 1 c/deg condition, and from 80.90 ms to 119.35 ms in the 5 c/deg condition. Moreover, a two-tailed $t$-test on the RT data confirmed that there was a difference in processing speed between spatial frequency conditions. Mean RTs are shown in Figure 2c. The results indicated that the RT was slower in the 5 c/deg than in the 1 c/deg condition ($t (8) = 2.82, p < .05, d = .30$).

4. Discussion

In the present study, we examined the effects of spatial frequency on synchrony perception of visual and auditory stimuli. The results indicated that there was a larger difference between the PSS values and 0 ms for high spatial frequency (5 c/deg) versus low spatial frequency (1 c/deg) stimuli in the SJ task. Moreover, the difference in processing speed between visual stimuli with different spatial frequencies was confirmed by measuring RT. If
subjective synchrony perception can be determined by physical synchrony between visual and auditory inputs, the PSS value becomes 0 ms. In general, the PSS values were shifted in the positive direction, indicating that simultaneity was maximally perceived if visual inputs slightly preceded auditory input (e.g., [7, 25, 26, 28, 29]). For example, Zampini et al. [25] have shown that the PSS shifts to within 30 ms. Fujisaki et al. [28] have revealed that the synchrony-asynchrony discrimination threshold (index near to PSS) was approximately 70 ms. This PSS shift reflects the difference in neural latency between visual and auditory neurons [30]; neural latency is slower for visual versus auditory neurons [31]. On the other hand, the PSS values shifted approximately 100 ms for both high and low spatial frequency stimuli in this experiment. The PSS shift was larger in the present study compared with previous studies (e.g., [7, 25, 26, 28, 29]). This difference may be attributed to the blurred edge of the Gabor patch. Moreover, PSS values differed significantly between high and low spatial frequency stimuli. Therefore, the present study clearly indicates that temporal synchrony perception of visual and auditory stimuli is affected by the spatial frequency of the visual stimulus. In the present study, RT was longer for high compared to low spatial frequency Gabor patches. This is consistent with previous studies showing that high spatial frequency stimuli are preferentially processed by sustained channels, while low spatial frequency stimuli are preferentially processed by transient channels [e.g., 32–34]. Because conduction speed is slower for neural fibers in sustained versus transient channels [35], visual processing speed is slower for high versus low spatial frequency stimuli [19]. Thus, the present results also suggest that the slow processing speed of visual stimuli affects integration of visual and auditory information [15, 16].

The present data showed that the sensory processing speed affected simultaneous judgment only in visual precedence conditions (i.e., +233 and +350 ms SOA). This asymmetry effect would be attributed to attunement toward the natural situation. In a natural scene, a visual stimulus (light) always precedes the auditory stimulus (sound). The situation when sounds precede light could not occur in a natural situation. Thus, simultaneous judgment would be more sensitive to sensory processing speed in visual precedence conditions than in auditory precedence conditions.

It is well known that spatial frequency information is processed at early visual cortex (V1). The previous study has shown that the left temporal parietal junction (TPJ) activations are modulated by the audio-visual synchrony

Fig. 2. Experimental results. (a) Mean percentage of simultaneous responses. The average fitting functions are plotted for 1 c/deg and 5 c/deg Gabor patch data. (b) Mean point of subjective synchrony. Error bars show 95% confidence interval (n = 9). (c) Results of response time experiment. Error bars show standard errors of the mean.
percept [36]. Furthermore, the connectivity between the left inferior parietal lobe (IPL) and the right dorsolateral prefrontal cortex (DLPFC), and the right TPJ are considered to be related to timing percept [36]. Spatial frequency information is conveyed to these brain areas and used to discriminate the audio-visual synchrony percept. Alternatively, the difference in conduction speed between visual and auditory signals to the TPJ would be used to this synchronous discrimination. These brain areas’ activities would conduct modifications for audio-visual timing percept according to various factors (e.g., distance from information resources [37], the difference in response latency [31]). The difference in response time would not be reflected directly in the difference in PSS between low and high spatial frequency stimuli because these modifications are conducted in audio-visual simultaneous judgment processes.

In our previous studies, audio-visual interaction was affected by visual processing speed based on complexity [15] and object quantity [16]. The complexity of visual stimuli is processed in the anterior inferotemporal cortex, posterior inferotemporal cortex, and V4 [38]. Moreover, the effects of object quantity likely occur during encoding [16], which is related to inferior parietal sulcus activity [39,40]. In contrast, it is well known that spatial frequency information is processed in the early visual cortex (V1). Unfortunately, it remains a possibility that the difference of contrast may affect the audio-visual synchrony judgment according to insufficient contrast control between low and high spatial frequency stimuli. However, it is also well known that contrast information is processed in early visual areas. In other words, both spatial frequency and contrast processing of visual stimuli precede in complexity processing of visual stimuli. Therefore, the present study clearly indicates that the effects of processing speed on audio-visual interaction occur at lower-order (i.e., responsible for processing spatial frequency information) in addition to higher-order (i.e., responsible for stimulus complexity) levels of processing.

Acknowledgments

We are grateful to the participants for their time and contribution. This research was supported by the Japanese Society for the Promotion of Science KAKENHI (Grant-in-Aid for Fellows: No. 24-4354).

REFERENCES


