Optimal Timing of Audio-Visual Text Presentation:

The Role of Attention.

Maiko Takahashi 1,2, Akihiro Tanaka 2

1 Japan Society for the Promotion of Science
2 Tokyo Woman’s Christian University

maiko_tk@lab.twcu.ac.jp, akihiro@lab.twcu.ac.jp

Abstract

This study investigates the optimal timing of audio-visual presentation for text comprehension. In Experiment 1, participants were asked to read and/or hear the expository passages in three conditions; visually presented, auditorily presented, and audio-visually presented condition. Comprehension performance in audio-visual presentation condition was not different from that in visual or auditory condition, raising the possibility that cognitive load on processing both visual and auditory information negated the positive effect of multi-modal presentation. To reduce the load of processing dual information simultaneously, we proposed to delay the timing of one of the two modality presentation and to direct participants’ attention to one of the two modality information during audio-visual text presentation. In Experiment 2 to 4, passages were presented audio-visually in three conditions; auditorily preceding, simultaneous, and visually preceding conditions. Participants were instructed to direct their attention to whole information (Exp. 2), visual information (Exp. 3), and auditory information (Exp. 4). The results showed that comprehension performance was higher in the visually preceding condition when their attention was directed whole or visual information. Based on the results, the integration process of audio-visually presented text information was discussed.

Index Terms: reading comprehension, listening comprehension, audio-visual integration, multimedia learning

1. Introduction

Multimedia material is being increasingly used as a tool to convey language information. We become not just to read, but "listen to" the text with the digital books. The question here is whether and how audio-visual presentation of the text facilitates the text comprehension. For speech perception, listeners use visual information from the speaker’s mouth as well as auditory speech [1] [2]. In the same way as for speech perception, do readers use both auditory and visual information for audio-visual presented text comprehension?

Several studies reported that the memory or comprehension performance was improved when the text was presented audio-visually [3] [4], suggesting that information from two modalities complement mutually to facilitate the text comprehension. In contrast to those findings, there are some studies that have reported the negative effect of audio-visual presentation on the text comprehension [5] [6]. Jamet and Le Bohec interpreted the results as the overload caused by dual channel inputs [5]. These inconclusive findings suggest that audio-visual text presentation have both positive and negative effect for text comprehension. To utilize multimedia material effectively, it is necessary to propose a way of preventing the negative aspect of audio-visual presentation. Given the interpretation of cognitive overload by simultaneous inputs from dual channels [5], we can raise the possibility that the cognitive load is reduced by manipulating presentation timing of auditory and visual information. If audio-visual text information is presented sequentially, we may attend to and utilize preceding information mainly and following information secondarily. This process would prevent the overflow of dual channel information and achieve effective text comprehension performance. Focusing on the situation of natural oral reading, we process visual information first, then read out the text and auditory feedback is given to ourselves. Kondo & Mazuka (1996) found that readers’ eye fixation points were ahead of uttering point [7]. Given the situation of natural oral reading in which visual input is ahead of auditory one, it is possible that text comprehension is well facilitated when visual text information precedes auditory one during audio-visual presentation.

The purpose of this study is to demonstrate the optimal audio-visual presentation timing for text comprehension. To prevent audio-visually presented information from overflowing, we suppose the way to present audio-visual text information sequentially. It is predicted that comprehension performance is well facilitated when visual information precedes auditory information, since the situation is similar to natural oral reading. In addition, we instruct the participants to direct their attention to information presented in only one modality (visual or auditory). Attending to one modality information, participants may use the attended information mainly and another one secondarily, leading to prevent the overflow of information.

In Experiment 1, we compared the performance of audio-visual presentation with single modality presentation. Using expository passages, the effect of audio-visual presentation on text memory and comprehension was investigated. In Experiment 2, we manipulated timing of audio-visual presentation; auditorily preceding, simultaneous, and visually preceding, and examined the text comprehension. When we read some text orally, we tend to look ahead of uttering point [7]. It means that we process different information from visual and auditory input simultaneously. As in the case of oral reading, we presented different phrases simultaneously in auditorily preceding and visually preceding conditions. In experiment 3 and 4, we instructed participants to direct their attention to only one modality information during audio-visual presentation.
presentation, and examine the effect of presentation timing on the text comprehension.

Figure 1 presents the overview of the stimulus presentation in the experiments of this study.

![Figure 1: Overview of the stimulus presentation in the experiments.](image)

2. Experiment 1

In this experiment, we compared the memory and comprehension performance of audio-visual presentation with that of visual or auditory presentation.

2.1. Method

2.1.1. Participants

A total of 27 college students, each reporting normal hearing and normal or corrected-to-normal vision, participated in the Experiment 1. All were native Japanese speakers.

2.1.2. Material

The experimental stimulus for each trial consisted of expository passages and questions about literal memory and text comprehension of the passages. Eighteen expository passages about biology, history, home economics, and everyday trivia, consisting about 200 Japanese letters involving Kanji and Kana characters were constructed for the experiment. For each passage, four questions about literal memory and text comprehension were prepared. Verbatim memory for a word or a sentence in the passage was tested with the literal memory questions, whereas the propositional representation constructed was examined by the text comprehension questions. The questions required either “True” or “False” response from the participants. An example of the passage and questions are shown in Table 1.

Each expository passage read aloud by a female experienced speaker at a normal speech rate was recorded and edited to each single phrase by Sound Forge (Sony Co. Ltd.). The average number of phrases in the expository passages was 51.94 (SD = 6.07, Min = 40, Max = 65). The mean duration of the phrase was 584.49ms (SD = 270.49, Min = 223, Max = 1400), and the mean number of characters involved in the phrases was 4.01 (SD = 1.88, Min = 2, Max = 10). The mean number of morae in the phrases was 4.61 (SD = 1.95, Min = 1, Max = 9).

Table 1. Example of the passage and questions.

**Expository passage:**

Cats can see and recognize objects even in the dark. They need only one-seventh of light as compared to human. Cats’ eyes gleam with reflected light in the dark because cats’ eyes have reflector “tapetum”. The little light from the retina is doubled by the reflection to recognize objects in sharp focus in the dark. Another secret of the cats’ eyes is their size. Cats have big eyes that are not proportional to their body size. Their pupils are also big to take more light.

Questions about literal memory:


- ネコ Gör[目/は/大きい/が]、[反射値]で[反射値]をつけています。

Questions about text comprehension:

- Cats can see clearly by the reflectors that double the amount of light. (true)
- Cats’ eyes have reflector “tappetum”. (false)

Note. Slash marks in Japanese passage represent the boundary of phrase.

2.1.3. Procedure

Each participant was tested individually. They were asked to read and/or hear the expository passages in three conditions; visually presented, auditorily presented, and audio-visually presented condition. One of the lists which consists of six passages was assigned to each condition. In visually presented condition, written passages were presented phrase-by-phrase at the center of the computer screen, and participants were instructed to read them silently. In auditorily presented condition, spoken passages were presented through headphones. In audio-visually presented condition, written and spoken passages were presented simultaneously phrase-by-phrase, and participants were told to attend to both information.
presentation rate in the three conditions was identical to the speech rate of the phrase files. After the passage presentation, participants were asked to answer four questions about literal memory and another four questions about text comprehension of the passages. These questions were presented visually in all three conditions. The presentation of the experimental stimulus and recording of participants' responses were controlled by Super Lab 4.5 (Cedrus Corporation).

2.2. Results and Discussion

We calculated the mean percentages of correct responses to the questions about literal memory and test sentences, for each participant, under each of the 3 conditions. Figure 2 presents the mean percentages of correct responses from the data collected. The data were analyzed with a one-way ANOVA. Both in the responses to the questions about literal memory and text comprehension, no significant effects of presentation conditions were found. \( F(2,52) = 1.22, \text{MSE} = 64.54, n.s., F(2,52) = 0.12, \text{MSE} = 45.50, n.s. \)

The performance in the audio-visually presented conditions did not exceed that in the other unimodal conditions, which is consistent with the idea that the load on processing both text and speech information simultaneously negated positive effect of the multi-modal presentation. This cognitive overload may be reduced when audio-visual information is presented sequentially. In Experiment 2, we manipulate the timing of audio-visual presentation and examine whether the overload is reduced by the manipulation.

3. Experiment 2

In Experiment 2, we manipulate the timing of audio-visual presentation; auditorily preceding, simultaneous, and visually preceding and examine whether the manipulation reduce the overload of multi-modal input.

3.1. Method

Except as noted, the method was identical to that of Experiment 1.

3.1.1. Participants

A total of 24 college students, each reporting normal hearing and normal or corrected-to-normal vision, participated in the Experiment 2. All were native Japanese speakers. None had participated in Experiment 1.

3.1.2. Procedure

In auditorily preceding condition, an audio file of each phrase was presented ahead of visual presentation. On the other hand, visually presented phrase was presented ahead of auditory presentation in visually preceding condition, which means that different phrases were presented simultaneously (Figure 1). The presentation rate was adjusted to longer phrase between the two modality presentations. We checked the audio file of visually presented phrase, compared its duration with simultaneously presented auditory phrase, and set the duration at longer one. The duration of the audio file of each phrase was about 600ms in average, which was sufficient for reading one phrase. Simultaneous condition in this experiment was identical to the audio-visual condition in Experiment 1.

3.2. Results and Discussion

We calculated the mean percentages of correct responses in the same manner as Experiment 1. Figure 3 shows mean percentages of correct responses from the data. A one-way ANOVA revealed there was no significant effect of presentation conditions in the literal memory task, \( F(2,46) = 2.26, \text{MSE} = 57.97, n.s. \). On the other hand, in text comprehension task, the effect of presentation condition was marginally significant, \( F(2,46) = 2.68, \text{MSE} = 57.92, p < .08 \). Ryan’s method for multiple comparison showed the performance in visually preceding condition is superior to that in auditorily preceding condition, \( t(46) = 2.29, p < .05 \). The performance in simultaneous condition was not significantly different from that in the other two conditions; comparison with auditorily preceding condition, \( t(46) = 1.42, n.s., \) comparison with visually preceding condition, \( t(46) = 0.87, n.s. \)

As predicted, the performance in the visually preceding condition achieved highest among the three conditions. Since the situation in visually preceding condition is similar to natural oral reading, participants might be able to process audio-visual information without any difficulty in this condition. However, the performance in visually preceding condition was not significantly higher than that in simultaneous condition. This result points the possibility that cognitive overload caused by dual channel inputs was not reduced sufficiently only by manipulating the presentation timing. When reading orally, visual information is not only presented ahead, but is attended. Thus, it is possible that superiority of visual preceding presentation may be shown when participant is instructed explicitly to attend to visual information. In following experiments, we examine the possibility to prevent the cognitive overload of audio-visual information processing by directing participants' attention to single information during audio-visual presentation.
4. Experiment 3

In Experiment 3, we instruct participants to direct their attention to only visual information, and examine whether the instruction prevent the overload caused by dual channel inputs.

4.1. Method

Except as noted, the method was identical to that of Experiment 2.

4.1.1. Participants

A total of 24 college students, each reporting normal hearing and normal or corrected-to-normal vision, participated in the Experiment 3. All were native Japanese speakers. None had participated in any of the earlier experiments.

4.1.2. Procedure

The participants were instructed to direct their attention to the visual information during expository passage presentation.

4.2. Results and Discussion

We calculated the mean percentages of correct responses in the same manner as the earlier experiments. Figure 4 shows mean percentages of correct responses from the data. No significant effects were found in both literal memory and text comprehension tasks, $F(2,46) = 1.02$, $MSE = 45.53$, n.s., $F(2,46) = 0.06$, $MSE = 79.56$, n.s.

In contrast to the result in Experiment 2 and 3, the superiority of the visually preceding presentation did not occur when participants were instructed to direct their attention to the auditory information. By attending to the auditory information, performance in the auditorily preceding condition achieved equivalent level of that in the other two conditions, however, it was not superior to the other. Unlike in the case of visually preceding presentation, we are unfamiliar with auditorily preceding situation. Therefore the superiority of auditorily preceding presentation might not occur even if participants attended to auditory information.
6. General Discussion

Development of the multimedia material such as digital books have provided huge benefit to our learning environment. Using the technology, we can choose whether “read” or “listen to” the learning material. To utilize technology effectively, it is necessary to clarify the mechanism for the process of audio-visual text comprehension.

The purpose of this study was to investigate the optimal timing of audio-visual presentation for text comprehension. In Experiment 1, the comprehension score of audio-visual presentation was not different form that of uni-modal presentation, raising the possibility that the positive effect of multi-modal input was negated by cognitive load of processing the dual information simultaneously. To reduce the cognitive load, we manipulated the presentation timing of auditory and visual information and directed the participants’ attention to one of the two modality information. The results in Experiment 2 to 4 suggested when the visual text presentation preceded the auditory one in one phrase, and participants attended to the visual information during audio-visual text presentation, they could process audio-visual text information effectively.

The situation of visually preceding condition in Experiment 3 is similar to our oral reading condition. When we read some text orally, we look ahead before uttering [7], which means visual information precedes and is mainly attended whereas auditory input is used secondarily. In this situation, we are exposed to different information from visual and auditory sources simultaneously, but we can process text by actively attending to the visual information. As in the case of oral reading, this study showed text comprehension is well facilitated when visual information precedes auditory one, even though different information from two sources is presented simultaneously. This line of study can contribute to develop better reading interfaces (e.g., digital book software).

For speech perception, lipreading is useful for identifying what is said [1] [2]. This means visual information is used secondarily for processing auditory information. On the other hand, this study have shown that participants used visual information mainly for audio-visual text comprehension. Comprehension performance was well facilitated when visual information was presented ahead and attended whereas the text comprehension was not facilitated when the auditory information preceded visual information, nor participants attended to auditory information. This visual information superiority might occur because participants in our study were proficient at reading. For the adult participants, visually presented text might be processed more quickly than the same text presented auditorily. Therefore they may tend to focus on processing visual information mainly during audio-visual text. Future studies should investigate the audio-visual text processing in children who have not much experience of reading.

Our results that the performance in visually preceding condition was highest are not inconsistent with the findings in the previous studies of speech perception that the temporal window is around 200ms when an auditory signal lags a visual signal [8] [9]. In the natural environment, auditory signal usually lags visual signal because sound velocity is much slower than light velocity. Thus, participants might not feel unusual in the visually preceding condition. In our study, an auditory information lagged about 600ms, which was the mean duration of one phrase. We can point out several factors, such as task (speech perception or text comprehension) or type of visual information (speaker’s mouth or text), might contribute to the difference in the size of temporal window. It is worth investigating these factors to explain the process of audio-visual integration during speech perception and text comprehension comprehensively in a future study.

In this study, we found the difference in the performance of the text comprehension task. However, there was no effect of presentation mode nor timing in the performance of literal memory question in all experiments. More research is needed to distinguish the memory and comprehension process and demonstrate the effect of presentation timing more persuasively.

7. Acknowledgements

This study was supported by Grant-in-Aid for JSPS Fellows (24 • 7995).

8. References