ORIGINAL ARTICLES

INFLUENCE OF AUDITORY INFORMATION ON READING SPEED AND EYE MOVEMENT CONTROL IN READING

Kazumichi MATSUMIYA*, Hideki SUGIYAMA**, Satoshi SHIOIRI* and Ichiro KURIKI*

* Research Institute of Electrical Communication, Tohoku University, 2-1-1 Katahira, Aoba-ku, Sendai 980-8577, Japan **Miyagiken Technical High School, 3-2-1 Komegafukuro, Aoba-ku, Sendai 980-0813, Japan

Abstract: We investigated how auditory information affected the speed of silent reading, using sentences extracted from Japanese novels. Subjects were instructed to read the sentences silently presented on a display while hearing the voice of reading through headphones and we measured the reading time and eye movements. The speed of the reading voice varied over the range of a factor of eight to examine whether auditory information affects reading speed. The results showed that the speed of reading and saccade length increased approximately in proportion to voice speed and the number of fixations decreased with the increase of voice speed. The increase of saccade length can be interpreted as the increase in the range of characters to be processed during a fixation (or effective visual field size). From these results, we suggest that the voice information may increase the effective visual field size for reading. **Keywords:** *Vision, reading, Voice, Eye movements, Effective visual field, Multimodal, Psychophysics*

1. INTRODUCTION

School teachers usually use visual and auditory information to teach in classes. Indeed, teachers are writing if they are not speaking in their class. For good teaching, it is important to provide the visual and auditory information effectively. How to provide effective visual and auditory information largely depends perhaps on individuals' experience. There is little systematic knowledge about how to increase the efficiency of visual and auditory information presentation [1].

Psychophysical studies have demonstrated that visual information interacts with auditory information [2-7]. 'McGurk effect' [5] is an example. When the person in the display makes the move of the mouth for speaking 'ga' and a speaker makes the sound of 'ba' at the same time, the subject hears the sound of 'da,' which is the intermediate sound between 'ga' and 'ba.' The brain integrates visual information with auditory information. Another example that is more relevance to the efficient use of visual and auditory information for teaching is the study of Kondo and Kakei (1993). They showed that the detection of a Japanese syllable in the noisy sound is improved by displaying the letter at the same time when making the sound of the syllable [7]. This suggests that the visual processing of a letter also interacts with sound processing. Although these studies showed the interaction between visual and auditory information, they were limited to syllables. Our purpose is to investigate how visual and auditory information interact in reading sentences, for example, on a screen or on a blackboard.

The pattern of eye movements provides important information of the ability of reading. People read sentences by making saccadic eye movements between the fixations of several hundreds milliseconds at a location on words [8-11]. Saccade length is known to affect reading speed and often related to the field size within which information is processed at each fixation (effective or functional visual field) [8-12]. In this report, we examined how auditory information influenced reading, and analyzed the relationship between the effect of auditory information on reading and the eye movement pattern. For the purpose, we measured the reading speed of sentences and the eyemovements with reading voice of various speeds for the same sentences.

2. METHODS

Visual stimuli were presented on a CRT display controlled by a visual stimulus generator ViSaGe (Cambridge Research Systems Inc., U.K.) and a computer. The visual stimuli consisted of sentences extracted from three Japanese novels: 'Yodaka-no-hoshi written by Kenji Miyazawa', 'Majutsu written by Ryunosuke Akutagawa' and 'Sanshiro written by Souseki Natsume'. We used forty-eight sentences of 69.5 letters on average. The display size was 52 deg in width and 34 deg in height in visual angle. One character subtended about 2 deg in width. Figure 1 shows an example of visual stimuli. Auditory stimulus was the voice of reading the sentence presented on the display and the voice started with the start of visual stimulus presentation. Voice data were taken from a web site that releases the voice data of books for



free [18]. We changed the reading speed of the voices, using a software named STRAIGHT [13]. This software was appropriate here because it can change the reading speed without large change in voice characteristics (simple change of the voice speed also changes the stimulus frequency and faster voice becomes more feminine and slower voice becomes more masculine). Four variations in reading speed were chosen: four times, two times and half of the original voice data in addition to the original. With the manipulation, the average speeds of the reading voice were 2.56, 5.13, 10.4, and 15.6 characters/sec on average. We labeled them as 2.5, 5, 10, and 15 characters/sec conditions for simplicity. Auditory stimuli were presented through headphones.

The subjects were instructed to read a sentence presented on the display silently at a speed that was best for his/her comprehension, and to ignore auditory stimulus. To keep the level of comprehending the sentences, the experimenter told the subjects that he would ask several questions related to the content of the sentences at the end of the experiment. When the subject pressed a button to start a trial, the visual and auditory stimuli were presented, and the trial ended when the subject pressed a button again to indicate the finish of reading. Auditory stimulus was not terminated until the end of the sound data. We defined reading time as the time from the initial button press to the next press. Since the numbers of letters were not identical among sentences, we calculated a reading rate by dividing the number of characters in each stimulus by the measured reading time. After all trials were finished, the subjects answered questions made up from each of arbitrarily chosen ten sentences, which were presented on a sheet.

Using an eye tracker (Cambridge Research Systems Inc., U.K.), we measured eye-movements during reading. We analyzed the saccade length, the fixation duration and the number of fixations.

Twelve naïve subjects with normal or corrected-tonormal visual acuity participated in the experiment. Twelve subjects were divided into four groups of three subjects. Each subject performed 48 trials with different 48 sentences. We divided the 48 sentences into four sets. Each set was used for each of four voice speed conditions. The sentence sets and the voice speed conditions were counterbalanced across subjects so that all sentences were used in the all four speed conditions. Subjects in one subject group ran the experiment with a certain combination of sentence sets and voice speed conditions. Subjects in other groups ran the experiment with different combinations. For each subject, voice conditions were presented in random order.

3. RESULTS AND DISCUSSION

Figure 2 shows examples of eye-movement trace of one subject at the voice speeds of 2.5 and 15 characters/sec. In Fig. 2, the solid circle represents the fixation point, and the solid line represents the trace of saccadic eye movement. The subject read the sentences repeating saccadic eye movements and the fixations as has been known. The difference in saccade length (the distance between adjacent fixations) can be seen between the two conditions shown in Fig.2. Saccade length is larger with the voice speed of 15 characters/sec than with that of 2.5 characters/ sec (the effect is clearly shown particularly in the first line).

Figure 3 shows the reading speed averaged over 12 subjects as a function of voice speed. The solid symbols represent the results when the subjects read sentences with voices. The open symbol represents the average reading speed of four subjects who read the sentences without



Figure 2: Eye-movement traces during reading for the same sentence but different subjects. (a)The voice speed of 2.5 characters/sec. (b)The voice speed of 15 characters/sec.



Figure 3: Reading speed as a function of voice speed. The solid symbols represent the average of the twelve subjects when they read sentences with voices. The open symbol represents the average of the four subjects when they read sentences without voices. Error bars indicate standard error of mean.

voices after the experimental session with voice stimuli. An ANOVA revealed a significant effect of voice speed (F(3, 44) = 6.84, p < 0.001). The reading speed increased with the voice speed. However, the change in reading speed was much shallower than the physical voice speeds. Reading speed was faster than the voice speed at the voice speeds slower than 10 characters/sec, and slower at the voice speeds faster than 10 characters/sec. Obviously, the subjects did not simply follow the voice during silent reading.

The reading speed without voice was approximately the same as the speed with the second fastest voice speed.

Considering this speed as a base line, the results suggest that reading speeds up by fast voice and slows down by slow voice. However, the reading speed without voice might be overestimated because the subjects read the same texts in the experimental condition with voice. In this case, we could still claim that fast voice hastens reading because the reading speed in the fastest voice condition is faster than the reading speed without voice. We also believe that slow voice delays reading because we found the difference in saccade length between the early and late stage of reading sentences in the slow voice conditions (see below). It is clear that reading was influenced by voice information.

The numbers of correct answers to the questions of the contents varied among the subjects. We found no correlation between the numbers of the correct answers and the reading speeds. The most of the questions required memory of some details, which were not usually memorized when one would understand a sentence. For example, one of the questions was "What did Sanshiro throw from the train window?" We simply assume here that each subject read all sentences with making a similar effort to understand all sentences. This is confirmed by the fact that individual reading speeds showed the similar tendencies to that for the average shown in Fig. 3 despite the difference in correct rate of questions related to some of sentences.

Figure 4 shows the eye-movement data averaged over 144 trials (12 sentences of 12 subjects) in each voice speed condition. The numbers of fixations were normalized as the number for each letter (fixation number was divided by the letter number of the sentence in each trial). The saccade length increased with the voice speed (Fig. 4a) similarly to the reading speed. We did not analyze saccade duration because it is relatively short (< 50 ms) and information during saccades is little processed [19, 20]. An



Voice speed (characters/sec)

Figure 4: Results of the behavior of eye movements for the average of the twelve subjects. (a) Saccade length as a function of voice speed. (b) Fixation number as a function of voice speed. The numbers of fixations were scaled as the number-per letters. (c) Fixation duration as a function of voice speed. Error bars plot S.E.

ANOVA showed a significant effect of voice speed (F(3, 44) = 4.24, p < 0.05). The number of fixations decreased with the increase of the voice speed (Fig. 4b). An ANOVA revealed a significant effect of voice speed (F(3, 44) = 5.56, p < 0.005). The fixation duration, on the other hand, was approximately constant over different voice speeds (Fig. 4c). An ANOVA showed no significant effect of voice speed (F(3, 44) = 0.84, p = 0.48 ns).

We analyzed the frequency distribution of fixation durations for each voice speed in order to examine how the configuration of the distribution of fixation durations changes depending on voice speed. Stimulus difference may change the distribution function in some cases [21]. The frequency distribution of fixation durations is shown in Fig. 5. Fixation durations had high frequencies at the period between 100 ms and 250 ms for all voice speeds. Not only the mean fixation duration, but also its distribution were influenced little by the voice speeds. The eye movements results described so far suggest that fixation number and saccade length are main factors to change reading speed.

We next examined the temporal dynamics of the effect of auditory information. At the beginning of each trial, they heard the voice of the sentences that they were reading. However, they heard the different part of the sentence at the end of each trial unless their reading speed was the same as the voice speed. Therefore, the effect of auditory information could vary with time because subjects could have heard the voice of the text very different depending on the part of the sentence. Although reading speed cannot be estimated at a certain time during reading, the temporal dynamics of the effect of voice on reading can be assessed from eye movement data. To examine whether the effect



Fixation duration (msec)

Figure 5: Frequency distributions of fixation durations. The solid circles, open squares, solid diamonds and open triangles represent 2.5, 5.0, 10 and 15 characters/sec, respectively.

of voice vary with time in a trial, we compared the mean saccade length at an early stage of trials (0 to 2 sec from the start of trials) with that at a late stage of trials (5 to 7 sec from the start of trials). Figure 6 shows the mean saccade length as a function of voice speed for the twelve subjects. The solid and open symbols represent the period of 0-2 sec and the period of 5-7 sec, respectively. The effect of voice speed on saccade length is significant for 0-2 sec period (F(3, 44) = 3.65, p < 0.05) while that is not significant for 5-7 sec period (F(3, 44) = 0.65, p = 0.59ns). However, the results of the ANOVA have nontrivial influence of the individual variations and the variations may hide the influence of the voice speed. Indeed, individual results show a strong tendency of increase in saccade length with voice speed even for the later periods. To examine whether there is the correlation between the voice speed and saccade length, we calculated the correlation coefficient for each period of each subject. We found that the significant positive correlation between saccade length and voice speed for both periods. (the mean correlation coefficient = 0.73, t(11) = 7.63, p < 0.0001 for 0-2 sec period; the mean correlation coefficient = 0.68, t(11) = 10.75, p < 0.0001 for 5-7 sec period). This analysis reveals the significant effect of voice speed even for the late period, although ANOVA shows no significant effect of voice speed. Note that these two analyses are not inconsistent



Figure 6: Comparison of saccade lengths between the period of 0-2 sec (solid symbols) and the period of 5-7 sec (open symbols). The graph shows the mean saccade length for the twelve subjects as a function of voice speed. Error bars plot S.E.

because showing no significant difference does not indicate that there exists no difference.

To compare the amount of contribution of voice information to saccade length between the two periods, we calculated the slope of saccade length with voice speed for each period. The mean slopes of saccade length across the subjects for the early and late periods were 0.088 and 0.028, respectively. The slope was significantly larger in the early period than in the late period (t(11) = 4.00, p < 0.005). This suggests that the effect of voice was larger at the beginning than at the end of trials, although there was the influence of voice speed in both periods.

The difference in saccade length between the early and late periods indicates that the voice effect is dynamic. However, the difference might come from the presence and absence of voice sounds. The performances of the reading speeds show that the reading voices were presented for most of the period of silent reading except for the voice speed of 15 characters/sec. The voices were presented for about 28, 14 and 7 sec for the voice speeds of 2.5, 5.0 and 10 characters/sec, and the average reading times were about 10, 9.0 and 7.2 sec for their voice speeds. For the voice speed of 15 characters/sec, the reading voice was presented for about 4.7 sec while the mean reading time for the twelve subjects was about 6.7 sec. Therefore, the reading voice was presented for the early period (0-2 sec) but it was not presented for the late period (5-7 sec). Nevertheless, saccade length was similar between the early and late periods. A statistical test showed that there is no significant difference in saccade length between the early and late periods (F(1, 22) = 0.23, p = 0.63 ns). This suggests that the influence of auditory information was held even at the late period as well as at the early period, indicating that the effect of auditory information on reading is not due to simply following the voice of the text.

4. GENERAL DISCUSSION

The present study revealed that the increase of the voice speed influenced both reading speed and eye movement patterns. With the increase of voice speed, saccade length increased and fixation numbers decreased. Fixation duration, in contrast, was approximately constant across different voice speed conditions. These results suggest that the decrease of fixation number is the primary factor of the increase of the reading speed as the voice speed increases.

Assuming that fixation number is crucial for the change in reading speed, we discuss the reasons why fixation number changed with the change of the voice speed. Firstly, the eye may have followed the letter that voice indicates at each moment. At least at the beginning of each sentence, subjects were likely able to find the letter of voice easily and their fixation may have been attracted to the letter. However, it is clear that this eye following effect cannot explain the reading speed obtained in the experiment because the reading speed of the subjects was very different from the speed of voice in each condition. In other words, the subjects often finished reading either long before or long after the voice. There should be another factor in addition to the eye following effect.

Second reason can be changes in the size of the effective or functional visual field. Several psychophysical studies estimated the range of characters to process during each fixation in reading [8-10], which is referred to as an effective or functional visual field. The size of an effective visual field is related to saccade length, although the effective visual field is usually wider than the saccade length [8, 9, 12]. The visual system likely chooses the next point of fixation based on the information that is processed on the periphery at a current fixation. If the visual system can process information from a larger area, the saccade length should increase, reducing the number of fixation. Consequently, reading time becomes shorter. It is possible, therefore, that the increase in saccade length reflects the increase of effective visual field size. Here, we define the effective visual field as a visual field within which certain visual information can be processed. Following the definition, we assume that effective visual field for identifying letter or word increases functionally when saccade length becomes long. Subjects can process more characters during a fixation when fast reading voice is given and that they can process less characters during a fixation when slow reading voice is given. In the case of faster voice presentations, the voice information is presented before reading the corresponding text at most of the time, and the voice information could help to recognize the letters and/ or words in periphery. In the case of slower voice presentations, the voice information is presented after reading the corresponding text at most of the time, and the voice might interfere with recognizing letters and/or words. Perhaps, the influence of voice is at the level of letter or word recognition and the same visual inputs from peripheral visual field can be processed differently depending on conditions. When a word can be recognized with a help of voice information presented before, the visual field is effectively larger than when it cannot be recognized.

Alternatively, it is possible that the increase in saccade length might reflect a process that is different from the letter or word recognition process. In the case of faster voice presentations, for example, the voice was presented before reading the corresponding parts at the most of the time. This might have made the subject guess some of the content of the sentences before they actually read and they could have skipped some parts of the sentences (skimming). This makes saccade length longer on average. This is different from the change of the effective visual field size, which assumes the change in visual processing. It is worth to note, however, that some visual information is required even in this case. To make a saccade over some parts which content can be guessed, the subjects have to know where they jump over and where they fixate next. It

know where they jump over and where they fixate next. It is hard to believe that the visual system can program a saccade without any information from the peripheral vision. In this sense, we could argue that the size of effective visual field in our definition may change for reading text.

The influences of auditory information may be caused by attention and practicing, which have been suggested to change the size of an effective visual field dynamically [14-17]. However, attention and practice are not likely involved in the present experiment. In the present study, the subjects were instructed to ignore the voice, and there was no reason why the auditory information would impose any explicit attentional load. The texts were different among trials, and there was no reason why the subjects would obtain any special skill, either. Therefore, the effect of auditory information should be different from either attention or practice.

Although our results suggest that the auditory information influences letter or word perception, we did not investigated which aspect of the auditory information. In the present study, the contents of auditory information were always consistent with those of visual information, while the speed of the voice was changed. We do not have any information about how the speed of different voice, such as the voice of reading different sentences or meaningless voice-like sound, influences silent reading. Considering the effects of attention on detecting auditory information might provide some insight into the issue about the effects of the voice contents. In the dairy life, we hear a variety of sounds but we can ignore the sound that we are not interested in. We speculate that readers might be able to ignore auditory information if it contains only irrelevant contents. However, the future work has to be done to resolve this issue.

In conclusion, the present study indicated that auditory information affects the reading speed and the behavior of eye movements in reading. This suggests that providing auditory information may cause change in the size of the effective visual field. As a consequence, the presentation of auditory information of a text may increase the number of characters processed during a fixation, improving the reading performance.

ACKNOWLEDGEMENTS

We thank Katsuya Fujita for help in data analysis. This work was supported by a Grant-in-Aid for Scientific Research from the Japanese Ministry of Education, Culture, Sports, Science and Technology, Scientific Research on Priority Areas No. 18047005 to SS and Young Scientists (B) No. 18700253 to KM.

REFERENCES

- T. Sato; Introduction to Educational Information Technology; Corona Publishing Co. Ltd., Tokyo, pp.20-47 (1989) [in Japanese].
- C. Jackson; Visual factors in auditory localization, Quarterly J. Exp. Psychol.; 5, pp.52-65 (1953).
- 3. D. H. Warren; Intermodal interactions in spatial localization, Cognit. Psychol.; 1, pp.114-133 (1970).
- B. Jones; Spatial perception in the blind, Br. J. Psychol.; 66, pp.461-472 (1975).
- 5. H. McGurk, J. MacDonald; Hearing lips and seeing voices, Nature; 264, pp.746-748 (1976).
- B. R. Shelton, C. L. Searle; The influence of vision on the absolute identification of sound-source position, Percept. Psychophys.; 28, pp.589-596 (1980).
- K. Kondo, K. Kakei; Effects of simultaneously presented character on auditory phoneme perception in a syllable, Technical Report of IEICE; SP93-43, pp.65-72 (1993).
- G. W. McConkie, K. Rayner; The span of the effective stimulus during a fixation in reading, Percept. Psychophys.; 17, pp.578-586 (1975).
- 9. M. Ikeda, S. Saida; Span of recognition in reading, Vision Res.; 18, pp.83-88 (1978).
- N. Osaka; Effect of peripheral visual field size upon eye movements during Japanese text processing. In; J. K. O'Regan and A. Levy-Schoen (Eds.) Eye movements: From physiology to cognition; North-Holland, Amsterdam, (1987)
- S. Saida; Eye movements in reading. In; R. Osaka, S. Nakamizo and K. Koga (Eds.) Experimental Psychology of Eye Movements; Nagoya University Press, Nagoya, pp.167-197 (1993) [in Japanese].
- S. Shioiri, M. Ikeda; Useful resolution for picture perception as a function of eccentricity, Perception; 18, pp.347-361 (1989).
- H. Kawahara; STRAIGHT, exploitation of the other aspect of VOCODER: Perceptually isomorphic decomposition of speech sounds, Acoust. Sci. & Tech.; 27, pp.349-353 (2006).
- M. Ikeda, T. Takeuchi; Influence of foveal load on the functional visual field, Percept. Psychophys.; 18, pp.255-260 (1975).

- 15. K. Matsubara, S. Shioiri, H. Yaguchi; Spatial spread of visual attention while tracking a moving object, Opt. Rev.; 14, pp.57-63 (2007).
- K. Matsumiya, K. Uchikawa, I. Kuriki; Change in Contrast Sensitivity in the Peripheral Visual Field by Central Visual Task and Observation under the Condition with Eye Movements, Journal of the Institute of Image Information and Television Engineers; 52, pp.565-570 (1998) [in Japanese].
- 17. T. Miura; Information processing in piano playing: An eye movement study, Proc. 1st Int. Conf. Music Percept. Cognit.; pp.295-298 (1989).
- Koe-no-hanataba; Japan Philanthropic Association., Tokyo, <u>http://www.koetaba.net/3book/</u>
- 19. S. Shioiri, P. Cavanagh; Saccadic suppression of lowlevel motion, Vision Res.; 29, pp.915-928 (1989).
- 20. F. C. Volkmann; Human visual suppression, Vision Res.; 26, pp.1401-1416 (1986).
- S. Shioiri, Postsaccadic processing of the retinal image during picture scanning, Percept Psychophys; 53, pp.305-314 (1993).



Kazumichi MATSUMIYA

Kazumichi Matsumiya received Ph.D. degree from Tokyo Institute of Technology in 2000. Then, he was a postdoctoral researcher at York University in Canada until December, 2001. From January, 2002 to December 2003, he was a postdoctoral researcher at Imaging Science and Engineering Laboratory,

Tokyo Institute of Technology. From January, 2004 to March, 2005, he was a research fellow at Human Information Sciences Laboratory, ATR. He moved to Research Institute of Electrical Communication, Tohoku University as a research associate in April, 2005. He is currently working there as an assistant professor. His research interests are in visual psychophysics, multimodal integration, motion perception, eye movements.



Hideki SUGIYAMA

Hideki Sugiyama graduated from Department of Communication Engineering, Tohoku Institute of Technology and received B. Eng in 1990. Then, he worked at Kahoku Shimpo co. inc. From 1992, he worked as a teacher at Kesennuma Koyo High School, and lectured on Electricity and Electronics

and Wireless Communication. He was also the director of the committee for guiding the course of students. He moved to Miyagiken Technical High School in 2002, and has lectured on Information Technology and Communication Technology. He is a permanent member of Miyagiken High School Culture League, and the director of Miyagiken High School Photograph League. In 2006, he was a trainee at Research Institute of Electrical Communication, Tohoku University, and he studied visual psychophysics at Shioiri-Kuriki Laboratory. His research interests are in the relation between vision and audition, and eye movements.



Satoshi SHIOIRI

Professor Shioiri graduated Tokyo Institute of Technology and received Dr. Eng in 1986. Then, he was a postdoctoral researcher at University of Montreal until May of 1989. From June of 1989 to April of 1990, he was a research fellow at Auditory and Visual Perception Laboratories of Advanced

Telecommunications Research Institute. He moved to Chiba University at May of 1990, where he spent 15 years as an assistant professor, an associate professor, and a professor of Department of Image Sciences Department of Image, Information Sciences and Department of Medical Systems. In 2005, he moved to Tohoku University. Since then, he has been a professor of Research Institute of Electrical Communication of Tohoku University. His research interests include motion perception, depth perception, color vision, mechanisms of visual attention and eye movements, and modeling of visual functions.



Ichiro KURIKI

Dr. Kuriki received Ph.D. degree from Tokyo Institute of Technology in 1996. After then, he worked at Imaging Science and Engineering Laboratory, Tokyo Institute of Technology as a research associate until October, 1999. He worked as a research associate at the Department of Mathematical Engi-

neering and Information Physics, Graduate School of Engineering, the University of Tokyo until March, 2001. He worked as a researcher in Communication Science Laboratories of NTT Corporation until December, 2005. He moved to the Research Institute of Electrical Communication, Tohoku University as an Associate Professor in January, 2006. His main interest of research is the mechanisms of human color vision. He uses non-invasive brain-imaging methods, case studies on brain-damaged patients and computational models in addition to psychophysical methods. He achieved best paper awards from the Optical Society of Japan in 2000 and Illumination Engineering Society of Japan in 1998.