Investigation of Speech-planning mechanism based on eye movement

HUANG Jinfeng (s1610067)

School of Information Science, JAIST, lllhjf@jaist.ac.jp

Extended Abstract

A major issue in speech production research is how speakers do speech-planning, and how long is taken. Reference with to the process of speech production, the triggering conditions for generating speech planning can be summarized as two ways. In the case of conversation, a speaker conducts a speech planning based on oneself speech intentions first, and then execute the planned to control motor commands articulatory movements, and finally generate speech sound. In the case of reading, speakers first obtain the text via their eyes, and comprehend the sematic information, and then carry out the same procedures as those in conversation.

According to the consideration above, some studies investigated the process of speech planning during reading isolated words by measuring the response time from the word presenting to onset of the uttering. They found the frequency and the length of the words would significantly affect the average latent time of speech planning. However, the results obtained from the isolated words are difficult to investigate the speech planning for real situation in which utterance is continuous but not isolated.

Some other studies combined the eye movement with acoustic measure, where they believed that eye movements can reflect speech planning processes during reading. Meyer et al., used the time difference between the onset of the eye leaving a word and the onset of uttering the word as the response latency. In uttering a sentence continuously, the flows of obtaining the text word by word, planning it and uttering it are executed simultaneously. There is no evidence to show that the onset of the eye leaving a word is the onset for planning. Furthermore, we found that in the latter part of a sentence the onset of uttering a word is earlier than the onset of the eye leaving the word. In the criterion of Meyer et al., the latent time will be a negative value. It is not reasonable. For this reason, we define the period from the onset of eye reaching a word to the onset of uttering the word as “the latent time”, which was used in the latent time measurement by acoustics only. According to Rayner, three standard eye movement parameters are selected and used in this study, which are the first fixation if there are multiple fixations on a target word, mean fixation duration that is the average of all effective fixations in a sentence, and total viewing time that is the summation of all fixations duration on the sentence.

In this paper, we investigated the mechanism of speech planning in the continuous speech by combining eye movement and speech sound. The major tasks

of this study are analyzing variations of the latent times of speech planning, and the unit used in speech planning in continuous speech.

Most of the previous studies using isolated words to measure the latent time of speech planning, the general conclusion based on those studies was that the latent time was related to the word length, the longer the words and the longer the latent time. In this study, we used different types of sentences for such an investigation. Our result did not show significant difference caused by the word length in the latent time. Instead of, we found that the latent time is heavily dependent on the location of the words in a sentence, where the latent time is linearly reduced as the word is located in the latter part of the sentence. When the sentence is a complex sentence, the tendency of the latent time decreased gradually along the time axis within each sub-sentence, while the tendency was reset in the boundary of two sub-sentences.

Based on these results, we can speculate that the latent time consists of two parts: semantic comprehension and motor command design. Among the two parts, the time consumed by semantic comprehension is reduced because the semantic information provided by preceding words shortened the comprehension time of the following words. To prove our hypothesis, we constructed a number of non-meaning sentences by rearranging the order of words in the normal sentence to exclude the semantic support for the following words. The results support our hypothesis.

Although this study did not ask any comprehension task, the result implies that semantic comprehension automatically takes place in the uttering task, and the semantic comprehension is executed in a sentence level or sub-sentence level. Context is important for semantic comprehension. The result from the complex sentence indicates that semantic discontinuity reset the activity of the semantic comprehension. In the experiment with non-meaning sentences, the average latent time did not change with word location, but its variation is much larger than that from the normal sentence. That is because that some subjects try to comprehend the semantic information for the non-meaning sentences.