JAIST Repository
https://dspace.jaist.ac.jp/

Title	遅延聴覚フィードバックが日本語文章読み上げ課題における発話リズムに与える影響
Author(s)	姫子松, 洋介
Citation	
Issue Date	2024-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/18892
Rights	
Description	Supervisor: 鵜木 祐史, 先端科学技術研究科, 修士(情報科学)

Japan Advanced Institute of Science and Technology

Abstract

In spoken communication, the speaker uses speech to express his intention, and the listener listens to the speech to understand the speaker's intention. However, as it is said that "the speaker is also the listener," the speaker's speech is heard by two routes: the listener and the speaker himself. In this study, we focus on the speaker's own listening route. We listen to speech with our own ears, and if there is an error in the speech, we can adjust speech production based on the error. This is called auditory feedback, and humans perform this behavior unconsciously. In addition to verbal information, speech also contains linguistic and paralinguistic information (such as the speaker's emotion), and auditory feedback of this information enables us to understand speech. In everyday life, when communicating with others, the speaker can listen to his or her own speech and adjust it to make sure others can better understand it. Smooth adjustment and communication are possible when auditory feedback works well. However, there are some people for whose auditory feedback may not function well. Examples include speech impediments such as stuttering and pronunciation errors that occur when learning another language. Stuttering is defined as a speech style that disrupts fluency through repetition of sounds, stuttering, and pauses in speech. The relationship between stuttering and auditory feedback has been studied extensively, and it is hoped that methods to improve speech production will be established soon. Regarding pronunciation errors that occur when learning other languages, there are many languages other than Japanese, such as English, German, French, etc., and humans are not able to speak all of them. Linguistic speech contains phonetic information in the form of phonemes, syllables, mora, and other phonetic units. All languages are known to have isochrony of speech units, and the phenomenon of isochrony speech units is defined as the rhythm of speech production. In this study, the rhythm of speech production is defined as the rhythm of speech production. It is known that the world's languages can be divided into three speech rhythms, and Japanese is classified as a mora rhythm. Thus, it is necessary to understand and perceive vocalization and the speech rhythm of a language in order to acquire the speech of another language, but it is thought that misperception of the speech rhythm can lead to pronunciation errors. To solve these problems, speech improvement methods using delayed auditory feedback (DAF) have been studied. has been studied. It is known that when normal speakers speak under DAF, the larger the amount of delay, the more unstable the speech motor control becomes, such as repetition of sounds similar to stuttering symptoms. Assuming that speech motor control based on auditory feedback information is not fully functional, this study aims to realize a speech improvement method using delayed auditory feedback by examining the relationship between the delay time of auditory feedback

information and speech motion on a speech unit basis.

There have been several studies on DAF and speech motor. Yamamoto and Kawabata found that the larger the feedback delay, the longer the average mora duration and the slower the speech rate in a reading task to adapt the DAF (adaptation-in-progress task). In particular, the mean mora duration was longest when the delay was 200 ms. In the post-adaptation task, speech rate increased with a delay of 200 ms, and the mean mora duration was shorter than in the adaptation task. This suggests that the speech motor itself adapts to the DAF and is capable of producing somewhat fluent speech. Ichinose and Noda also found that speech rate decreased the most (was most affected by DAF) under DAF with a delay of 200 ms when significant and meaningless sentences were read aloud. These results indicate that speech produced by reading Japanese sentences aloud under DAF is affected by the speech rate and the mean mora duration, which is the speech unit. However, it is not clear whether the mora duration of the entire utterance or a portion of the utterance is lengthened, and in either case, the mean mora duration is lengthened. Furthermore, they do not mention whether the isochrony of mora changes as the average mora unit lengthens. Dellwo and Wagner used DeltaC, the standard deviation of consonant intervals, and %V, the ratio of vowels to total utterance time, as rhythm indices to quantify speech rhythm, a phenomenon in which speech units are isochronous. An inter-linguistic analysis was conducted to classify rhythms based on differences in speech rate among English, German, and French. Intra-linguistic analysis was also conducted to classify rhythms based on differences in speech rate within the same language, including individual differences. The results showed that DeltaC increased with increasing speech rate in all languages for intra-linguistic analysis, indicating that rhythm classification is possible even with the inclusion of individual differences. On the other hand, no significant change in %V was observed with increasing speech rate. In the inter-lingual analysis, the data for each speech rate were plotted on a graph with %V on the horizontal axis and DeltaC on the vertical axis, and English and German were plotted at positions where %V was small and DeltaC was large, while French was plotted at positions where %V was large and DeltaC was small, indicating that rhythm classification is possible. In a similar experiment, Dellwo analyzed VarcoC, the coefficient of variation of consonant intervals, to look at the relative variability of speech. The results showed that in the intra-linguistic analysis, VarcoC values differed from language to language as speech rate increased, and that some languages could be classified as containing individual differences while others could not. The inter-linguistic analysis showed that the plots were in the same position as DeltaC, indicating that rhythm classification was possible. However, since the speech rhythms of these three languages are classified into strong and syllabic beats, it has not been mentioned whether the rhythmic index of mora beat rhythm changes depending on individual differences such as speech rate.

In this study, we will conduct a speech experiment in a Japanese sentence reading task under DAF in order to observe how the rhythmic indexes of the above phonemes and mora beats change with the amount of feedback delay. also, Phonemes and mora are extracted from the speech data obtained in the experiment, and their respective rhythm index are obtained. In the speech experiment, 10 sentences of approximately 30 mora were read at four delay conditions (0 ms, 100 ms, 200 ms, and 300 ms) and the speech was recorded. The recorded speech was used for speech rhythm analysis. The first step in the analysis is annotation, which extracts phonemes and morae. Next, the annotated data is used to obtain a rhythmic index for each participant. Next, a intra-participant analysis is performed to see how each participant's rhythmic index varies with the amount of delay in the DAF. The rhythm index is then averaged for each delay amount and evaluated using a inter-participant analysis and statistical tests to see if there is a difference in the mean rhythm index for all delay amounts and if there is a difference between the mean value for delay amount 0 ms and the mean value for other delay amounts. The analysis yielded the following findings. First, the mean durations of consonants, vowels, and moras tended to increase with increasing delay time, and the two experimental participants tended to vary significantly. However, there were large differences in the mean duration of consonants and moras across all conditions, but no differences in vowels across conditions. Second, DeltaC, DeltaV, and DeltaM tended to increase with increasing latency, and two experimental participants tended to vary significantly. These three rhythm indices differed significantly among all conditions. Third, VarcoC, VarcoV, and VarcoM tended to increase with increasing latency, with the majority of VarcoC tending to vary significantly. Half of the VarcoV indicators also tended to change significantly. However, VarcoC and VarcoV differed significantly among all conditions, while VarcoM did not differ significantly among all conditions.

The results of the speech rhythm analysis yielded the following findings. First, the average duration between consonants and moras was found to increase with feedback delay. Second, the variability between consonants, vowels, and moras was found to increase with feedback delay time. Third, feedback delay time was found to increase the relative variability of consonant and vowel durations in speech compared to speech without feedback delay.