Precise Understandig of Reading Activities
Sight, Aural and Page Turning

Kyota Aoki, Shuichi Tashiro and Shu Aoki
Graduate School of Engineering, Utsunomiya University, 7-1-2 Yoto, Utsunomiya, Japan

Keywords: Reading Activity, Automatic Measurement, Objective Measurement, Multiple Features.

Abstract: In Japanese public elementary schools, every pupil may use an ICT device individually and simultaneously. In the cases, a few teachers must teach all pupils. For being welcomed from a teacher, the ICT devices must help pupils to use the ICT devices by itself for effective usage, and it must help the teacher to use the ICT devices in a class. For help the user, the ICT devices must understand the state of the user. For help the teacher, it must precisely understand the users' reading activities. This paper proposes a method to recognize the precise reading activity of a user with read aloud voices and facial images. This paper discusses the relation between reading activity and features caught with a voice and a facial image, and proposes the method to implement the precise understanding of reading activity.

1 INTRODUCTION

In Japan, if a pupil shows two years delay of reading ability, we say that the pupil has a reading difficulty. Some Japanese normal public elementary schools have about 20% of pupils with a light reading difficulty. Of course, there are pupils with a heavy reading difficulty. The pupils with a heavy reading difficulty attend special support education classes or schools.

Recently, ICT devices are spreading in Japanese elementary schools. In a near future, there is an ICT device for every pupil in a normal class. We will cover the easy problems about the usage of an ICT devices with the ICT device itself. In Japan, a normal class includes about 32 pupils. About 20% of pupils have some problems about using ICT devices. We will cover the 80% of the problems with the ICT device itself. In the case, the teachers can treat only two pupils that have the problems not covered by the ICT device itself.

To help a pupil, an ICT device must understand the activity of a pupil precisely. A human teacher can observe and understand not only the activity but also the inner state of a pupil. However, it need a huge computation power and a huge measuring system. In this paper, we will propose the method to understand the activity precisely with the feasible ICT devices in a near future. The understanding of an activity is the start point of understanding of the inner state of a pupil.

In a near future, the personal ICT device will have the power of a personal computer now. So, our goal must be achieved with a personal computer. Now, a personal computer has a camera, a microphone, a keyboard and a touch panel to input.

We have developed Japanese text presentation system to help the pupils to read Japanese texts (Aoki et al., 2014, Aoki and Murayama, 2012). To the system, we will add the ability to understand the precise reading activity of a user. Already, the system has the ability to recognize the rough reading activity. In this paper, we improve the ability to understand the more precise reading activity of a user.

Frist, we discuss the precise reading activity. Then, we discuss the relation between the measurable actions and reading activity. Next, we show the method to understand reading activities with images and sounds. Then, we conclude this work.

2 READING ACTIVITIES

2.1 Japanese Texts

First, we must discuss the structure of Japanese texts. Japanese texts include mainly three types of characters. Two types of characters are Hiragana and Katakana. They are phonogram as alphabet. The other is Kanji. Kanji is ideogram. There is no word spacing in Japanese texts. We can easily recognize

Aoki, K., Tashiro, S. and Aoki, S.
Precise Understandig of Reading Activities - Sight, Aural and Page Turning.
In *Proceedings of the 8th International Conference on Computer Supported Education (CSEDU 2016) - Volume 1*, pages 464-469
ISBN: 978-989-758-179-3

word chunks with the help of boundary between a Kanji character and Hiragana character. A sequence of Katakana character makes one word that represents the phonetic representation of a foreign word.

Japanese sentence ends by a punctuation mark. We can easily find a sentence in a sequence of characters. In a sentence, we can find a word chunk starting from a Kanji character and ending at the last Hiragana character in a sequence of Hiragana characters. There may be a word chunk only including Hiragana character. In the case, we have some difficulty to find a ward chunk.

2.2 Change of Japanese Text in Elementary School Ages

In Japanese elementary schools, pupils start to learn Japanese characters. In Japan, many infants learn Hiragana before elementary school ages. However, an elementary school is the first step of compulsory education in Japan.

In six years of an elementary school, pupils learn Hiragana, Katakana, and Kanji characters. In Japan, if a pupil shows two years delay of reading ability, we say that the pupil has a reading difficulty. Some Japanese normal public elementary schools have about 20% of pupils with a light reading difficulty. Of course, there are pupils with a heavy reading difficulty. The pupils with a heavy reading difficulty attend special support education classes or schools.

Teachers want to help pupils with reading difficulties. However, it is difficult to find pupils with light reading difficulties in first and second year in an elementary school. If we can understand the precise reading activities, we can find a tiny sign of reading difficulties in very first stage. Teachers can help the pupils in very first stage of reading difficulties. The fast guidance may prevent the increase of reading difficulties. In many cases, a fast guidance is more effective than a late guidance.

Table 1: Number of Kanji characters to learn.

School year	#Kanji to learn
1	80
2	160
3	200
4	200
5	185
6	181
Total	976

Table 1 shows the number of Kanji to learn in a school year. In s first year of an elementary school, there are only 80 Kanji characters learned. So, the text for a pupil at the start of second year only includes about 80 Kanji at most. In a second year, texts have no word spacing as normal Japanese texts. At this stage, some pupils show reading difficulty about recognizing word chunks in a sentence. However, they can read the sentence as written by Hiragana and small number of Kanji. Their reading aloud voice has a features that can be detected by experienced teachers.

In older pupils, there is a problem about Kanji. Some pupils do not remember enough number of Kanji. Some pupils do not remember the phenomes representing the Kanji. In the case, a teacher easily finds the problem. However, there needs long time for checking all pupils in a class.

Our Japanese text presentation system enables to check all pupils in a class simultaneously. This enables to repeat the test in a short interval.

2.3 Word Chunk

In Japanese texts, a word chunk forms the sequence of characters starting from Kanji, and ending to Hiragana. Of cause, in a very first year in elementary school life, almost all word chunk is formed only by Hiragana. In the texts, a word chunk is separated from other chunks with a space.

Our Japanese text presentation system presents a text with three levels of masking and high-lighting. With the high-lighting, a user can easily find a word chunk.

The standard length of a high-lighted part expands with the development of reading ability. In the long high-lighted part, a pupil finds basic word chunks and recognize the relations among word chunks. In older pupils, there is a problem about this function.

In the text for older pupils, there are many Kanji characters. So, it is easy to find a basic word chunks in a sentence. However, in a long high-lighted part, there are complex relations among word chunks.

Some older pupils with reading difficulties have problems about recognizing the relations among word chunks. Experienced teachers can find this problem easily. This problem appears in a long sentence that enables to include complex relations of word chunks. Using a long text for checking this kind of reading difficulties, the time for checking must increase. As a result, it is difficult to check all pupils in a class. In this case, our Japanese text presentation system can help a teacher with the precise understanding of pupils' reading activities.

3 MEASURABLE ACTIVITIES

3.1 Reading Environments and Activities

In normal class room, there are 40 pupils at most in Japan. A class room is well lighted and has windows at south side. There is no heavy noise.

There are two types of reading activities. One is reading aloud, and the other is a silent reading. In silent readings, there is no aural activities. We cannot estimate the precise place of readings. In reading aloud, we can estimate the place of readings.

Our long goal is the understanding reading activities reading aloud and silent reading. However, our next step is understanding the precise reading activities in reading aloud.

The reading activity in reading aloud has many sub-activities. They are looking at a text, looking at a sentence, following a sequence of word, recognizing word chunks, recognizing the relations among word chunks, understanding a sentence, constructing a sequence of vocal sounds, and uttering aloud the sequence of vocal sounds. There are observable actions and un-observable actions. Figure 1 shows the structure of the activity of reading aloud. In Figure. 1, doublet rectangles are observable sub-actions. Other are un-observable actions. Figure 2 shows the relation between observable actions and measured features. User's observable reading actions makes four kinds of measureable body actions. They are key operations, eyes' movements, mouth movements, and voice utterances.

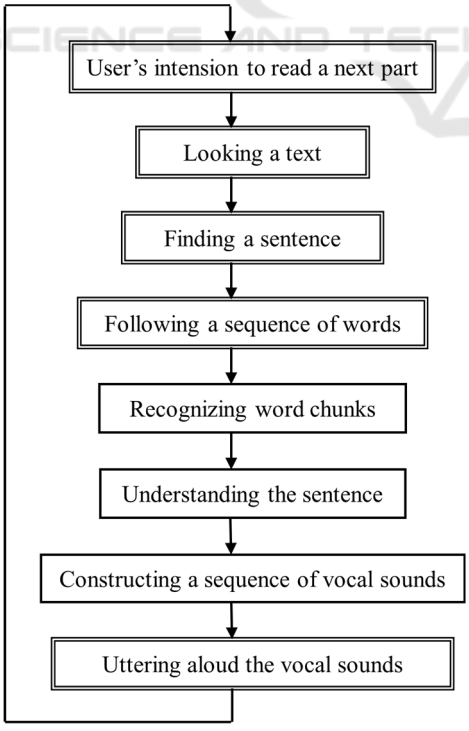

Figure 1: Activities about Reading aloud.

3.2 Effective Sensors

Now, a personal computer has a camera, microphones, a touch-panel, and a key-board for input. A camera takes full-HD images. A camera takes a user's facial images. In full-HD images, we can recognize eyes and irises. Of cause, we can recognize a mouth.

A microphone of a personal computer is not best for distinguish a voice of a user among others' voice and noises. In a near future, a personal computer can have an array of microphones. However, now, a personal computer's microphone does not construct an array. However, with the help of a user's mouth movements, we can distinguish the voice of a user among others' voices and noises.

In our Japanese text presentation system, a touch-panel has no role. Key-inputs are clear presentations of user's intension about reading texts. Observable user's reading activities are expressed through the movements of muscles. Figure 3 shows the relation between observable actions and resulting features.

3.3 Relation between Sensors and Activities

A camera catches the facial image of a user. In facial images, there are eyes and a mouth. In reading activities, sight takes important role. Eyes are only sensors supporting sight. The motion of eyes represents the reading activity directly.

The movement of a mouth is also caught by a camera. The movement of a mouth represents the reading aloud action itself.

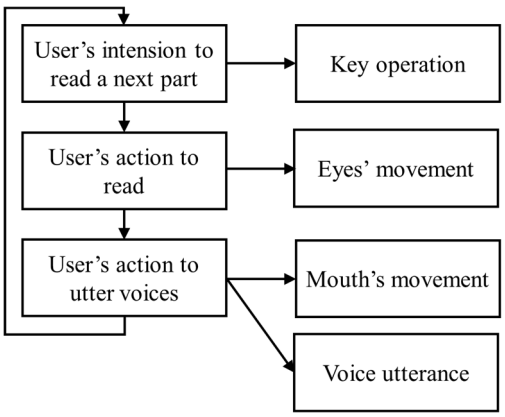

Figure 2: Relation between actions and measured features.

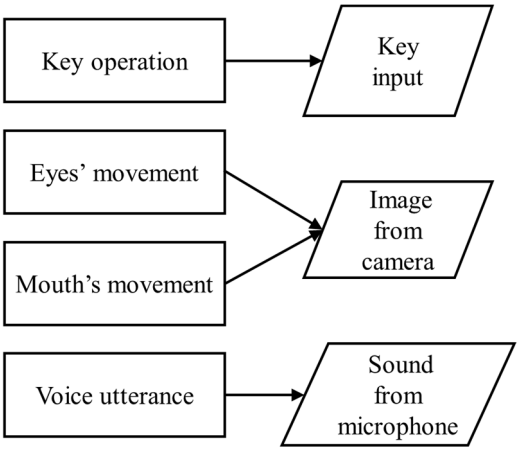

Figure 3: Relations between features and sensors.

A microphone catches the reading aloud voices. In reading aloud, a voice is the direct expression of reading aloud. A key operation is the only expression of the intension to proceed the next word chunk to read, not an expression of reading activities.

4 PROCESSINGS ABOUT IMAGE AND SOUND

4.1 System Overview

We built our previous Japanese text presentation system with Python, Pyglet, Julius, Mecab, and OpenCV. The previous system already utilizes the benefits of multi-processing (Julius, 2016) (Mecab, 2016) (OpenCV, 2016) (Python, 2016) (Pyglet, 2016). Mecab is a Japanese part-of-speech and morphological analyser. Now, a personal computer's processor can handle two or more process simultaneously. Our system utilizes this benefit. Constructing a system based on multi-processing, it is easy to make many of real-time measurements without depending each other.

Python is a programming language powerful enough to include all those features. Pyglet is a real-time library only depending Python itself. This feature keeps portability. In Japan, public schools' ICT devices are decided by Education Board of each city or town. A distractive change of ICT devices may occur. In the case, portability of our system helps to survive.

4.2 Key Operations

Our Japanese text presentation system does not turn page. The Japanese text presentation system puts the high-lighted part forward with user's key-operations. In reading activities, the key-input to put the high-lighted part forward is the only key-operation. Our Japanese text presentation system understand user's intension to proceed the reading part.

4.3 Image Processing

4.3.1 Mouth Movements

Figure 4 shows the relations among three detectors and three measurements of features. The camera of a personal computer catches facial images. With a facial image, we can have a movements of a mouth. In reading aloud, the user's mouth must move. The motion of a mouth is easily measured while the user's face shows no motions. However, in some cases, a user's face moves. We finds the base-point in a face image. In our processing method, a nose and two eyes make base-points. With the base-points, we measure the relative motion of a mouth. Figure 5 shows the precise structure of the measurements of mouth's movements.

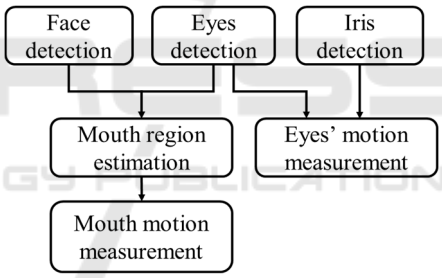

Figure 4: The relations among detectors and features.

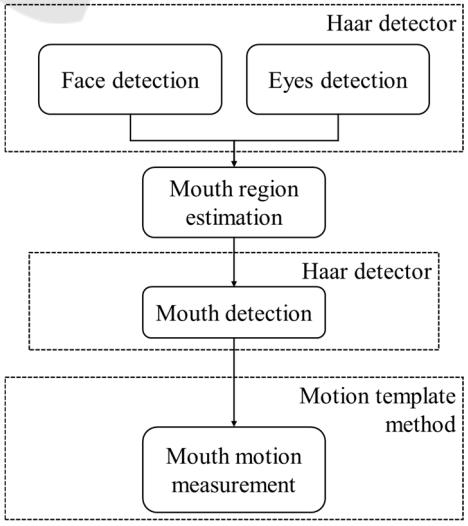

Figure 5: Processing about images.

In many cases, the movement of a mouth relates the action of reading aloud directly. Some pupils move mouth without reading aloud. In the case, there is no direct relation between the movement of a mouth and reading aloud. The multiscale detections decrease the errors about mouth detections. Figure 6 shows an example of parts detections on a face image. The rectangles shows the results of a face detection, eyes detections, a nose detection and a mouth detection.

4.3.2 Eye Movements

A normal personal computer camera is a normal colour camera. With a normal colour camera, it is difficult to catch the precise gazing position. However, we can estimate a left-right movement of eyes with the normal colour images. We can measure the relative position between irises and eyes. The left-right movements of eyes is important in understanding the process following the word chunks in a text. A normal Japanese text book about language learnings places texts in vertical lines. However, our Japanese text presentation system can present Japanese text in vertical lines and horizontal lines. In normal usages, our Japanese text presentation system shows texts in horizontal lines, because the display in our experimental environments is wide profile. In the case, left-right movements of eyes carries much information about the reading activities.

4.4 Sound Processing

The microphone of a personal computer catches the voice of a user. We can have a reading aloud voice. However, in a normal class room, there are many pupils. Multiple pupils read aloud simultaneously. In the case, a cheap microphone of a personal computer catches voices of many pupils. It is not easy to distinguish the voices of other pupils.

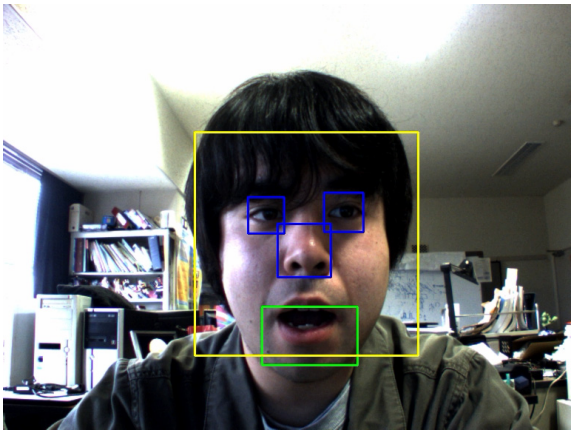

Figure 6: An example of parts detections in an image.

If we have a time while a user is speaking, we can catch the feature of a user. When we have the feature of the voice of a user, we can distinguish the voice of the user from other voices.

4.5 Corporation between Measurements of Mouth Movement and Measurement of Voices.

Using both of a mouth's movements and voices, we can distinguish the reading aloud activity more precisely.

4.5.1 Voice Distinction with the Help of Mouth Movement

Voice distinction is difficult using only a cheap microphone. With the help of the measurement of mouth's movements, we can have a timing of voice utterance. With the timing of voice utterance, we can distinguish the voice of a user from others' voices and noises. Once we have the reliable samples of a user's voice, we can distinguish the user's voices from others' voices more easily. Figure 7 shows the relations among mouth's movements and utterance detections.

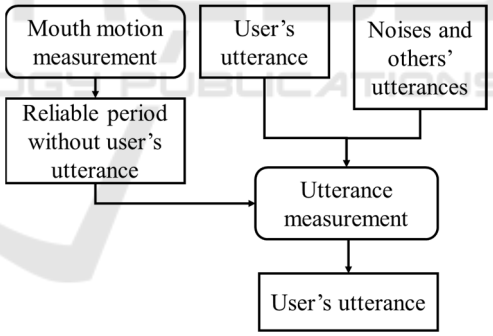

Figure 7: Voice distinction with the help of mouth movements.

4.5.2 Personal Mouth Movement

A personal difference of mouth a movement is important to understand the reading activity of a user. Our preliminary experiments show that some peoples move their mouths without making utterances. With the help of voice measurements, we can distinguish the personal type of user's mouth movements in reading activities. Figure 8 shows the relations among voice measurements and mouth's motion measurements. After understanding the personal type

of reading activity, we can precisely understand the users' intension about reading activity.

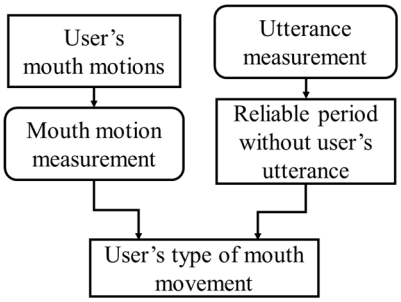

Figure 8: User's type estimation with the help of voice's measurement.

4.5.3 Precise Understanding of Reading Activity

Once our measurement system make to be able to distinguish the voice of a user from other voices and noises. We can easily find the part that a user is reading aloud. We use JULIUS for recognizing phonemes (Julius, 2016). JULIUS has the ability to recognize voices and makes the Japanese text representing the voice. However, in uncontrolled environments its performance is not good. Our system uses only the phenome recognition part of JULIUS.

The other hand, our measuring system understands the personal type of mouth movement of a user. The system can find troubles about reading activities. In some cases where a user has a trouble, he starts to read aloud and stops the activity before generating a voice. At the time, his mouth starts to move for making a voice, but he does not speak aloud.

5 CONCLUSIONS

With a cooperative measurements of audio and video, the proposed system can understand the precise reading activities of a user. Our previous works only uses the total reading time of a sentence for understanding the types of reading activities. However, there is a limitation for understanding the reading activity. Our new cooperative measurements of audio and video enables to understand the reading activities based on the word chunk uttered by the user. This understanding of a user's reading activity enables to recognize the problems of the user. Some of the recognized problems relate about the usage of the Japanese text presentation system itself, and others do about the reading ability of the user.

With the recognized problems about the usage of the system, we will expand the function to help a user. About the reading ability of a user, we will expand the functions to fit the presentation for the user's reading ability, and the functions to report the problems about the user's reading ability to a teacher.

The report to the teacher must include the profiles of a user's reading ability and the problems about the usage of the Japanese text presentation system that are not solved by the system itself. With these expand functions, teachers will introduce and use the Japanese text presentation system.

ACKNOWLEDGEMENTS

This work is partially supported with JSPS25330405.

REFERENCES

Aoki, K., Murayama, S., Harada, K.: Automatic Objective Assessments of Japanese Reading Difficulty with the Operation Records on Japanese Text Presentation System. CSEDU2014, vol. 2, pp.139-146, Barcelona, Spain, (2014).

Aoki, K., Murayama, S., 2012. Japanese Text Presentation System for Persons With Reading Difficulty -Design and Implementation-. CSEDU2012, vol.1, pp. 123-128, Porto, Portugal. (2012).

Julius. https://github.com/julius-speech/julius. (2016).

Mecab. http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html?sess=3f6a4f9896295ef2480fa2482de521f6. (2016).

OpenCV. http://opencv.org/, (2016).

Pyglet. http://pyget.com/about.html. (2016).

Python. https://www.python.org/. (2016).