# Precise Estimation of Reading Activities with Face Image and Read Aloud Voice

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 primary schools, every pupil may use an ICT device individually and simultaneously. In normal primary school, a few teachers must teach all pupils in a class. It is difficult to help all pupils to use an ICT device. For using ICT devices individually in a normal class, the ICT device help its' user by itself. To help the user, the ICT device must understand the state of the user. To help a teacher, it must precisely understand the users' reading activities. Reading ability is the base of all subjects. It is important that pupils acqure reading ability. This paper proposes a method to recognize the precise reading activity of a user with read aloud voices and facial images, shows its implementation, and experimental results. With the cooperative analysis of a read aloud voice from a microphone and a movement of a mouth from a camera, our implementation enables to estimate the action of read aloud much more precisely. The timing of a read aloud action is estimated in phrase by phrase manner. In-vitro experiments confirm the performance of our implementation.

## 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 primary 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. In primary school years, girls show about two years progress than boys in their development. This difference makes it difficult to teach boys and girls in a same class.

In Japan, Information and Communication Technology (ICT) devices is spreading. In a near future, there is an ICT device for every pupil in a normal class. A personal ICT device may help to overcome the personal difference. However, the number of ICT devices in a class makes new problems. Some pupils need a help to use ICT devices. Some ICT devices may fail. In the case where is one ICT device in a class, a teacher can treat the problems. However, the number of ICT devices make it difficult that a teacher processes all the problems.

To introduce a personal ICT device in a normal class, we must decrease the problems about an ICT device drastically. However, there must be problems to use ICT devices. For this reason, we will cover the easy problems about the usage of 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 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 and implement the method to understand the activity precisely with the feasible ICT devices in a near future. The understanding of an activity is a start point of understanding of the inner state of a pupil.

In a near future, the personal ICT device will have a power of a powerful personal computer now. Therefore, 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.

Aoki, K., Tashiro, S. and Aoki, S.
Precise Estimation of Reading Activities with Face Image and Read Aloud Voice.
DOI: 10.5220/0006315603150322
In *Proceedings of the 9th International Conference on Computer Supported Education (CSEDU 2017) - Volume 1*, pages 315-322
ISBN: 978-989-758-239-4

We have developed Japanese text presentation system to help the pupils to read Japanese texts (Aoki, Murayama and Harada, 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 an ability to recognize the rough reading activity (Aoki, Murayama, Aoki and Tashiro, 2015). And, we show our project to enable to estimate much more precise reading activity (Aoki, Tashiro and Aoki, 2016). This paper shows the implementation of the project and experiments to estimate the ability to understand the 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, this paper proposes the implementation of the proposed method and experimental results. And last, 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 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 Primary School Ages

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

In six years of an primary 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 primary 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 a primary 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.

In the first year of an primary school, there are only 80 Kanji characters learned. Therefore, the text for a pupil at the start of second year only includes about 80 Kanji at most. Texts have word spacing. 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 features that can be detected by experienced teachers.

In elder 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, most of word chunks form the sequence of characters starting from Kanji, and ending to Hiragana. Of cause, in a very first year in primary 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 (Aoki, Murayama and Harada, 2014; Aoki and Murayama, 2012). 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 elder pupils, there is a problem about this function.

In the text for elder pupils, there are many Kanji characters. Therefore, 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 elder 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 length 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 words, 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. Observable actions are eyes' movement, mouth's movement, utter voice and key operation. Figure 1 shows the relation between observable actions and measured features.

## 3.2 Effective Sensors

Now, a PC 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 PC is not best for distinguish a voice of a user among others' voices and noises. In a near future, a PC can have an array of microphones. However, now, a PC'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. Figure 2 shows the relation between measured features and sensors.

## 3.3 Relation between Sensors and Activities

Figure 2 shows the relation between a reading activity and a sensor that can catch the activity. A camera catches the facial image of a user. In facial images, there are eyes and a mouth. In reading activities, sight

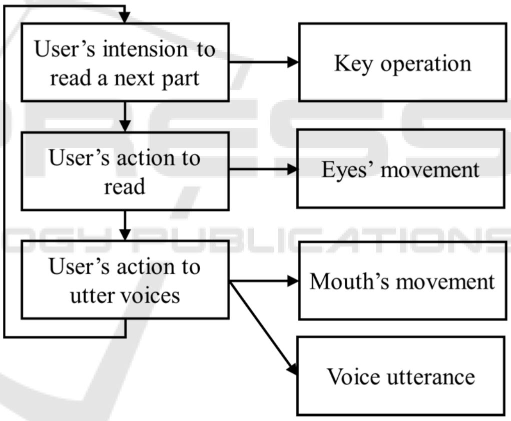

Figure 1:Relation between actions and measured features.

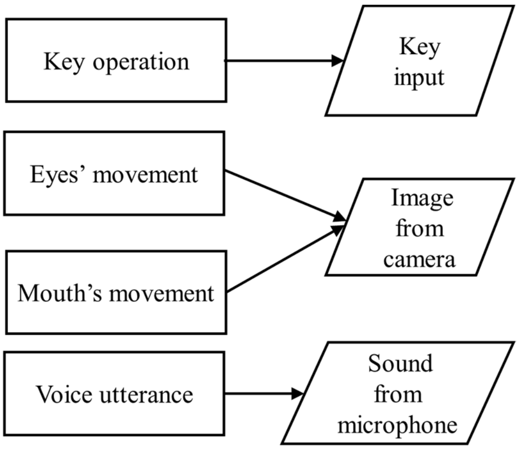

Figure 2: Relations between features and sensors.

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 itself. It also represents the preparation for uttering voices.

A microphone catches the reading aloud voice. In reading aloud, voice is the direct expression of reading aloud.

A key operation is the only expression of the intension to proceed the next word chunk, not an expression of reading activity.

# 4 PROCESSINGS ABOUT IMAGE AND SOUND

## 4.1 System Overview

The authors 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 PC'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 the education board of each city or town. A drastical 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

The camera of a PC catches facial images. With a facial image, we can have a motion 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 move. However, in some cases, a user's face moves. We find the base-point in a face image. In our processing method, a nose is a base-point. With the base-point, we measure the motion of a mouth. To take a motion, we need at least two frames. Therefore, at first, one frame is get and processed to find the region of interest.

In many cases, the movement of a mouth relates the action of reading aloud directly. However, in precise measurement, there are mouth movements just before an utterance. We need to make a proper form of mouth to make a proper sound. Some pupils move mouth without reading aloud action. In the case, there is no direct relation between the movement of a mouth and reading aloud activity.

If we know the type of a reader, we can treat this problem. However, with video images only, it is difficult to decide the type of a reader. With the help of a sound processing, we treat this problem. A video of a PC is properly fixed to the user. So, there is little possibility to mixed with other face images. However, sound is not restricted to the user. In a normal class room, the distance between pupils is short. The sound of other pupils must mix into the sound detected by a microphone of a PC.

### 4.3.2 Method and Implementation

A face is not fixed in a video frame. We need to compensate the movement of a face. We need to measure the movements of a mouth in a face. So, we need to the movement of a mouth based on the place which is not move in a face. There may be fixed eyes and a nose in a face. There are a face detector, a nose detector, a mouth detector, and an eye detector in Opencv. However, these detectors are not powerful enough to detect their objects without error detections. The large regions as a face is detected with less error detections than the small regions as an eye. So, we restrict the region searched to detect an object. However, to restrict the region searched, we need to know the region. So, we first detect the largest region. The largest region is a face. A face includes eyes, a nose and a mouth. Figure 3 shows the relations among sub-process of image processings.

In experiments, the upper three fifth of the face region detected is a candidate region to detect a region of eyes. If the detection of eyes has a single detection result, our system decides the regions of a nose and a mouth using the face region detected and the eyes region.

The detected regions of a face and eyes are not strictly defined. As a result, small amount of movements is not correct enough to estimate the motion of a mouth. We estimate the precise movements of a face and a mouth using a block matching method.

With the region detected as a nose, we decide the region of interest for block-matching. The block size and the block placement depends on the image size. Block-matching needs much computations. In the detected nose region, we make block matching to estimate the precise face movement. The block size is eight by eight pixels square in our experiments. The blocks are arranged as a tile arrangement. There is no overlap and no gap.

In a nose region, there is a little deformation. So, a small number of blocks can work well. However, in a mouth region, there are relatively large deformations. However, we select eight by eight square as a block size, and do a similar arrangement of blocks as a nose region.

We define the mouth region with the face region and eyes regions detected. With the coordinate of the detected regions, a mouth region is defined as (1) and (2). (1) is the upper left corner of the region. (2) is a width of the region. (3) is a height of the region.

$$(\mathrm{cx} - \mathrm{int}(\mathrm{mr}\ \times 0.4), \mathrm{cy} + \mathrm{int}(\mathrm{d} * 1.1)) \quad (1)$$

$$\mathrm{int}(\mathrm{mr} \times 0.9) \quad (2)$$

In these equations, mr is the half of the width of a face region detected. d is the distance between the center of a left eye and one of the right eye detected.

A nose region is defined as (4), (5) and (6).

$$\mathrm{int}(\mathrm{mr} \times 0.55) \quad (3)$$

(4) defines the upper left corner of the region. (5) defines the width of the region. (6) defines the height of the region.

In (6), (cx, cy) is the center of two eyes as (1). In (5) and (6), mr is the half of the width of a face detected.

For example, in our experiments, the size of a nose region is 77x77 pixels. The size of a mouth region is 126x77 pixels.

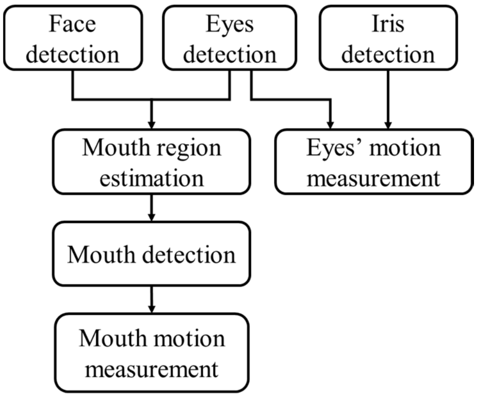

Figure 3: Image Processing relations.

$$(\mathrm{cx} - \mathrm{int}(\mathrm{mr} \times 0.25), \mathrm{cy} + \mathrm{int}(\mathrm{mr} \times 0.15)) \quad (4)$$

$$\mathrm{int}(\mathrm{mr} \times 0.55) \quad (5)$$

$$\mathrm{int}(\mathrm{mr} \times 0.55) \quad (6)$$

The region of interest of a moth region is set as a lower part of a real mouth region. Based on the place of a nose, upper region of a mouth does not move much. The movements of an upper region of a mouth is mainly a deformation.

The lower region of a mouth moves largely vertically with the open and close of a mouth. Of cause, with the vertical movement, there are deformation also. Our face skins move continuously.

We define the move of a mouth as the relative motion of a lower part of a mouth based on the nose position. With block-matching, we have large number of motion vectors. Our system ignores the horizontal motions, So, we have the set of vertical motions. The position deference of a nose is defined as the average of vertical difference of block-matchings.

The motion of a mouth is defined as the average difference between the motion of a mouth and the motion of a nose. A nose is not move in a face, so the motion of a nose represents the motion of a face.

The result of block matching is a set of two-dimensional vector that represent the motion of each block. The motion of a nose is defined with the mode of the result of block matching.

The decision about the motion of a mouth is done with the threshold defined with the width of the face region detected.

There are 6 to 12 years old pupils in primary schools. There is a difference of the distance between a camera and a face. As a result, there must be a change of a scale. For robustness about the scale change, the threshold changes rationally with the scale of a face.

### 4.3.3 Experiments about Mouse Motion Detection

We have three experiments about mouse motion detection. Figure 4 shows the results of mouth region detections. In the figure,a yellow box is detected face region. Blue boxes show the eyes and a nose. A green box represents the mouth region.

Table 1 shows length of mouth movement in each experiment. Table 2 shows the performances of the three experiments about mouth movement detection. The F-values of all three experiments are larger than 0.8. The performance of mouth movement detection is good enough.

## 4.4 Reading Aloud Detection

From the sound processing, we have the utterance period. If the user in front of a camera really read aloud, there must be a motion of the mouth of the user and the utterance of the user. Therefore, the system decides that the read aloud activity is in the period when both of the mouse movements and the utterance are detected.

The detection of a read aloud voice is simple. When a microphone catches a sound that is louder than a threshold, our system decides that there is a sound of read aloud voice.

Table 1: Length of mouth movement.

| Experiment # | Length (seconds) | Length of mouth movements (seconds) |
|---|---|---|
| 1 | 19.93 | 9.250 |
| 2 | 22.40 | 12.00 |
| 3 | 18.07 | 5.33 |

Table 2: Performance of mouth movement detections.

| Experiment # | Precision | Recall | F-value |
|---|---|---|---|
| 1 | 0.926 | 0.758 | 0.833 |
| 2 | 0.889 | 0.865 | 0.877 |
| 3 | 0.875 | 0.824 | 0.848 |

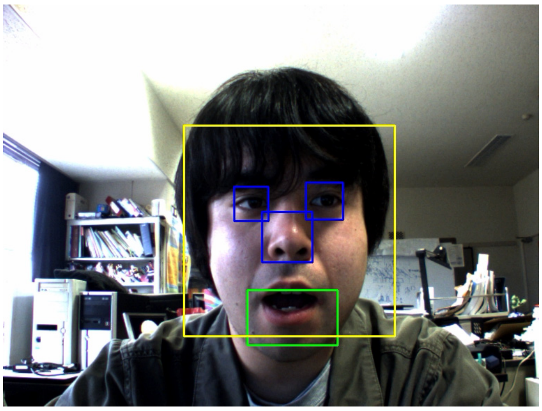

Figure 4: Mouth and face detection.

### 4.4.1 Experiments

In vivo, experiments to estimate the system's performance must be large. Therefore, it is difficult to execute. This paper makes the experiments in vitro. In the experiments, there are a user in front of a camera and a microphone, and the others interfere with their voices at the side of the user outside of the sight of the camera and near enough the microphone.

There are three types of experiments. One experiment has no interfere. Only a user read aloud a text. In another experiment, a user read aloud only the odd-numbered parts and an interfere read only the even-numbered parts. In another experiment, a user read aloud only the odd-numbered parts, and only moves his mouth without voice at the even-numbered parts.

Figure 5 shows the results of mouth movement detection and sound detection without interfere. In the figure, horizontal scal is time in second. Vertical scale represent Yes/No. Large waves represent the result of sound detection. Small waves represent the result of mouth movements. The dots do the timing of key operations.

Figure 6 shows the results in the experiment with interfere. In the odd-numbered parts, both of sound and mouth movements are detected. In the even-numbered parts, only sounds are detected.

Figure 7 shows the results with irregular mouth movements. In the odd-numbered parts, both of sound and mouth movements are detected. In the even-numbered parts, there is no detection of sounds.

In all these three experiments, the decision about read aloud activities are correct. There is no fail decision.

## 4.5 Sound Processing

The microphone of a PC catches the voice of a user. We can have a reading aloud voice. However, in a

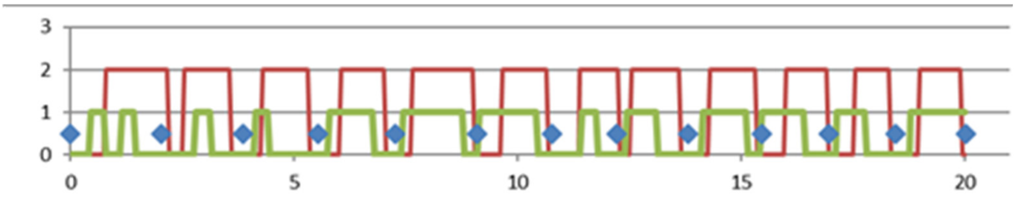

Figure 5: Result of mouth movements detection and sound detection.

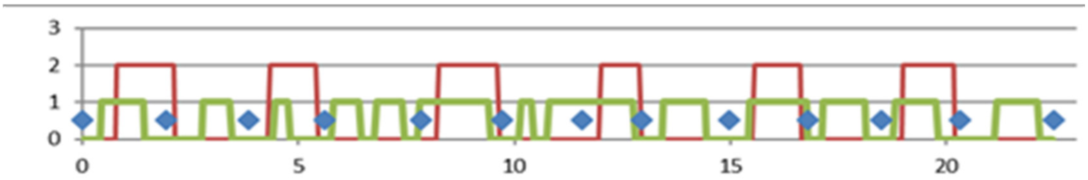

Figure 6: Result of mouth movements detection and sound detection.

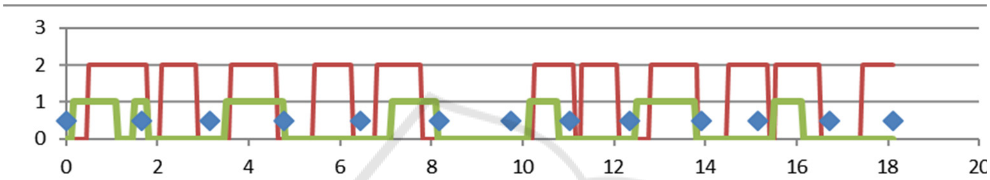

Figure 7: Result of mouth movements detection and sound detection.

normal class room, there are many pupils. Multiple pupils read aloud simultaneously. In the case, a cheap microphone of a PC catches voices of many pupils. It is not easy to distinguish the voices of other pupils.

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.

### 4.5.1 Precise Reading Activities from Sound

When we have a reading aloud voice, we can estimate the timing of reading activity precisely. However, a user can make some mistakes about pronunciations. Japanese texts have Hiragana, Katakana, and Kanji characters. The pronunciation of Hiragana and Katakana is strictly defined. There is a little mistake about the pronunciations of Hiragana and Katakana. Kanji characters have multiple pronunciations. They are two types of pronunciations. One type is Kunyomi, and the other is Onyomi. The Kunyomi is the old Japanese word that represents the same meaning of the Kanji character. The onyomi is the pronunciation of the Kanji character in old china.

Some users make wrong pronunciations about Kanji characters. It is difficult to recognize the pronunciations. We treat a voice as the sequence of phonemes. Of cause, there are errors in a sequence of phonemes. We must find the correspondence between the sequence of phonemes of read aloud voice and the sequence of phonemes that represents the proper read aloud voice of a text.

We use weighted edit distance to find the proper correspondence between the sequence of phonemes of read aloud voice and the sequence of phonemes that represents the proper read aloud voice of a text (Levenshtein, 1966).

Proposed system calculates weighted Levenshtein distance and finds the correspondence between the sequence of phonemes of read aloud voice and the sequence of phonemes that represents the proper read aloud voice of a text. If there is a few reading errors, the system finds the proper correspondence. In the case, the system finds the part of reading errors and the types of the reading errors.

There are four types of reading errors that we treat. One is a simple pronunciation error. This type of an error leads a substitution of phonemes between a read aloud voice and a text. Another one is a skip over of a part of a text. In this case, the skipped part is deleted from a text. The last one is a rereading of a part of a text. In the case, the rereaded part are inserted into a text.

The proposed system estimates these types of errors in user's reading activities. This helps to understand user's reading activities precisely.

Table 3: Correspondence between the text and the read aloudvoice with some pronousation errors.

| Process | C | C | C | C | I | I | C | C | C | C | C | C | O | O | C | C | C | C | C | C | C |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Phonemes of a text. | m | u | sh | i | * | * | s | a | g | a | sh | i | o | sh | i | m | a | sh | i | t | a |
| Recognized Phonemes | m | u | sh | i | r | o | s | a | g | a | sh | i | * | * | * | m | a | sh | i | t | a |
| similarity | o | o | o | o | x | x | o | o | o | o | o | o | x | x | x | o | o | o | o | o | o |

### 4.5.2 Experiments about Precise Understanding about Read Aloud Voice

For measuring the performance of the proposed system, we make some experiments about read aloud voice processings. The experiments are four types. One type is that a user reread some parts of a text correctly. Another one is that a user read aloud some parts of a text with wrong pronunciations. Another one is that a user skips over some parts of a text. The last one is that the system cannot process user's pronunciations with noises or other causes.

We show one of four experimental results in a table. Table 3 shows the case that has pronunciation errors. In table 3, there are some insertions and deletions. Insertions are represented as '*' in a text. Deletions are represented as '*' in a sequence of phonemes of a read aloud voice.

These experiments shows that the system can understand properly the relation between the parts of a text and the parts of read aloud voices. With this relations, we can understand the reading activities of a user much more precisely. The understanding includes the timing of read aloud and the correctness of read aloud voices.

## 5 CONCLUSION

With a cooperative measurement of audio and video, the proposed system can estimate the precise reading activities of a user. Our previous works only uses the total reading time 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 that is used for processing a text. The key operation that represents the intension of reading next part. The read aloud timing represents the types of reading activities. The precise understanding of reading activities can help to find proper assisting method for pupils with reading difficulties.

We need to measure reading activities of many pupils in next steps.

## ACKNOWLEDGEMENTS

This work is supported with JSPS16K01057 and JSPS16K04818.

## REFERENCES

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

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

Aoki, K., Murayama S., Aoki, S. and Tashiro, S. (2016). Recognition of Reading Activities and Reading Profile of User on Japanese Text Presentation System, Computer Supported Education, 7th International Conference, CSEDU 2015, Lisbon, Portugal, May 23-25, 2015, Revised Selected Papers, pp. 57-80. Springer.

Aoki, K., Tashiro, S. and Aoki, S. (2016). PRECISE UNDERSTANDIG OF READING ACTIVITIES - Sight, Aural, and Page turning-", 8th International Conference on Computer Supported Education, Rome, Italy, April.

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

Levenshtein A. (1966). Binary Codes Capable of Correcting Deletions, Insertions and Reversals, Soviet Physics Doklady, vol. 10, no. 8, pp. 707-710.

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

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

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

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