

Role of a Subjective Difficulty Rating in Using a System for Practicing English Speaking

Harumi Kashiwagi, Kobe University, Japan
Min Kang, Kobe University, Japan
Kazuhiro Ohtsuki, Kobe University, Japan

The Barcelona Conference on Education 2020
Official Conference Proceedings

Abstract

We have been developing a system that checks and provides information about the words and sentences that learners use when they practice speaking English. In this study, we investigated the role of a subjective difficulty rating to identify sentences that were problematic for the learners by using our system. In the experiment, 72 Japanese university students were provided with 47 Japanese sentences and their English translations in advance of review quizzes. We instructed them to practice speaking the English translations without looking at the text information. This system, then asked them to translate the Japanese sentences into English. Then a self-reflective questionnaire was administered and participants rated the difficulty of each sentence on a five-point Likert scale. The average difficulty ratings of the 31 sentences that were answered correctly by more than 80% of the students varied from 1.7 to 4.2. Even though most students answered the questions correctly, they did not regard all the correctly answered sentences as easy. The standard deviation scores of the difficulty ratings of each sentence varied from 0.9 to 1.4. The difficulty ratings of some sentences were different for individual students. These results suggest that a subjective difficulty rating could play a role in observing how students actually feel about the difficulty levels of the sentences in the system, and could identify their individual weaknesses. Incorporating this kind of subjective difficulty rating into the system could help generate useful information for selecting the sentences within the system that suit individual student needs.

Keywords: Language Learning System, Subjective Difficulty Rating, Speaking Practice, CG Characters

iafor

The International Academic Forum
www.iafor.org

Introduction

The Ministry of Education, Culture, Sports, Science and Technology published the English Education Reform Plan (2014) that prioritizes improving Japanese students' oral English proficiency. However, many students report they lack confidence in speaking English (Kashiwagi, Kang, & Ohtsuki, 2018). A lack of opportunities for speaking English may also exacerbate this problem. Students need a practice environment where they can develop their English-speaking skills, and also some studies on reducing EFL learners' unwillingness to speak English are needed as reported by Isoda in a 2009 study. When considering how best to facilitate their English-speaking practice, we need to observe which language items they find difficult to handle and identify the specific words and phrases that pose challenges.

We have been developing a system that checks and provides information about the words and sentences that learners use when they practice their English speaking. In this study we investigated the role of a subjective difficulty rating that identifies sentences that are problematic for learners by using our system. With this system and a questionnaire using a five-point Likert scale, we conducted an experiment with first-year Japanese university students to investigate the following research questions:

1. What difficulty level did students perceive for the sentences they answered correctly? What difficulty level did students perceive for the sentences they answered incorrectly?
2. Are the difficulty ratings of the sentences different depending on the individual student? What is the response variance for each sentence?
3. Could a subjective difficulty rating play a role in providing information about sentences that learners can use to practice English?

First, we describe the system, followed by the experiment and its results, and then discuss the findings. Finally, we give our conclusions on the role of a subjective difficulty rating in our system.

Structure of the System

This section introduces the structure of the system in Figure 1, and provides an example of its use. The structure consists of a question setting tool, "MINI BASIC," and AnimeViewer. AnimeViewer is a viewer tool for generating and displaying CG content from a script written in TV Program Making Language (hereinafter called TVML). To produce CG content in the system, we use TVML Player X. TVML is a text-based scripting language that automatically generates television programs (Hayashi, 1999). TVML Player X is the software used to read a TVML script and it automatically generates the program's video and audio.

For the teacher to easily create TVML scripts by selecting prepared question data, we developed a question setting tool and an application tool called "MINI BASIC" (Kashiwagi, Kang, & Ohtsuki, 2020). Japanese and English sentences are organized into situational categories, including "Cleaning and Washing," "Illness and Injury," and "Commuting" in the question setting tool, and related question data are prepared in a CSV file such as "Qdata file." When question data are selected, these are sent to "MINI BASIC."

In “MINI BASIC,” a template file in TVML format is prepared to create a TVML format question file. A TVML format question file is created in the tool by obtaining question sentence text data from the question setting tool and by inputting the data into the template file in TVML format. In this manner, when a teacher selects a question in the question setting tool, corresponding question sentence data are obtained, and a TVML format question file is automatically generated. The TVML format question file is then sent to the AnimeViewer and the CG content is produced.

After the teacher checks student’s answers, the results of the evaluation are saved in the question setting tool. To exchange data between the question setting tool and “MINI BASIC,” virtual communication ports are used.

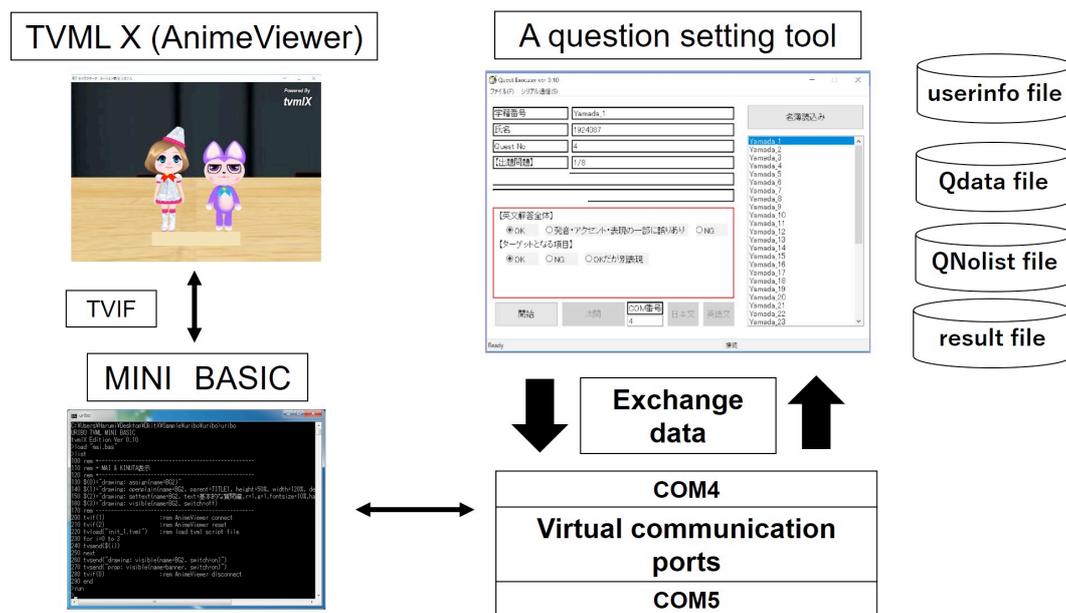


Figure 1: System overview.

An Example of the System in Use

An example of how the system operates is shown in Figures 2 and 3.

1. When a teacher presses the button, the phrase “Read in the user information” appears on the top right side of the screen shown in Figure 2 (1), and user information is read in. Usernames are then displayed on the right side of the screen (2) in the tool.
2. When the teacher selects a username (3), corresponding question data are presented in the middle of the screen (4). For example, in Figure 2, the username “Yamada_1” is selected, and the information from the first question is displayed, in this case, the Japanese sentence “*Sentakumono wo sotoni hoso*” and the English sentence “Let’s hang the laundry outside.”
3. When the teacher presses the “Start” button (5) on the bottom left of the screen, related question data are sent to MINI BASIC and converted into a TVML file. The converted TVML file is then displayed on the AnimeViewer. In this example, CG characters appear on the screen of the AnimeViewer, as shown in Figure 3. The learner is then given instructions such as “Please translate the following Japanese sentence into English.” The Japanese sentence from the first question “*Sentakumono*

wo sotoni hoso” is then displayed on the AnimeViewer in Figure 3 (6). At the same time, a synthetic voice is generated and speaks the Japanese sentence for the learner.



Figure 2: Example screen from the question setting tool.



Figure 3: Example screen of AnimeViewer.

4. The learner orally translates the Japanese sentence into English.
5. While listening to the learner’s answer, the teacher checks the items shown in Figure 2 (7). Two evaluation items are prepared by the system. The first item is used to check whether the answer is correct as a whole sentence including vocabulary, grammar, and pronunciation. Three evaluation levels are set for the item: “correct,” “some mistakes,” and “incorrect.” When there are some mistakes in the answer, such as mistakes in the pronunciation of a word, vocabulary, or grammar, we evaluated it as “some mistakes.” The second item is used to check whether the target word or phrase is correct. Here, the target words or phrases are shown in blue in Figure 4. For

this item, we set three evaluation options, “correct,” “incorrect,” and, “correct, but another expression.”

6. Navigation to the next question is accomplished by pressing the “Next” button shown in Figure 2 (8).

7. When the system finishes providing questions, and after the answers are checked, the results are saved in the “result file” (Figure 1) in the question setting tool.

Experiment

Participants

Participants were 72 first-year students who were learning English at a university in Japan. They took review quizzes in their English language classes using the system. They also answered a post-practice questionnaire rating the difficulty of the questions in the quizzes on a five-point Likert scale.

Procedures

We first provided students with Japanese sentences and their English translations in advance of each of three review quizzes. We instructed them to practice speaking the English sentences without looking at the text information. All 47 sentences are shown in Figure 4. These sentences are expressions related to “Cleaning and Washing,” “Illness and Injury,” “Commuting,” and other daily life situations.

Review quizzes	English sentences	
	Yeah, the laundry has piled up .	I have several loads of clothes to wash.
	OK. Where is the laundry detergent ?	Let's hang the laundry outside.
	Yes, and please fold the laundry.	I have to vacuum the floor.
1	I veg out at home.	It's warm and comfortable .
	There's a 60 percent chance of rain .	Lately the weather forecast has proven right .
	Could you cut it short all around ?	Just trim it overall, please.
	Oh, It has just slipped my mind .	It's on the tip of my tongue .
	I feel dizzy .	I have the chills .
	I feel like throwing up .	I have diarrhea .
	I've got hives on my back.	I sprained my ankle.
	It's just constipation .	You have a seasonal nasal allergy .
2	I'll prescribe a painkiller for you.	I'll prescribe antibiotics for you.
	You might have pneumonia , so I'll take an X-ray for you.	It's tight around my waist.
	I think I need the next size up .	It's in the back of the store.
	It's down the hall .	
	No, I'm not up to it. Go ahead without me.	You're a strong drinker. It doesn't show on your face at all.
	Shall we order sea urchin and fatty tuna sushi?	We should get out of the way of the car.
	Yes, the car is splashing water from a puddle .	I didn't notice the bicycle passing me from behind .
	I can tell we won't make it across the crosswalk from here.	You lose your sense of direction walking around in underground malls.
3	Walking there, I found that it's not that far .	It's just before the third intersection .
	Is there anything that would make a good landmark ?	I'm sorry to have made you call so many times.
	It's all garbled .	My unread e-mail is piling up .
	Unfamiliar attachment files carry viruses .	I forward an e-mail from our client to our boss.
	I made proposal materials in Power Point.	Sure, I import the Excel graphs and photos into Power Point.

Figure 4: English sentences provided to students in the review quizzes.

Next, we administered the respective review quizzes one at a time. Students were asked to translate the Japanese sentences into English using the system. Each review

quiz has three or four versions, and the number of participants are shown in Figure 5. In addition to 47 sentences, six sentences that were not shown to the students in advance were included in the review quizzes, these unannounced sentences are highlighted in yellow in Figure 5.

Quiz version	Version1	Number of participants	Version2	Number of participants	Version3	Number of participants	Version4	Number of participants
Review quizzes	Yeah, the laundry has piled up.		I have several loads of clothes to wash.		Yeah, the laundry has piled up.			
	Let's hang the laundry outside.		Yes, and please fold the laundry.		Yes, and please fold the laundry.			
	OK. Where is the laundry detergent?		It's warm and comfortable.		OK. Where is the laundry detergent?			
	My clothes smell damp.	22	I put the folded clothes away in the drawers.	25	I hear the washing machine spin-drying.	25	-	-
	There's a 60 percent chance of rain.		Lately the weather forecast has proven right.		Lately the weather forecast has proven right.			
	I have to vacuum the floor.		I veg out at home.		I have to vacuum the floor.			
	Could you cut it short all around?		Just trim it overall, please.		Just trim it overall, please.			
	Oh, It has just slipped my mind.		It's on the tip of my tongue.		Oh, It has just slipped my mind.			
2	I feel dizzy.		I have the chills.		I feel like throwing up.			
	I have diarrhea.		I have diarrhea.		It's just constipation.			
	I sprained my ankle.		I've got hives on my back.		I've got hives on my back.			
	You might have pneumonia, so I'll take an X-ray for you.	22	You have a seasonal nasal allergy.	25	You might have pneumonia, so I'll take an X-ray for you.	25	-	-
	I'll prescribe a painkiller for you.		I'll prescribe antibiotics for you.		I'll prescribe a painkiller for you.			
	I can't decide what to wear.		I don't have a bag that goes with the clothes I'm wearing today.		Doesn't this skirt look childish?			
	It's tight around my waist.		I think I need the next size up.		It's tight around my waist.			
3	It's down the hall.		It's in the back of the store.		It's down the hall.			
	No, I'm not up to it. Go ahead without me.		You're a strong drinker. It doesn't show on your face at all.		You're a strong drinker. It doesn't show on your face at all.		Shall we order sea urchin and fatty tuna sushi?	
	We should get out of the way of the car.		Yes, the car is splashing water from a puddle.		I didn't notice the bicycle passing me from behind.		We should get out of the way of the car.	
	I can tell we won't make it across the crosswalk from here.		You lose your sense of direction walking around in underground malls.		I can tell we won't make it across the crosswalk from here.		I didn't notice the bicycle passing me from behind.	
	Walking there, I found that it's not that far.	16	Is there anything that would make a good landmark?	16	You lose your sense of direction walking around in underground malls.	20	It's just before the third intersection.	20
	I'm sorry to have made you call so many times.		I'm sorry to have made you call so many times.		My unread e-mail is piling up.		I'm sorry to have made you call so many times.	
	My unread e-mail is piling up.		Unfamiliar attachment files carry viruses.		Unfamiliar attachment files carry viruses.		It's all garbled.	
3	I forward an e-mail from our client to our boss.		I made proposal materials in Power Point.		I made proposal materials in Power Point.		Sure, I import the Excel graphs and photos into Power Point.	

Figure 5: Quiz patterns of respective review quizzes.

A doctoral student evaluated the students' answers by checking the items mentioned in the previous section. We calculated the percentage of correct answers using the results of the first evaluation item and whether the answer was correct as a whole sentence. Only the results of the evaluation option "correct" were used in the calculation. The options, "some mistakes" and "incorrect" were excluded.

During the review quizzes, a CG character gave students instructions, such as "Please translate the following Japanese sentence into English." The first Japanese sentence was then displayed on the screen, as shown in Figure 3. At the same time, a synthetic voice was generated which read the Japanese sentence to the student out loud.

After the quizzes, a self-reflective feedback questionnaire was administered to gather responses regarding perceived difficulty ratings for the sentences. A five-point Likert scale (one = least difficult to five = most difficult) was used.

Results and Discussion

We calculated the percentage of correct answers and the average values of the difficulty ratings for the respective sentences to address our research questions.

RQ1: What difficulty level did students perceive for the sentences they answered correctly? What difficulty level did students perceive for the sentences they answered incorrectly?

First, we analyzed how students rated the difficulty levels, using the five-point Likert scale, of the sentences that most answered correctly. Figure 6 shows 31 sentences that were answered correctly by more than 80% of the students. The average subjective difficulty ratings for these sentences are also shown. Additionally, we calculated the standard deviation scores of the difficulty ratings for each sentence to provide a rough indication.

As illustrated by the figure, the results indicate that average subjective difficulty ratings for sentences answered correctly by 80% or more of students varied from 1.7 to 4.2. When we look at the detailed results, we find that some words such as “prescribe antibiotics,” “constipation,” “hives,” and “garbled” were included in the sentences with average subjective difficulty ratings higher than 4.0. We assume that these words were unfamiliar and difficult for most students to use. The results suggest that students do not regard all sentences with a high percentage of correct answers as easy, even when they gave correct answers themselves. By using subjective difficulty ratings, we were able to better observe how the students felt about the difficulty levels of these sentences.

Next, we analyzed how students rate the difficulty of the sentences which were translated correctly by less than 50% (see Figure 7) of the participants. The average subjective difficulty ratings for those sentences and their respective standard deviation scores are also shown.

As this figure shows, the average difficulty ratings for eight of the ten sentences were greater than three. This indicates that most students found it difficult to work with the sentences that less than 50% of the participants translated correctly. With most sentences, the percentage of correct answers coincides with the average values of the subjective difficulty ratings.

This also holds true for the sentences that 10% of the students answered correctly. The average difficulty rating of those three sentences was 4.6. Most students found it very difficult to use these sentences easily. We found that these sentences included words such as “damp,” “spin-drying,” and “put away,” that were probably unfamiliar or unknown to the students. This is likely true for words such as “prescribe” “antibiotics” “constipation,” “hives,” and “garbled” that were found in other sentences rated as difficult. These results suggest that students need a way to become more familiar with terms and phrases related to daily life.

English sentences	Percentage of correct answers	Average values of the difficulty ratings	Standard deviation scores of the difficulty ratings
It's tight around my waist.	95.7	2.6	1.0
Let's hang the laundry outside.	95.5	2.8	1.2
I forward an e-mail from our client to our boss.	93.8	3.4	1.1
No, I'm not up to it. Go ahead without me.	93.8	3.2	1.2
I didn't notice the bicycle passing me from behind.	92.5	3.0	1.1
It's warm and comfortable.	92.0	1.7	0.9
Could you cut it short all around?	90.9	3.5	1.1
Sure. I import the Excel graphs and photos into Power Point.	90.0	3.6	0.9
It's just before the third intersection.	90.0	2.7	1.1
I have to vacuum the floor.	89.4	2.5	1.2
I'll prescribe a painkiller for you.	89.4	3.8	1.1
We should get out of the way of the car.	88.9	3.2	1.0
I'm sorry to have made you call so many times.	88.5	2.8	1.1
I veg out at home.	88.0	3.6	1.4
I'll prescribe antibiotics for you.	88	4.1	1.1
OK. Where is the laundry detergent?	87.2	3.3	1.3
There's a 60 percent chance of rain.	86.4	3.3	1.1
I made proposal materials in Power Point.	86.1	3.4	1.1
It's just constipation.	84.0	4.2	1.0
I think I need the next size up.	84.0	3.2	1.0
Yes, and please fold the laundry.	84.0	2.7	1.2
I have the chills.	84.0	2.9	1.3
It's down the hall.	83.0	3.4	1.1
I've got hives on my back.	82.0	4.2	0.9
Yes, the car is splashing water from a puddle.	81.3	3.8	0.9
Is there anything that would make a good landmark?	81.3	3.3	1.1
It's in the back of the store.	80.0	3.1	1.1
Just trim it overall, please.	80.0	3.9	1.2
I feel like throwing up.	80.0	3.2	1.2
It's all garbled.	80.0	4.0	1.2
It's on the tip of my tongue.	80.0	3.9	1.2

Figure 6: Sentences that more than 80% of students answered correctly.

English sentences	Percentage of correct answers	Average values of the difficulty ratings	Standard deviation scores of the difficulty ratings
I can't decide what to wear.	45.5	2.6	1.1
I don't have a bag that goes with the clothes I'm wearing today.	44.0	3.6	1.0
You have a seasonal nasal allergy.	36.0	4.1	0.9
Unfamiliar attachment files carry viruses.	33.3	3.6	1.0
My unread e-mail is piling up.	33.3	2.8	1.2
Doesn't this skirt look childish?	28.0	3.9	0.9
Shall we order sea urchin and fatty tuna sushi?	20.0	3.6	1.1
My clothes smell damp.	9.1	4.6	0.8
I put the folded clothes away in the drawers.	4.0	4.5	0.8
I hear the washing machine spin-drying.	0.0	4.7	0.7

Figure 7: Sentences that less than 50% of students answered correctly.

It should be noted that all six of the unannounced sentences highlighted in yellow were among the sentences that less than 50% of students answered correctly. It seems that most students find it difficult to work with sentences they had not seen before, and they need practice that will help them smoothly generate appropriate English terms and phrases as they encounter them.

RQ2: Are the difficulty ratings of the sentences different depending on the individual student? What is the response variance for each sentence?

We analyzed the standard deviation scores of the difficulty ratings of the sentences discussed regarding RQ1 to observe the response variance. For the sentences answered correctly by more than 80% of the students, the standard deviation scores of the difficulty ratings varied from 0.9 to 1.4. The difficulty ratings of some sentences were different depending on the individual student.

Among the sentences that less than 50% of students translated correctly, notable response variance was only observed in the sentence with a standard deviation score of 1.2. Most students felt that these sentences were difficult to use. This is especially apparent when we look at the sentences that 10% of students answered correctly; the average values of difficulty ratings for these three sentences were more than 4.5. Their standard deviation scores for these sentences were lower than 1.0. The difficulty ratings of these three sentences were similar for almost all the students.

These results suggest that the standard deviation scores of the difficulty ratings allow us to observe students' response variance for individual sentences, and it can help us understand the difficulty levels of each sentence on an individual student level.

RQ3: Could a subjective difficulty rating play a role in providing information about sentences that learners can use to practice English?

For the third research question, students used a five-point Likert scale to subjectively rate the difficulty levels of the English sentences. The subjective difficulty ratings revealed several important points, as described above in relation to RQ1 and RQ2. First, even though students correctly answered certain sentences, they found those sentences difficult to work with. The subjective difficulty rating allowed us to observe what an analysis of correct and incorrect answers alone could not by showing how the students felt about the difficulty levels. Second, the standard deviation scores of the difficulty ratings allowed us to observe the response variance among students for each sentence. We believe that the subjective difficulty rating can assist educators in understanding students' performances and identifying their individual weaknesses to provide tailored support alongside the objective review quizzes. Incorporating this kind of subjective difficulty rating as part of the system, can offer useful guidance for selecting sentences according to individual student needs.

However, a limitation of the current study is that the sentences were rated only after the review quizzes were taken. The difficulty of using some sentences could change between the period before students practiced the sentences and took the quizzes, and after they practiced them, while the difficulty levels of other sentences will likely never change regardless of how much the student practices. By conducting pre- and post-practice questionnaires, we might observe changes in the difficulty ratings of

some sentences. We believe that we can best observe the detailed differences in the difficulty levels of each sentence by analyzing the results of both questionnaires. In future research, we hope to investigate both pre- and post-practice subjective difficulty ratings.

Conclusion

We have been developing a system that helps learners practice English speaking by preparing the sentences and the evaluation items to check the words and sentences they can use. We conducted our study with 72 Japanese university students who used the system. They then answered a questionnaire with a five-point Likert scale that helped us to investigate the role of a subjective difficulty rating in identifying sentences in the system that were problematic for the learners.

The results indicate that average difficulty ratings for the 31 sentences that 80% or more of students answered correctly varied from 1.7 to 4.2. Even though most students gave correct answers, they did not rate all these sentences as easy. We also found that the standard deviation scores of the difficulty ratings for these sentences varied from 0.9 to 1.4. This indicates that some students rated some sentences differently than the average. The detailed results suggest that students need to become familiar with terms and phrases related to daily life. Problematic words included “spin-drying,” “constipation,” and “garbled.” From the results of the experiment, we suggest that a subjective difficulty rating could play a role in observing students’ feelings regarding the difficulty levels of the sentences, and in learning how to support individual students’ weaknesses. If we incorporate this kind of subjective difficulty rating into the system, it could offer useful information for selecting the sentences within the system that suit the needs of individual students.

As a continuation of this study, we hope to observe the detailed differences of the difficulty levels of respective sentences by conducting a questionnaire both before and after they practice the sentences. We believe that we could observe the process of students’ practicing by analyzing the results of the two questionnaires.

Acknowledgement

This work was supported by JSPS Grant-in-Aid for Scientific Research Number JP18K02822.

References

- Hayashi, M. (1999). TVML (TV program Making Language) - Automatic TV program generation from text-based script -. *Proceedings of Imagina '99*, 119–133. doi:10.11485/itetr.23.4.0_1
- Isoda, T. (2009). Reducing EFL learners' unwillingness to speak English. *JACET Journal*, 48, 53–66. (In Japanese)
- Japan Ministry of Education, Culture, Sports, Science and Technology. (2014). *English Education Reform Plan corresponding to Globalization*. Retrieved from http://www.mext.go.jp/en/news/topics/detail/___icsFiles/afieldfile/2014/01/23/1343591_1.pdf
- Kashiwagi, H., Kang, M., & Ohtsuki, K. (2018). A basic study on the conformity of Japanese university students in language communication activities. *Official Conference Proceedings of The Asian Conference on Education & International Development 2018*, Kobe, Japan, 299–309.
- Kashiwagi, H., Kang, M., & Ohtsuki, K. (2020). A study on a method for dynamic control and use of TVML contents. *Educational technology research*, 43(Suppl.), 9–12. (In Japanese)
- TVML Home Page. Retrieved from <http://www.nhk.or.jp/str/tvml/index.html>. (In Japanese)

Contact email: kasiwagi@kobe-u.ac.jp