Comparison of unknown word estimation performance between Japanese and French people when reading English comics aloud

by

Hayato Seki (1211201068)

A Bachelor's Thesis
Submitted to
Osaka Prefecture University
in Partial Fulfillment of the Requirements
for the Degree of
Bachelor of Engineering

Academic Year February 2025

Advisor: Motoi Iwata Associate Professor

Department of Computer Science
School of Electrical and Electronic Engineering
College of Engineering
Osaka Prefecture University

Comparison of unknown word estimation performance between Japanese and French people when reading English comics aloud

Group 3 Hayato Seki

1. Introduction

Globalization makes English learning crucial in information gathering and communication increasingly. Reading aloud is one of the English learning methods and enable us to improve reading, listening, speaking and writing skills. However, if learners find words they do not know the meanings of (hereinafter called "unknown words") when reading aloud, they are concerned about unknown words and read aloud more inefficiently. Therefore, we propose a system which automatically estimates unknown words when learners read aloud. Then, learners can read aloud without their concern for unknown words and can also review unknown words later. The proposed system uses comics, and learners can enjoy because of its entertainment and remember unknown words more easily with its story and images.

This paper proposes a method estimating unknown words by learners' speech and eye gaze when learners read comics aloud. In this situation, eye gaze cannot be divided into words, but can be divided into speech balloons. Therefore, In our research, we estimate speech balloons including unknown words using Takaike et al.'s method [1] and extract unknown words from the estimated speech balloons. In order to consider accuracy improvement methods, we used Takaike et al.'s method for French people which have different cultural and linguistic background from Japanese people and analyzed the difference between Japanese and French people. It should be noted that this research has been approved by the Osaka Metropolitan University Graduate School of Engineering Ethics Committee.

2. Experiment

We conducted an data measurement experiment recording speech and eye gaze information for French people, and estimated speech balloons by using Takaike et al.'s method [1] for the obtained data in the experiment. Then, We compared the estimation using French data with that using Japanese data.

2.1. Data Measurement

We asked 13 French students to read English comics aloud using a PC and recorded their speech and eye gaze information. The 13 participants read pages specified in 3 English comics aloud and recorded unknown words every time they finished reading one English comics aloud. The measurement time is around 60 minutes including experiment explanation.

2.2. Evaluation

We performed user-independent leave-one-user-out cross validation and evaluated based on AUPR, which is the AUC of the PR curve. In this paper, the estimation using only text information was used as the baseline.

2.3. Result

Figure 1 shows AUPR averages of the estimations for Japanese data, for French data and for both data, which are expressed as JP, FR and JP+FR, respectively. As shown in Fig. 1, the estimation using speech, eye gaze, and text information exceeded the baseline in all estimation. Also, since the AUPR average of FR is lower than that of JP and JP+FR, it is possible that Takaike et al.'s method is more effective for Japanese people than for French people.

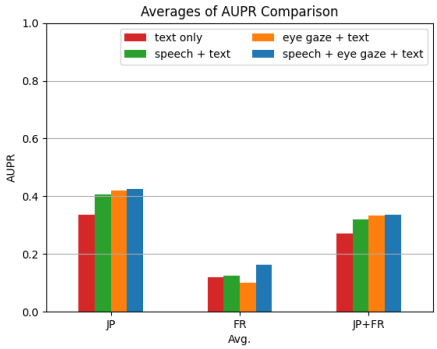

Figure 1: AUPR averages of estimation using Japanese and French data

In addition, We checked the selected percentage of 9 speech features regarding reading time and confidence for IBM Watson Speech to Text, and 12 eye gaze features, in each estimation for Japanese and French data. In this paper, the selected percentage of feature is calculated by dividing the number of times the feature is selected by the total number of Japanese or French participants. In terms of speech features regarding reading time and confidence for IBM Watson Speech to Text, the selected percentage in FR is higher than that in JP for 7 out of 9 features. Confidence for IBM Watson Speech to Text is the accuracy of pronunciation. Therefore, it is possible that speech features regarding reading time and the accuracy of pronunciation are effective for French people. In terms of eye gaze features, the rate of JP is higher than the rate of FR for 10 out of 12 features, so it is possible that eye gaze features are effective for Japanese people.

3. Conclusion

This paper compared unknown word estimation performance for French participants with that for Japanese participants, and confirmed that the estimation performed better for Japanese ones than for French ones. In addition, This paper confirmed that speech features regarding reading time and the accuracy of pronunciation are effective for French participants and that eye gaze features are effective for Japanese participants.

References

[1] T. Takaike, M. Iwata, and K. Kise. Estimation of unknown words using speech and eye gaze when reading aloud comics. In *Pattern Recognition, Computer Vision, and Image Processing. ICPR 2022 International Workshops and Challenges*, pp. 91–106, 2023.

Contents

List of Figures

List of Tables

Chapter 1
Introduction

The importance of English has increased with recent globalization. In order to obtain the latest information from around the world, it is necessary to read texts written in English, so English reading comprehension is becoming particularly essential. In Japanese English education, many lessons are taught using a grammar-translation-reading style, in which English sentences are translated into Japanese and understood [1]. As a result of this, Japanese students have a habit of translating English into Japanese and understanding it. It has become a problem that some English learners are able to translate English texts, but when they read the text at a certain speed, they become incomprehensible because of the process of translating it into Japanese [2].

Reading aloud is an effective way to eliminate this process of translating into Japanese. Unlike silent reading, Reading aloud has the constraints of having to understand the content while vocalizing it, making students realize that they do not have time to translate it into Japanese. Furthermore, through vocalization, connections are formed between words and their sounds. Therefore, it is confirmed that learners who read aloud change the way they read English sentences to one that allows them to understand the English sentences without using Japanese language [2]. As a result, it has been shown that reading aloud affects not only reading comprehension but also listening and speaking ability [3].

One of the problems in learning to read aloud is that while reading aloud, students sometimes encounter words that they do not know the meanings of (hereinafter called "unknown words"). When encountering unknown words, the efficiency of reading aloud may be reduced by having to look up the meaning of the unknown words or being unable to concentrate on reading aloud due to worrying about the meaning. For this reason, we propose a system in which a computer automatically records unknown words. With such a system, learners can read aloud without worrying about them due to the sense of security that the unknown words are recorded and acquire new

vocabulary by reviewing the unknown words.

Although various materials such as articles and novels can be used as reading-aloud materials, we focus on English comics in this study. Comics are subjects that attracts the interest of many people, and it is expected that learners will enjoy being exposed to English. Furthermore, "edutainment" is a learning method that combines entertainment, and edutainment is expected to enhance the continuity of learning. In addition, illustrations make it easier to understand the situation of the story and to comprehend the English sentences [4]. Furthermore, illustrations of comics stories, and scenes are effective because they help reinforce the retention of unknown words during review [5]. Therefore, the proposed system uses English comics will be used as read-aloud materials.

The ultimate goal of this study is to automatically estimate unknown words from the learner's behavior while reading English comics aloud. Specifically, this study estimates speech balloons including unknown words in English comics with Takaike et al.'s method [6] which is introduced in Chapter 2, and then extracts unknown words from the estimated speech balloons.

Takaike et al.'s method has problems with its accuracy and has been used only for Japanese native speakers. In order to consider accuracy improvement methods, this paper performed user-independent estimation with Takaike et al.'s method for French native speakers who have different cultural and linguistic backgrounds from Japanese native speakers, and then, investigated the difference in the estimation performance between Japanese speakers and French native speakers, and discuss the difference between Japanese native speakers and French native speakers based on the estimation results.

In the following chapters, we describe the related work in Chapter 2, the experiment in Chapter 3, the discussion in Chapter 4, and the conclusion and future work in Chapter 5.

Chapter 2
Related Work

2.1 Effectiveness of speech and eye gaze on unknown word estimation

The estimation of unknown words is closely related to learners' internal states, such as understanding and confidence. Sabu et al. estimated children's confidence from their speech while reading stories aloud [7]. In this study, pause, speech rate, pitch, intensity, voice quality, and enunciation were calculated as speech features. Based on the obtained speech features, the confidence was estimated using a random forest. Hasegawa-Johnson et al. estimated whether children were confident, confused, or hesitant based on their speech while performing a Lego task and talking to their tutors [8]. In this study, various speech analyses such as lexicon, prosody, spectrum, and syntax were performed. The results of these analyses were used to estimate the emotion by a robust classification and regression tree (CART). Thus, speech has been used to estimate confidence, and speech is expected to be effective in estimating unknown words.

Eye gaze has been also used to estimate people's internal states: Martínez-Gómez et al. estimated learners' language proficiency from their eye gaze while reading educational texts [9]. Ishimaru et al. classified learners' language proficiency based on their eye gaze while answering questions about a physics textbook that they had read once [10]. Garain et al. estimated reader-specific difficult words in a document and showed that eye gaze is related to comprehension [11].

2.2 Estimation of internal states when reading English comics

There are some studies that estimate the comprehension level of English learners when reading English comics. Daiku et al. estimated the comprehension level during reading English comics in the form of speech balloons based on eye gaze, micro-expression features, and the number of words in the speech balloon [12]. Eye gaze was measured with an eye tracker device, and facial expression features were extracted from facial images recorded with a high-speed camera. Takahashi et al. used eye gaze to estimate learners' comprehension of each page in English comics [13]. In these two studies, fixation and saccade were used as eye gaze, and it was shown that eye gaze is also effective in estimating comprehension while reading English comics. Fixation refers to the point where a gaze stops for a certain period of time, and saccade refers to the quick movement of gazes between fixations.

2.3 Existing unknown word estimation based on behavior when reading English comics aloud

Takaike et al.'s method [6] has been used to estimate speech balloons including unknown words when learners read English comics aloud. Takaike et al. asked 20 Japanese participants to read English comics aloud, and estimated speech balloons including unknown words based on speech, eye gaze, and text information. The speech was recorded using a headset, and the eye gaze was measured using an eye tracker device. MFCC and confidence of IBM Watson Speech to Text[1] were used as features for the speech information, and fixation and saccade were used as features for the eye gaze information. As features for the text information, word frequency and the number of words in the speech balloons were used. Takaike et al. evaluated user-dependent estimation by performing leave-one-episode-out cross-validation, using one episode of data as test data and the rest as training data. The results showed that the average AUPRs of all 20 participants in the experiment when the estimation used the speech, eye gaze, and text information were higher than the average AUPRs when the estimation used the text information alone, indicating that the estimation using the speech, eye gaze, and text information performed better than the estimation using the text information alone. However, the average AUPR of the estimation using speech, eye gaze, and text information is 0.325, and the estimation has problems with its

[1] `https://www.ibm.com/products/speech-to-text`(accessed on January 22, 2024)

accuracy.

Chapter 3

Experiment

We conducted a data measurement experiment recording speech and eye gaze information for French native speakers, and estimated speech balloons by using Takaike et al.'s method [6] for the obtained data in the experiment. Then, We compared the estimation performance using French data with that using Japanese data. Hereinafter, the dataset obtained from the data measurement experiment conducted in Japan from December 2021 to January 2022 by Takaike et al. will be referred to as "Japanese Dataset A". The dataset obtained from the data measurement experiment conducted in Japan from December 2022 to March 2023 by Takaike et al. will be referred to as "Japanese Dataset B". The combined dataset of Japanese Dataset A and B will be referred to as "Japanese Dataset". Also, the dataset obtained from the data measurement experiment conducted in France in June 2024 by us will be referred to as "French Dataset", and then, The combined dataset of Japanese Dataset and French Dataset will be referred to as "Mixed Dataset".

3.1 Data measurement

We asked 13 French native speakers to read three English comics aloud using a comics display application on the PCs, and their behavior during the reading aloud was recorded using a headset and eye tracker. The comics display application displayed a double-page spread of the English comics as a single image. The headset and eye tracker are AKG C544L and Tobii Pro 4C, respectively. Each comics title includes 26 pages.

Before the measurement, the participants were told how to read comics because they were not accustomed to reading comics, and then fixed their sitting position and head position. After that, they read three English comics aloud and recorded unknown words every time they finished one English comics. The participants of this experiment

Table 3.1: Label information of speech balloons in Japanese Dataset A

Participant No.	Number of balloons			Percentage of Positive
	Positive	Negative	Total	
A01	43	235	278	15.5%
A02	44	277	321	13.7%
A03	40	261	301	13.3%
A04	10	80	90	11.1%
A05	88	291	379	23.2%
A06	57	298	355	16.1%
A07	47	249	296	15.9%
A08	33	197	230	14.3%
A09	62	130	192	32.3%
A10	62	181	243	25.5%
A11	22	599	621	3.5%
A12	50	259	309	16.2%
A13	56	244	300	18.7%
A14	51	270	321	15.9%
A15	95	217	312	30.4%
A16	145	179	324	44.8%
A17	84	232	316	26.6%
A18	138	172	310	44.5%
A19	36	525	561	6.4%
A20	67	569	636	10.5%
Average	61.5	273.3	334.8	

consist of 13 French native speakers(11 males and 2 females, PhD students). The total duration of the experiment, including guidance and post-measurement questionnaires, was an hour, and each participant was paid 15 euros as an honorarium. The gender and age of the experimental participants are shown in Appendix A.

3.2 Details of used data

Tables 3.1 and 3.2 show the label information of speech balloons in Japanese Dataset A and B, respectively. Table 3.3 shows the label information of speech balloons in French Dataset. Note that we counted only speech balloons with speech and eye gaze as the number of speech balloons. In Tabs. 3.1, 3.2 and 3.3, the number of speech balloons including unknown words is denoted as “Positive”, the number of speech balloons not including unknown words as “Negative”, the total number of speech balloons as “Total”, and the percentage of “Positive” in “Total” as “Percentage of Positive”. In Japanese Dataset A, the average number of speech balloons was 334.8

Table 3.2: Label information of speech balloons in Japanese Dataset B

Participant No.	Number of balloons			Percentage of Positive
	Positive	Negative	Total	
B01	40	243	283	14.1%
B02	33	277	310	10.6%
B03	67	247	314	21.3%
B04	81	366	447	18.1%
B05	22	235	257	8.6%
B06	64	348	412	15.5%
B07	53	222	275	19.3%
B08	35	323	358	9.8%
B09	87	228	315	27.6%
B10	48	316	364	13.2%
B11	36	223	259	13.9%
B12	49	196	245	20.0%
B13	24	99	123	19.5%
B14	43	485	528	8.1%
B15	25	230	255	9.8%
B16	37	414	451	8.2%
B17	62	290	352	17.6%
B18	30	305	335	9.0%
B19	35	282	317	11.0%
B20	51	446	497	10.3%
Average	46.1	288.6	334.9	

Table 3.3: Label information of speech balloons in French dataset

Participant No.	Number of balloons			Percentage of Positive
	Positive	Negative	Total	
P01	0	94	94	0.0%
P02	5	103	108	4.6%
P03	1	111	112	0.9%
P04	5	82	87	5.8%
P05	3	75	78	3.9%
P06	4	107	111	3.6%
P07	0	91	91	0.0%
P08	0	74	74	0.0%
P09	1	83	84	1.2%
P10	0	99	99	0.0%
P11	1	89	90	1.1%
P12	0	73	73	0.0%
P13	1	105	106	0.9%
Average	1.6	91.2	92.8	

with the range [90, 636], and the percentages of Positive speech balloons ranged from [3.5%, 44.8%]. In Japanese Dataset B, the average number of speech balloons was 334.9 with the range [123, 528], and the percentages of Positive speech balloons ranged from [8.1%, 27.6%]. In French Dataset, the average number of speech balloons was 92.8 with the range [73, 112], and the percentages of Positive speech balloons ranged from [0.0%, 5.8%]. The number of Positive speech balloons and the number of Negative speech balloons were unevenly distributed among all participants, and the number of Positive speech balloons was generally smaller than that of Negative ones. In French Dataset, The percentage of Positive speech balloons was generally lower than in Japanese Dataset.

3.3 Evaluation manner

User-independent estimations were performed on Japanese Dataset, French Dataset, and Mixed Dataset. User-independent estimation was evaluated by performing leave-one-user-out cross-validation, where user-independent estimation means that the estimator is trained using the data of participants other than the participant being estimated. Because the data were highly biased toward labels, we oversampled the training data using SMOTE [14]. Feature selection by SBFS using the training data was performed for each cross-validation. The AUPR, the AUC of the Precision-Recall curve, was used as the evaluation metric.

We will refer to the user-independent estimation using Japanese Dataset as “JP”, the user-independent estimation using French Dataset as “FR”, and the user-independent estimation using Mixed Dataset as “JP+FR”.

Table 3.4 shows features in Takaike et al.’s method. Estimation using only text information is denoted as “text only”, and we assume the estimation is the baseline. Estimation using speech and text information is denoted as “speech + text”, estimation using eye gaze and text information is denoted as “eye gaze + text”, and estimation using speech, eye gaze, and text information is denoted as “speech + eye gaze + text”.

3.4 Result

Figure 3.1 shows the average AUPRs for all participants for user-independent estimation with all options for Japanese Dataset, French Dataset, and Mixed Dataset, respectively. The average AUPRs of FR are lower than those of JP and JP+FR. This shows the unknown word estimation with Takaike et al.’s method is not effective for

Table 3.4: Features in Takaike et al.'s method

Category	Type of features	Features
Speech	Frequency feature	Avg. of MFCC (20 dimensions, time-frequency)
	Features using IBM Watson Speech to Text API	Confidence of transcription per speech balloon
		Max., min., avg. of confidence of transcription per word
		Max., min., avg. of confidence of transcription assuming the transcription is same as the text
	Reading aloud time	Reading aloud time per speech balloon
		Reading aloud time per word
Eye gaze	Fixation	Number of fixations
		Max., min., avg. of duration of fixations
	Saccade	Max., min., avg. of the length of saccades
		Max., min., avg. of the speed of saccades
	Gazing time	Gazing time per speech balloon
		Gazing time per word
Speech and Eye gaze	Time difference between reading aloud and gazing	Time difference in start time
		Time difference in end time
Text	Number of words	Number of words in speech balloon
	Word frequency	Max., min., avg. of word frequency

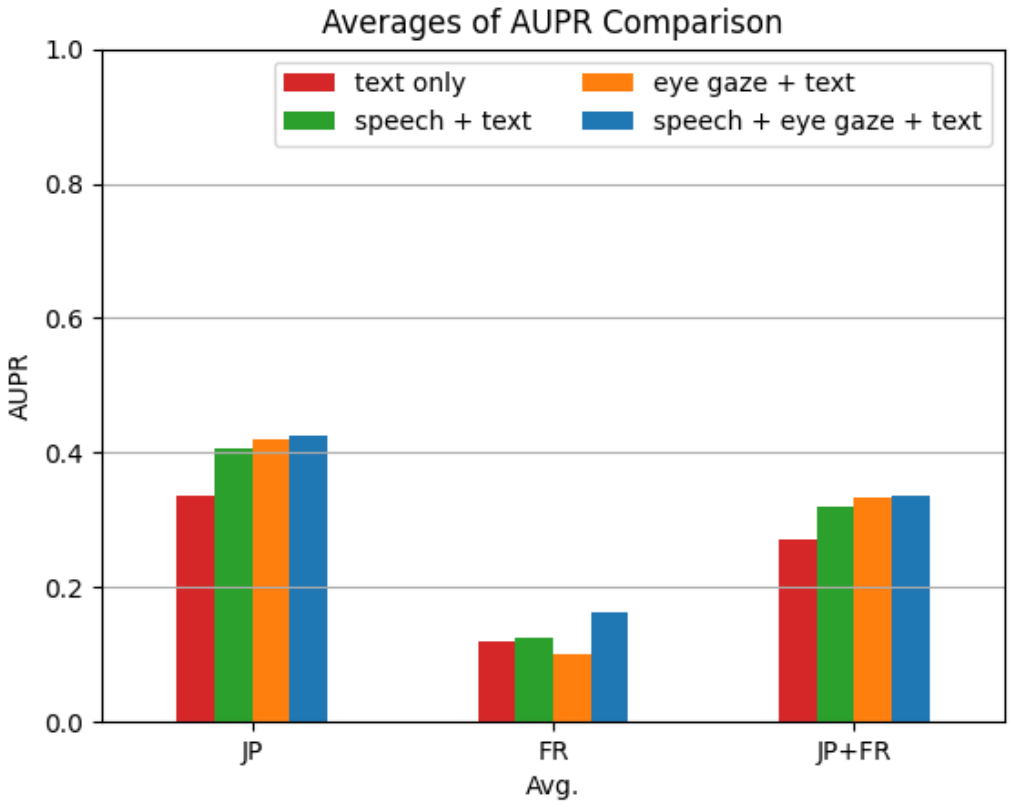

Figure 3.1: Averages of AUPR Comparision

French participants. About the average AUPR of FR being less than 0.2, there are few

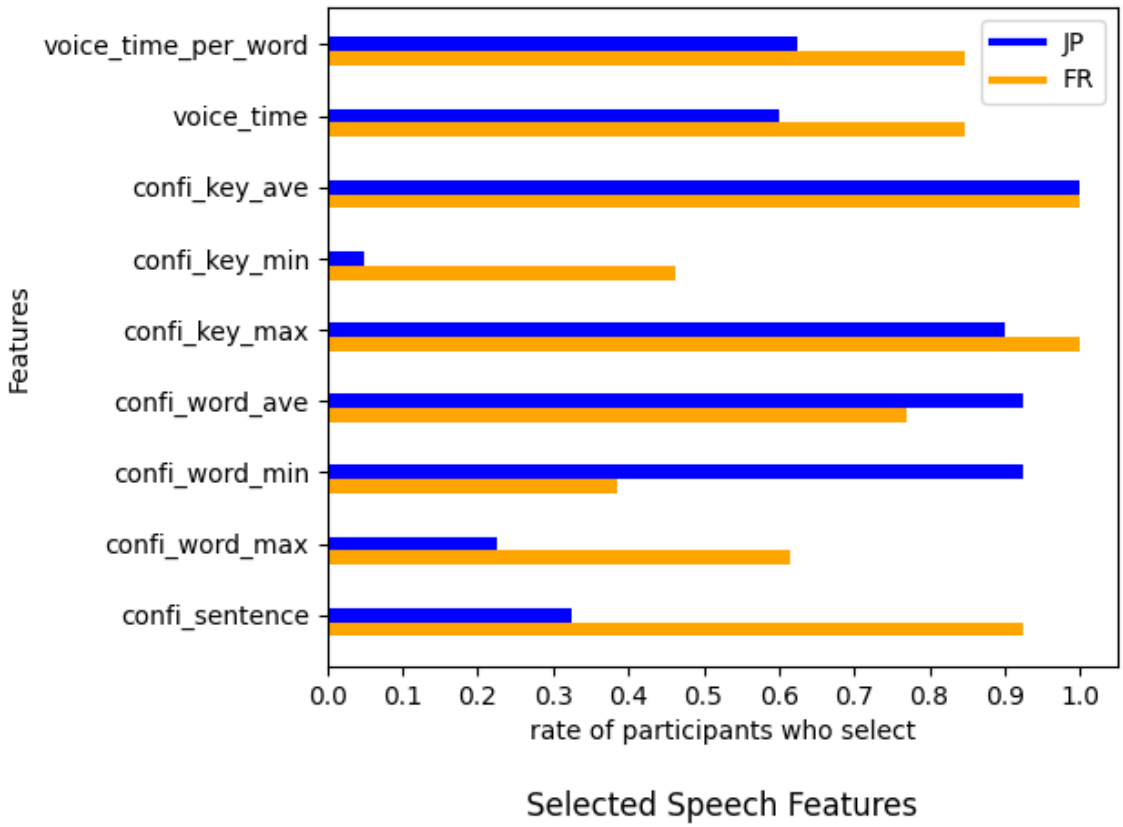

Figure 3.2: Selected percentages of speech features regarding reading time and confidence for IBM Watson Speech to Text

unknown words in the French data, and it is possible that the model is insufficiently trained. Moreover, in FR, the AUPR average of the estimation using eye gaze and text information is lower than that using only text information, and it is possible that eye gaze features are not effective for French participants.

Figures 3.2 and 3.3 show the selected percentages of speech features regarding reading time and confidence for IBM Watson Speech to Text and eye gaze features, respectively, where the selected percentage of features is calculated by dividing the number of times the feature is selected by the total number of Japanese or French participants. In Fig 3.2, the selected percentage in FR is higher than that in JP for 7 out of 9 features. Confidence for IBM Watson Speech to Text is the accuracy of pronunciation, which means how close it is to native pronunciation. Therefore, This shows that speech features regarding reading time and the accuracy of pronunciation are effective for French participants. In Fig 3.3, the selected percentage in JP is higher than the that in FR for 10 out of 12 features, and this shows that eye gaze features are effective

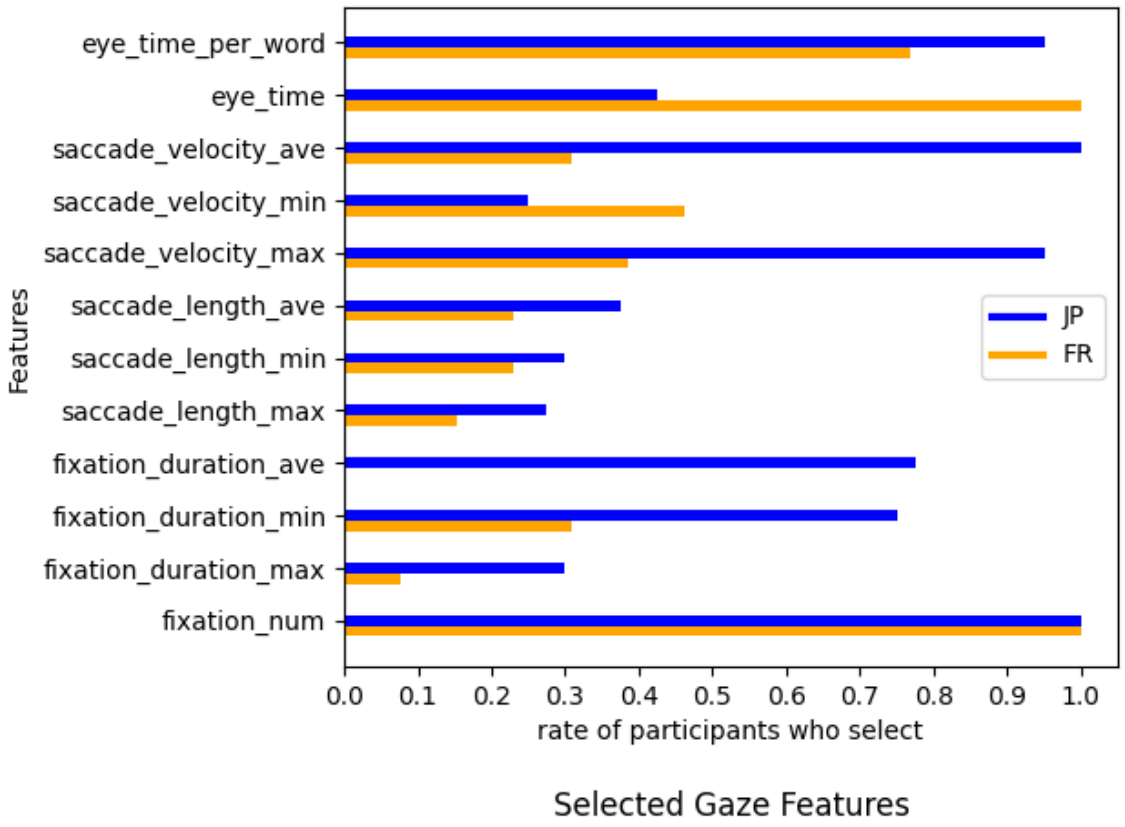

Figure 3.3: Selected percentages of eye gaze features

for Japanese participants.

Chapter 4
Discussion

This chapter discusses the difference between Japanese and French participants in frequently selected features in Figs 3.2 and 3.3. Specifically, we examined the difference between these features of Positive and Negative speech balloons in JP or FR. In order to see frequently selected features for the estimation, we focus on features selected more than 60 % in both JP and FR. Tables 4.1 and 4.2 show the averages, standard deviations, standardized mean difference of the speech features and eye gaze features, respectively, of Positive or Negative balloons in JP or FR.

In Tab. 4.1, almost all features regarding confidence of Positive balloons are lower than those of Negative balloons in JP and FR except for the minimum of confidence of transcription per word in FR. This shows that both Japanese and French participants pronounce speech balloons including unknown words in a more ambiguous or incorrect way that would be difficult for the transcription model to recognize correctly. Also, features regarding reading time of Positive balloons are higher than those of Negative balloons in JP. In contrast, those of Positive balloons are lower than those of Negative balloons in FR. This shows that Japanese participants read more slowly and French participants read faster when finding unknown words.

In Tab. 4.2, the number of fixations and gazing time of Positive balloons are higher than that of Negative balloons in JP. In contrast, those of Positive balloons are slightly lower than those of Negative balloons in FR. Also, gazing time per word of Positive balloons is higher than that of Negative balloons in JP. In contrast, that of Positive balloons is slightly lower than that of Negative balloons in FR. This shows that Japanese participants gaze more at speech balloons including unknown words, and French participants gaze slightly less at speech balloons including unknown words.

In Tab. 4.1 and 4.2, features where the standardized mean difference between Positive and Negative balloons is less than 0.2 are the average of confidence of transcription per word, max of confidence of transcription assuming the transcription is same as the

text, gazing time per word. These features are generally difficult to classify, but frequently selected in the estimation. One possibility is that these features become effective as a result of interaction with other features.

Table 4.1: Averages, standard deviations and standardized mean differences of effective speech features regarding reading time and confidence for IBM Watson Speech to Text

Features	**JP or FR**	**Positive or Negative**	**Avg. ± std.**	**SMD**
Avg. of confidence of transcription per word	JP	Positive	0.74 ± 0.15	0.13
		Negative	0.76 ± 0.16	
	FR	Positive	0.78 ± 0.15	0.14
		Negative	0.80 ± 0.14	
Avg. of confidence of transcription assuming the transcription is same as the text	JP	Positive	0.23 ± 0.23	0.44
		Negative	0.36 ± 0.31	
	FR	Positive	0.035 ± 0.054	0.21
		Negative	0.066 ± 0.15	
Max of confidence of transcription assuming the transcription is same as the text	JP	Positive	0.64 ± 0.42	0.17
		Negative	0.71 ± 0.41	
	FR	Positive	0.27 ± 0.41	0.048
		Negative	0.29 ± 0.42	
Reading aloud time per speech balloon	JP	Positive	6.27 ± 5.48	0.70
		Negative	3.56 ± 3.48	
	FR	Positive	2.56 ± 1.33	0.30
		Negative	3.51 ± 3.16	
Reading aloud time per word	JP	Positive	0.68 ± 0.50	0.28
		Negative	0.52 ± 0.58	
	FR	Positive	0.41 ± 0.23	0.25
		Negative	0.62 ± 0.84	

Table 4.2: Averages, standard deviations and standardized mean differences of effective eye gaze features

Features	**JP or FR**	**Positive or Negative**	**Avg. ± std.**	**SMD**
Number of fixations	JP	Positive	4.16 ± 3.48	0.62
		Negative	2.70 ± 2.07	
	FR	Positive	1.48 ± 2.11	0.33
		Negative	1.87 ± 1.15	
Gazing time per word	JP	Positive	1.27 ± 1.48	0.14
		Negative	1.00 ± 2.06	
	FR	Positive	1.22 ± 0.60	0.0091
		Negative	1.26 ± 4.45	

Chapter 5
Conclusion

In this paper, We performed user-independent estimation using sensor data from 40 Japanese participants and 13 French participants and compared the unknown word estimation performance for French participants with that for Japanese participants. We confirmed that the estimation performed better for Japanese ones than for French ones. In addition, we confirmed that speech features regarding reading time and the accuracy of pronunciation are effective for French participants and that eye gaze features are effective for Japanese participants. Future work is to propose a method for identifying actual unknown words from speech balloons including unknown words.

acknowledgment

I would like to express my deepest gratitude to Associate Professor Motoi Iwata for their direct guidance and advice on the content of my research, the writing of this thesis, presentation methods, and the preparation of materials. I sincerely appreciate Professor Koichi Kise and Associate Professor Andrew Vargo for their guidance on the direction of my research during our learning group meetings. I am thankful to Associate Professor Masakazu Iwamura and Lecturer Yuzuko Utsumi for their advice during the research presentations. I would like to thank French student Sofiya Kobylyanskaya for her help in conducting the data collection experiment in France. I would like to thank all of the participants for their willingness to participate in the experiment. In addition, I would like to express my sincere gratitude to the administrative staff of the Intelligent Media Processing Research Group for their administrative procedures, and to the students of the Intelligent Media Processing Research Group for their support and cooperation.

February 2025

Bibliography

[1] 渡辺英雄 and 長坂勇太郎. 用法に焦点を当てた文法指導は学習者のビリーフと学習方略を変えるのか. **武蔵野教育學論集**, (14):85–94, 2023.

[2] 西田晴美. 音読と黙読学習の実践において学習者の内容理解に見られる変化に生じる相違. **跡見学園女子大学人文学フォーラム**, (16):196–175, 2018.

[3] 直子 橋本 and 義訓 東原. ポートフォリオ評価を取り入れた英語科における音読学習. **信州大学教育学部附属教育実践総合センター紀要 教育実践研究**, 3:151–160, 07 2002. ISSN 1345-8868. URL `https://cir.nii.ac.jp/crid/1050564288866562560`.

[4] Mukhamad Efendi. The use of pictures as media to improve students' reading comprehension. *Journal of English Teaching, Literature, and Applied Linguistics*, 2(2):84–86, 2021. ISSN 2614-5871. doi: 10.30587/jetlal.v2i2.2467. URL `https://journal.umg.ac.id/index.php/jetlal/article/view/2467`.

[5] May Ali Abdul-Ameer. Improving vocabulary learning through digital stories with iraqi young learners of english at the primary level. *Journal of Studies in Social Sciences*, 8(2):197–214, 2014. ISSN 2201-4624.

[6] Taro Takaike, Motoi Iwata, and Koichi Kise. Estimation of unknown words using speech and eye gaze when reading aloud comics. In *Pattern Recognition, Computer Vision, and Image Processing. ICPR 2022 International Workshops and Challenges*, pages 91–106, 2023.

[7] Kamini Sabu and Preeti Rao. Automatic prediction of confidence level from children's oral reading recordings. In *INTERSPEECH*, pages 3141–3145, 2020.

[8] Mark Hasegawa-Johnson, Stephen Levinson, and Tong Zhang. Children's emotion recognition in an intelligent tutoring scenario. pages 1441–1444, 10 2004. doi: 10.21437/Interspeech.2004-552.

[9] Pascual Martínez-Gómez and Akiko Aizawa. Recognition of understanding level and language skill using measurements of reading behavior. In *Proceedings of the 19th international conference on Intelligent User Interfaces*, pages 95–104, 2014.

[10] Shoya Ishimaru, Syed Saqib Bukhari, Carina Heisel, Jochen Kuhn, and Andreas Dengel. Towards an intelligent textbook: eye gaze based attention extraction on materials for learning and instruction in physics. In *Proceedings of the 2016 ACM international joint conference on pervasive and ubiquitous computing: Adjunct*, pages 1041–1045, 2016.

[11] Utpal Garain, Onkar Pandit, Olivier Augereau, Ayano Okoso, and Koichi Kise. Identification of reader specific difficult words by analyzing eye gaze and document content. In *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, volume 1, pages 1346–1351. IEEE, 2017.

[12] Yuki Daiku. Estimation of understanding and interest based on reading behavior in extensive reading with japanese comics translated in english. In *Master Thesis of Osaka Prefecture University*. 2019. 36 pages.

[13] Ryota Takahashi. Estimation of understanding based on eye information in extensive reading with japanese comics translated in english. In *Graduation Thesis of Osaka Prefecture University*. 2021. 21 pages.

[14] Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16:321–357, 2002.

Appendix

A French participants information

Table 5.1 shows the gender and age of French participants. "-"in Age means no answer.

Table 5.1: Information of French participants

Participant ID	Gender	Age
P01	Male	-
P02	Male	-
P03	Male	-
P04	Male	-
P05	Female	22
P06	Male	29
P07	Male	26
P08	Male	22
P09	Male	23
P10	Male	27
P11	Male	24
P12	Male	19
P13	Female	26