Title	A model for identifying and creating appropriate same language subtitles for Japanese language learning
Sub Title	
Author	Shum, Kit Shing Cody(Kishi, Hiroyuki) 岸, 博幸
Publisher	慶應義塾大学大学院メディアデザイン研究科
Publication year	2019
Jtitle	
JaLC DOI	
Abstract	
Notes	修士学位論文. 2019年度メディアデザイン学 第715号
Genre	Thesis or Dissertation
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=KO40001001-00002019-0715

Master's Thesis
Academic Year 2019

A Model for Identifying and Creating
Appropriate Same Language Subtitles for
Japanese Language Learning

Keio University
Graduate School of Media Design

Shum Kit Shing Cody

A Master's Thesis
submitted to Keio University Graduate School of Media Design
in partial fulfillment of the requirements for the degree of
Master of Media Design

Shum Kit Shing Cody

Master's Thesis Advisory Committee:
      Professor Hiro Kishi   (Main Research Supervisor)
      Professor Akira Kato   (Sub Research Supervisor)
      Project Senior Assistant Professor
         MHD Yamen Saraiji   (Sub Research Supervisor)

Master's Thesis Review Committee:
      Professor Hiro Kishi   (Co-Reviewer)
      Professor Akira Kato   (Co-Reviewer)
      Project Senior Assistant Professor
         MHD Yamen Saraiji   (Co-Reviewer)

Abstract of Master's Thesis of Academic Year 2019

A Model for Identifying and Creating Appropriate Same Language Subtitles for Japanese Language Learning

Category: Design

Summary

Media and information are widespread in today's interconnected society. It is easy to find media from sources outside one's own country and native language. This study investigated the feasibility of using these sources as a possible teaching resource, as well as provide possible models to most effectively make use of these resources as a second language learning tool. The model is tested across 14 different learners, and both quantitative and qualitative data were taken in order to assess the feasibility of the model. The model supports a mathematical approach to determine difficulty of subtitles and what to create as learning materials. Many autonomy features such as adaptation to the user language level, the ability to change focus on different aspects of language learning, and constant calibration to the users true language levels are also included and discussed within the study.

Keywords:

Design Thinking, Education, Subtitles, Captions

Keio University Graduate School of Media Design

Shum Kit Shing Cody

Contents

List of Figures

List of Tables

Acknowledgements

Chapter 1

Introduction

1.1. Background

In contemporary society, the number of second language learners are increasing. While there are ongoing efforts to test the efficiency of using subtitles as a learning tool, research on a system or model that can use these subtitles effectively have been minimal. Textbooks have been the staple resource for budding foreign language learners, yet there exists an unlimited amount of untapped resources that are available for language learners to make use of. This paper explores the possibility of creating a subtitling model such that the process of using subtitles for learning can be streamlined from the creation side in a way that users are able to effectively use subtitles to learn foreign language.

1.2. Research Goals

- Create an environment where learners can focus more on learning when compared to L2 subtitles

- Create an environment where learners can acquire second language knowledge using their own preferred subjects

- Create a model or models where suitable subtitles can be chosen to transfer second language knowledge to learners

- Create an automated model through the use of computers or other techniques to extract useful learning materials from different video media sources

- Create an automated model to adapt the material to the user level

- Minimize the time required to create these materials

1.3. Overview

The paper is split into 6 chapters. The second chapter will go through the relevant topics that have been studied up until now, as well as their merits and the challenges that they face. The third chapter will talk about the methodology for creating tools that will attempt to mitigate these challenges. The fourth chapter will delve into the design of the trial experiment, the challenges from the trial experiment, the lessons learned to create the final version of the experiment, and a final small experiment on the feasibility of implementing cognitive measurement for the model. The fifth chapter will discuss the results of the experiment. The final chapter will be the conclusion of the paper, as well as the future implications of the system.

Chapter 2

Literature Review

2.1. Current Caption Implementation

Captioning started with the purpose of providing a simultaneous and mostly accurate transcript for viewers of television broadcasts in the 1970s (Allen 2015). The purpose of its creation was to make television broadcasts more accessible to those that are deaf or hard at hearing. Over time, more and more situations arise where other user can also make use of these captions in order to improve their understanding of these multimedia works. Over time, its purpose also transformed to include that of providing for those in a position where aural information may not be sufficient for understanding the multimedia in question. Even in situations such as busy environments that are not suitable for accurate hearing, subtitles can provide visual aids in which viewers that are not able to receive the aural information a way to perceive the information of the multimedia.

Viewers of foreign language multimedia that may not be fluent in the language can be thought of in a similar manner as those who cannot perceive the aural information of these multimedia clearly. Captions are able to fill in the gaps for information that these viewers may not be able to understand clearly, while providing confirmation for materials that the viewer understands. To this end, subtitles can be prepared in the language that the viewer is most familiar with, such as in the native language of the viewer. This helps with the understanding of the foreign language material to the extend that even without voice dubbing, which is the act of replacing original voices of the actors and actresses within the show to ones in a different language, viewers are able to make use of the visual information in order to understand the multimedia material, removing the need of the words entirely.

It is important to clearly define the words associated with the two types of

subtitles before continuing. The paper will use the following words precisely for these types of subtitles:

Table 2.1 The different modes of subtitles available

Subtitle Mode	Audio Language	Subtitle Language	Learning Mode
L1 subtitles	L2	L1	Incidental
L2 subtitles, captions	L2	L2	Intentional
Reverse subtitles	L1	L2	Intentional
Translated	L1	L1	Incidental

L1 represents the native language of the learner, while L2 represents the foreign language that is being learned. The audio language of the TV program or commercial multimedia can be originally L1 or L2. The audio language for reverse subtitles can be replaced with an L1 language version.

Nowadays, these subtitles are widely available in many different multimedia, such as television broadcasts, commercial entertainment products, and even social multimedia content. Some places, such as in North America[1] and the UK, even mandate television broadcasts (Myers 2017) or even internet streaming businesses (Enamorado 2019) to include closed captioning by law. There is no shortage of material with a wide variety of genres that can accommodate the learners' preferences, although regional restrictions may mitigate some of their choices.

Closed captioning[2] is one of the most widespread use of captioning method for television broadcasts. Viewers are able to choose whether to enable the captioning text that plays alongside the multimedia in real time depending on their needs.

---

1   The US Law requires closed captions to be included in all multimedia in TV broadcasts.

2   Closed captioning (CC) and subtitling are both processes of displaying text on a television, video screen, or other visual display to provide additional or interpretive information. Both are typically used as a transcription of the audio portion of a program as it occurs (either verbatim or in edited form), sometimes including descriptions of non-speech elements. Other uses have been to provide a textual alternative language translation of a presentation's primary audio language that is usually burned-in (or "open") to the video and cannot be disabled.

2.2. Related Works

Research as early as 1983 have shown that captioning have benefits for viewers learning the language of the multimedia. The groundbreaking work on the research of subtitles by Karen Price in 1983, which proved that captions provided non-native speakers more accessibility to otherwise incomprehensible foreign language multimedia or films (Price 1983), have paved the way for more insightful research and understanding on the benefits of both subtitle and captions as a language teaching medium.

In general, L1 subtitles, or subtitles in another language, facilitates incidental learning, where viewers indirectly pick up words and phrases in the spoken language without the script of the foreign language. L2 subtitles help with vocabulary and listening comprehension in a more direct manner (Danan 2015).

2.2.1 Incidental Learning through L1 Subtitles

Watching L1 subtitles are also beneficial to learning, although not in a direct manner. A study was conducted by Koolstra and Beentjes (Kuppens 2010) on the degree of incidental learning for watching a foreign multimedia through L1 subtitles. The study was conducted on Dutch fourth graders and sixth graders to test short-term acquisition of English language vocabulary. The group was split into three and were asked to watch TV shows that differed from language and subtitles. One group was asked to watch only Dutch TV programs, while the other groups watch English programs. The group watching the English program was further split into 2 to watch with and without subtitles. Results show that the groups watching English programs reported a higher vocabulary and word recognition scores. The students that watched with L1 subtitles (in this case, in Dutch) reported an even higher statistically significant score on the tests, which suggests the benefit of actual long-term foreign vocabulary acquisition.

A study conducted by Lommel, Laenen, and d'Ydewalle (Marzà and Torralba 2015) confirms that the cause of improved incidental learning can be mainly attributed to the L1 subtitles, as opposed to that of just watching the foreign multimedia. In the study, both Dutch children and teenagers are shown a Esparanto film with Dutch subtitles, and using different instruction rules, determine the

effects of subtitles and language acquisition. While this type of presentation enhanced the performance of secondary school students more than primary school students, no acquisition of language was found by only watching the movie by itself.

These results corroborate with past studies that were conducted with adult viewers (d'Ydewalle and Pavakanun 1995), and results indicate there are also mild improvements in grammar and writing skills in adults, in addition to short term and long term grammar and vocabulary acquisition.

2.2.2 Improving Speech Segmentation through L2 Subtitles

For foreign language learners, a difficult process of learning through traditional method is improving speech segmentation[3]. A study conducted by Goh (Goh 2000) found that second language learners are aware that they have difficulty in recognizing words when they are spoken to, even if they already have knowledge of the individual words. Further experiments by Charles and Trenkic (Charles and Trenkic 2015) revealed that a sample of 10 international students accepted in a UK university based on their IELTS score only managed to repeat 71% of words correctly drawn from a pool of excerpts, compared to native L1 speakers who managed 100%.

While written words have indicators to differentiate between boundaries, there may only be subtle indicators for word segmentation during continuous speech. Rather than physical boundaries, spoken word boundaries lie within language specific lexical segmentation methods according to research by Culter and Butterfield (Cutler and Butterfield 1992), such as that of stressed syllables in English. Since speech segmentation strategies differ from language to language, foreign language learners may find themselves applying strategies from their own language to foreign languages, increasing the difficulty of speech segmentation.

As such, research on bi-modal input[4], or same language subtitling, have in-

---

3   Speech segmentation is a term for a mental process used to describe identifying the boundaries between individual words, syllables or phonemes in a continuous stream of speech.

4   Bi-modal input is a term to describe the simultaneous presentation of matching aural and orthographic simulation.

creased our understanding of how to improve a foreign language learners' speech segmentation ability. Another experiment by Charles and Trenkic (Charles and Trenkic 2015) tested the efficiency of language acquisition through different forms of multimedia. Using a sample size of 12 same nationality students roughly similar in average IELTS score and time spent abroad in the UK studying, they were shown videos with L2 subtitles and sound, L2 subtitles on without sound, and sound with no subtitles. While initial scores were lower for the group with subtitles and sound pre-test, they scored consistently higher over the course of the four week test where they were shown videos with the subtitles and sound. The test seem to find better improvements in language acquisition with the use of watching multimedia with same language subtitles.

2.2.3 Reverse Subtitles

The process of using reverse subtitling as a means of language education is to have learners translate L1 materials into L2 subtitles. A study conducted by Paulina Burcyńska (Burczyńska 2015) found that there are many distinct advantages of the reverse subtitling method compared to learning through watching regular L1 or L2 subtitled multimedia. The study was conducted in a classroom environment, and its purpose was to see if the use of reverse subtitling would lead to a decrease in grammatical and spelling mistakes, as well as the terminology that students are able to acquire.

The pilot study involving people of ages ranging from 20 to 40 was split into three groups. The participants were shown selected scenes of a Polish comedy show and were tasked to create a group translation. Translations were then debated between groups for the most accurate translation. The show contains many cultural elements and specific terminology on different register levels, proverbs and idioms to ensure discourse of the selected elements. Both qualitative and quantitative tests were performed to evaluation the method. In order to test if students improved more using reverse subtitling, participants were required to submit a 200-word written assignment before and after the test for comparison. The material is then evaluated based on the mistakes made and vocabulary used.

Before pilot test, the written assignments from the experimental group and the control group were compared. The experimental reverse subtitles group made an

Table 2.2 Evaluation of written production BEFORE study (Burczyńska 2015)

	Control Group	Experimental Group	Difference(%)
Grammar mistakes	45	42	7%
Spelling mistakes	33	32	3%
Advanced vocabulary	25	27	8%

average of 7% and 3% less mistakes on grammar and spelling tests respectively. Participants of the experimental group also scored 8% higher in tests involving the use of advanced vocabulary[5] compared to the control group. Overall, the difference is not significant and the results of the groups should be comparable to each other for future comparisons after the test.

Table 2.3 Evaluation of written production AFTER study (Burczyńska 2015)

	Control Group	Experimental Group	Difference(%)
Grammar mistakes	38	28.5	25%
Spelling mistakes	28	25	10.5%
Advanced vocabulary	32	39	17.9%

After the pilot test, there is a significant increase of ability for the experimental group using the reverse subtitles compared to the control groups in classes. The difference of making less grammatical mistakes grew to 25%, while the difference between spelling mistakes grew by 10.5%. The experimental group also shows a stronger command over advanced vocabulary when compared to the control group. Overall, it would seem that the results quantitatively supports the hypothesis of reverse subtitling being a better learning method than traditional textbook environments.

The questionnaires for the experiment show the attitude of the student towards reverse subtitles. More students in the experimental group found reverse subtitling to be more interesting to learn through. The students also felt that reverse subtitling helped them practice their writing more, and had more interesting and

---

5   Advanced vocabulary in the test is defined as various fixed expressions, idioms, synonyms, to advanced verbs etc.

new vocabulary to learn compared to traditional textbook methods. The ability to learn grammar, however, seem to be similar for both the control and the experimental group.

Overall, students find subtitling to be more rewarding and enjoyable compared to traditional methods, and believed that the use of subtitling helped in their writing and translation skills. Since these aspects may affect the motivation of the student. Since motivation may the efficiency of the learning, it may also be a useful factor to consider measuring.

Since the exercise also requires learners to make up the subtitles using their own knowledge, a complete translation of all the dialogue found within the multimedia may be difficult or even impossible to those not possessing advanced knowledge of the foreign language (Zanón 2006). However, as they have to come up with their own translations, the interpretation and summary work of subtitling L1 multimedia using L2 subtitles can potentially increase the awareness of the user. The target use of this user test is mainly advanced English students in university. Similar confident results may not be achievable with all levels of learner using reverse subtitles. In addition, the knowledge learned through the process may derive from the group work discussion period, rather than the reverse subtitling period. It is unclear how much of the knowledge learned directly comes from the subtitling process.

2.3. Limitations of Current Methods

While there is an ever-growing library of captioned materials both online and offline, accessibility of these captions are also an issue. Due to the fact that original language subtitles are targeted for audiences that speak the language, the availability of these original language subtitles is limited to foreign language learners that may wish to make use of these materials. Since closed captions are broadcasted within the country, learners that are not living in these countries are often only offered subtitled version in their native language. The option for a lot of multimedia simply does not exist across borders, with multimedia sold overseas have captions removed and replaced with foreign subtitles due to licensing issues and fees. As long as the idea of using captions as a learning tool in addition to a

hearing aid are not within common knowledge of the people, L2 foreign subtitles will have a difficult time to propagate (Vanderplank 2015).

A limitation of using the captioning technique for foreign language learning is the additional work required in order to focus the learners' attention (Vanderplank 2015). The purpose of learning using the caption is that the caption acts as an aid to the overall dialogue that learners ' have difficulty in understanding the aural information. Therefore, it is key that learning be performed by reading the caption without focus on dialogue from the multimedia (Lertola 2015). In order to focus the attention of the viewer, one must prepare tasks beforehand for the viewer, and without tutor guidance, preparation of tasks beforehand may be impossible due to the different level of learning required for different individuals. Using normal subtitles require additional self-discipline from the learner and effort from teacher to prepare suitable tasks that must also refer from the video material itself. Another challenge that learners might face when using captions is the difficulty of breaking the habit of reading captions. Learners that read captions may become used to reading captions that they instinctively read all lines of captions even if the learner may already have the ability to understand without the use of such caption (Lertola 2015), in some some cases may become a liability in learning the foreign language.

Chapter 3
Methodology

3.1. Overview

This paper proposes a new set of rules for a captioning system specifically designed for educational purposes. The method aims to provide both teachers and learners improved efficiency over the use of commercial media and captions currently present in the market. The following sections will detail on the design of the system, as well as the method to evaluate such method in order to find its efficacy.

3.1.1 Design Goals

The new subtitles should provide more accurate transcription of subtitle scripts for less confident viewers. Current captioning systems have been designed with the expectation of the situation where the subtitles can only be viewed once, such as during television programs or public viewing. As such, there may be instances where caption accuracy may be sacrificed in favor of readability for the viewer. For this reason, the subtitles may have conflicting words or sentence structures when compared to the spoken script. While fluent viewers may find such changes reasonable, less learned viewers may find the differences in the script confusing, or may even misidentify the spoken script as the edited subtitle script. With the implementation of the new subtitles, accuracy of transcription can be made a priority to provide learners a more confident understanding and positive reinforcement of the learning material.

The new subtitles can provide both the merits of L1 and L2 subtitling without sacrificing the others. Current implementations and studies of captioning have only considered allowing one of L1 or L2 captioning. However, since L1 and L2

subtitles both have their own unique benefits to viewers, learners of a foreign language may find use in each of the above captioning methods. Carefully consideration and planning of the uses of L1 and L2 subtitling can provide different learning opportunities for the viewer, and reap the benefits of both captioning methods at once. Thus, the new experimental subtitles can potentially provide for a more well-balanced curriculum for foreign language learner to be created.

The method should also be beneficial in that the learner would be put in a learning environment without taking them out of the entertainment of a commercial product. Although there exists commercial digital products, such as Rosetta Stone, that can provide an environment with audio and visual aids, the program is academic in nature and does not provide an entertainment experience. Individuals have many reasons to learn foreign language, and these products are only gateways to access other foreign content such as TV shows or films. By using the new subtitles, learners should be more motivated to learn the language while also learning to consume the media in the process.

The subtitle script can be adapted for different level of users, such that certain materials do not have to be limited to a certain level of learners. Since the difficulty of words and phrases differ from show to show, it presents a challenges to some learners that may lack the language ability required to watch the show. Even with commercial L2 subtitles, learners may be discouraged to watch the media that they wish to enjoy due to the fear or fatigue of not understanding. While it is true that the materials in itself has a its own difficultly level depending on the type of show, and the target audience, the materials can be adapted in which the focus of the learning material can be changed to that appropriate for the user. Media of difficult topics such as politics, documentary, comedy etc. which by using traditional subtitle methods, may be out of reach for learners, can be adapted such that even difficult figures of speech can have words and phrases that can be of focus for the learner of the current level. By designing the the new subtitles to be adaptable, difficulty of the material and level of the learner can be decoupled, and learners can enjoy both learning and entertainment without sacrificing the other.

3.1.2 Challenges

The factor of an additional available track of information on top of conventional subtitles on display would mean that while there are more possibilities of creating a subtitle track for the purpose of educating the viewers, there are also pitfalls to the system.

If both the subtitles and the media source are not well-optimized by hand, the learner may be overwhelmed by the overload of information present within the two subtitle tracks and the visual-audio information from the medium itself. According to a study by Romero-Fresco, the usual televised speech rate does not match the usual reading rate of an average person (Romero-Fresco 2009). This is especially true for learners of the foreign language, as they may not have ability to read as quick as the average native speaker of the that language.

The manual labor involved in the creation of these subtitles will also be increased compared to traditional captioning methods. While the creation of the L1 and L2 subtitles will most likely only take a similar amount of time compared to conventional subtitles, transforming these subtitles into the appropriate form may require more time. A key element of designing the subtitles is to mitigate the additional time spent creating the materials using the new method.

3.2. Captioning Strategy

3.2.1 Objectives

In order to reach the design goals and reduce the negative impact brought by the above challenges, we can begin by defining a set of rules to follow such that the process can be streamlined and make use of different strategies to fulfill the goals.

3.2.2 Transcription

The transcription process is straightforward. First, the whole source material will be transcribed from the original language into L2 subtitles. Long sentences should be broken into phrases by their connectives if necessary. While there is no strict guideline on when a sentence should be broken up, a sentence should not

take more than two lines of subtitles, as a overly long sentence may confuse the learner.

Each of these lines that appear will be classified as units of speech. Each unit will then be assessed based on its difficulty, as well as the foreign language ability of the user, in order to determine the best possible way to use these captions to maximize the learning potential and efficiency of the user while simultaneously minimizing confusion.

3.2.3 Translation Strategy

The translation of the materials into L1 subtitles can be approached from two different ways.

Direction translations Direct translations are how commercial L1 subtitles are created. Since the goal of L1 translations is to portray the speech in a native manner that is most comparable to the foreign language, L1 translation may not use the same sentence structure or even the same words as the speech. In the context of the this study, a direction translation is to translate a unit of speech completely without considering the matching of L1 and L2 word order. A direct translation may be favourable in situations where whole expressions or phrases are difficult to break apart, but should be used sparingly and only at critical moments as it may potentially double the amount of characters on screen.
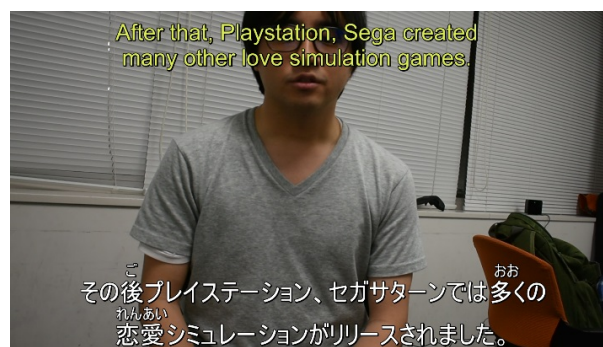
Figure 3.1 Example of full translations on screen from test trail.

Partial translations For the case of normal Japanese sentences to English trans-
   lation, a less direct approach may be favored since the sentence structure
   is different for each language. The constitution of a complete Japanese
   sentence is different compared to English, and a less grammatically correct
   English translation, or even a broken up translation, may actually be a more
   effective way of making learners read and understand the text quickly, and
   pair up with audio information in real time[1]. In addition, since there may
   be increased amount of characters on screen at a time for the learner to
   read, a partial translation technique may be more favourable to reduce the
   amount of characters on screen to facilitate easier reading and better focus
   of critical words.

Figure 3.2 Example of partial translations on screen from test trail.

A peculiarity of the Japanese language must also be considered for learners.
The Japanese language makes frequent use of overloaded adaptations of the Chi-
nese logographic script[2] in its writing called the kanji. In contemporary written
Japanese, kanji is almost entirely orthograhic, in which words represent meaning.
A Japanese word or morpheme[3] may be represented in writing by a Chinese lo-

---

1   The Japanese language uses a sentence structure of SOV: subject-object-verb, while the
    English language uses a sentence structure of SVO: subject-verb-object.

2   Logographic script represents written characters, or logograms, that itself carry meaning of
    words and phrases. Japanese Kanji, Chinese characters, Egyption hieroglyphs are examples
    of logographic writing systems.

3   A morpheme is a unit of meaning which cannot be divided further.

gographic character, with no regard to the original sound value of the Chinese character itself (Frellesvig 2010).

This adds another challenge to learners, and gives an advantage to those who are already familiar with the Chinese logographic script. An implication of this is that learners' level in the Japanese vocabulary may not entirely be representative of their reading ability. Since kanji represents invidual words, the system can adapt to the kanji knowledge of the learner in order to output appropriate either the phonographic script of its kanji equivalent. While learning kanji is not the focus of the experiment, furigana will be used to accomodate to people that are not well versed in using kanji.

Figure 3.3   Furigana implementation for the experiment.

There are exceptions to the above translation strategy. Often, wordings of idioms and common expressions are not completely shared amongst languages, or may even be exclusive to the foreign language from the users' native language.

In such cases, a direct translation may fail to confer the meaning of the idiom accurately, creating additional confusion for the understanding and learning process of the viewer. Therefore, idioms should be considered a special item of its own, and clearly shown to the viewer that the phrase in question is a foreign language idiom or expression. Extra care should be taken by the tutor in order to accurately explain the purpose of said idiom or expression, and the context that it is used in[45].

---

4   Similar example: 出る釘は打たれる, the nail that sticks out gets hammered down.

5   Dissimilar example: 眉毛を読まれる, to be read like an open book

3.2.4 Materials Used

Since the scope of the paper resides mostly on the performance of the proposed subtitling technique on beginner to intermediate learners, materials that will provide more learning opportunities for these users are preferred. In order to discern whether the materials are useful for the experiment, a rough estimate can be given by the choice of the type of the video. Intuitively, materials that are aimed for younger people of the native speakers should indicate a tendency for suitability for earlier learners of the foreign language.

Table 3.1 Vocabulary distribution for a children's cartoon show

JLPT Level	Count	Common	Common %
1	4	4	100.00%
2	11	11	100.00%
3	18	18	100.00%
4	27	27	100.00%
5	60	60	100.00%
Unclassified	53	35	66.04%
Total	173	138	79.77%

The main element for consideration of the materials are the distribution of N5-N1 vocabulary of the material. Since the vocabulary's distribution of difficult directly determines the amount of the material that will be of use to the learner, the material should not be biased against the learner's level. In the case of the experiment, since the participants will mostly be at a beginner to intermediate level, the material used should have the majority of vocabulary difficulty distribution towards N5 to N3.

However, not all words are classified into the N5 to N1 system. In order to determine the usefulness for unit of speeches that contain unclassified words, the factor of commonality can be considered. Words that are considered common are more useful to the learner in real world applications, and can be therefore seen as more valuable for the learner to learn. To determine if a word is common or not,

the experiment will use the EDRDG's[6] Japanese dictionary project JMDict[7] in order to determine the commonality of the project. Word entries in the JMDict dictionary project are considered common if they are within the top 20% of the search results within the dictionary. Further information regarding the method for classification of words to JLPT levels can be found within Chapter 4.

3.2.5 Playback

While there is no standardized reading rate for creating subtitles, many studies have been done regarding how many characters a person can read at a given time. The reading rates of adults and children are different, since adults have less difficulty correlating the audio information with the orthographic information of the media. For viewers reading L1 subtitles, studies by scholars such as Karamitroglou (Karamitroglou 1997) indicate that the average reading speed of adults are at 150-180 words per minute, while children could only read slower at 90-120 words per minute. When reading subtitles and correlating information with the media, the ability of learners that are less familiar with the language are similar to that of young children. Additional time is required to take in and decode the information of the orthographic text. As their ability improve, so does that of their speed of reading (Minucci and Cárnio 2010).

The 6 second rule seems to be the number most scholars agree upon as a target to be set for standard subtitles aimed at adults (Karamitroglou 1997). The 6 second rule establishes that a subtitle of two lines long should allow for no shorter than 6 seconds of reading time. Foreign language learners, however, are similar to children in that they are still developing their reading skill, and will require longer reading times. However, unlike the adult 6 second rule, no general rules for children have been well established and agreed upon. A guide by Lorenzo indicates a modified 8 second rule, while programs such as Screen Subtitling Systems Ltd.

---

6   The Electronic Dictionary Research and Development Group was established in 2000 within the Faculty of Information Technology, Monash University.

7   The JMdict (Japanese-Multilingual Dictionary) project has at its aim the compilation of a multilingual lexical database with Japanese as the pivot language. The project began in 1999 as an offshoot of the EDICT Japanese-English Electronic Dictionary project.

or WIN2020 recommends 9 characters per second, a 50% reduction compared to adults. By matching the subtitle rates, which are measured in characters per second, to the reading rate of the learner, we can gauge the most suitable rate of playback required for efficient learning. The study is mainly conducted with early beginner to intermediate users, thus a slower playback rate is required to make sure users are comfortable with reading the subtitles while also being to take in visual feedback from the media.

3.3. Assessment

As individuals have their own learning modes and speeds, it is difficult to compare the results of two separate groups of learners taking the control and the experimental test separately. Thus, the approach of the paper is to determine the effectiveness of the method by testing participants in different orders. Both qualitative and quantitative approaches will be used.

Table 3.2 Method to test effectiveness

	Batch 1	Batch 2
First Half	Control	Test
Second Half	Test	Control

Chapter 4

Prototype

4.1. Trial Implementation

A trial test was performed including the elements that were included within the last section in order to assess the real life efficiency of the combination of different methods.

4.1.1 Video

Two shows in two different topics will be produced. The learner can then choose between the topic that they are more interested in. The shows will be captioned according to the principles as outlined in the previous chapter, and the transcript of the subtitles will also be produced for analysis.

The length of the short videos will be aimed at two to four minutes, and there will be minimal scripts and cues for the actors involved, with just enough to facilitate discussion. The video will then be recorded and subtitled, and chosen randomly as to create both the control and experimental videos.

As the objective of the research is to prove that commercial programs with the proposed academic subtitles create a better learning environment, it is crucial that the video emulates an authentic television show with no pretense of use in education as much as possible.

The actors recruited should either be a native speaker in Japanese, or deemed fluent and have the qualification to teach Japanese to ensure that if used as education material, the aural information quality, such as the pronunciation, grammar, and intonation are correct. The actors were told that the production of the video is strictly for entertainment for learners of Japanese. The actors were requested to speak in a clear manner for the show, and to avoid over-use of colloquial language.

although the use is not forbidden, nor encouraged to be ignored completely.

Colloquial language represents the way casual conversation occurs, and may sometimes not follow standard language conventions. Although differing from formal speech and writing, it may be just as valuable, if not more depending on the medium, for the learner to understand the differences and nuances between colloquial and written language.

The difficulty of the conversation is also an important factor to control. If the topic of the speech is not controlled well, the topics may veer the conversation to that of a complex one, which may be difficult for the audience of the scope of the paper to understand. To avoid this situation, easier to understand themes and topics are chosen for the video.

4.1.2 Pre-Test

A test is first prepared for candidate selection. The questions in the test will be categorized from N5-N1, based on the difficulty level of the grammatical item on the JLPT test. Although the JLPT does not provide an explicit list of which grammatical items belong to which level, there are many resources both online and offline that can provide a good estimation for each item. The questionnaire is created based on the free online resource grammar[1]provided by Jonathan Waller and other offline textbook guides as reference resources in order to provide the most accurate sampling of candidates. Since the research concerns only beginner to intermediate level learners, only questions from N5-N3 are sampled and used for creating the questionnaire.

The answers from the candidates are collected, and the answers are sorted by level. N5 questions are evaluated first, then N4, and finally N3 to estimate the level of the learner. The criteria for passing a level is the number of grammatical items that the candidate is familiar with compared to all questions. If the candidate is confident in more than 75% of the questions for that level, the candidate is deemed to have passed the N5 level, and the next level N4 is evaluated. Should the candidate fail to score a passing grade, the learner will then be allocated to the level that the candidate failed to pass. Adaptations to the subtitle will be

---

1 `http://www.tanos.co.uk/jlpt/skills/grammar/`

based on this level.

4.1.3 Type Categorization

Once the materials have been transcribed, we can then move onto creating the trial experiment subtitles. To start, we can first look at the units of speech of the subtitle script from the perspective of a learner in order to gain insight of how to effectively design a system for learning. From the perspective of the learner, the different units of the subtitle script can be separated into three distinct categories.

Table 4.1 Rules for categorization

Scenario	Purpose
Type A: Units that are **BELOW** the learners' level	Reinforcement of learned knowledge.
Type B: Units that are **SIMILAR** to learners' level	Acquisition of new knowledge.
Type C: Units that are **ABOVE** the learners' level	Provision of context for entire materials. No knowledge acquisition required.

All units of speech, no matter its difficulty, should fall into either one of these categories. Of course, the learner cannot be expected to create and categorize their own subtitles. Therefore, there must be a set of created rules in order to interpret all the units of phrases into the above created categories.

At its simplest form, a sentence is formed by individual vocabulary items given context by the grammar. Therefore, we can crosscheck the sentence elements with reputable sources in order to determine the difficulty of each element within the unit of speeches, and can therefore determine which category the unit of speech falls under. The method used in the paper measures difficulty through The Japanese Language Proficiency Test[2].

---

2   The Japanese Language Proficiency Test is a test offered by Japan Foundation and Japan Educational Exchanges and Services, and is the largest-scale Japanese-language test in the world. The test is divided into five levels, with N5 being the lowest, while N1 being the highest one can achieve in the test.

Since the Japanese Language Proficiency Test is divided into N5 beginner to N1 mastery of the Japanese language, the system can therefore be sub-divided into five different levels for adapting to the progress of individual learners. By crosschecking the vocabulary items with the Japanese Language Proficiency Test vocabulary level and the grammatical items used with the Japanese Language Proficiency Test level, we can roughly understand how to categorize each of these unit phrases into the above mentioned categories. To clearly understand the process of categorization, figure 4.1 is an excerpt of a transcription of a Japanese cartoon.

N5	N4
この嫌な気配は...	この嫌な気配は...
やっぱり　お前か	やっぱり　お前か
オラのテレパシー感じてくれた　かずまくんって...	オラのテレパシー感じてくれた　かずまくんって...
やっぱりオラのこと愛してくれてるんのねん～	やっぱりオラのこと愛してくれてるんのねん～
鼻息が項にかかっているや　だれだって分かるよ	鼻息が項にかかっているや　だれだって分かるよ
う～　なるほど	う～　なるほど
どこ行くの？	どこ行くの？
関係ないだろう	関係ないだろう
もう～　冷たい～	もう～　冷たい～
じゃあ　当ててみようか	じゃあ　当ててみようか
ママに頼まれてお買い物に行く	ママに頼まれてお買い物に行く
（なんで分かるんだ）	（なんで分かるんだ）
あまやかされ子のかずまくんがお使いなんて珍しい	あまやかされ子のかずまくんがお使いなんて珍しい

Figure 4.1 The adaptability of the system for two different level of learners

The three different color highlights indicate the different categories of the unit of speeches. Green denotes *Type A*, yellow denotes *Type B* orange denotes the relevant vocabulary for the level of the learner, and red denotes *Type C*. The strategies used for each of the three types of units can thus be discussed.

Table 4.2 Summary for strategy for captions

Purpose	Strategy
Type A (Reinforcement)	L2 subtitles only
Type B (Acquisition)	L1 and L2 subtitles
Type C (Context)	L1 subtitles only

Type A2 The purpose of Type A subtitles are to reinforce likely learned knowl-
edge from the learner. In order to achieve this, L1 subtitles that the learner
is well familiar with should be minimized or removed altogether. This will
encourage learners to try to make out speech using mostly aural information.
Leaving the L2 subtitles will provide necessary fallback as well as positive
reinforcement for the learner in order to increase their confidence in their
ability to process foreign aural speech patterns.
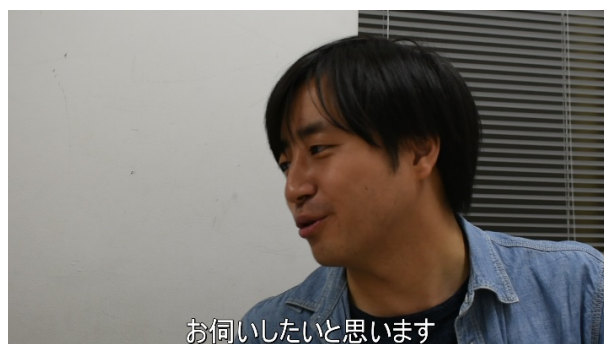
Figure 4.2 Example of a way to display Type A2 units of speech

Type A1 Alternatively, the L1 and L2 subtitles can be removed altogether in
order to create an environment where the learner must rely solely on aural
and visual information in order to interpret the scene without the aid of
textual information. The purpose of such strategy is to remove the reliance
of subtitles without requiring extra aids such as encouragement from the
users. This strategy may be more useful compared to the above if the
knowledge of the learner is stages above the unit of speech in question. In
addition, the subtitles can be designed to hinder or obscure the textual
information without outright removal of the subtitles. This should provide
the learner opportunities to try to understand the script without removing
the option to fallback should the learner fail to grasp the aural information
of the media.

Type B Type B is the category that should challenge the user and build up the
users' knowledge on vocabulary. Since the focus of learning is within Type
B unit of speeches, tasks at the end of the experiment should be designed

with the majority of the questions challenging the new knowledge learned in this area. A method to challenge the aural perception ability of the foreign language for the learner is to blank out the relevant vocabulary. A translation aid can displayed alongside the blanked out word to encourage learners to listen and identify the correct aural information.

Figure 4.3 Example of a way to display Type B units of speech

Type C Type C deals with units of speech with difficulty levels above what the learner is required to learn at the current state, or requires other previous knowledge in order for the learner to comprehend. As such, the function of the units of speech of this level is mainly to act as a transition and bridge other units of speech together in a manner which the learner is able to understand the whole dialogue, or provide context for future conversations. A direct translation is favored for Type C style unit of speeches in order to relay the necessary information for the learner to continue with consumption of media.

The tool used to add subtitles to the video is the free and open software Aegisubs[3]. The experimental subtitles are made according to the above principles. The transcript of the video is used to create the experimental subtitle test materials. The whole transcript is split into units of speech. All the vocabulary within these units are analyzed and given a JLPT rating. The rating will then be used to rate the units of speech on the JLPT N5 to N1 scale. ($Unit of Speech Difficulty =$

---

3 http://www.aegisub.org/

Figure 4.4 Same dialogue as above, but in Type C format

$f(MaxVocabDiff, CommonWord)$) The rating is then compared to the level of
the material, which in the case of the experiment is N5 and N4, and categorized
according to the above listed types. From table 4.5, we can see that different
JLPT levels will have different tasks for most of the unit of speeches, and will
have different styles of subtitles displayed.

Text	JLPT Level	n5	n4
はい	5	a2	a1
なるほど	5	a2	a1
なんか、こう他に暑いところっていうと	4	b	a2
どんなところを連想しますか？	3	c	c
暑いところだと、アフリカとか	5	b	a2
アフリカ、アフリカか	5	a2	a1
砂漠みたい	4	c	b
砂漠、サハラ砂漠とか	3	c	c
はい	5	a2	a1
行ったことありますか？	5	b	a2
ないですね	5	a2	a1
ないですか	5	a2	a1
なんか、足ん...靴も入らなくて そのまま砂の上立ちが	4	c	b
あ〜いいねいいね	5	a2	a1
砂漠か	3	c	b

Figure 4.5 Excerpt of rated subtitles for video A

4.1.4 Preparation Summary

Table 4.3 and 4.4 shows the amount of time each task takes to create the above
two videos

Another flaw with this method is the experimental subtitle creation process.
With two different difficulty levels in mind, the time that it takes to create the
appropriate subtitles increases linearly. With only 2 sets of experimental subtitles,

Table 4.3 Time taken for each task for video A

Video A (150s)		Time Taken
	Scripting	0.5
Video	Recording	0.5
	Captioning	1
Analysis	Type Cat.	1.5
	Furigana	0.5
Edit	Subtitle N5	1.5
	Subtitle N4	1.5

Table 4.4 Time taken for each task for video B

Video B (240s)		Time Taken
	Scripting	0.5
Video	Recording	0.5
	Captioning	2
Analysis	Type Cat.	3
	Furigana	1
Edit	Subtitle N5	2
	Subtitle N4	2

the time that is taken to prepare for the experiment when compared to just captioning and creating the video is an approximate 250% and 233% difference for Video A and Video B respectively. Should the whole range of N5 to N1 subtitles be created, it would take an even higher amount of time to create, reducing the efficiency of creating the subtitles massively.

4.1.5 Trial Experiment

The playback device chosen for the test is the free and open-source software, VLC media player[4]. The media player supports functions such as media playback speed modifier, subtitle tracks, and plug-in support that would aid in the playback of the video the participants.

Since the tool allows for a variable rate of playback, such function can be used to allow for more adaptability to the needs of the individual. Beginner learners who have trouble with the rate of speech for a certain program may choose to reduce the speed of playback for separate videos or even difficult sections if it is necessary. While such function is also available to use for normal commercial captioned media, it alleviates much more pressure for the increased amount of information that the academic subtitles this paper proposes provide to the learners. In addition, the participant can also stop at anytime, and stop points should be created within the video in order to allow for replaying scenes should the participant require it.

The materials will be divided into two parts of equal length, with one part being subtitled in the conventional manner, and the other with the method described in this chapter. The participants are first requested to finish a test to assess their JLPT level. Since the trial focuses on beginner to intermediate level learners, should the participants exceed JLPT level 4, the test will terminate. For the other participants, they will then be divided into two groups. For the first group, the participants will first be shown the control video. Afterwards, they will be shown the video with the trial experimental subtitles. The other group will also perform the same tasks, but in reverse order. For both videos, the participants will be asked to learn vocabulary through the video as much as possible. The participant can ask the instructor any questions they have about the language

---

4 https://www.videolan.org/vlc/index.html

problems regarding the video dialogue. After each video, they will be asked to complete a questionnaire to assess the experiment. The videos will be initially set to 80% playback speed, and the user can freely change the speed of playback to their liking.

Table 4.5   Trial experiment process for the two batch of participants

Batch 1	Batch 2
Experimental Video	Control Video
Questionnaire	
Control Video	Experimental Video
Questionnaire	

4.1.6 Trial Results

Participants often showed vocabulary knowledge that is different from their grammatical knowledge. Since the pre-test assessment using tests the participants' grammatical knowledge, and then assumes their vocabulary knowledge to be on a similar level, it is not uncommon for participants to find the subtitle content and the way it is presented to be either too easy or too hard. The method used in the trial experiment to assess the vocabulary ability of the participant may be too random and inaccurate.

Since the subtitles are constantly cycling through different types, participants also noted that this constant change of subtitle display method creates confusion on what they were supposed to do. Rather than improve focus on learning, the participants easily lost focus and interest on learning through the subtitles, and spent more energy on deciding what the subtitles were supposed to mean.

Overall, the trial results were disappointing. Participants did not show signs of learning more than the control test in the experiment. Questionnaires completed by the participants also show that participants did not seem too interested in learning using the trial experiment subtitles over conventional subtitles.

With the lessons learned from the trial in mind, the next section describes a more streamlined design for the subtitles can be created from the feedback and the results of the trial experiment.

4.2. Prototype

The idea of the new model is to use more mathematical models in order to increase the automation of the process of determining the difficulty and reduce the amount of guesswork required from instructors. In addition, explicit tasks are created using the data from the model in order to better focus the student and improve upon the use of the subtitles as a learning tool.

In this new model, the focus will be to create a new difficulty system decoupled from the JLPT test scores. Each individual is given a score level, and depending on if the user is able to complete tasks that are similar to his own level, the score changes to better reflect their ability. The score is ever changing as the individual completes more tasks. Thus, the most important factor is to create a specific model that assesses the difficulty of each unit of speech used in subtitled videos.

The ability of the learner to understand the dialogue spoken in the video can be classified into two categories: Listening, and vocabulary knowledge.

$$Unit of Speech Difficulty = f(Listening, Vocabulary)$$

We can then design the system around calculating the values necessary for these two factors in order to create a system to find the difficulty value of each unit of speech. The difficulty value can then be used to create environments that will aid in improving the language skill of the learner.

4.2.1 Listening

For listening, the difficulty of processing aural and subtitle information mainly comes from the speed of the unit as well as its length. These factors can be measured in terms of characters per second [5], as well as the unit of speech length respectively.

$$Listening Difficulty = f(CPS^6, Char^7)$$

---

5   Refer to Chapter 2 for more information on research for characters per second

6   Characters per second

7   Character length of unit of speech

Character length is simply the number of characters displayed onscreen for each unit of speech. CPS can be found by dividing the character length of the subtitle by the time that the subtitle is displayed.

$$CPS = \frac{CharacterLength}{SubtitleTime(s)}$$

Using this method, each unit of speech will have their own CPS and $Char$ variables, which can be later used to determine the difficulty of the listening portion for each unit of speech.

Time Start(s)	Time End(s)	CPS	Chars	#	Lines
24.71	26.50	3.34	6.00	10	という体験は
28.03	31.13	3.87	12.00	12	非常に新鮮に写りました。
31.13	34.95	4.97	19.00	13	110万本の売り上げを記録した本作は、
48.58	53.24	4.29	20.00	18	卒業の日に校庭の外れに植えてある木の下で
53.24	57.64	5.00	22.00	19	結ばれたカップルは永遠に幸せになれると聞いた
57.64	62.50	4.11	20.00	20	主人公が卒業までに奮闘するという話です。
63.31	66.55	5.55	18.00	21	PC エンジンでリリースされた本作は
66.55	72.74	4.36	27.00	22	開発販売を担当したコナミの予想をはるか上回る結果を残し
72.74	77.55	4.79	23.00	23	一時は品切れになりプレミアム価格になりました。

Figure 4.6   Excerpt of unit of speeches and its CPS and $Char$ variables

4.2.2 Vocabulary

A unit of speech may consist of many vocabulary items. To understand the difficulty of the whole unit of speech, we can first break down the unit of speech into individual vocabulary. Then, for each vocabulary, the difficulty rating for can be calculated by $Difficulty = 6 - JLPTLevel$. (The difficulty rating will range from 1-5, with 5 being the hardest.)

8	大きな衝撃を与えました。				
	vocab	JLPT Level	Difficulty Rating (6 - JLPT Rating)	Occurance	Final Rating
	大きな	5	1	1	5
	衝撃	1	5	Max Rating	
	与える	3	3	5	

Figure 4.7   Hardest vocabulary class determines the difficulty rating

Once we have a list of vocabulary for each unit of speech, we can then choose the most difficult vocabulary as the most significant component that will affect

31

the difficulty of the unit of speech. The difficulty rating can then be determined by:

$$MaxRating = Max(DifficultyRating)$$

In addition, the difficulty level should also be influenced by the number of most difficult vocabulary, such that units of speech that contain more difficult words will be rated higher than those with less.

12 非常に新鮮に写りました。					
	vocab	JLPT Level	Difficulty Rating (6 - JLPT Rating)	Occurance	Final Rating
	非常	3	3		2	3.1
	新鮮	3	3	Max Rating	
	写る	4	2		3

Figure 4.8 Two occurrences of the highest rated vocabulary in a unit of speech

With two occurrences of the highest rated vocabulary, the difficulty of the unit of phrases should increase. For this experiment, a linear scaling function is used. The final rating for vocabulary for each unit of speech is determined by:

$$VocabularyRating = MaxRating + (Occurence - 1) \cdot Min(OccW, OccWMax)$$

Where $OccW$ represents the weight of the occurence. For the experiment and the above example, the weight is set to 0.1, with a $OccWMax$ maximum weight of 0.9. The final vocabulary rating is used to compare to other phrases in order to determine how difficult a unit of speech is.

Time Start(s)	Time End(s)	Vocab	#	Lines
18.72	21.90	3	8	大きな衝撃を与えました。
21.90	24.71	3	9	ゲームの中の女の子に恋をする
24.71	26.50	5	10	という体験は
26.50	28.03	5	11	ゲーマーたちの目に
28.03	31.13	3.1	12	非常に新鮮に写りました。

Figure 4.9 Excerpt of unit of speeches and its $Vocab$ variables

4.2.3 Final Difficulty

Although we have a method to derive CPS, $Char$ and $Vocab$ from unit of speeches, more work is required in order to transform these variables such that

32

they are comparable to each other in order to create further optimization and some degree of adaptability to the needs and ability of individuals.

The method used in the experiment is to first normalize the above variables, since they are measuring different types of units. The variables CPS, $Char$ and $Vocab$ of each unit of speeches are first subtracted by the minimum of a sample of unit of speeches. The values are not used with a minimum of 0 as subtitles must contain at least some characters and thus must have some minimum value CPS abd $Char$ above 0. In the case of the experiment, the video used in the experiment is used as the sample. With the updated variables, they are then normalized by linearly scaling the resulting values to a range of 0 to 1, with 0 being the lowest difficulty and 1 being the highest. The resulting variables $CPSD$, $CharD$ and $VocabD$ represents the normalized values for the above difficulty variables.

Vocab	Vocab-V	VocabD	CPS	CPS-CPS	CPSD	Chars	Chars-I	CharsD
3	2	0.5	5.34	3.77	0.60	17.00	11.00	0.28
3	2	0.5	3.00	1.43	0.23	6.00	0.00	0.00
4	3	0.75	3.30	1.73	0.28	10.00	4.00	0.10
3	2	0.5	4.29	2.72	0.44	20.00	14.00	0.36
3	2	0.5	5.00	3.43	0.55	22.00	16.00	0.41
5	4	1	4.11	2.54	0.41	20.00	14.00	0.36
3	2	0.5	5.55	3.98	0.64	18.00	12.00	0.31
5	4	1	4.36	2.79	0.45	27.00	21.00	0.54
3	2	0.5	4.79	3.22	0.52	23.00	17.00	0.44
5	4	1	7.55	5.98	0.96	45.00	39.00	1.00
5	4	1	5.17	3.60	0.58	32.00	26.00	0.67
4	3	0.75	6.33	4.76	0.76	26.00	20.00	0.51
5	4	1	5.00	3.43	0.55	25.00	19.00	0.49

Figure 4.10 Normalized variables $CPSD$, $CharD$ and $VocabD$.

We can then use the above normalized parameters to then find the final difficulty level of the units of speech within the video. A simple method to determine the difficulty level is through finding the average of the three variables.

$$FinalDifficultyUnoptimized = \frac{CPSD + CharD + VocabD}{3}$$

A linear scaling method can then be used to calculate the final difficulty level.

Table 4.6 Variance-covariance matrix of the three variables

ρ	Vocab	CPS	Char
Vocab	1		
CPS	-0.0758	1	
Char	0.2027	**0.7689**	1

However, if we check the correlation between CPS and $Char$, in which the variables $CPSD$ and $CharD$ are derived from, it can be seen that these two factors are highly dependant. Although there may be a combination of reasons, such as the combination of longer phrases are often spoken faster, and consistent pause time between phrases, using an average of the three variables would mean that difficulty level will be biased towards the listening difficulty of the units of speech. As such, additional weights $CPSW$, $CharW$ and $VocabW$ can be added in order to control and optimize the difficulty level. The coefficients from the weights are calculated through the below formula:

$$Coef = \frac{VariableWeight}{CPSW + CharW + VocabW}$$

Such that the sum of the coefficients will be equal to 1, such that the final difficulty rating will be in the range of [0,1]

$$CPSCoef + CharCoef + VocabCoef = 1$$

Thus, we can finally derive the final difficulty for every units of speech by scaling each factor by its corresponding coefficients.

$$FinalDifficulty =$$

$$CPSD \cdot CPSWCoef + CharD \cdot CharWCoef + VocabD \cdot VocabWCoef$$

4.2.4 Tasks

The use of the resulting difficulty levels can be used to prepare appropriate tasks for learners to used. Since difficulty level for a unit of speech have the range of [0,1], the level of difficulty can be split into equal parts to provide the learner more

Table 4.7 Coefficients used in this experiment

	Coefficient
$VocabWCoef$	0.4118
$CPSWCoef$	0.4118
$CharWCoef$	0.1765

Final	#	Lines
0.30	16	本作の目的は
0.44	17	同級生の早乙女好雄に
0.45	18	卒業の日に校庭の外れに植えてある木の下で
0.50	19	結ばれたカップルは永遠に幸せになれると聞いた
0.64	20	主人公が卒業までに奮闘するという話です。
0.52	21	PC エンジンでリリースされた本作は
0.69	22	開発販売を担当したコナミの予想をはるか上回る結果を残し
0.50	23	一時は品切れになりプレミアム価格になりました。
0.98	24	その後プレイステーション、セガサターンでは多くの恋愛シミュレーションがリリースされました。

Figure 4.11 Excerpt of lines 16 - 24 and their final difficulty levels

optimized exercises. For the experiment, three different tasks have been designed to make use of the aural and orthographic simulation of captioned videos.

Fill-in-the-blank Exercise The exercise requires students to listen to missing vocabulary that is removed from the subtitles. They are to fill in the subtitles, and once corrected, try to speak the words correctly three times.

```
Example:
結ばれるまでの_____を_____
Complete the sentence
Repeat the missing vocabulary three times.
```

Transcription Exercise A more difficult version of the above exercise. The student is required to fill in the entire unit of speech, or units of speech. They are also required to correctly translate the sentence afterwards.

```
Example:
------------------------------------------------------------
Transcribe the sentence.
```

Speaking Exercise The exercise focuses on the speaking aspect of the sentence. The sentence is written on the paper, and the student is required to speak aloud the sentence correctly. Special emphasis is placed on the intonation of the learner's speech.

```
Example:
それまで常識だった
Translate the sentence.
Repeat the above sentence three times.
```

The questions will be embedded within the video through the use of the subtitling software. Questions will be colored in red, and instructions will be written on the task sheet that is given to the learner. The level of the task sheet and the questions will change depending on the difficulty level chosen. Once the learner notices the red subtitles on the screen, they can be made aware that there they must finish a task before continuing.

Figure 4.12 Example of red text indicating a task to complete

4.2.5 Preparation Summary

Table 4.8 summarizes the process of preparing for the method in general, as well as the amount of time used to create the subtitles used for the test. The video for the test is 2 minutes and 30 seconds long. The video creation process is shared amongst both the experimental and control set-up, while the analysis and caption editing portion is exclusive to the experimental method. Overall, the method

requires an approximately 66.7% extra time taken over traditional same language subtitles.

Table 4.8 Time taken for each task

		Time Taken(s)
	Scripting	1.5
Video	Recording	0.5
	Captioning	1
	CPSD	0
Analysis	VocabD	1
	CharsD	0
Caption	Furigana	0.5
edit	Tasks	0.5

For the experiment, the vocabulary difficulty rating process requires each individual vocabulary from each unit of speech to be checked manually by hand. A process can be introduced such that vocabulary items can be processed through computers in order to automatically find the parameters required to find the difficulty level. The large number of homonyms, conjugations and no clear delimiters may cause the identification for each individual vocabulary for the Japanese language to be difficult and prone to error.

4.2.6 Experiment

For the experiment, the difficulty is split into 10 levels. Then, two levels of exercises, level 4 and level 7, have been chosen. As such, the targets for the units of speech will be around those level[8].

In a similar manner to the first experiment, participants will be split into two groups without their knowledge. The first group will perform the control test first, then the experimental test second. The second will have this procedure switched. The control video will be videos reused from the first experiment.

The participant will be given the instructions before the video begins. During

---

8   See Appendix for more information on the actual exercise used

Final	#	Lines
0.47890	3	結ばれるまでの過程を描いた
0.37832	8	大きな衝撃を与えました。
0.52893	10	という体験は
0.50472	15	数々のジャンル名を生み出しました。
0.30006	16	本作の目的は
0.34603	35	それまで常識だった
0.45816	39	彼女とともに永遠に過ごす事ができるという
0.35499	41	現在は恋愛シミュレーションは

Figure 4.13 Level 4 units of speech chosen for tasks

Final	#	Lines	
0.67088	1	恋愛シミュレーションとは	2
0.71072	11	ゲーマーたちの目に	2
0.59070	12	非常に新鮮に写りました。	2
0.74199	14	恋愛シミュレーション、ギャルゲー、美少女ゲームなど	2
0.69106	22	開発販売を担当したコナミの予想をはるか上回る結果を残し	2
0.76703	25	To Heartや痕などストーリーを重視した作品が発売され。。。	1
0.72971	28	恋愛シミュレーションは、据え置き型ゲーム機を中心に	1

Figure 4.14 Level 7 units of speech chosen for tasks

the video, the participant may pause the video at any time, rewind the video, and ask any questions that they might have regarding any language knowledge within the subtitles text. For the control test, the student is asked to learn from their video in their own way as much as possible, as if they are watching a commercial video with subtitles. For the experimental set-up, the student is ask to answer all the questions that was prepared using the materials as illustrated in either figures 4.13 or 4.14. They are allowed to choose the level of the test at the beginning, and are given the appropriate task sheet[9]. The participants are allowed once to restart the test if they feel the questions are too overwhelming for their current ability.

At the end of each video, they are asked five questions based on the content of the video, testing their newly acquired vocabulary knowledge. They will also be asked on what they thought about the learning method through a small series of questions. For the experimental set-up, both questions focused on vocabulary from the task and outside the task was asked, and the correctness of their answer was recorded. During the whole process, the participant and the control of the playback device are video recorded to allow for further observation and analysis.

Table 4.9 Test process for the two batch of participants

Batch 1	Batch 2
Control Video	Experimental Video
Vocab Test and Questionnaire	
Experimental Video	Control Video
Vocab Test and Questionnaire	

4.3. Cognitive Load Analysis

While the above mentioned parameters $Char$, CPS, and $Vocab$ are useful in determining the difficulty of the each units of speech, it is only a part of the equation for the system created. On the side of the user, it is important to

---

9   Refer to the appendices section for the contents of both level 4 and level 7 task sheets.

find ways to adapt the difficulty setting to the individual such that the most appropriate units of speech can be used alongside the method as explained in the above experiment. One possibility to determine the appropriateness of the current set difficulty level, and thus the units of speech used for the exercises, is to investigate the effects of the cognitive load of the user for different levels of difficulty as mentioned in the above section.

4.3.1 Objective

The cognitive load of a user watching subtitles can be shown by their eye movements on the displayed subtitles. Intuitively, an overly difficult subtitle for the user will results in a cognitive load that may cause a loss in focus of the reading of the subtitle in question. On the other hand, easy subtitles may not provide a high enough cognitive load for the user to provide an engaging learning experience. As such, cognitive load may be a useful indicator in determining if the appropriate subtitles are being used as tasks, and if such is not the case, the appropriate actions can be taken to change the behaviour of the model in order to attain more accurate results.

4.3.2 Experiment Format

An experiment was designed in order to investigate if the difficulty levels of the units of speech determined by the system have a correlation between the cognitive load of the user. An entirely new batch of participants will be requested to watch the videos that were used in the experiment section.

The participants will first be asked to watch the L2 subtitled video, while holding a pen to point at where their eye movement are during the whole process. The participants will be graded according to their ability to follow the subtitles accurately throughout the test. The participants are not told the true purpose of the task, and are asked to answer simple questions pertaining to the video after they finished watching. Unlike the above experiment, the user will not be allowed to replay previously watched sections, and are only able to pause the video between each subtitle. During the entire process, including the pointing movement of the participant, will be recorded for analyzed for the results of the experiment.

For analysis, the units of speech used in the above section will be split into 3 categories: Easy, normal and hard. The categories will be filled until the number of lines are equal for each category. Starting from the easiest, lines that have the lowest $FinalDifficulty$ rating will be included in the easy category. Similarly, starting from the lines that were not included in the easy difficulty category, the next easiest lines are grouped into the normal category, and then finally for the hard category.

Each time a participant incorrectly follows a displayed subtitle, that subtitle will be determined incorrect, and thus indicates an overly high cognitive load for the participants. The ratio of incorrect answers to total number of lines indicate that the difficulty setting may be ill-suited for the participants.

4.3.3 Results

A total of 13 people participated in the experiment. The percentage of incorrect answers for each category for each person was averaged out and the following was the result of the experiment.

Table 4.10 Ratio of incorrect readings over three categories

Difficulty	Incorrect Ratio
Easy	33.8
Normal	46.3
Hard	70.6

Overall, there seems to be a trend that indicates are user's level of understanding may be correlated to the difficulty of the subtitle that the individual perceives. However, using eye-tracking may not be a comprehensive enough test to determine a person's cognitively load during the test. As the format is in a video, it is different from purely read materials in which the participants may have different approaches to absorbing the visual information from the subtitles. A user does not need to read the text information according to the exact moment of the audio information, and there is no evidence to suggest that doing so would have any benefit over methods that do not do so. Users may not focus on the subtitles if they are confident in understanding the material, or users may quickly skim over

the text information before using their ears to absorb the aural information. It is difficult to guarantee that eye movement away from the subtitles are an absolute sign of cognitive load.

Nevertheless, since there is still a trend, rather than using the above information as a direct determinant for adapting the user level, it can be used as a sign to further pursue the question from the user if they have difficulty with the subtitles, and if so, the appropriate changes can be made to the system. The possible changes are to be further discussed in the discussion section of Chapter 5.

Chapter 5

Results and Discussion

5.1. Results

A total of 14 tests were conducted. The 14 participants were groups into 2 batches. The group that was shown the control experiment video first will be referred to as Batch A, while the other group that was shown the experimental video first will be referred to as Batch B. For the qualitative tests, questions were asked based on what the subjective opinions on the tests. In total, 6 questions were asked during the entire test, with Batch A participants and Batch B participants being handed different set of questions due to a different process.

5.1.1 Questionnaire Part 1

For the first part of the questionnaire, 3 questions were asked based on: How much they enjoyed using the method to learn, how easy it was to learn, and how much they thought they had learned using the subtitles.

The below charts illustrate the results. The questions in the chart were modified such that it will be easier for the reader to understand the chart. Whether the subtitles belonged to either control and experimental set-up were not disclosed to the participants during the test process.

Figure 5.1 shows the result when participants were asked about how they thought about learning though subtitles. The general sentiment of using subtitles as a learning tool is positive. Over 75% of the participants thought that the subtitles were at least a good alternative as a teaching method compared to traditional textbook methods.

The ease of use data from participants were also asked, and the results are displayed on figure 5.2. In general, the response for the ease of use for subtitles
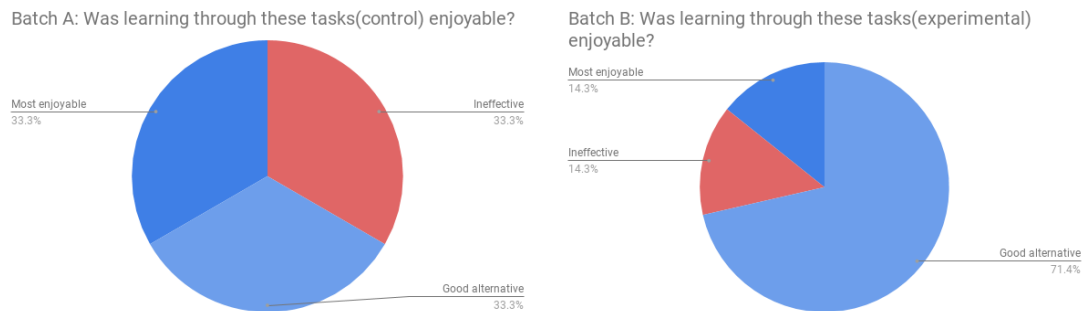
Figure 5.1 Enjoyment of participants for using subtitles
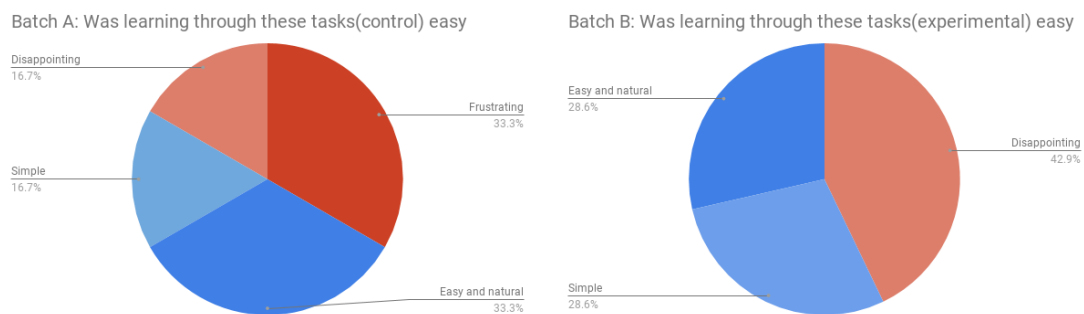
Figure 5.2 Ease of use of subtitles for learning purpose

for learning was neutral. With only a little over 50% of the people having positive feelings about the use of ease of use of subtitles, it seems the effort for learning across both subtitles and traditional teaching methods are similar.

5.1.2 Questionnaire Part 2

The next section of the questionnaire was asked when the participants had finished the second video. This part of the questionnaire also had 3 questions, with 2 focusing on what they felt about the second video relative to the first one. The comparison data is useful in confirming whether the new experimental set-up is successful in the eyes of the participants.

Batch A learners switching from the control video to the experimental video responded positively across the board. With no negative feeling towards the new method, it would seem that the experimental set-up has a positive effect on

44

Batch A: After control subtitles, how did you feel about experimental subtitles

No difference
16.7%

More enjoyable
83.3%
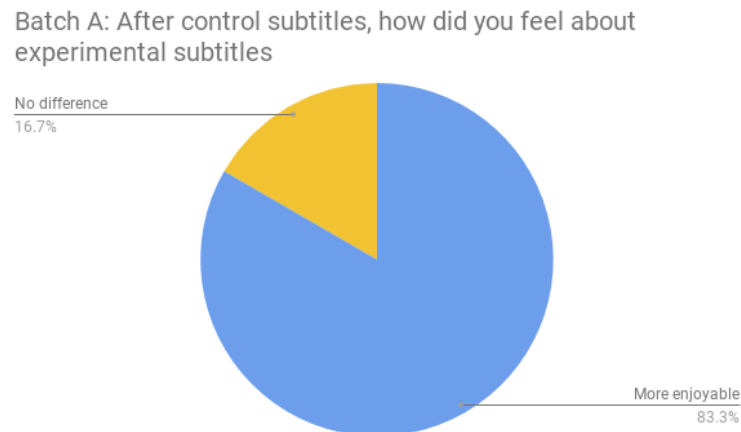
Figure 5.3 Batch A learners on the feeling of switching to experimental setup

encouraging and motivating people to learn on this platform.

Again, the questions have been modified in the paper in order to allow for the ease of understanding.

Batch B: After experimental subtitles, how did you feel about control subtitles

Much more annoying
14.3%

No difference
14.3%

More enjoyable
28.6%

More annoying
42.9%
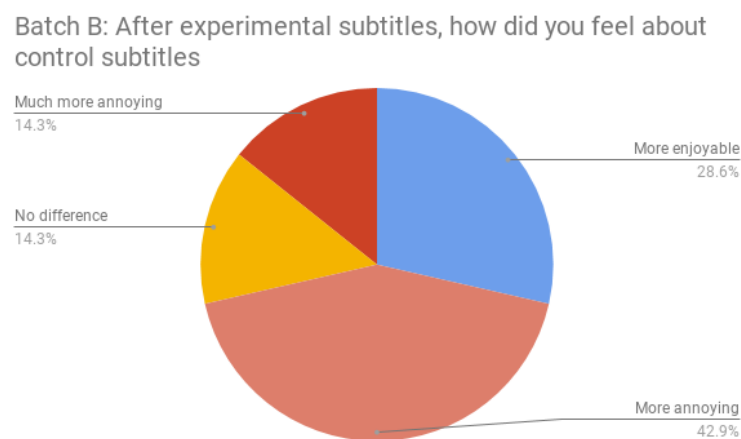
Figure 5.4 Batch B learners on the feeling of switching to control setup

Batch B learners switching from the experimental video to control video generally have negative experiences. Only a little over 25% of the participants agreed that they find the control subtitles more enjoyable, and that over 50% over the people agree they would prefer the experimental setup over the conventional sub-

titles.


Figure 5.5 Ease of use comparison for the second set-up for both batches

To reinforce the sentiment mentioned in figure 5.2, the data in figure 5.5 shows that for the second video, there is no consensus on whether the control subtitles or the experimental subtitles were easier to use. Opinions on the ease of use of the two different subtitles are split among users.

The amount that the student felt that they had learned were also recorded through the question of "How much did you think you learned through these subtitles?", with an answer range between 0 to 8. Both batches were asked this question twice, once after each test. The follow data is the result of the combination of the following procedure: For batch A, the first score is subtracted by the second score. For batch B, the second score is subtract by the first score. The distribution of figure 5.6 shows the sentiment of using the experimental subtitles over conventional L2 subtitles, with a positive score meaning that the participant thought that experimental subtitles are relatively more useful for learning new vocabulary, and a negative score being the opposite.

Table 5.1 Additional data on the subjective learning score distribution

Mean μ	St.Dev. σ_s
0.928571	0.828742

According to the the distribution and the derive data on table 5.1, and assuming normal distribution of scores, a random person would have **86.87%** chance[1]

---

1   z-score at sentiment $= 0$ is $\frac{0-0.928571}{0.828742} = -1.12046$

Sentiment for the experimental subtitles
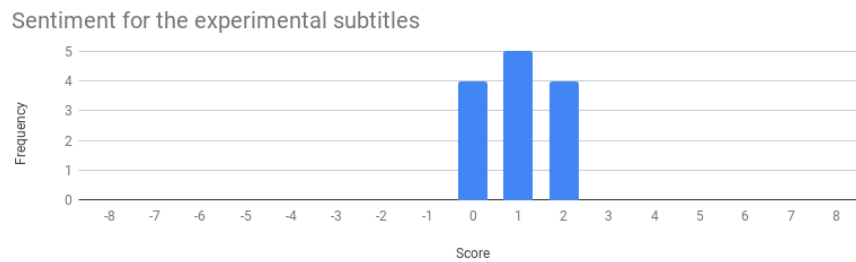
Figure 5.6 Distribution of the difference between learning

of considering the experimental subtitles to be more useful in learning when compared to conventional L2 subtitle models.

5.1.3 Test Scores

After each test, the participants were quizzed on the knowledge that was presented to them through the videos. Table 5.7 shows the distribution of the scores after the experimental test compared to the control test. A higher score means a better performance from the experimental test.

Scoring difference betweeen control and experimental (experimental - control)
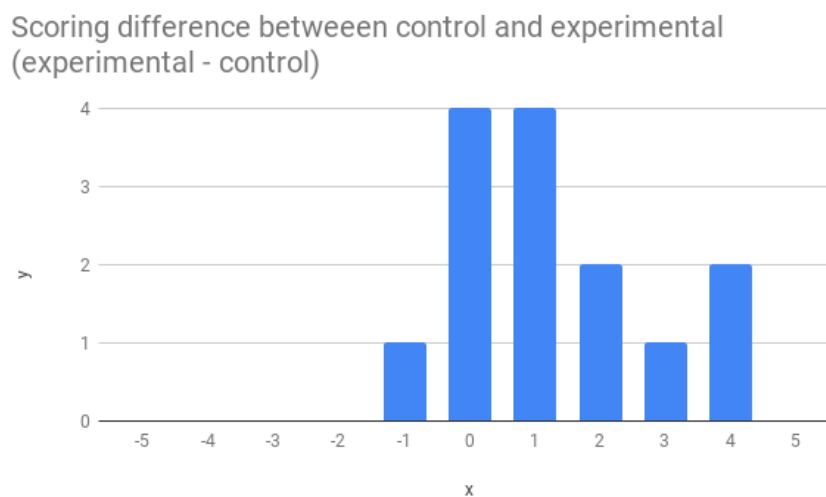
Figure 5.7   Distribution of the difference between scores for experimental and control

Assuming the distribution to be normally distributed, a random student taking the experimental subtitles will have a 82.86% chance[2] of scoring higher compared to the normal test.

Table 5.2 Additional data on the test score distribution

| Mean μ	St.Dev. σ_s
1.461538	1.540658

5.1.4 Conclusion

Generally, the experimental subtitles were well received compared to the conventional L2 subtitles, and there were notable improvements of the performance of the subtitles from the participants. The improvements were not at the cost of ease of use, compared to the trial experiment. Overall, many aspects, including autonomy, were build upon and improved from the trial experiment. The follow section will continue to showcase the flexibility of the model beyond what the learner can see.

5.2. Discussion

5.2.1 Sample Accuracy

With the experiment, the sample size in determining the difficult is 44 lines. Since the final difficulty level[3] are relative to combination the components of the system $CPSD$, $CharsD$, and $VocabD$, a larger sample size will allow the system to become more accurate. If multiple videos and their subtitles use the system, the difficulty level across the units of speech within these videos will be graded more consistently for the model. For the users, it would not be necessary to calculate their own language rating again, allowing more accurate results across videos.

---

2   z-score at score = 0 is $\frac{0-1.461538}{1.540658} = -0.948645$

3   Refer to Chapter 4 for more information for the formulas for calculating the final difficulty level.

5.2.2 Self-Correcting User Rating

While the experiments conducted currently only provides two level, Level 4 and Level 7, to participants, each individual can be given their own score through the model. An initial input of the score can be based on their JLPT level, but the a model can be designed such that system can automatically tune the individual's score towards the true user score relatively to the system.

A user that is rated and consistently doing well for all the tasks can have their rating increased until they are at a level that the tasks become appropriate for the user. The model can change the individual rating based on the accuracy of your answer. If a learner successfully completes tasks that are higher rated than users, they are then promoted to a higher ranking, and the system assigns harder tasks for the user. Should a learner fail tasks that are rated lower than their individual rating, their rating can then be tuned back down to provide tasks involving simpler units of speech.

A simple model to change to difficulty would be to find the sum of the user rating and the difficulty rating of the unit of speech, and divide by two to get the new score. If the user rating is lower than the unit of speech, and the user answers correctly, their score gets increased. If the user rating is higher than that of the unit of speech, but the user answers incorrectly, their score gets reduced.

$$UserRatingNew = U, UserRatingCurrent = U_0, UnitofSpeechDifficulty = D$$

$$AveragedScore \begin{cases} U = \frac{U_0+D}{2} & U_0 > D, UserCorrect = 0 \\ U = \frac{U_0+D}{2} & U_0 \leq D, UserCorrect = 1 \end{cases}$$

However, since the rating system cannot be 100% accurate. A large increase or decrease due to an inaccurate result may cause the system to fluctuate such that the rated user score will become further away from the true user score. To mitigate large increases, we can apply the follow systems to the equation such that dramatic changes in numbers can be reduced.

$$ArbitraryWeight = a[1, \infty)$$

$$LinearScaling \begin{cases} U = \frac{a \cdot U_0 + D}{a+1} & U_0 > D, UserCorrect = 0 \\ U = \frac{a \cdot U_0 + D}{a+1} & U_0 \leq D, UserCorrect = 1 \end{cases}$$

<div align="center">Or</div>

$$LogScaling \begin{cases} U = \frac{U_0 + D}{2} & U_0 > D, U_0 - D \leq 1, UserCorrect = 0 \\ U = U_0 - \sqrt{U_0 - D} & U_0 > D, U_0 - D > 1, UserCorrect = 0 \\ U = \frac{U_0 + D}{2} & U_0 \leq D, D - U_0 \leq 1, UserCorrect = 1 \\ U = U_0 + \sqrt{d - U_0} & U_0 \leq D, D - U_0 > 1, UserCorrect = 1 \end{cases}$$

5.2.3 Adaptability

One phenomenon observed during the test was that students already possessing abundant knowledge of kanji characters did very well during tasks involving reading the subtitles, but had trouble with the listening tasks. These types of learners are able to handle sections where aural and orthographic information were presented to them together (such as that of L2 subtitles). But when presented with tasks that challenged their aural reception, these learners spent more time on figuring out vocabulary and phrases that were not present in the subtitles, and often could not correctly complete the tasks. For these students, the system can be adapted such that there is more focus on listening tasks.

For tasks, priority could be given to create exercises that challenges the users listening ability. Once the system initially picks the appropriate difficulty subtitles, another pass could be done to specifically filter in subtitles with a higher CPS and $Char$, while eliminating those units of speech with difficulty rating highly influenced by $Vocab$[4]. For the choice of unit of speech, in addition to the condition that $IndividualLevel + a > UnitofSpeechDifficulty > IndividualLevel - a$ (where a is the range of acceptable units from the individual level) for a unit of speech to be chosen, we can also add the following condition:

$$CPS \cdot CPSCoef + Chars \cdot CharsCoef - Vocab \cdot VocabCoef > b[-1, 1]$$

---

4 $ListeningDifficulty = f(CPS, Char). Refer to Chapter 4 for more information.$

Where the variable b is the discrepancy of the level of listening and vocabulary skills. The larger the variable b is, the harder the listening tasks will be chosen. A negative number would mean that units with tougher vocabulary will be chosen instead.

Another approach is to modify the weights directly, thus affecting the whole difficulty of scoring system. By increasing the weights of $CPSW$ and $CharsW$, the whole difficulty choice will be become biased towards selecting tasks with higher CPS and $Chars$. However, since the individual's level remains constant, by using this method, an individual's level may be changed relative to the system, and may need be re-calibrated again through usage of the system in order to acquire the true score for the individual. An extreme change in the weight may initially provide tasks that are unsuitable for the user.

Since manual changes involve the decision making of the user or a supervisor, which may be biased or contain error, a self-correcting model can also be implemented for the weights. Similar to the above self-correcting system for the user level, a model can be implemented such that small changes can be applied to the weights in order to change the behaviour of the selection of units of speech as exercises such that the user score can remain relevant while slowing correcting the behaviour of the selection process to the most appropriate level. Overall, the steps to implementing such model would be to first have a method to identify if the user has failed either listening or vocabulary tasks, then have a method to adjust the weights, and repeat the process until a balance of difficult listening and vocabulary tasks to the user can be chosen by the system.

For the identification process, every time a student fails an exercise, the difficulty composition of the unit of speech can be compared in order to find the difficulty of the unit of speech. Similar to the above equation, the difference between the derived difficulty levels of $CPS + Char$ and $Vocab$ can then be used to adjust the weights.

$$ListeningBias = CPS \cdot CPSCoef + Chars \cdot CharsCoef - Vocab \cdot VocabCoef$$

A positive value for $ListeningBias$ indicates that the task the student incorrectly answered have a high difficulty towards listening elements. A slight adjustment to the weights can then be performed in other to increase the difficulty

number generated by the CPS and $Char$ values of a unit of speech such that the system will perceive listening elements as more difficult for the user, and thus choose easier listening tasks for the user to complete. A small increment to the weights $CPSCoef$ and $CharCoef$ and decrement to the weight $VocabCoef$ can be done to achieve such the above effect. The opposite is be true if the value for $ListeningBias$ is negative. We can also make use of the value of $ListeningBias$ and an arbitrary weight b in order to adjust the weights in small values.

$$ArbitraryWeight = b$$

$$ListeningBias > 0 \begin{cases} CPSCoef = CPSCoef_0 + b \cdot ListeningBias \\ CharCoef = CharCoef_0 + b \cdot ListeningBias \\ VocabCoef = CPSCoef_0 - b \cdot ListeningBias \end{cases}$$

$$ListeningBias < 0 \begin{cases} CPSCoef = CPSCoef_0 - b \cdot ListeningBias \\ CharCoef = CharCoef_0 - b \cdot ListeningBias \\ VocabCoef = CPSCoef_0 + b \cdot ListeningBias \end{cases}$$

Chapter 6
Conclusion

6.1. Summary

Learning through subtitles can be improved from conventional methods. Through thorough investigation and analysis of conventional subtitling techniques, and innovating with newer perspectives and implementation of new techniques, the model created in the paper provides an alternative to both conventional learning methods as well as L2 subtitles teaching methods that implements the research objectives as stated within the introduction chapter.

When correctly implemented and used, the model described within the paper allows for useful and appropriate learning material to be extracted within any type of media, and can be adapted and changed according to the needs of the individuals. Learners no longer need to worry about using appropriate material, and can instead focus on the learning through the material itself. Educators also do not need to worry about taking too much time to create materials from the video source, as the model allows for automatic identification of which materials can be used for good educational effect. From both perspectives, the model described within this paper creates a good alternative with many benefits at a relatively low cost in time and effort.

The method also seems to leave a similar impression for people that have tested both conventional subtitling techniques and the method described within this paper. For learners that prefer a more dynamic environment, the method may pique the interests of these learners without some of the downsides of conventional L2 subtitling methods, with the possibility of also improving learning efficiency over conventional methods if used correctly.

6.2. Future Works

Since learning language is a complex matter involving many different factors, the scope of the paper is only limited and focused on building a model for automation and adaptiveness to individuals. For the Japanese specific traits, factors such as kanji characters and grammar can also be considered to create a more accurate rating system. For all languages, the feasibility of implementing a similar system can also be considered, in addition to all the other traits of languages.

For adapting to the individual needs, the system could be tested long term on different individuals as a qualitative test in order to work out the most appropriate method of changing the different factors that rate the system, as well as implementing a global and shared database of different words in order to rank vocabulary levels that are specific to the system, rather than being dependent on human rating through JLPT materials and references as detailed by the method in the this paper. Creating a global database of different lines for different videos can help with accurate pinpoint the real difficulty of different units of speech through using the collective data to adjust individual subtitles.

References

Allen, Scott (2015) "A Brief History of Closed Captioning," *Mental Floss*.

Burczyńska, Paulina (2015) ""Reversed Subtitles as a Powerful Didactic Tool in SLA "," *Subtitles and Language Learning. Bern: Peter Lang*, pp. 221–244.

Charles, Tendai and Danijela Trenkic (2015) "Speech segmentation in a second language: The role of Bi-modal input," *Subtitles and language learning*, pp. 173–197.

Cutler, Anne and Sally Butterfield (1992) "Rhythmic cues to speech segmentation: Evidence from juncture misperception," *Journal of Memory and Language*, Vol. 31, No. 2, pp. 218–236.

Danan, Martine (2015) "Subtitling as a Language Learning Tool: Past Findings, Current Applications, and Future Paths," *InSubtitles and Language Learning. Principles, Strategies and Practical Experiences, ed. by Yves Gambier, Annamaria Caimi, and Cristina Mariotti*, pp. 41–61.

d᾽Ydewalle, Géry and Ubolwanna Pavakanun (1995) "Acquisition of a second/foreign language by viewing a television program," in *Psychology of Media in Europe*: Springer, pp. 51–64.

Enamorado, Sofia (2019) "U.S. Laws for Video Accessibility: ADA, Section 508, CVAA, and FCC Mandates," *3PlayMedia*.

Frellesvig, Bjarke (2010) *A history of the Japanese language*: Cambridge University Press.

Goh, Christine CM (2000) "A cognitive perspective on language learners' listening comprehension problems," *System*, Vol. 28, No. 1, pp. 55–75.

Karamitroglou, Fotios (1997) "A Set of Proposed Subtitling Standards in Europe," *en línia]. Dins:¡ http://www. accurapid. com/journal/04stndrd. htm¿[Consulta: 15 gener 2010].*

Kuppens, An H (2010) "Incidental foreign language acquisition from media exposure," *Learning, Media and Technology*, Vol. 35, No. 1, pp. 65–85.

Lertola, Jennifer (2015) "Subtitling in language teaching: Suggestions for language teachers," *Subtitles and language learning*, pp. 245–267.

Marzà, Anna and Gloria Torralba (2015) "Incidental language learning through subtitled cartoons: is it possible in a dubbing country," *Subtitles and Language Learning*, pp. 199–220.

Minucci, Michele Viana and Maria Silvia Cárnio (2010) "Movie subtitles reading skills of elementary school children," *Pró-Fono Revista de Atualização Científica*, Vol. 22, No. 3, pp. 227–232.

Myers, Erin (2017) "When Are Closed Captions Required by Law? Here's What You Need to Know."

Price, Karen (1983) "Closed-captioned TV: An untapped resource," *Matsol Newsletter*, Vol. 12, No. 2, pp. 1–8.

Romero-Fresco, Pablo (2009) "More haste less speed: Edited versus verbatim respoken subtitles.," *Vigo International Journal of Applied Linguistics*, Vol. 6.

Vanderplank, Robert (2015) "Thirty years of research into captions/same language subtitles and second/foreign language learning: Distinguishing between 'Effects of' subtitles and 'Effects with' subtitles for future research," *Subtitles and language learning*, pp. 19–40.

Zanón, Noa Talaván (2006) "Using subtitles to enhance foreign language learning," *Porta Linguarum: revista internacional de didáctica de las lenguas extranjeras*, No. 6, p. 4.

Appendices

A. Transcripts for the Videos Used in the Experiments

```
START    END      TEXT
11.83    15.12    こんにちは、A と いいます
15.12    18.54    こんにちは、B と いいます
19.12    21.83    A さん、旅行は好きですか？
21.83    23.77    旅行は好きですよ
23.77    26.09    どんなところに行かれますか？
26.34    28.43    最近は北海道に行った
28.10    29.56    北海道？
28.82    30.21    はい 行きました
30.21    31.23    いいですね
31.23    32.25    いいですね
32.25    33.45    広い大地に
33.45    34.51    広い大地
34.58    35.65    いいですね
35.65    39.33    沖縄とか暑いところとか\Nって好きですか？
39.03    40.76    沖縄は好きですね
40.76    40.83    本当にですか
40.83    44.24    うん。。。なんか
44.24    46.94    島っぽいじゃないですか そして、
46.94    49.14    あんまり日本の感じがしない
49.14    49.88    あ〜確かに
49.88    53.59    別文化だから、なんだっけ
53.59    55.95    シーサーとか
55.95    57.20    シーサーとかね
57.20    57.92    はい
```

57.92	60.23	なるほど
60.28	65.65	なんか、こう他に暑いところっていうと\n どんなところを連想しますか？
65.65	68.82	暑いところだと、アフリカとか
68.82	71.37	アフリカ、アフリカか
71.37	72.59	砂漠みたい
72.59	74.65	砂漠、サハラ砂漠とか
74.65	75.49	はい
75.65	76.94	行ったことありますか？
76.99	78.22	ないですね
78.22	79.03	ないですか
79.03	85.76	なんか、足ん... 靴も入らなくて そのまま砂の上立ちたいね
85.76	87.82	あ〜いいねいいね
87.82	89.51	砂漠か
89.51	93.13	砂漠一回以前、あの、エジプトに行ったことがあって
93.13	94.51	え、どうだった？
94.56	100.63	あのね、砂漠はね、すごかった でもなんも、なんも、焼けるような思いがして
100.63	101.37	えへへ
101.37	102.78	暑くて
102.78	104.00	皮が抜けそう
104.00	105.56	そうそう、皮が抜けそう
105.56	106.60	皮膚が抜けそう
106.60	107.87	そう、だから
107.87	110.97	エジプトの人たちみんな白い服でおお
110.97	113.17	おお、覆うように服着たから
113.17	115.97	実はそんなになんていうか
116.02	118.47	えっと
118.47	120.07	少ない服？
120.07	125.05	ぬ、布の少ない服を着ないほうがいい
125.05	126.48	っていう説がある
126.48	127.45	そうそう、そうそうそう
127.45	129.72	結構なんていうか全部着て
129.72	131.74	でもなんか風が通るように
131.74	135.09	スカスカの感じの服が方が
135.09	138.61	良さそうって聞いたことがあるんだけど

```
138.61   143.96   なんか、い、一回エジプトの旅行の人に聞いたら、
143.96   147.08   そう、なんか同じようなこと言ってました\N
147.08   152.96   なんか東南アジアだったら割とサンダルとか短パンとか
152.96   155.74   Tシャツとか履いてるイメージなんだけど、
155.74   157.43   本当に暑い時、
157.43   160.86   ち、ところは逆だったというね
160.86   164.72   そうそうね、そんなそんなことしてらんないっていうね
164.72   169.28   え～、やけどにしちゃうみたいな
```

```
START    END      TEXT
4.58     9.81     え～Aさん、 次のところに行っていい、\n参りたいと思うんですが
9.81     12.43    次は休日について
12.43    13.87    お伺いしたいと思います
13.87    16.48    休日ですが。。。{\c&HH0000FF&}はい
16.18    22.06    Aさん、普段休日はどういった風に過ごされますか？
22.06    24.14    どういった風に？
24.14    25.28    何をされてますか？
25.28    28.84    最近は、寝るの方が多いですね
28.84    29.81    なるほど
29.81    30.79    はい
30.79    33.15    それはやはり疲れてるからでしょうか？
33.15    37.22    疲れるもありますし
37.22    44.14    あとは、うんなんか、外に出たら何をするのか
44.14    47.15    うんと、わからなくて
47.15    51.60    あと人多いところも行きたくない\N
51.60    53.29    あ～わかります
53.29    55.39    私もあの人混みが苦手なので
55.39    56.76    そうなんですよね
56.76    58.31    はい、あの。。。
58.31    63.03    うちの、私の使ってる沿線から
71.13    75.97    私の使ってる沿線から、横浜駅が近いんですけど
75.97    79.24    横浜駅の中央口に出ると
79.24    80.95    ものすごく人がいて
80.95    84.42    小さい頃から苦手でして
84.42    91.20    人混みをこううまく遠ざける方法はないんだろうか
91.20    94.49    とつい疑問に思っております
```

94.49   97.64   まさ、まさにその通りで

97.64   100.00   本当になんというか

100.00   104.70   最近友達がよく日本にきて旅行にっとかちょっと、

104.70   109.40   案内を色々しているので

109.40   112.52   浅草のなんか

112.52   113.43   あ〜、浅草寺

113.43   118.29   浅草寺とか行ったら、もう本当に

118.29   119.68   人混みで

119.68   120.67   そうですよね

120.67   124.88   あえてちっちゃい道に入る。。。

124.88   126.55   （入る）場合もありますね

126.55   127.94   なるほど

126.55   127.94   はい

127.94   129.51   わざとその

129.51   133.98   人混みのなんていうんでしょう、\N ショッピング、商店街とか

133.98   136.97   商店街とか、ショッピングの店とかに

136.97   140.21   入りたくないのと思ってて

142.01   146.30   それはあの非常に理解できますが

146.53   151.27   わざわざ日本に来て、わざわざ\N なぜ人混みを楽しむのかっていう

151.27   156.46   なぜ全部外国人が集まっている場所に

156.46   158.19   行くんのですかね

158.19   161.23   そういう考えもちょっと

161.23   164.28   そうですよね、ありますよね

164.63   169.10   私も、まぁ、外国人の友人と出かける際に

169.10   173.22   まぁ、彼らがここに行きたいあそこに行きたい言うんで

173.22   174.26   連れて行くんですけど

174.26   176.11   実際に行ってみると

176.11   177.82   外国人だらけだと言う

177.82   178.66   そうなんですね

178.66   180.67   日本は一体どこにいるのかって

180.67   183.36   えっとね、この間、たまたま、あっと

183.36   188.22   ゴールデンウィークで行ってましたので

188.22   191.46   なんか多分地方とかの

191.46   196.13   お客さんもいっぱい入ってたから

196.13   200.90   まぁそれそれなりには日本人の方も多いです

200.90   202.25   なるほど

203.29	207.78	やはり、あの中国人の方は結構 日本に来られますか
207.78	211.37	世界のあっちこっちは全部中国人です
211.74	214.72	世界の隅まで中国人です
214.72	216.90	おっしゃる通りです、おっしゃる通りです
216.90	221.99	あの２年ほど前にイギリスロンドンに行ったんですけど
221.99	224.21	やはり中国の方だらけでした
225.60	229.28	そうですよね、まぁ中国人口は世界のなん
229.28	233.98	四分の１？五分の１？ぐらいですね
233.98	234.63	確かにそう
234.49	238.89	４人の中は一人一人中国人
238.89	240.37	なるほど
240.51	242.62	これから公用語とかって
242.62	245.86	やはり中国語になるって話が出てますけど
245.86	247.08	どう\N 思われますか？
247.08	247.82	こうよう？
247.82	248.82	公用語
248.82	251.13	公用語？
251.13	252.43	世界で共通に話されてる
252.43	255.02	今英語が話される
255.02	260.32	うんと、 中国語は難しいなので
260.32	264.14	もう無理やり世界のみんな一緒に
264.14	267.25	中国語を勉強しましょうみたいな
267.25	270.35	自分は中国人としても言わないですね
270.35	272.66	ハハ、はるほど、はい
272.66	276.94	そうなんですね、わかりました、ありがとうございます

START	END	TEXT
1.76	4.12	恋愛シミュレーションとは
4.12	6.56	意中の彼女に告白し
6.56	9.05	結ばれるまでの過程を描いた
9.05	11.52	シミュレーションゲームです。
11.52	14.76	１９９４年に発売された
14.76	16.90	「ときめきメモリアル」は
16.90	18.72	日本のゲーム業界に
18.72	21.90	大きな衝撃を与えました。

21.90	24.71	ゲームの中の女の子に恋をする
24.71	26.50	という体験は
26.50	28.03	ゲーマーたちの目に
28.03	31.13	非常に新鮮に写りました。
31.13	34.95	１１０万本の売り上げを記録した本作は、
34.95	39.70	恋愛シミュレーション、ギャルゲー、美少女ゲームなど
39.70	42.88	数々のジャンル名を生み出しました。
43.54	45.54	本作の目的は
45.54	48.58	同級生の早乙女好雄に
48.58	53.24	卒業の日に校庭の外れに植えてある木の下で
53.24	57.64	結ばれたカップルは永遠に幸せになれると聞いた
57.64	62.50	主人公が卒業までに奮闘するという話です。
63.31	66.55	PC エンジンでリリースされた本作は
66.55	72.74	開発販売を担当したコナミの予想をはるか上回る結果を残し
72.74	77.55	一時は品切れになりプレミアム価格になりました。
78.61	84.57	その後プレイステーション、セガサターンでは多くの恋愛シミュレーションがリリースされました。
85.47	91.67	To Heart や痕などストーリーを重視した作品が発売され。。。
91.67	95.78	始めたことをきっかけに、恋愛シミュレーションではなく
95.78	100.78	恋愛アドベンチャーゲームと呼ばれるものもあります。
102.05	106.97	恋愛シミュレーションは、据え置き型ゲーム機を中心に
106.97	109.28	多く発売されましたが
109.28	111.71	２０００年代後半に入ると
111.71	113.80	ニンテンドー DS や
113.80	119.18	Playstation Portable などの携帯機でも発売されるようになりました。
120.16	122.63	中でもニンテンドー DS で
122.63	126.18	２００９年に発売されたラブプラスは
126.18	128.76	それまで常識だった
128.76	134.80	告白して終了するという固定概念を打ち砕きました
135.84	140.28	女の子に告白されるだけでは終わらず
140.28	142.88	その後の高校生活を
142.88	147.40	彼女とともに永遠に過ごす事ができるという
147.40	151.41	まさに終わりがない夢のようなソフトです。
151.41	154.31	現在は恋愛シミュレーションは
154.31	157.00	シミュレーションゲーム中心ではなく

157.00   161.48   デジタルノベルのような形式のものが多く
161.48   165.30   を占めます。

Table A.1 Subtitles Statistics

	Time	VocabLevel	CPS	CharacterTotal
Mean	2.96	1.96	3.42	11.25
StDev	0.45	0.45	1.98	6.47
Var	0.21	0.21	3.94	41.89

B. Task Sheet Example

Name:

Level 4 Task Sheet
Read the instructions carefully before watching the video. A video of 3 minute length w

Q1. Fill in the blanks.
結ばれるまでの_____を_____.
Complete the sentence
Repeat the missing vocabulary three times.

Q2. Speaking Practice
大きな衝撃を与えました。
Translate the sentence.
Repeat the above sentence three times.

Q3. Transcription Practice

Transcribe the sentence.

Q4.Fill in the blanks.
_____のジャンル名を_____。
Complete the sentence
Repeat the missing vocabulary three times.

63

```
Q5.Transcription Practice
------------------------------------------------------------
Transcribe the sentence.


Q6.Speaking Practice
それまで常識だった
Translate the sentence.
Repeat the above sentence three times.


Q7.Fill in the blanks.
彼女とともに_____事ができるという
Complete the sentence
Repeat the missing vocabulary three times.


Q8.Speaking Practice
現在は恋愛シミュレーションは
Translate the sentence.
Repeat the above sentence three times.


Name:


Level 7 Task Sheet
Read the instructions carefully before watching the video. A video of 3 minute length w


Q1. Transcription Practice
------------------------------------------------------------
Transcribe the sentence.


Q2. Transcription Practice
------------------------------------------------------------
Transcribe the sentence.


Q3. Transcription Practice
------------------------------------------------------------
Transcribe the sentence.
```

```
Q4. Transcription Practice
------------------------------------------------------------
Transcribe the sentence.
```

```
Q5. Fill in the blanks.
```
To Heart や痕などストーリーを_____が_____。。。
```
Complete the sentence
Repeat the missing vocabulary three times.
```

```
Q6. Fill in the blanks.
```
恋愛シミュレーションは、_____を_____に
```
Complete the sentence
Repeat the missing vocabulary three times.
```