

University of Tartu
Faculty of Social Sciences
Institute of Education
Curriculum of Educational Technology

Shintaro Horikawa
Effects of Using Audio-visual Materials on the Listening Performance of Secondary EFL
Learners in Japan
Master's thesis

Supervisor: Associate Prof. Katrin Saks

Tartu, 2024

Abstract

Effects of Using Audio-visual Materials on the Listening Performance of Secondary EFL Learners in Japan

This study investigates the impact of audio-visual materials on the listening performance of Japanese secondary EFL learners. Using a quasi-experimental design, 81 first-year high school students were divided into three groups, each exposed to different stimuli: hypervideo, video, and audio. Pre- and post-tests measured the effectiveness of these materials. Results showed no statistically significant improvement in listening comprehension for any group, though the audio group demonstrated the closest approach to significance. These findings suggest that while audio-visual materials may not significantly enhance short-term listening skills, audio-only resources might be more effective for test preparation. The study highlights the need for further research on long-term effects and multimodal integration in EFL education.

Keywords: audio-visual materials, listening performance, EFL learners, multimodal learning

Table of contents

Abstract.....	2
Introduction.....	4
Theoretical Overview.....	5
1. Multimodality in Language Learning and Teaching.....	5
2. Multimodal Approach to Teaching Listening Skills.....	7
3. Audio-Visual Materials in EFL Studies.....	9
4. Relevance to the Japanese Secondary Education Context.....	10
Methods.....	12
1. Research Design.....	12
2. Participants.....	13
3. Materials and Instruments.....	14
3.1. The pre-and post-tests.....	14
3.2. The interventions.....	15
3.2.1. Video.....	16
3.2.2. Hypervideo.....	16
3.2.3. Audio.....	16
4. Procedure.....	17
5. Data Collection and Data Analysis.....	17
6. Ethical Considerations.....	18
Results.....	19
1. Results of the Pre-test for all three groups.....	19
2. Results of the Post-test for all three groups.....	20
3. Comparison of the Pre-and Post-tests of Group A (Hypervideo).....	21
4. Comparison of the Pre-and Post-tests of Group B (Video).....	22
5. Comparison of the Pre-and Post-tests of Group C (Audio).....	23
Discussion.....	24
1. Overview of Findings.....	24
2. Hypervideo Group.....	24
3. Video Group.....	24
4. Audio Group.....	25
5. Comparison with Expectations.....	25
6. Implications for EFL Education.....	25
6.1. Test and Training Alignment.....	25
6.2. Real-World Listening Skills.....	26
6.3. Multimodal Integration in Testing and Training.....	26
6.4. Future Research Directions.....	26
Acknowledgments.....	27
Author's declaration.....	28
References.....	29
Appendices.....	31

Introduction

In the evolving landscape of global communication, proficiency in the English language has become a pivotal asset for individuals worldwide, particularly in non-English speaking countries like Japan. With the advent of technological advancements, there is a compelling opportunity to revolutionize English as a Foreign Language (EFL) education. Mobile and digital technologies offer unprecedented possibilities for enhancing learning methodologies and outcomes, underscoring technology's potential to enrich EFL education significantly.

Despite this potential, the application of technology within Japanese secondary education, especially in English language learning, remains limited. Traditional methods have predominantly focused on audio and written materials, overlooking the enriched learning environment that multimedia resources can offer. Mayer's (2009) cognitive theory of multimedia learning provides robust evidence that audio-visual materials, which combine visual and auditory stimuli, can significantly facilitate deeper learner engagement and comprehension. However, the utilization of such resources in Japanese EFL contexts has been underexplored.

Existing literature extensively documents the benefits of multimedia in language learning, highlighting its positive impact on listening comprehension. Yet, there is a conspicuous gap in understanding the specific effects of various types of audio-visual materials on the listening performance of EFL learners in Japan. This gap is notable given the increasing accessibility of technology and multimedia resources in educational settings globally. The lack of empirical evidence on the comparative effectiveness of audio, video, and hypervideo materials in enhancing listening comprehension among Japanese secondary school students points to a significant area for exploration.

This research endeavors to bridge this gap by investigating the effects of different types of audio-visual materials on the listening performance of secondary EFL learners in Japan. By examining the distinct impacts of audio, video, and hypervideo formats, the study aims to find out the most effective methodologies for incorporating multimedia in English language instruction, thereby enhancing the overall quality and efficacy of EFL education in the country.

Research Questions

Which type of audio-visual is the most effective in improving EFL learners' listening skills?

- A. Audio
- B. Video
- C. Hypervideo

The outcomes of this study are anticipated to have broad implications for a wide array of stakeholders in the Japanese EFL domain, including learners, educators, and material developers. For learners, understanding the differential impact of multimedia types on listening performance can guide more personalized and effective learning strategies. Educators and curriculum developers might use these insights to design more engaging and pedagogically sound instructional materials, ultimately fostering a more enriching learning environment for EFL students. Moreover, this research could spur further academic inquiry into the integration of technology in language learning, paving the way for innovative educational practices that resonate with the digital generation's learning preferences.

Theoretical Overview

1. Multimodality in Language Learning and Teaching

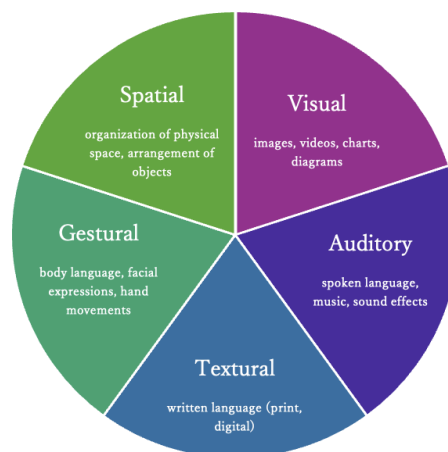
When considering the use of multimedia in language learning, delving into the concept of multimodality is inevitable. The term refers to the use of multiple modes or methods of communication and expression within a single interaction or text. In the context of language learning and teaching, it encompasses various forms of representation and expression beyond just written and spoken language. These modes include visual, auditory, gestural, spatial, and textual elements, which work together to create a richer and more effective learning experience. (Kress, 2003)

In language learning and teaching, multimodality is defined as the integration of different semiotic resources to enhance communication and comprehension. This approach acknowledges that language learners make meaning not only through linguistic input but also through visual, auditory, and kinesthetic cues. For example, when teaching vocabulary, a teacher might use images, sounds, and physical gestures to help students grasp the meaning and usage of new words. This multimodal approach aligns with contemporary

educational practices that recognize the diverse ways in which students learn and interact with information (Rosewell & Collier, 2017).

The theoretical frameworks underpinning the concept of multimodality include social semiotics and the pedagogy of multiliteracies. Social semiotics, developed by Michael Halliday, emphasizes that meaning-making involves a variety of semiotic resources beyond language, such as images, gestures, and spatial arrangements. This theory suggests that all forms of communication are multimodal because they draw on multiple resources to convey meaning (Halliday, 1978). The pedagogy of multiliteracies, introduced by the New London Group, extends this idea by highlighting the need for educational practices to embrace multiple modes of literacy in order to address the diverse ways students interact with and produce knowledge in a digital and globalized world (New London Group, 1996). This framework advocated for a shift from traditional, text-based teaching methods to approaches that integrate visual, auditory, and kinesthetic modes of learning (Jewitt, 2008). Different modes involved in multimodal learning include visual, auditory, textual, gestural, and spatial elements, as illustrated in Figure 1. Visual modes might involve the use of images, videos, charts, and diagrams to represent information. Auditory modes include spoken language, music, and sound effects, which can enhance the understanding of concepts through listening. Textual modes involve written language, including both print and digital text, which remains a fundamental component of literacy. Gestural modes include body language, facial expressions, and hand movements, which can support communication and comprehension, especially in language-learning contexts. Spatial modes involve the organization of physical space and the arrangement of objects, which can impact how information is perceived and understood (Kress, 2003; Jewitt, 2008)

Figure 1. *The five modes of multimodal learning*



The interaction of these modes enhances learning by providing multiple entry points for understanding and engaging with content. For instance, combining visual and auditory information can create a more comprehensive learning experience, as students can see and hear the information simultaneously, reinforcing their comprehension through dual coding. Moreover, the use of gestures and spatial arrangements can help learners internalize concepts through physical movement and spatial reasoning. This multimodal interaction supports deeper cognitive processing, as learners are able to connect and integrate information across different modes, leading to improved retention and understanding (Moreno & Mayer, 2007). However, it is important to note that while multimodal learning can enhance comprehension through the integration of various sensory inputs, it can also introduce extraneous cognitive processing if not carefully designed. Extraneous cognitive processing refers to the mental effort required to deal with elements that do not directly contribute to learning, often caused by poorly designed instructional materials. According to Mayer (2009), this type of cognitive load can hinder learning by diverting cognitive resources away from essential and generative processing (Mayer, 2009).

In addition to the interaction of different modes, it is crucial to consider the modality principle as posited by Mayer (2009). The modality principle states that people learn better when information is presented through both auditory and visual channels rather than overloading a single channel. For instance, presenting words as spoken text alongside visuals, rather than as printed text, can reduce cognitive overload and enhance learning outcomes.

In summary, multimodality in language learning and teaching involves the use of various semiotic resources to create richer and more effective educational experiences. Grounded in social semiotics and the pedagogy of multiliteracies, this approach recognizes the importance of integrating visual, auditory, textual, gestural, and spatial modes to enhance communication and comprehension. By leveraging the interaction of these modes, educators can support deeper learning and cater to the diverse needs of students in today's digital and globalized world (Bouchev, Castek, & Thygeson, 2021).

2. Multimodal Approach to Teaching Listening Skills

The multimodal approach in teaching listening skills involves integrating various modes of communication, such as visual, auditory, textual, and gestural elements, to enhance learners' understanding and engagement. This approach acknowledges that listening is an auditory activity and involves cues, contextual information, and interactive elements. In a

classroom setting, this might involve using videos, interactive activities, visual aids like charts and diagrams, and even body language to support and enhance the listening process. The goal is to create a richer, more immersive learning environment that caters to different learning styles and helps students develop their listening skills more effectively (Kress, 2003).

The multimodal approach differs significantly from traditional listening instruction methods, which often focus solely on listening to audio recordings or live speech without additional sensory inputs. Traditional methods typically involve listening to dialogues, lectures, or other spoken materials followed by comprehension questions or discussions. In contrast, the multimodal approach integrates visual aids, interactive activities, and sometimes even physical movement to engage learners more holistically (Guichon & McLoran, 2008). For example, using videos, interactive whiteboards, and multimedia presentations can provide context and visual support that traditional audio-only methods lack, making the learning process more dynamic and effective (Mayer, 2009).

The benefits of using a multimodal approach for listening skills development are manifold. First, it caters to diverse learning styles, ensuring that all students have the opportunity to engage with the material in a way that suits them best. Visual learners, for instance, benefit from seeing images or videos related to the audio content, while kinesthetic learners benefit from interactive activities that involve movement or hands-on tasks. Additionally, the multimodal approach helps to reinforce learning as students are exposed to the same information through different channels, which aids in retention and understanding. This approach also promotes critical thinking and problem-solving as students learn to interpret and integrate information from various sources, making them more adept at processing complex information in real-world scenarios (Jewitt, 2008).

Theoretical and empirical evidence supports the effectiveness of multimodal approaches in language learning. The Cognitive Theory of Multimedia Learning (CTML) posits that learners construct knowledge more effectively when information is presented in both verbal and visual forms (Mayer, 2009). Empirical studies have also demonstrated the positive impact of multimodal instruction on listening comprehension and overall language proficiency. For example, Al-Seghayer (2001) found that learners using multimedia annotations while listening to foreign language texts showed significantly better comprehension than those relying solely on audio input. Similarly, research by Jones and Plass (2002) highlighted the effectiveness of multimedia glosses in improving listening skills and vocabulary acquisition. These studies underscore the benefits of integrating

multimedia elements, enhancing listening instruction's effectiveness, and supporting robust language learning outcomes.

In conclusion, the multimodal approach to teaching listening skills offers a comprehensive and engaging alternative to traditional listening instruction methods. Integrating various sensory inputs and media caters to diverse learning styles, enhances comprehension and retention, and is supported by theoretical frameworks and empirical research. This approach not only makes listening activities more dynamic and interactive but also provides a robust framework for developing listening skills in a more inclusive and effective manner.

3. Audio-Visual Materials in EFL Studies

Audio-visual materials in EFL (English as a Foreign Language) contexts are resources that combine sound and visual components to enhance language learning. These materials include videos, films, television programs, podcasts, and multimedia presentations. They are designed to provide learners with authentic language input and to simulate real-life communication situations. Audio-visual materials are valued for their ability to present language in context, showcasing not only the spoken word but also non-verbal cues such as gestures, facial expressions, and body language, which are critical for understanding meaning in communication (Bajrami & Ismaili, 2016).

Various types of audio-visual materials are commonly employed in EFL classrooms to support language learning. These include educational videos, documentaries, TV shows, movies, and animated content. Interactive multimedia, such as language learning software and online platforms that integrate audio and visual elements are also widely used. Teachers often utilize songs, podcasts, and video clips from platforms like YouTube to engage students. Hypervideos, which are interactive videos that allow learners to control navigation and access additional materials through hyperlinks, have also been gaining traction for their ability to provide a more dynamic learning experience (Bal-Gezegin, 2014; Sauli, Cattaneo, & van der Meij, 2018)

Audio-visual materials support language learning by providing rich, contextualized input that can enhance listening skills. These materials allow learners to hear authentic language use, including various accents, intonations, and speech speeds, which are essential for developing listening comprehension. For example, videos can illustrate conversational norms, cultural references, and pragmatic aspects of language that are difficult to convey through text alone. The multimodal nature of these materials—combining auditory and visual stimuli—helps learners to process information more effectively, as they can see and

hear the language being used simultaneously, which aids in retention and understanding. Hypervideos, in particular, offer interactive features such as annotations and quizzes that facilitate deeper engagement and reflection on the content (Akbari & Razavi, 2016; Sauli, Cattaneo, & van der Meij, 2018)

The use of audio-visual materials in EFL education offers several advantages. They can increase student motivation and engagement by making lessons more enjoyable and relevant. Audio-visual materials also provide opportunities for learners to experience authentic language use, which can improve their communicative competence and cultural awareness. These materials can also cater to different learning styles, supporting visual and auditory learners. However, there are potential drawbacks, such as the need for significant preparation time and resources to select and integrate appropriate materials. Moreover, some learners might find it challenging to keep up with the pace of authentic audio-visual content, leading to frustration. Teachers must, therefore, carefully select and tailor these materials to suit their students' proficiency levels and learning needs (Akbari & Razavi, 2016; Bajrami & Ismaili, 2016)

By integrating audio-visual materials thoughtfully, educators can enhance the language learning experience, making it more interactive and effective. These materials help bridge the gap between classroom learning and real-world language use, preparing students to communicate effectively in diverse contexts.

4. Relevance to the Japanese Secondary Education Context

The integration of audio-visual materials in Japanese secondary education aligns seamlessly with the curriculum's focus on developing comprehensive language skills—listening, speaking, reading, and writing. The Ministry of Education, Culture, Sports, Science and Technology (MEXT) has been promoting the use of Information and Communication Technology (ICT) in education, emphasizing its importance in creating dynamic and interactive learning environments. The new Course of Study, revised in 2020, underscores the need to cultivate “information utilization skills,” which involves effectively using ICT tools across various subjects. By exposing students to authentic language input and real-life communication contexts, audio-visual materials support these educational goals, enhancing students' listening and speaking abilities within diverse linguistic and cultural contexts. (MEXT, 2018)

Japanese secondary EFL learners face specific needs and challenges that audio-visual materials can address effectively. One significant challenge is the limited exposure to

English outside the classroom, which hampers students' listening and speaking proficiency (Yashima, 2019). Audio-visual materials can provide continuous exposure to native and fluent English speakers, helping students familiarize themselves with different accents, intonations, and speech patterns. Additionally, Japanese students frequently exhibit high levels of anxiety and low confidence when speaking English. This is largely due to their limited opportunities for practice, which exacerbates their fears of making mistakes and being misunderstood (Elliott & Vasquez, 2021). The use of videos and interactive media can create a more engaging and less intimidating environment for practicing listening and speaking skills, thereby boosting students' confidence and motivation.

Audio-visual materials have been implemented in Japanese EFL classrooms in several innovative ways, reflecting both teacher creativity and institutional support. Teachers utilize various tools, such as videos and interactive software, to enhance the learning experience. For instance, free software like Audacity for audio editing and VideoStudio for video editing are used to create and modify teaching materials (Akase, 2021). Schools have integrated interactive whiteboards and language labs with multimedia capabilities, providing a more engaging and immersive environment for students (MEXT, 2020). Moreover, blended learning models have been adopted, combining traditional classroom instruction with online platforms that offer audio-visual content and interactive exercises. This approach not only supplements textbook material but also provides contextually rich language input, making learning more relevant and effective (MEXT, 2020). Overall, the implementation of audio-visual materials in Japanese EFL classrooms demonstrates a growing recognition of the benefits of multimedia in enhancing language learning outcomes (Akase, 2021; MEXT, 2020).

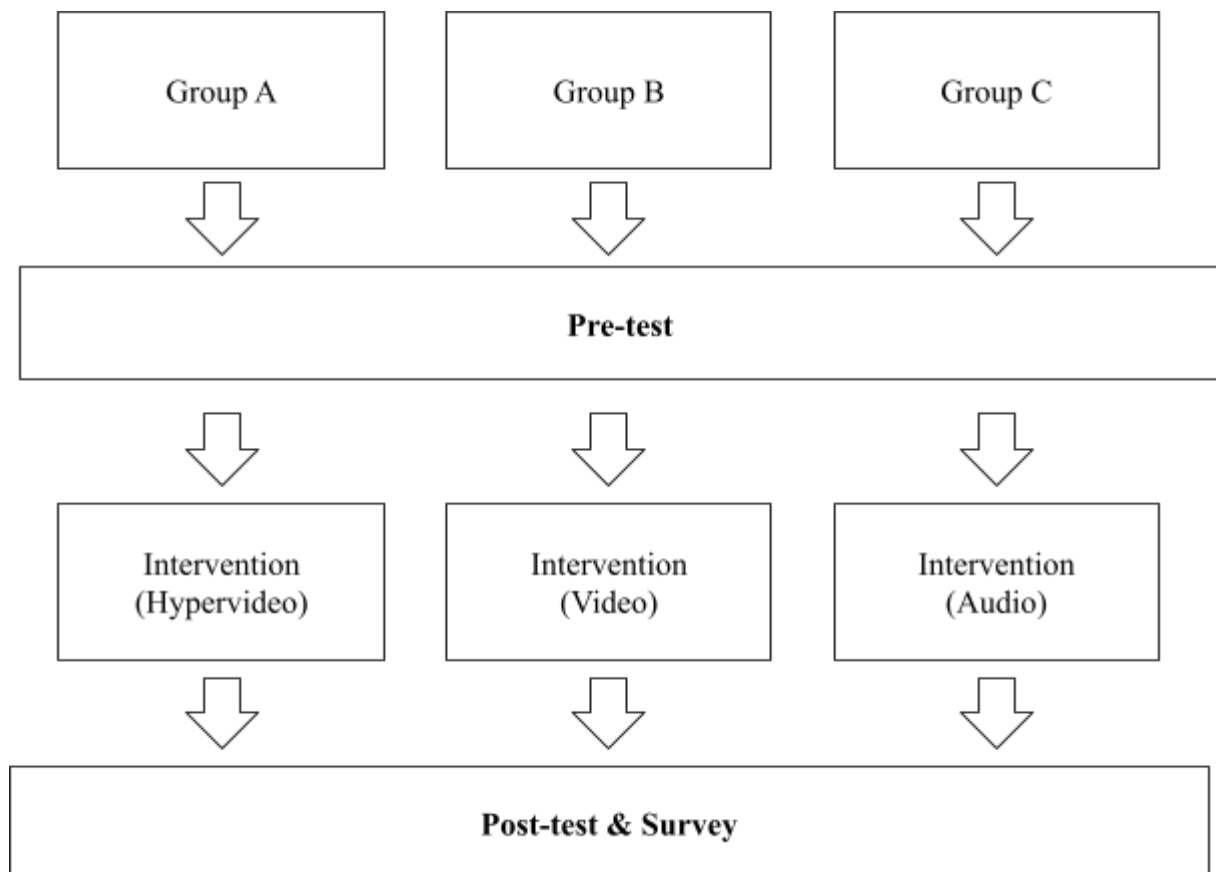
This research's focus on the effects of different types of audio-visual materials on listening performance in Japanese secondary EFL learners is timely and relevant, given the current educational reforms and shifts in pedagogical approaches. By identifying the most effective audio-visual resources for enhancing listening skills, the study can provide empirical evidence to support further integration of multimedia in the EFL curriculum. The findings could inform curriculum development, teaching strategies, and educational policy, contributing to the broader goal of improving English language proficiency and intercultural communication skills among Japanese students.

Methods

1. Research Design

The primary objective of this research is to meticulously investigate the influence of audio-visual materials on the listening performance of secondary EFL learners in Japan. To achieve this goal, this study employs a quasi-experimental research design to examine the impact of audio-visual materials. Specifically, the investigation focuses on the effectiveness of different types of audio-visuals — hypervideo, regular video, and audio — in enhancing listening performance. As shown in Figure 2, The participants are divided into three groups, each exposed to one of the three different types of stimuli. A pre-test and post-test framework is used to measure improvements in listening comprehension.

Figure 2. The Research Design

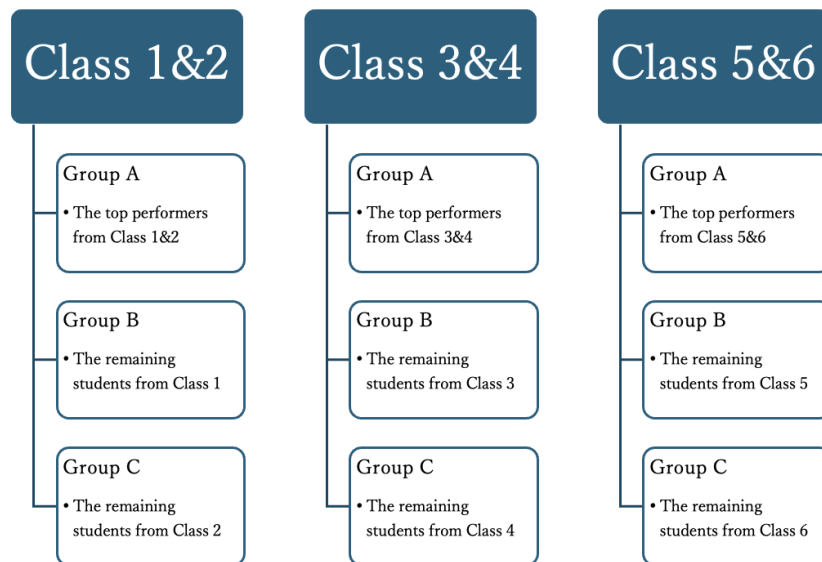


2. Participants

The participants comprised 81 first-year senior high school students aged 15 to 16 from a public high school in Japan who attended the school where the researcher teaches.

In the first grade, students are divided into six homeroom classes. For English lessons, these six classes are further divided into nine groups. Based on their high school entrance exam scores, the top-performing group from each set of two classes is designated as Group A, while the remaining students are assigned to Groups B and C, respectively. As illustrated in Figure 3, the first-grade cohort is thus divided into nine groups according to their entrance exam scores. For the purposes of this study, the focus will be on one Group A, one Group B, and one Group C, which are taught by the researcher.

Figure 3. The 6 classes and the 9 groups of first-year students



Although students are divided into Group As and other groups based on their entrance exam scores, their English proficiency is considerably low compared to the average Japanese high school students. Generally, public high schools in Japan require an entrance selection examination, which includes written tests in five subjects: Japanese, English, Mathematics, Social Studies, and Science. Admission is determined by the scores of these exams, along with the student's middle school grades and interview performance. Consequently, it is no exaggeration to say that the school one attends is determined by their academic abilities. This particular school has a deviation score of 37 (36 being the lowest, 50 being the average, and 67 being the highest in the prefecture) on the high school entrance exam, making it one of the lowest-performing schools in the region.

According to the researcher's observation, although there is a difference in academic ability between Group A and Groups B and C, all students are approximately at the CEFR A1 level. Ideally, I would have liked to select samples randomly to ensure equal academic ability across groups. However, this was challenging, so I decided to use the three first-grade classes I am currently teaching as samples. The number of students in each group is shown in Table 1.

Table 1. The number of participants

	Male	Female	Total
Group A	27	3	30
Group B	17	8	25
Group C	17	9	26

3. Materials and Instruments

3.1. The pre-and post-tests

For the English test administered during the second-term final exams on November 22, 2023, the listening section was used as the pre-test (Appendix 1), and the listening section from the end-of-year test on February 14, 2024, was used as the post-test (Appendix 2). Each test required students to listen to unfamiliar English passages with similar sentence structures and topics as those studied in their textbooks (Appendix 3) and answer questions. The audio recordings were performed by the researcher, who is from Japan, and an assistant language teacher, who is from the United States. Both tests consisted of 10 questions involving either dialogues or monologues, with scanning questions requiring students to listen for specific details. The English passages included both scripted and semi-scripted elements. Four types of passages were prepared for each test. The details of the questions are shown in Figure 5. All questions were multiple-choice.

Table 2. *Descriptions of Pre-test and Post-test Listening Comprehension Activities***Pre-test**

	Topic	Speech types	Degree of scripting	Question types
Question 1	Conversation about a photo	Dialogue	Scripted	3 scanning questions
Question 2	Conversation about a video	Dialogue	Scripted	2 scanning questions
Question 3	Conversation about endangered species	Monologue	Scripted	3 scanning questions
Question 4	Conversation about favorite animals	Dialogue	Semi-scripted	2 scanning questions

Post-test

	Topic	Speech types	Degree of scripting	Question types
Question 1	Explanation about the history of tea	Monologue	Scripted	3 scanning questions
Question 2	Explanation about the history of steam engines	Monologue	Scripted	2 scanning questions
Question 3	Explanation about the boat school in Bangladesh	Monologue	Scripted	3 scanning questions
Question 4	Conversation about chocolate	Dialogue	Semi-scripted	2 scanning questions

3.2. The interventions

Five lessons focusing on listening were conducted as an intervention for each group between the pre-test and post-test. The contents of the lessons were the same across all groups, but different media were used: hypervideo for Group A (Appendix 4), video for Group B (Appendix 5), and audio for Group C (Appendix 6). Students used Chromebooks and headsets the school provided to listen to and watch the materials. Group A answered questions within the hypervideo, while Groups B and C entered their answers using Google Forms. Each session for the listening tasks was set at 10 minutes, during which students could replay or rewatch the materials as many times as needed.

3.2.1. Video

The videos (Appendix 5) were recorded by the researcher based on the content of the textbook (Appendix 3). Using Keynote, slides were created with photos and illustrations to aid in understanding the content, and the recording was done via Zoom. The photos and illustrations were either sourced from the textbook or generated using DALL-E 3. The completed videos were uploaded as unlisted videos on YouTube and embedded in Google Forms. During the lesson, students accessed the Google Form, watched the video, and worked on the tasks provided.

3.2.2. Hypervideo

Hypervideos (Appendix 4) were created using a platform called Edpuzzle. The reasons for selecting this platform included its free availability, simple operation, and compatibility with Microsoft Teams, which students were already using in school. The same videos used for Group B were utilized for the hypervideos. The questions used were identical to those used in the video group, but with hypervideo's interactive features, the questions appeared immediately after viewing the necessary parts, and feedback on whether the answers were correct was provided instantly. This approach was employed to deliberately enhance the listening strategy of selective listening, which involves focusing attention on necessary information. While selective listening was explicitly taught to all groups, the hypervideo group was subjected to semi-compulsory training to focus on necessary information through its interactive features. This method aimed to improve listening skills, as suggested by previous research (Boitsova Bogach, 2019).

3.2.3. Audio

The audios (Appendix 6) were created based on the videos made for the video and hypervideo groups. Instead of the slides shown to the video and hypervideo groups, only the lesson title is displayed. This shuts down visual information, forcing students to derive answers solely

from the audio. As with the video group, the created videos were published on YouTube as unlisted and embedded in a Google Form.

4. Procedure

The study is conducted over two months, from January to mid-February, with each group attending three English classes per week. During this period, interventions were conducted five times as part of Listening Practice in all English classes. In each listening practice session, the following procedure is adhered to:

Lead-in

- A question related to the text is presented to engage students.
- A discussion about the question is facilitated among students.

Pre-listening Task

- Students speculate on the answers to some of the questions or predict the content of the listening task.
- Instruction is provided for several 'blocking words' to ensure vocabulary does not hinder comprehension.

Listening Task

- Students are given 10 minutes to complete the task using their computers.
- They are allowed to replay the audio/video/hypervideo as many times as needed within the allocated time.
- Students' responses are automatically collected for analysis.

Each group was subjected to the intervention five times during the study period. The pre-test was administered before the intervention to establish a baseline for each group's listening comprehension. A similar test was then used as a post-test to measure any improvements after exposure to the intervention.

5. Data Collection and Data Analysis

The pre-test and post-test scores were calculated by assigning one point per question. From the pre-test results, the assumptions for using ANCOVA were checked. SPSS was used for the analysis, and the Univariate General Linear Model indicated a significant difference between the intervention groups based on the pre-test results, as shown in Table 3.

Table 3. *The Results of the Tests of Between-Subjects Effects*

Dependant Variable: pretest

Source	Type of III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	94.749 ^a	2	47.375	14.311	<.001
Intercept	3341.776	1	3341.776	1009.498	<.001
Interv	94,749	2	47.375	14.311	<.001
Error	244.965	74	3.310		
Total	3811.000	77			
Corrected Total	339.714	76			

a. R Squared = ,279 (Adjusted R Squared = ,259)

The significant differences found between the intervention groups indicate that the assumption of homogeneity is violated, and thus, ANCOVA could not be used. Also, from the results of the normality tests, Table 4 shows that the data are not normally distributed.

Table 4. *The Results of the Tests of Normality*

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
pretest	.131	77	.002	.954	77	.007
posttest	.140	77	<.001	.957	77	.011

a. Lilliefors Significance Correction

In addition to the previously mentioned analysis, due to the relatively small sample sizes of the three groups (N=30; N=25; N=26), the researcher opted to compare the means for each group using the non-parametric Wilcoxon Signed Ranks test.

6. Ethical Considerations

Prior to the commencement of the research, all participants were provided with detailed information about the study's purpose, procedures, potential risks, and benefits. Consent forms were distributed and collected, ensuring that participation was entirely voluntary. Also, to protect the privacy of the participants, strict measures were taken to ensure anonymity and confidentiality. All data collected were anonymized, with unique numbers assigned to each participant instead of using personal identifiers. Personal information that could identify participants was not included in any publications.

Results

1. Results of the Pre-test for all three groups

During the pre-test, one student from Group A (Hypervideo) and two students from Group B (Video) were absent for an extended period. Thus, the test was administered only to the remaining students. The test results are presented in Table 5 and Figure 4.

Table 5. *The Results of the Pre-test (Descriptive Statistics)*

	Listening Section Score		
	Audio	Hypervideo	Video
Valid	26	29	23
Missing	0	1	2
Mean	5.654	8.138	5.957
Std. Deviation	1.979	1.432	2.099
Minimum	1.000	5.000	2.000
Maximum	9.000	10.000	10.000

Figure 4. *The Results of the Pre-test (Distribution Plots)*

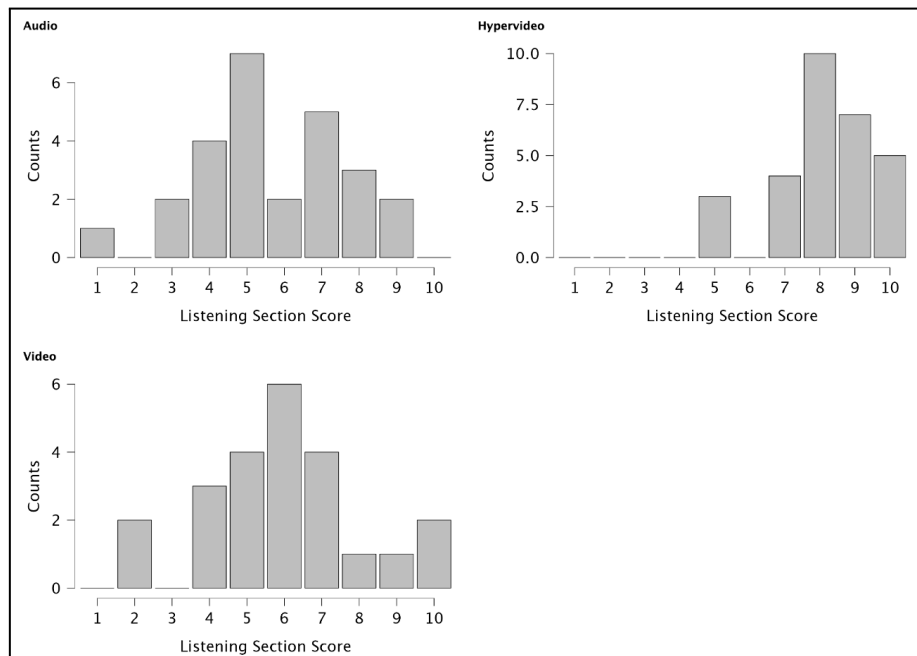


Table 5 shows that the mean of the hypervideo group is relatively higher than that of the other groups. As mentioned in the previous chapter, Figure 4 indicates that the data are not normally distributed.

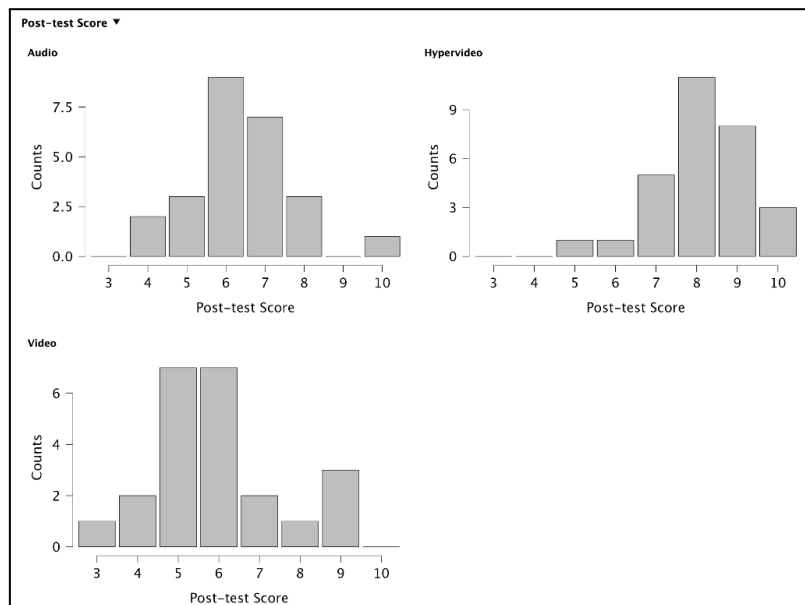
2. Results of the Post-test for all three groups

In the post-test, in addition to the three students who were absent during the post-test, one more student from the audio group had a prolonged absence. Therefore, the test was administered to the remaining students. The results are presented in Table 6 and Figure 5.

Table 6. *The Results of the Post-test (Descriptive Statistics)*

Descriptive Statistics	Post-test Score		
	Audio	Hypervideo	Video
Valid	25	29	23
Missing	1	1	2
Mean	6.400	8.138	5.957
Std. Deviation	1.323	1.156	1.609
Minimum	4.000	5.000	3.000
Maximum	10.000	10.000	9.000

Figure 5. *The Results of the Post-test (Distribution Plots)*



Similar to the pre-test results, Table 6 indicates that the mean score for the Hypervideo group is higher. However, since the Hypervideo group originally had a higher average, it is impossible to determine the intervention's effect based on this alone. Figure 5 demonstrates that the data is not normally distributed.

3. Comparison of the Pre-and Post-tests of Group A (Hypervideo)

The following Tables 7 and 8 present a comparison of the pre-test and post-test results for the Hypervideo group. As shown in Table 8, the difference between the pre-and post-test is not statistically significant.

Table 7. Comparison of the Pre-and Post-tests of the hypervideo group (Descriptive Statistics)

	N	Mean	Std. Deviation	Minimum	Maximum
pretest	29	8.1379	1.43238	5.00	10.00
posttest	29	8.1379	1.15648	5.00	10.00

Table 8. Results of the Wilcoxon Signed Ranks Test (Hypervideo group)

		Ranks		
		N	Mean Rank	Sum of Ranks
posttest - pretest	Negative Ranks	9 ^a	7.50	67.50
	Positive Ranks	7 ^b	9.79	68.50
	Ties	13 ^c		
	Total	29		

- a. posttest < pretest
- b. posttest > pretest
- c. posttest = pretest

Test Statistics^a

		posttest - pretest
Z		-.027 ^b
Asymp. Sig. (2-tailed)		.978

- a. Wilcoxon Signed Ranks Test
- b. Based on negative ranks.

4. Comparison of the Pre-and Post-tests of Group B (Video)

The subsequent Tables 9 and 10 provide a comparison of the pre-test and post-test results for the Video group. The results indicate a greater number of positive ranks compared to negative ranks (Table 10). Nevertheless, the difference between the pre-test and post-test results is not statistically significant.

Table 9. Comparison of the Pre-and Post-tests of the video group (Descriptive Statistics)

	N	Mean	Std. Deviation	Minimum	Maximum
pretest	23	5.9565	2.09931	2.00	10.00
posttest	23	5.9565	1.60902	3.00	9.00

Table 10. Results of the Wilcoxon Signed Ranks Test (Video group)

		Ranks		
		N	Mean Rank	Sum of Ranks
posttest - pretest	Negative Ranks	7 ^a	10.07	70.50
	Positive Ranks	10 ^b	8.25	82.50
	Ties	6 ^c		
	Total	23		

a. posttest < pretest

b. posttest > pretest

c. posttest = pretest

Test Statistics^a

		posttest - pretest
Z		-.290 ^b
Asymp. Sig. (2-tailed)		.772

a. Wilcoxon Signed Ranks Test

b. Based on negative ranks.

5. Comparison of the Pre-and Post-tests of Group C (Audio)

The subsequent Tables 11 and 12 present a comparison of the pre-test and post-test results for the Audio group. Due to an additional absence during the post-test, the analysis was conducted on the 25 students who completed both tests. The results indicate a greater number of positive ranks than negative ranks (Table 12), and the mean values are higher in the post-test compared to the pre-test (Table 11). However, the difference between the pre-test and post-test results is not statistically significant (Table 12).

Table 11. Comparison of the Pre-and Post-tests of the audio group (Descriptive Statistics)

	N	Mean	Std. Deviation	Minimum	Maximum
pretest	25	5.7600	1.94251	1.00	9.00
posttest	25	6.4000	1.32288	4.00	10.00

Table 12. Results of the Wilcoxon Signed Ranks Test (Audio group)

		Ranks		
		N	Mean Rank	Sum of Ranks
posttest - pretest	Negative Ranks	8 ^a	9.00	72.00
	Positive Ranks	13 ^b	12.23	159.00
	Ties	4 ^c		
	Total	25		

a. posttest < pretest

b. posttest > pretest

c. posttest = pretest

Test Statistics^a

posttest - pretest	
Z	-1.557 ^b
Asymp. Sig. (2-tailed)	.119

a. Wilcoxon Signed Ranks Test

b. Based on negative ranks.

Overall, the results indicate that none of the three groups showed a statistically significant improvement in listening comprehension. The results challenge the initial expectation that visual aids (in video and hypervideo formats) would enhance listening skills more than audio-only materials.

Discussion

1. Overview of Findings

The results of this study offer important insights into the impact of different types of audio-visual materials on the listening performance of secondary EFL learners in Japan. Specifically, the study investigated the effectiveness of hypervideo, video, and audio-only materials. Despite the intuitive expectations that visual aids in video and hypervideo formats would enhance listening comprehension more than audio alone, the results did not support this hypothesis.

2. Hypervideo Group

For the hypervideo group, the difference between the pre-test and post-test scores was not statistically significant (Table 8). This indicates that the use of hypervideo did not lead to a measurable improvement in the learners' listening skills. This finding is somewhat unexpected given the interactive nature of hypervideo, which allows learners to control the pacing of comprehension. It suggests that while hypervideo may offer an engaging format, it may not necessarily translate into better listening comprehension in the short term. This finding could be related to Mayer's (2009) cognitive theory of multimedia learning, which emphasizes that not all multimedia enhancements lead to improved learning outcomes. While hypervideo may offer an engaging format, it may introduce extraneous cognitive load, which can hinder rather than help comprehension (Mayer, 2009).

3. Video Group

In the video group, although there were more positive ranks than negative ranks (Table 10), the mean values for the pre-test and post-test remained the same (Table 9). Therefore, using video did not have a significant effect on the learners' listening skills. This result suggests that the inclusion of visual elements, such as images and gestures, does not necessarily enhance listening comprehension beyond the capability of audio alone. It is possible that the visual stimuli in the videos might have served as distractions rather than aids, particularly for learners who are not accustomed to processing multimodal inputs simultaneously.

4. Audio Group

The audio group showed more positive ranks than negative ranks (Table 12), and the mean values were higher in the post-test compared to the pre-test (Table 11). However, the difference was not statistically significant. Despite the lack of statistical significance, this group demonstrated the closest results to achieving a significant improvement. This outcome implies that audio-only materials might still be effective for listening practice, potentially due to their alignment with the format of the listening tests, which were also audio-only. Mayer's (2009) modality principle posits that people learn better when information is presented using both visual and auditory channels rather than just one. However, a related concept is that learning and assessment modalities should be aligned for optimal performance. This means that if an assessment is in an audio-only format, practicing with audio-only materials can be more effective because it reduces extraneous cognitive load and aligns the learning process with the assessment conditions.

5. Comparison with Expectations

The initial hypothesis posited that video and hypervideo, with their visual aids, would improve listening skills more than audio alone. However, the results contradicted this expectation. One plausible explanation for this discrepancy is the duration of the intervention. The relatively short experimental period may not have been sufficient for learners to fully adapt to and benefit from the multimodal materials. Additionally, since both the pre-tests and post-tests were conducted using audio-only formats, it is reasonable to surmise that training with audio-only materials provided a more direct preparation for the test conditions.

6. Implications for EFL Education

The findings of this study have several implications for EFL education, particularly in the context of Japanese secondary schools.

6.1. Test and Training Alignment

The results suggest that training that utilizes audio-only materials may be more effective when listening assessments are conducted using audio-only formats. This is relevant given the nature of high-stakes English language tests that many English learners take in Japan, such as university entrance exams, Eiken,

TOEIC, TOEFL, and IELTS, which predominantly use audio-only listening sections.

6.2. Real-World Listening Skills

Although the study found no significant advantage of video and hypervideo for short-term listening comprehension, it is important to consider the development of real-world listening skills. Listening often involves multimodal processing in everyday situations, except in specific contexts like telephone conversations. Therefore, while audio-only training might be effective for test preparation, incorporating multimodal materials could still be beneficial for enhancing overall communicative competence in real-life scenarios.

6.3. Multimodal Integration in Testing and Training

To better reflect real-world listening demands, it may be advantageous to consider integrating multimodal elements into listening tests and corresponding training programs. This approach could provide learners with a more comprehensive skill set, preparing them for academic and practical communication challenges.

6.4. Future Research Directions

Future research with extended intervention periods and a variety of testing formats (audio, video, and hypervideo) could provide deeper insights into the long-term benefits and potential of multimodal materials in EFL education. Additionally, exploring learner preferences and cognitive load during different types of multimedia exposure could inform more effective instructional design.

In conclusion, while the immediate impact of video and hypervideo on listening performance was not statistically significant, the nuanced results highlight the complexity of multimedia learning. Educators should consider both the specific demands of listening tests and the broader goals of learning education when designing and implementing instructional materials.

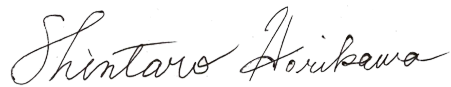
Acknowledgments

I would like to express my deepest gratitude to all those who supported me in completing this thesis. First and foremost, I extend my heartfelt thanks to my thesis supervisor, Associate Professor Katrin Saks, for her invaluable guidance, insightful feedback, and unwavering support throughout this research. Her expertise and encouragement have been instrumental in the successful completion of this study. I am also deeply grateful to Emanuele Bardone, the Program Director of Educational Technology, whose leadership, warm support, and cheerfulness have been pivotal during my time in the program. His commitment to fostering an enriching educational environment has greatly enhanced my learning experience. I would like to acknowledge all the other lecturers of the Master's program for their dedication and the knowledge they have imparted. Their teaching and mentorship have significantly contributed to my academic growth and development. My sincere appreciation goes to my groupmates in the Master's program as well. Their camaraderie, collaboration, and encouragement have been a source of motivation and inspiration throughout this program. It is such an honor for me to share this wonderful and fruitful journey with them. Lastly, I wish to express my profound gratitude to my wife for her unwavering support and understanding. Balancing work, childcare, and household responsibilities while writing this thesis was made possible only through her steadfast support and patience. Her belief in me has been a constant source of strength.

Thank you all for your invaluable contributions to this accomplishment. I have come this far all because of you.

Author's declaration

I hereby declare that I have written this thesis independently and that all contributions of other authors and supporters have been referenced. The thesis has been written in accordance with the requirements for graduation theses of the Institute of Education of the University of Tartu and is in compliance with good academic practices.



Shintaro Horikawa
01/06/2024

References

- Akase, M. (2021). Application of audiovisual media tools in teaching English as a foreign language. *長野工業高等専門学校紀要*, 55, 2-5.
- Akbari, O., & Razavi, A. (2016). Using authentic materials in the foreign language classrooms: Teachers' perspectives in EFL classes. *International Journal of Research Studies in Education*, 5(2), 105-116.
- Al-Seghayer, K. (2001). The effect of multimedia annotation modes on L2 vocabulary acquisition: A comparative study. *Language Learning & Technology*, 5(1), 202-232.
- Bajrami, L., & Ismaili, M. (2016). The role of video materials in EFL classrooms. *Procedia - Social and Behavioral Sciences*, 232, 502-506.
- Bal-Gezegin, B. (2014). An investigation of using video vs. audio for teaching vocabulary. *Procedia - Social and Behavioral Sciences*, 143, 450-457.
- Boitsova, E., & Bogach, N. (2019). The strategy of L2 selective listening in ELT classroom. *The European Proceedings of Social & Behavioural Sciences*, 19, 678-688.
- Elliott, L., & Vasquez, M. G. (2021). Speaking anxiety: Japanese students in the EFL environment. *The Asian Conference on Education 2021 Official Conference Proceedings*. Nagoya University of Foreign Studies, Japan.
- Guichon, N., & McLoran, K. (2008). The effects of multimodal text on L2 learners: Implications for CALL resource design. *System*, 36(1), 85-93.
- Halliday, M. A. K. (1978). *Language as social semiotic: The social interpretation of language and meaning*. Edward Arnold.
- Jewitt, C. (2008). *Technology, literacy, and learning: A multimodal approach*. Routledge.
- Jones, L. C., & Plass, J. L. (2002). Supporting listening comprehension and vocabulary acquisition in French with multimedia annotations. *The Modern Language Journal*, 86(4), 546-561.
- Kress, G. (2003). *Literacy in the new media age*. Routledge.
- Mayer, R. E. (2009). *Multimedia learning (2nd ed.)*. Cambridge University Press.
- New London Group. (1996). A pedagogy of multiliteracies: Designing social futures. *Harvard Educational Review*, 66(1), 60-92.
- Rowsell, J., & Collier, D. R. (2017). Researching multimodality in language and education. In K. A. King, Y. J. Lai, & S. May (Eds.), *Research methods in language and education* (pp. 311-324). Springer.

- Sauli, F., Cattaneo, A., & van der Meij, H. (2018). Hypervideo for educational purposes: A literature review on a multifaceted technological tool. *Technology, Pedagogy and Education*, 27(1), 115-134.
- Yashima, T. (2019). L2 motivation and willingness to communicate. In M. Lamb et al. (Eds.), *The Palgrave Handbook of Motivation for Language Learning* (pp. 203-223). Palgrave Macmillan.
- 文部科学省. (2018). 文部科学白書2018: ICT活用の推進. Retrieved from <https://www.mext.go.jp>
- 文部科学省. (2020). 外国語の指導におけるICTの活用について. Retrieved from <https://www.mext.go.jp>

Appendices

Appendix 1: The pre-test and its audio

Pre-test	https://drive.google.com/file/d/1Uh-lAIg8-o1Kt459b7XbHjKJOU4k8Tx_/view?usp=share_link
Audio	https://drive.google.com/file/d/1TEhJzHJ19w1HztG4_2wma_cthqeefnKh/view?usp=share_link
Script of the audio	https://drive.google.com/file/d/1TFV9iBItgxov9-4fIPKKnwmI5l4VfxZ6/view?usp=share_link

Appendix 2: The post-test and its audio

Post-test	https://drive.google.com/file/d/1T9Reg6SPOy1xfKTD2geVkvCTmHeJHdJu/view?usp=share_link
Audio	https://drive.google.com/file/d/1U8mwR3kqEreH-FarZAYrkADMyZ0x9guw/view?usp=share_link
Script of the audio	https://drive.google.com/file/d/1TPc8GAGvIJ1OVeq8TMmcdYzpCr1ktd45/view?usp=share_link

Appendix 3: Textbook used for the tests and the interventions

https://drive.google.com/file/d/1TUeVszPW1NBQldsFcron7dNc0dS0jUCI/view?usp=share_link

Test/Practice Stage	Details
Pre-test	Scope of the test: Lesson 3 and Lesson 4
Listening Practice 1	Lesson 5 Section 1 was used.
Listening Practice 2	Lesson 5 Section 2 was used.
Listening Practice 3	Lesson 6 Section 1 was used.
Listening Practice 4	Lesson 6 Section 2 was used.
Listening Practice 5	Lesson 7 Section 1 was used.
Post-test	Scope of the test: Lesson 5 and Lesson 6

Appendix 4: Interventions (Hypervideo)

Listening Practice 1	https://edpuzzle.com/media/657821b8d08aef4142516317
Listening Practice 2	https://edpuzzle.com/media/65a7742e6c5c0b1185dca687
Listening Practice 3	https://edpuzzle.com/media/65b769a0dcc741e532e6abcd
Listening Practice 4	https://edpuzzle.com/media/65b8fb5cc9391052ba087c45
Listening Practice 5	https://edpuzzle.com/media/65c209e95b521bc986c8f323

Appendix 5: Interventions (Video)

Listening Practice 1	https://forms.gle/W5ZtspY72BwgbmUc9
Listening Practice 2	https://forms.gle/fXtcKL2KEjH4eHBu6
Listening Practice 3	https://forms.gle/FqxyVconhL9yYSvf7
Listening Practice 4	https://forms.gle/HjdqrTDyebGLpEVM7
Listening Practice 5	https://forms.gle/B38As38BSrZ1SmWE6

Appendix 6: Interventions (Audio)

Listening Practice 1	https://forms.gle/29NVoyxjXWXfccvA8
Listening Practice 2	https://forms.gle/RiNNbivB2AfzgvBYA
Listening Practice 3	https://forms.gle/Bu17RbFcpF8oz3KX6
Listening Practice 4	https://forms.gle/dbQZU92QMV47ndy99
Listening Practice 5	https://forms.gle/gXNfJx7Vb2XMrTjv8

Non-exclusive licence to reproduce the thesis and make the thesis public

I, *Shintaro Horikawa*,

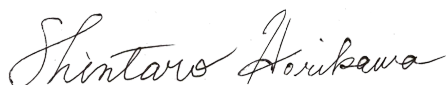
1. grant the University of Tartu a free permit (non-exclusive licence) to

reproduce, for the purpose of preservation, including for adding to the DSpace digital archives until the expiry of the term of copyright, my thesis

Effects of Using Audio-visual Materials on the Listening Performance of Secondary EFL Learners in Japan,

supervised by *Associate Prof. Katrin Saks*.

2. I grant the University of Tartu a permit to make the thesis specified in point 1 available to the public via the web environment of the University of Tartu, including via the DSpace digital archives, under the Creative Commons licence CC BY NC ND 4.0, which allows, by giving appropriate credit to the author, to reproduce, distribute the work and communicate it to the public, and prohibits the creation of derivative works and any commercial use of the work until the expiry of the term of copyright.
3. I am aware of the fact that the author retains the rights specified in points 1 and 2.
4. I confirm that granting the non-exclusive licence does not infringe other persons' intellectual property rights or rights arising from the personal data protection legislation.



Shintaro Horikawa
01/06/2024