AsiaCALL Online Journal
ISSN 1936-9859; https://asiacall.info/acoj
Vol. 12; No. 5; 2021
pp. 100-113

Received: 20/08/2021
Revision: 25/10/2021
Accepted: 25/10/2021
Online: 26/10/2021

An exploratory study on the use of interactive video via Netflix to improve second language aural vocabulary learning

Gilbert Dizon[1], John M. Gayed[2]

[1] Himeji Dokkyo University, Japan

[2] Tokyo Institute of Technology, Japan

Correspondence: Gilbert Dizon, Himeji Dokkyo University, Japan. E-mail: gdizon@gm.himeji-du.ac.jp

EOI: http://eoi.citefactor.org/10.11251/acoj.12.05.007

Abstract

While video streaming services have had a profound impact on how people view television programs and movies, very little is known about their use to promote second language (L2) development. To address this gap in L2 research, this paper details the results of an exploratory study that examined the combination of interactive video, a new form of entertainment that prompts viewers to make choices that affect the story, and Netflix to see if these tools could enhance different aspects of L2 aural development. Fourteen Japanese university EFL students participated in the study, which utilized a counterbalanced, pretest-posttest research design. The participants viewed two interactive episodes of a TV series via Netflix, each under one of the following conditions: non-interactive and interactive. Pre- and post-tests were administered to measure any gains that were made in aural vocabulary at two different levels (form recognition and meaning recall). Findings from the experiment indicated that interactive video did not significantly impact the incidental learning of aural vocabulary in terms of form recognition or meaning recall. While these results cast doubt on the possible benefits of interactive video, more studies are needed to investigate the potential of the technology for L2 learning.

Keywords: video streaming, vocabulary learning, EFL, L2 video

1. Introduction

As Van, Dang, Pham, Vo, and Pham (2021) note, the use of technology to learn a foreign language comes with many distinct advantages, including increased opportunities for self-directed language learning inside and outside of the classroom (Nguyen, 2021). One technology that has been a common feature of language classrooms for the past several decades is the video (Vanderplank, 2016). Montero Perez, Peters, and Desmet (2018) point out that the accessibility of video, particularly internet television (YouTube, Netflix, Amazon Prime, Hulu), has the capacity to facilitate greater exposure to the target language outside of the classroom. Captioned and subtitled videos (hereafter, the term "captioned video" is used throughout to minimize repeated and/or potentially confusing use of the two terms), which present written text synchronously with video and audio to support comprehension, have also made it possible for

CITATION | Dizon, G., & Gayed, J. M. (2021). An exploratory study on the use of interactive video via Netflix to improve second language aural vocabulary learning. *AsiaCALL Online Journal, 12*(5), 100-113. EOI: http://eoi.citefactor.org/10.11251/acoj.12.05.007

even beginner learners to make use of video to study a foreign language. While the use of video by L2 learners may present some downsides, for instance, the fact that students may associate them with leisure viewing rather than language learning (Vanderplank, 2010), as well as the possibility of learners using them as a crutch to aid comprehension (Winke, Gass, & Sydorenko, 2010), the benefits of textual aids when watching video are clear, specifically in terms of listening and vocabulary development. A meta-analysis by Montero Perez, Van den Noortgate and Desmet (2013) found that those who used L2 subtitles performed significantly better on listening and vocabulary tests than language learners who had no access to on-screen text.

Although extensive literature exists on the use of captioned video with L2 learners (see Vanderplank, 2016; Yeldham, 2018), there are still some areas that have received little to no attention. One such area is the use of video streaming services for foreign language learning. Because of their ubiquity in today's society, it is critical to examine the use of video streaming services to promote L2 development. However, to the best of the authors' knowledge, only two studies have looked at their use with students in the context of L2 learning (Dizon, 2018; Wang & Chen, 2019). Another technology that has not been studied in L2 research is the use of interactive video. Interactive video is a relatively new form of entertainment delivered via the Internet, which allows users to make choices that affect the outcome of the story, with companies such as Netflix, Amazon, YouTube, and the BBC currently in the process of developing interactive titles (Perez, 2019). Lastly, although incidental vocabulary learning is a well-studied topic in L2 video research, less attention has been paid to aural vocabulary learning with L2 video. Due to these gaps in the literature, an exploratory study was conducted to examine the use of interactive video through Netflix on the incidental learning of L2 English aural vocabulary.

2. Theoretical framework

This study is informed by the theory of multimedia learning by Mayer (1997, 2001). According to the theory, learners process information most effectively using visual and written aids. When faced with pictorial (visual) and written (verbal) information, students must select relevant input and organize this information into visual and verbal mental representations of said input. Learners can then build connections between the different forms of input, thereby allowing them to integrate this information into working memory (see Figure 1). Accordingly, a more fertile learning environment is created, thus leading to vocabulary, reading, and listening development. Mayer's (1997, 2001) theory of multimedia learning has been used extensively in the field of L2 learning, especially in terms of examining the role that annotations have on language learners. Jones (2009) found that beginner and advanced L2 learners benefited most from written and pictorial annotations, thereby supporting listening comprehension and vocabulary learning. In another study on vocabulary acquisition and listening comprehension, Jones and Plass (2002) discovered that students with access to both pictures and written captions better remembered word translations and recalled the listening passage more effectively than those who only used one form of annotation or who did not have access to aids. While no

significant difference was found in relation to reading comprehension, the results from Akbulut's (2007) study indicated that access to word definitions and visuals led to greater gains in vocabulary compared to when students were only given access to definitions. These findings illustrate the importance of visual and written information for L2 vocabulary and listening and highlight the positive impact that multimedia environments can have on foreign language learners. As Rodgers (2018) notes, combining input modes such as video and audio may enhance different features of the L2 learning process, including listening comprehension and vocabulary development. Considering captioned video incorporates multiple modes–visual in the form of the video, aural in the form of the L2 spoken language, and verbal in the form of the written captions–video streaming services may be a useful tool for language learning as they offer learners audio and video captioning options in multiple languages.

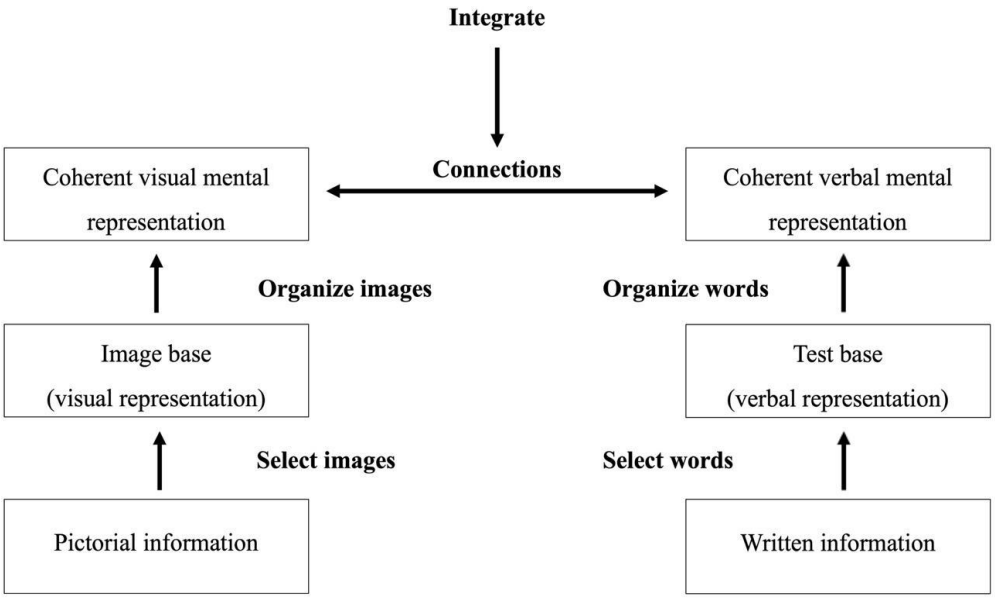

Figure 1. Mayer's (1997, 2001) theory of multimedia learning (Jones & Plass, 2002)

3. Literature review

3.1 Captioned video & L2 vocabulary learning

Much of the research on video for L2 learning has focused on the use of captions and their effects on vocabulary development, with a particular emphasis on comparing L1, L2, and non-captioning groups. The students in Bianchi and Ciabattoni's (2008) study who used L1 captions and L2 captions were found to make more significant improvements on long-term vocabulary acquisition than those who used no captions while viewing movies. However, language proficiency did have an impact, as beginner learners benefited least from L2 captions while advanced students benefited the most. In a year-long longitudinal study involving adolescent L2 learners, Pujadas and Muñoz (2019) investigated the role of caption type (L1 vs. L2), proficiency, and pre-teaching on vocabulary form recall and meaning recall. The researchers

found that both captioning groups were equally as effective in terms of vocabulary development, but that pre-teaching the target words and a higher level of language proficiency led to greater gains.

There seem to be conflicting results when it comes to other forms of captioning, such as keyword captions or glosses. In a study involving five different captioning groups, Hsieh (2019) found that the full captions with highlighted target words and L1 gloss condition best promoted word form and meaning learning. Not surprisingly, the no captions group and the full captions with no audio group were found to perform the worst in terms of vocabulary learning. In a recent study focusing on incidental vocabulary learning among primary L2 English learners, Teng (2019) investigated how different forms of captioning (full captions, keyword captions, no captions) affected word form/meaning recognition word recall of meaning. According to the results of the study, those in the full captioning group significantly outperformed the other two groups. Montero Perez, Peters, Clarebout, and Desmet (2014) also investigated vocabulary learning using four different types of tests: form recognition, clip association, recall, and meaning recognition. The results of their study revealed that the three caption groups outperformed the control group on three of the tests (form recognition, clip association, meaning recognition), while no significant differences were found in relation to the meaning recall. On the other hand, their findings also indicated that the form of captioning used (full captions, keyword captions, full captions with keywords highlighted) did not result in significant differences between the three groups, thereby illustrating that captioning, regardless of type, can make a positive impact on vocabulary development. In a similar study examining the same four aspects of vocabulary, Montero Perez, Peters, and Desmet (2015) found that students who used keyword captions outperformed full captions on the form recognition test. Lastly, in a follow-up study by Montero Perez et al. (2018) involving four L2 captioning conditions (full, keyword, glossed keyword with access to meaning, and no captions), the researchers found that the glossed keyword captions group outperformed the other forms of captioning on form recognition and meaning recall, which illustrates that access to meaning may benefit L2 learners when they watch the video.

As noted by Vidal (2003), little research has been done on vocabulary learning via aural input. Having said that, there have been a few studies that have investigated the effects of this input method on vocabulary learning. Winke et al. (2010), for instance, examined the use of target-language captioned video on written and aural vocabulary knowledge and found that the caption group outperformed the control group on both variables on the post-test. In another study, Peters, Heynen, and Puimège (2016) examined L1 and L2 captions in two exploratory experiments. The first experiment focused on form recognition and meaning recall of aural vocabulary. It was found that caption type did not predict meaning recall but that L1 captions better promoted form recognition. In their second experiment, which examined form recognition and meaning recall of written vocabulary, positive effects for captions were found in form recall but not form recognition. Similarly, Syodorendo (2010) looked at the effects of input modality on vocabulary acquisition. In a mixed-methods study of twenty-six Russian

English language learners, the researcher found that video supplemented with captions resulted in greater word form recall and word meaning recall, while non-captioned video improved recognition of aural word forms. Another important finding from the study was that some of the participants indicated the tendency to focus on the captions more than the audio if captions were provided, which is a complaint echoed by the participants in Winke et al. (2010). Moreover, all participants indicated the importance of having visual images as beneficial to their understanding of the spoken input, thus illustrating the importance of visual input alongside aural input.

3.2 Video streaming in L2 learning

Even though video streaming services have been mentioned in L2 literature as having the potential to be used for language learning (Godwin-Jones, 2018; Rosell-Aguilar, 2017), few studies have actually examined their use in the context of foreign language learning in empirical research. Only two studies have looked at video streaming and language learning to the best of the authors' knowledge. One is Dizon's (2018) case study on the use of Netflix with EFL learners. The learners were allowed to use the video streaming service however they desired and were under no obligation to use it specifically for language learning. Following the end of their 3-month membership, the students were interviewed to understand their opinions towards Netflix as a language learning tool. According to the interview data, several themes were identified. First, student comments suggested that learner effectiveness was improved through the use of L1/L2 captions. Secondly, the service enhanced motivation because it was a fun way to study English. Additionally, the learners indicated that they had better access to cultural and linguistic information by watching TV programs and films via Netflix. However, one disadvantage that the participants noted was a lack of convenience due to the large amount of data required to stream video on their mobile devices. Therefore, some of them limited their use of the service to when they had access to Wi-Fi. Another study that has specifically looked at video streaming in the context of L2 learning is Wang and Chen's (2019) research on the use of YouTube as an informal language learning resource. The researchers interviewed university students in Taiwan who had extensive experience in watching L2 English-learning YouTube videos to understand the affordances and limitations of the streaming service for foreign language learning. Based on these interviews, a few key advantages were discovered, namely, that learning English via YouTube was more flexible, interactive, and interesting for the L2 learners. That said, the students also felt that the streaming service was ineffective at improving L2 development. Although results from these two studies were largely positive, they involved a small number of participants and did not examine any language gains the students may have made. Thus, there is still a need to investigate if video streaming services can have an impact on the development of L2 skills.

In summary, the use of L1 and/or L2 captions seems to be beneficial for language learners in terms of vocabulary learning (Bianchi & Ciabattoni, 2008; Pujadas & Muñoz, 2019). Yet, when it comes to other caption types, namely, full captions vs. keyword captions, it is not clear what form best promotes vocabulary learning. While the aforementioned literature

highlights the positive effects of captioned video on L2 vocabulary learning, there are still gaps that need to be addressed. First, the use of interactive video has yet to be studied in the context of L2 learning. As an emerging technology that the researchers predict will only continue to grow, it is critical to evaluate the tool for language learning. Compared with traditional video in which a learner passively takes in input, interactive video has the potential to direct more attention to input, particularly written information in the form of on-screen text when students are prompted to make choices that impact the story (see Figure 1). Moreover, despite their widespread use and popularity, research on video streaming services in the context of language learning is scarce. Although video in the form of DVDs and downloadable content has been available for more than a decade, the advent of video streaming and the ubiquity of smartphones has made language learning via video much more flexible for learners (Wang & Chen, 2019). In other words, viewing video, TV shows, and movies in a foreign language has never been easier due to video streaming and mobile devices. Lastly, although some studies have investigated the effects of captions on aural vocabulary learning (Peters et al., 2016; Syodorenko, 2010; Winke et al., 2010), past L2 captioning research has generally focused on written vocabulary development and ignored aural vocabulary. This indicates that more studies ought to be conducted to see if captioned video can positively influence the incidental learning of aural or listening vocabulary.

3.3 Research Questions

Due to these gaps in L2 research, this study looked at the use of interactive video via Netflix to develop L2 English aural vocabulary. Specifically, the following research question was addressed: *Does the use of Netflix and interactive video have a significant effect on L2 English learners' aural vocabulary?*

4. Methods

4.1 Pedagogical Setting & Participants

Fourteen male and female students at a Japanese university participated in the study. These participants were chosen via convenience sampling. The participants were six first- and eight second-year students enrolled in separate communicative English classes, which were taught by the researchers in the fall 2019 semester. Their ages ranged from 18-21, and they were all native-Japanese speakers. Based on the participants' EIKEN scores, their English language ability ranged from A1 to A2 on the CEFR proficiency scale, which equates to a beginner level of English language proficiency.

4.2 Design of the Study

The present study incorporated a counterbalanced, pretest-posttest research design to examine if the use of interactive video via the streaming service Netflix promotes significant enhancements in aural vocabulary at two levels: form recognition and meaning recall. Form recognition refers to the ability to recognize whether or not one has been exposed to a word

(reading or listening), whereas meaning recall is the ability to produce an L1 definition of a target word. While listening comprehension is another commonly studied variable in L2 video research, this was not examined in this study due to the non-linear nature of the interactive video. In other words, each user sees different content depending on their choices, so a listening comprehension test based on the target videos was not feasible.

4.3 Data collection & analysis

The researchers developed an assessment and administered to assess any potential gains in aural vocabulary made by the participants. A total of 20 words were identified (10 per episode) as target vocabulary. All the words were above the 2k word frequency level according to the British National Corpus (BNC) and the Corpus of Contemporary American English (COCA). In other words, they were terms that the participants were less likely to know. Another reason for the focus on target words beyond the 2k word level was that hat the participants were studying the 1k and 2k word levels through a digital flashcard program at the time of the study. Therefore, the inclusion of target vocabulary from these high-frequency word levels might have resulted in test items that the participants largely knew.

The target words were unique to each episode that was shown, regardless of the prompts chosen by the participants. That said, the frequency of occurrence of the words could have differed depending on the individual prompts selected by each participant. However, it is important to stress that the participants were exposed to all the target words at least once, irrespective of their prompt choices. The words contained nouns, adjectives, and verbs (see Table 1 for the list of target vocabulary). The presentation of the target items was audio-recorded by a native British-English speaker, which mirrors the variety of English spoken in the interactive episodes. The speaker repeated each word twice, with a two-second interval between repetitions and a ten-second interval between individual items. In addition to the target words, four non-words were added to control for pre-test post-test learning effects (Nation & Webb, 2011) and were taken from http://lextutor.ca/freq/lists_download/pnwords.html. Four high-frequency words (deliver, rope, balance, village) which appeared in the episodes were also added to aid test motivation (Peters et al., 2016). Similar to Peters et al. (2016), participants were asked to tick "yes" or "no" if they recognized the word for each item. They were also asked to provide an L1 definition of each target word, i.e., meaning recall. The written presentation of the test was in L1 Japanese so that the assessment was clear and understandable for the participants. To be clear, the introduction of the vocabulary items was done aurally; only the instructions were provided in written form. Cronbach's alpha, which measures the internal consistency of a group of items, was used to assess the reliability of the two sections of the assessment. Values for the form recognition and meaning recall portions of the test were .62 and .80, which indicates slight and moderate reliability, respectively (Brown, 2014).

Table 1. Breakdown of target vocabulary

Target word	Part of speech	BNC-COCA frequency	Episode
ashore	noun	6k	1
canyon	noun	7k	1
bushwhack	verb	17k	1
abandoned	adjective	3k	1
critical	adjective	3k	1
fierce	adjective	3k	1
volcanic	adjective	4k	1
hacking	noun	5k	1
stalking	verb	5k	1
reptile	noun	7k	1
isolated	adjective	3k	2
straining	verb	3k	2
primitive	adjective	4k	2
ticking	verb	4k	2
improvise	verb	5k	2
perishable	adjective	6k	2
trolley	noun	6k	2
commando	noun	7k	2
budge	verb	8k	2
rappel	verb	17k	2

The aural vocabulary pretest was administered in December of 2019. Four weeks later, in January of 2020, the participants viewed the first two episodes of the interactive TV series *You vs. Wild* through Netflix in consecutive class periods. Due to the counterbalanced design, the first-year students watched episode one under the non-interactive condition and episode two under the interactive condition, whereas the second-year students watched episode one under the interactive condition and episode two under the non-interactive condition. In the non-interactive condition, episodes were displayed via class projector, and each prompt choice was determined by letting the selection time period run out, i.e., choices were selected randomly. In the interactive condition, participants watched the episodes individually on desktop PCs located in the classroom or with their individual smartphones and made prompt choices on their own (see Figure 1 for an example prompt from *You vs. Wild*). While the actual length of each episode depends on the choices made, the average run-time of each episode as listed by Netflix was 14 minutes and 20 minutes for episodes 1 and 2, respectively. Due to the participants' English proficiency levels and to enhance comprehension, L1 captions were displayed throughout both episodes, regardless of the condition. After viewing each episode, the participants immediately took the aural vocabulary post-test pertaining to the target words in that particular episode (ten

target words each).

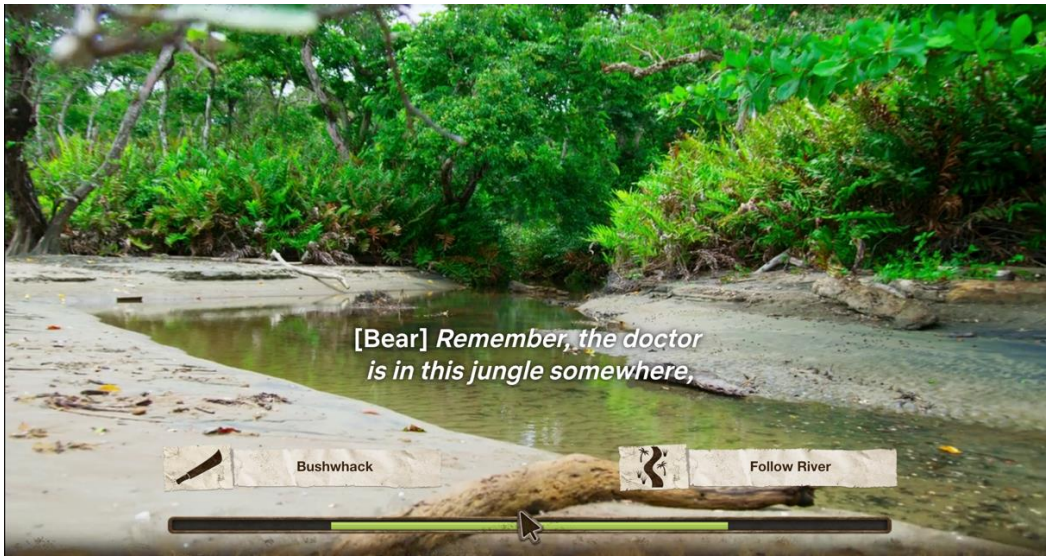

Figure 2. Interactive video prompt is *You vs. Wild*

Given the small sample size, non-parametric statistical tools were used to analyze the data. The Wilcoxon Signed-Rank test was used to analyze the gains made within each group with respect to aural vocabulary form recognition and meaning recall, and the Mann-Whitney U test was used to determine if there were significant differences between the relative gains the non-interactive and interactive conditions were able to make in aural vocabulary form recognition and meaning recall from the pre- to the post-test. Rather than solely examining the differences between pre-and post-test scores with respect to the variables studied, relative gains were calculated based on a formula [(post-test score – pre-test score) / (number of test items – pre-test score) × 100] outlined by Webb and Chang (2015). According to Horst, Cobb and Meara (1998), relative gains may be a more accurate measure of learner improvement compared to absolute gains as they take into account differing opportunities for growth. Descriptive statistics of the students' aural vocabulary test scores were also provided.

5. Results/Findings and discussion

Table 2 below illustrates the pre-and post-test results as it relates to the participants' aural form recognition and meaning recall scores. As the descriptive statistics show, the non-interactive and interactive conditions promoted little to no gain with respect to the variables studied. For the non-interactive condition, form recognition slightly decreased from the pre- to the post-test. Non-positive results were also found in the interactive condition; specifically, there was no difference between the mean pre-test and post-test form recognition results. Likewise, the non-interactive condition ($Z = 1.26$, $p = .20$) nor the interactive condition ($Z = 0.0$, $p = 1.0$) promoted significant improvements in form recognition from the pre-test to the post-test according to the Wilcoxon signed-rank test. When comparing between groups with the

Mann-Whitney U test, a significant difference was not found between the form recognition relative gains of the non-interactive condition and interactive conditions (U = 86.5, p = .61). There were similar results when examining meaning recall. Both the non-interactive and interactive conditions supported small gains in the variable. However, these modest within-group improvements were non-significant in the non-interactive condition (Z = 0.81, p = .41) and the interactive condition (Z = 1.0, p = .31). Similarly, there was not a significant difference between the non-interactive and interactive conditions when it came to relative gains in meaning recall (U = 96, p = .94).

Table 2. Pre- and post-test results

	Pre-test				Post-test			
	Form recognition		Meaning recall		Form recognition		Meaning recall	
	M	SD	M	SD	M	SD	M	SD
Non-interactive	4.79	1.53	0.79	1.12	4.50	2.03	1.07	1.38
Interactive	4.71	1.27	0.29	0.61	4.71	1.68	0.57	1.16

There are several possible explanations for the lack of improvements. First, the frequency of occurrence of the target words could have had an impact on the results. Prior research has indicated that more frequent exposures to a word lead to more successful learning (Teng, 2019). Yet, this variable could not be controlled in the study as interactive video is non-linear. Another plausible explanation for the non-gains may be the incorporation of low-frequency words, which likely led to a higher "learning burden" for the students (Laufer, 2005, p. 234). The target vocabulary words were chosen due to their difficulty and unfamiliarity, and this possibly made it too challenging for the students to recognize and learn these terms. Furthermore, it is difficult to infer the meaning of unfamiliar words in real-time (Buck, 2001), so the learners may have been unable to learn the words as they could have appeared only once, and the students could not replay portions of the video. The use of L1 instead of L2 captions is another factor that might have influenced the results. There were likely instances where there was a mismatch between what the learners heard in English versus what they simultaneously read in Japanese, and these inconsistencies could have had a negative influence on aural vocabulary learning.

6. Conclusion

Interactive video is an emerging technology and thus should be investigated for its language learning potential. Accordingly, this study was conducted to evaluate the efficacy of the technology to support aural developments in the target language, making it the first research to involve interactive video in the context of L2 learning. However, while Mayer's (1997, 2001) theory of multimedia learning posits that video watching affords L2 learners with listening

comprehension and vocabulary learning benefits, the present study resulted in no to limited gains in the variables studied. Participants in the interactive condition did not make significantly greater gains in form recognition or meaning recall than the non-interactive condition. The findings from this study suggest that interactive video via streaming services may not offer any aural vocabulary learning benefits over traditional video. Although interactivity has been viewed as a positive for L2 learning (e.g., Chapelle, 2005), its incorporation may have had a detrimental effect in that students could have paid more attention to the entertainment aspect of the videos rather than the English audio track. This mirrors the warning Vanderplank (2010) made concerning leisure viewing and how learners may associate movies and TV shows with entertainment, which in turn may restrict the potential of video to support L2 development. As a result, the researchers cannot recommend the use of interactive video in the language classroom at this time. However, it is important to stress the small sample size, which likely substantially influenced the results. In other words, the participants may not have been representative of the impact that interactive video can have on aural language learning. Therefore, it is important to conduct more studies on the use of interactive video with L2 learners to gain a more comprehensive understanding of its affordances, especially since technology-mediated L2 learning has been shown to bring about positive effects on all four language skills, that is, listening, speaking, reading, and writing (Van et al., 2021).

Other limitations of the study include the non-randomized research design. Because of this, future studies involving video streaming and/or interactive video should be taken from a randomized and much larger sample. Another limitation was that the type of on-screen text in the study was limited to L1 captions. Research involving textual aids with L2 video has largely revolved around examining the types of help options that best promote language learning. Thus, a worthwhile avenue of research would be to compare the efficacy of L1 and L2 captions when viewing the interactive video. Furthermore, the learners' views on video streaming and interactive video were not obtained. As a result, it would be valuable to evaluate L2 learner perceptions towards these emerging technologies through the use of surveys, interviews, and/or reflective reports. Lastly, although this study did not examine this variable, a future one could examine if interactive video affects L2 listening, given the affordances that technology provides in L2 listening development (Nguyen, 2021).

References

Akbulut, Y. (2007). Effects of multimedia annotations on incidental vocabulary learning and reading comprehension of advanced learners of English as a foreign language. *Instructional Science, 35*(6), 499-517. https://doi.org/10.1007/s11251-007-9016-7

Bianchi, F., & Ciabattoni, T. (2008). Captions and subtitles in EFL learning: An investigative study in a comprehensive computer environment. In A. Baldry, M. Pavesi, C. Taylor Torsello, & C. Taylor (Eds.), *From didactas to ecolingua: An ongoing research project on translation and corpus linguistics* (pp. 69-80). Trieste: Edizonioni Università Trieste.

Birulés-Muntané, J., & Soto-Faraco, S. (2016). Watching subtitled films can help learning foreign languages. *PLOS ONE*, *11*(6), 1-10. https://doi.org/10.1371/journal.pone.0158409

Brown, J. D. (2014). Classical theory reliability. In A. J. Kunnan (Ed.), *The companion to language assessment* (pp. 1165–1181). Oxford, UK: Wiley–Blackwell.

Buck, G. (2001). Assessing Listening. Cambridge, UK: Cambridge University Press.

Chapelle, C. (2005). Interactionist SLA theory in CALL research. In J. L. Egbert & G. M. Petrie (Eds.), *CALL research perspectives* (pp. 53-64). Mahwah, NJ: Lawrence Erlbaum.

Dizon, G. (2018). Netflix and L2 Learning: A case study. *The EuroCALL Review*, *26*(2), 30-40. https://doi.org/10.4995/eurocall.2018.9080

Jones, L. (2009). Supporting student differences in listening comprehension and vocabulary learning with multimedia annotations. *CALICO Journal*, *26*(2), 267-289. https://doi.org/10.1558/cj.v26i2.267-289

Jones, L., & Plass, J. (2002). Supporting listening comprehension and vocabulary acquisition in French with multimedia annotations. *The Modern Language Journal*, *86*(4), 546-561. https://doi.org/10.1111/1540-4781.00160

Godwin-Jones, (2018). Chasing the butterfly effect: Informal language learning online as a complex system. *Language Learning & Technology*, *22*(2), 8-27. https://doi.org/10125/44643

Horst, M., Cobb, T., & Meara, P. (1998). Beyond a Clockwork Orange: Acquiring second language vocabulary through reading. *Reading in a foreign language*, *11*(2), 207-223.

Hsieh, Y. (2019). Effects of video captioning on EFL vocabulary learning and listening comprehension. *Computer Assisted Language Learning*. https://doi.org/10.1080/09588221.2019.1577898

Laufer, B. (2005). Focus on form in second language vocabulary learning. *EUROSLA Yearbook*, *5*(1), 223–250.

Mayer, R. E. (1997). Multimedia learning: Are we asking the right questions? *Educational Psychologist*, *32(1)*, 1-19. https://doi.org/10.1207/s15326985ep3201_1

Mayer, R. E. (2001). *Multimedia learning*. Cambridge, MA: Cambridge.

Montero Perez, M., Peters, E., Clarebout, G., & Desmet, P. (2014). Effects of Captioning on Video Comprehension and Incidental Vocabulary Learning. *Language Learning & Technology*, *18*(1), 118–141. http://dx.doi.org/10125/44357

Montero Perez, M., Peters, E., & Desmet, P. (2015). Enhancing vocabulary learning through captioned video: An eye-tracking study. *The Modern Language Journal*, *99*(2), 308-328. https://doi.org/10.1111/modl.12215

Montero Perez, M., Peters, E., & Desmet, P. (2018). Vocabulary learning through viewing video: the effect of two enhancement techniques. *Computer Assisted Language Learning*, *32*(1-2), 1-26. https://doi.org/10.1080/09588221.2017.1375960

Montero Perez, M., Van Den Noortgate, W., & Desmet, P. (2013). Captioned video for L2 listening and vocabulary learning: A meta-analysis. *System*, *41*(3), 720-739. http://dx.doi.org/10.1016/j.system.2013.07.013.

Nation, I. S. P., & Webb, S. (2011). *Researching and analyzing vocabulary*. Boston: Heinle.

Nguyen, T. H. N. (2021). Developing Bottom-up Listening Skills in a Google Classroom-based EFL module. *AsiaCALL Online Journal*, *12*(3), 47-57. Retrieved from https://asiacall.info/acoj/index.php/journal/article/view/45

Nguyen, V. M. (2021). English language-learning environments in COVID-19 era: EFL contexts, English-language environments, technology-based approach, English language learning. *AsiaCALL Online Journal*, *12*(3), 39-46. Retrieved from https://asiacall.info/acoj/index.php/journal/article/view/21

Perez, S. (2019). Interactive content is coming to Walmart's Vudu & the BBC. Retrieved from https://techcrunch.com/2019/04/29/interactive-content-is-coming-to-walmarts-vudu-the-bbc/

Peters, E., Heynen, E., & Puimège, E. (2016). Learning vocabulary through audiovisual input: The differential effect of L1 subtitles and captions. *System*, *63*, 134-148. https://doi.org/10.1016/j.system.2016.10.002

Pujadas, G., & Muñoz, C. (2019). Extensive viewing of captioned and subtitled TV series: a study of L2 vocabulary learning by adolescents. *Language Learning Journal*, *47*(3), 1-18. https://doi.org/10.1080/09571736.2019.1616806

Rodgers, M. P. H. (2018). The images in television programs and the potential for learning unknown words. *ITL-International Journal of Applied Linguistics*, *169*(1), 191–211. https://doi.org/10.1075/itl.00012.rod

Rosell-Aguilar, F. (2017). State of the app: A taxonomy and framework for evaluation language learning mobile applications. *CALICO Journal*, *34*(2), 243-258. https://doi.org/10.1558/cj.27623

Syodorenko, T. (2010). Modality of input and vocabulary acquisition. *Language Learning & Technology, 14(2)*, 50-73. http://dx.doi.org/10125/44214

Teng. F (2019). Incidental vocabulary learning for primary school students: the effects of L2 caption type and word exposure frequency. *The Australian Educational Researcher*, *46*, 113–136. https://doi.org/10.1007/s13384-018-0279-6

Van, L. K., Dang, T. A., Pham, D. B. T., Vo, T. T. N., & Pham, V. P. H. (2021). The Effectiveness of Using Technology in Learning English. *AsiaCALL Online Journal*, *12*(2), 24-40. Retrieved from https://asiacall.info/acoj/index.php/journal/article/view/26

Vanderplank, R. (2010). Déjà vu? A decade of research on language laboratories, television and video in language learning. *Language Teaching*, *43*(1), 1-37. https://doi.org/10.1017/S0261444809990267

Vanderplank, R. (2016). *Captioned media in foreign language learning and teaching*. London: Palgrave Macmillan.

Vidal, K. (2003). Academic listening: A source of vocabulary acquisition? *Applied Linguistics*, *24*(1), 56-89. https://doi.org/10.1093/applin/24.1.56

Wang, H.-C., & Chen, C. W.-Y. (2019). Learning English from YouTubers: English L2 learners' self-regulated language learning on YouTube. *Innovation in Language Learning and Teaching*, 1-14. https://doi.org/10.1080/17501229.2019.1607356

Webb, S., & Chang, C.-S. (2015). Second language vocabulary learning through extensive reading with audio support: How do frequency and distribution of occurrence affect learning? *Language Teaching Research*, *19*(6), 667-686. https://doi.org/10.1177/1362168814559800

Winke, P., Gass, S., & Sydorenko, T. (2010). The effects of captioning videos used for foreign language listening activities. *Language Learning & Technology*, *14*(1), 65-86. http://dx.doi.org/10125/44203

Yeldham, M. (2018). Viewing L2 captioned videos: what's in it for the listener? *Computer Assisted Language Learning*, *31*(4), 367-389. https://doi.org/10.1080/09588221.2017.1406956

Biodata

Gilbert Dizon is an associate professor at Himeji Dokkyo University, Japan and a doctoral student in the Department of Literacy, Culture, and Language Education at Indiana University. His research interests lie in computer-assisted language learning, specifically, technology-mediated informal language learning and the use of artificial intelligence in language education.

John M. Gayed is a PhD student at the Tokyo Institute of Technology researching the potential to use an AI-based writing assistant for second language learners. The digital assistant will help English language learners overcome the various cognitive barriers they face when they attempt to produce written text in English.