杏林社会科学研究
第37巻3,4合併号
2022年3月

An Intermediate Japanese Learner's Difficulties with Comprehending Authentic YouTube Videos: A Case Study

Chunmei Huang

Listening is a complex activity that involves four types of processing: neurological, linguistic, semantic, and pragmatic (Rost, 2011). Neurological processing involves hearing, consciousness, and attention. Linguistic processing requires listeners to decode the linguistic input, which involves perceiving speech, identifying units of spoken language, using prosodic features in processing speech, recognizing words, and syntactic parsing. This is known as bottom-up processing. Semantic processing refers to the integration of memory and prior experience to reach comprehension. Activating appropriate schemata and social structures is necessary for listeners to build common ground with the speaker to achieve understanding. This requires inferencing and listeners' enrichment of the input which involves reasoning. This is known as top-down processing. Pragmatic processing requires listeners to infer speakers' intention by identifying both verbal and non-verbal cues in the input and inject their own interpretation into the process of constructing meaning. Comprehension requires focused effort from listeners. According to Rost (2011), neurolinguistic changes, occurring after an L1 is acquired, and subsequent changes in motivation to

learn a new language after acquiring an L1 are the key differences between L1 and L2 acquisition. Another difference concerns access to input and interaction, which is especially important for developing listening ability. The final difference is that learning an L2 can enrich learners' life at a personal, social, cultural, and professional level. This paper addresses the problems learners encounter due to the complex skills needed to simultaneously deal with both top-down and bottom-up processing. Carney (2020) and Goh's findings concerning students' listening comprehension difficulties shed light on the problems of limited memory capacity, word recognition, L1 phonological influence, English linked speech, and background knowledge. However, the listening materials used in their studies were not real-world materials. As I regularly use authentic videos in my classroom, I investigated problems learners have with listening comprehension when watching authentic YouTube videos.

Literature Review

Difficulties L2 Learners Have with L2 Listening Comprehension

In Goh's (2000) qualitative study, she found the most reported difficulty with listening comprehension was the ability to recall what had just been heard. This might have been an effect of limited working memory. The second difficulty was the inability to recognize known words, possibly because the sound-to-script matching has not been fully automatized. The third difficulty was learners could understand the words, but not the message the speaker intended to convey. This might have been caused by a lack of background knowledge.

Carney (2020) challenged the assumption that having more lexical knowledge leads to better listening comprehension by investigating 15

learners' difficulties in comprehending 10 sentences and a narrative with orthographically known lexis. The researcher found that the participants of three proficiency levels had difficulties understanding known words in the listening materials made up of high-frequency words. At the sentence level, the difficulties were caused mainly by incorrect phonological encoding such as confusion between /r/ and /l/ and the misinterpretation of phrases such as *all for* being heard as "offer". In addition, the results from the narrative test showed inaccurate interpretation of ergative verbs such as the camera "stopped working" caused misinterpretation of the listening text. The researcher concluded that factors influencing comprehension include L1 phonological transfer, English connected speech, and of the interpretation of top-down information. In sum, both bottom-up and top-down processing can cause difficulties for listening comprehension even when lexis is familiar.

Topical Knowledge in L2 Listening

Research indicates that schemata are used by L2 learners in the comprehension process (Long, 1990). Long explored the effect of background knowledge on learners' listening comprehension among university students who enrolled in a Spanish course. The students were asked to write summaries of two listening passages of equal length in their L1 (English) after listening to them twice, and then check statements on a checklist, which comprised true and false statements from the listening passages. The summaries were scored based on the total number of idea units supplied and the number of correct idea units identified. The results showed that schemata, which were measured using a background knowledge survey, could facilitate and hurt the learners' listening comprehension. Linguistic knowledge appeared to play a prominent role in comprehension when the listeners did not have the appropriate schemata. Participants were found to

rely on linguistic knowledge in completing the checklist when they had less background knowledge about the listening passage. On the other hand, linguistic knowledge seemed not to play as important a role in comprehension when participants possessed relevant schemata. They also found that good listeners seemed to abandon linguistic knowledge in favor of the familiar schema.

Phonemic Encoding in L2 Listening

It is well-known that the acquisition of the phonological system of a second language involves considerable challenges (Rost, 2011). Perceiving and mastering L2 pronunciation and intonation imposes a difficulty on L2 listeners (Hayes, 2004). An example of this phenomenon is seen in Broersma's (2012) study, which showed that difficulties of perceiving speech for L2 listeners leads to an increased number of activated lexical competitors. He investigated how inaccurate phoneme processing affects word recognition of partially onset-overlapping pairs such as *daffodil-deficit* and minimal pairs such as *flash-flesh* for Dutch listeners. The results of the first experiment showed that there were differences in lexical activation between English and Dutch listeners. Whereas Dutch listeners experienced more activation for mismatched primes than matched primes, English listeners experienced the opposite. The results of the second experiment revealed that lexical competition led to the efficient inhibition of mismatching competitors for English listeners. For Dutch listeners, this was found to be less efficient due to more activation of lexical competitors and a lack of inhibition of those competitors. As a result of the differences in lexical processing, word recognition slows down considerably for L2 listeners. Hui and Godfroid (2020) found that higher-level processing builds on lower-level processing. Phonological strings need to be processed by learners to reach the next level

of lexical representations, which in turn contributes to the formulation of propositional meanings. In terms of the role of processing speed, they found that the formulation of propositional meanings and general listening proficiency were predicted by lexical processing speed. Because learners tended to focus on meaning during listening, faster lexical processing helps with L2 listening as meaning is carried by lexical items. Learners who were faster at lexical processing were more efficient at higher level processing as they are more efficient at resolving difficulties at the lower-level processing.

Lexical Coverage in L2 Listening Comprehension

Van Zeeland and Schmitt (2012) aimed to identify the relationship between lexical coverage and listening comprehension and if the same lexical coverage for reading and listening comprehension should be the same. They used the same four texts and comprehension questions in both listening and written modalities. The texts were controlled for length and difficulty level by ensuring that all words fell under the 2,000-frequency band using the *Vocabulary Profiler (BNC-20 version)* (Cobb n.d.). Words outside the 2,000-word frequency band were replaced with non-words. In addition, all 76 participants took the vocabulary test revised by Schmitt et al. (2001) based on Nation's (1983) Vocabulary Levels Test. The test results of the multiple-choice test used to test comprehension showed that for an adequate listening comprehension, 95% lexical coverage for informal narratives was sufficient.

Speech Rate

Griffith (1990) provided evidence that using material controlled for speech rate not only facilitates listening comprehension, but also accustoms L2 learners gradually to faster speech. He further suggested controlling speech rate can be a practical way to approach authentic materials as it provides L2

learners support instead of confronting them instantly with authentic materials. McBride (2011) investigated listening comprehension by looking into the effects of speech rate and distributed practice. She found that the group that was assigned to slow speech rate (135 wpm) outperformed the fast speed group (180 wpm), the group that had choice of speed, and the pausing group. She concluded that the slow speed aided the participants develop their listening comprehension because it did not overload their working memory and allowed them to improve bottom-up processing skills, including lexical, grammatical and pronunciation in the L2 dialogues. The benefits of the training sessions were found to transfer to faster speed as the same group made greater gains at faster speeds despite being trained on slow speed.

Using Authentic Videos

Meyers (1999) used an authentic Spanish TV program *telenovela* with Spanish learners in New Mexico University. In his study, he provided the participants with a synopsis of the episodes to assist them understand the development of the storyline and the relationship between the characters. In addition, a list of basic vocabulary and colloquialisms for the first five episodes was given to the students to help them understand the story. During video viewing, the students were asked to answer a list of 10 comprehension questions to assist them in listening selectively, which allowed them to anticipate what they were going to hear (Omaggi Hadley, 1993). To measure the impact of using *telenovela*, a pretest and a posttest measuring listening comprehension and verbal production were administered. The results showed that the participants improved significantly in both listening comprehension and oral production. Furthermore, as the participants gained better understanding of native speakers' speech, they became more comfortable with the input, which consequently made them feel more confident in speaking

Spanish and more willing to make mistakes to communicate ideas.

As indicated by the studies mentioned above, there are many difficulties concerning both bottom-up and top-down processing that L2 learners face during listening. A better understanding of the problems learners encounter would help teachers design better instruction to facilitate students' development of listening skills. Based on research findings and methods used in these studies, this study aims to answer the following research question: What problems does a learner encounter when listening to authentic YouTube videos?

Methodology

This current study was conducted with one female participant and investigated the problems she had listening to authentic video excerpts. It is a mixed methods study involving collecting, analyzing, and integrating data from the learner's self-reports and scores of summaries from each listening, including the pretest and the posttest. An initial needs assessment was conducted to identify the participant's problems with listening, and to select materials that were of interest to the participant. At the end of the study, a short interview was conducted to collect feedback from the participant.

Participant

The participant Emi (a pseudonym) is a 57-year-old Japanese female who is of intermediate proficiency level. Emi had six years of English education in the junior and senior high school in the Japanese education system and majored in English literature at a two-year college. After she graduated from college, she studied English in the United States when she was in her late twenties. In addition, she has attended an English conversation circle since

2014. The members meet once a week for two hours of English conversation lessons taught by a native speaking teacher. In general, she has had opportunities to practice all four skills in this English circle. Emi also took one extra lesson per week at an English conversation school run by the British Council for one year in 2019. According to her, the lessons had a focus on grammar and speaking skills. Outside the English circle, Emi also actively participates in volunteer work at a local non-profit organization. Therefore, she has regular contact with both English and non-English speaking foreigners living in her city or neighboring cities. At home, she discussed English studies with her two children who are university students and helped them choose schools for studying abroad. Emi has never taken any standardized proficiency test, so a Listening Vocabulary Levels Test (LVLT) (McLean et al., 2015) was given after the pretest. Her test score was 119 out of 144. Overall, in English, Emi can explain events in adequate detail, maintain a conversation, and interact with other speakers in a small group with no problems.

Procedure

The study took place over the course of eight weeks. The researcher and the participant met once a week from a needs assessment (meeting one) to the review lesson 2 (meeting seven). The review lesson 3 and the immediate posttest (meeting eight) took place two weeks after the review lesson 2 (meeting seven) due to a schedule conflict. The procedure was as follows: the pretest and the LVLT test (McLean et al., 2015) were conducted in the first meeting, followed by three diagnostic lessons, then two review lessons, and finally the review lesson 3, a posttest and an informal semi-structured interview with the participant. In addition, before listening, Emi rated topic familiarity of each video from 1 (I have never heard of this person and I

know nothing about her career) to 5 (I know who she is, what her career is, and why she is successful).

For the pretest and the posttest, following the procedure in Long's (1990) study, Emi watched the video twice without captions then wrote a summary for each video in Japanese. In addition, asking the participant to write the summary in her L1 attempted to avoid the problem of assessing the participant's listening skills by means of her writing skills as noted by Meyers (1999). During listening, Emi could take notes on the test paper. Since the test was designed to test her listening comprehension, not memory, note-taking was allowed. Moreover each video was approximately three and half minutes long, it would be impossible for her to remember details of the video without notes.

The first diagnostic lesson was a practice lesson both for the researcher and the participant as this was the first time for them to work together. In the first diagnostic lesson, vocabulary was pre-taught and background information about the interviewee Christine Lagarde was explained in class before listening. During listening, Emi took notes of the key points. After listening, we discussed the content of the video using the questions prepared in advance. Summary writing was assigned as homework. After this first lesson, some adjustments were made based on Emi's feedback on the lesson and the need to collect data to assess her listening comprehension. Because the summary was written after the lesson, it was excluded from the data analysis. For the second and third diagnostic lessons, a list of vocabulary including multiword sequences was prepared and given to Emi two days before the lesson. Emi could study these vocabulary items and wrote Japanese translation on the list to help her learn these words. In the beginning of each lesson, the words on the list were reviewed and Emi was given time to ask questions about the vocabulary followed by an introduction

on the interviewee. After that, Emi took notes while listening to the video twice, then wrote a summary for each video in Japanese. After the first two listenings, Emi watched the video and paused every time she could not understand. I discussed with her the cause of the communication breakdown. In addition, I checked her comprehension of the video by asking her to summarize or translate parts of the video into Japanese. Both lessons ended with a short discussion to help her consolidate the information in the videos. The lessons were recorded, and the data were then transcribed and coded for further analysis.

The review lessons were designed based on the problems identified in the diagnostic lessons. The purpose of the three review lessons was to help Emi deepen her understanding of the videos and address the main problem she had with listening comprehension i.e., word recognition. In each lesson, we reviewed the words on the list before Emi watched the video. During listening, five comprehension questions were prepared to help her listen selectively. The video was played in three to four segments depending on the length. Emi could pause any time she wanted to rewatch certain parts of the video. After listening, we looked at her answers and discussed the parts that she did not understand. To help her focus on certain words and grammar structures that she had problems with during the diagnostic lessons, a cloze exercise was prepared using parts of the transcript. A second purpose of this listening while reading activity was to help her build sound-to-script relationships, which was a problem identified during the diagnostic lessons. At the end of each lesson, the entire transcript of each video was given to her as homework. Emi could listen and read simultaneously to help her make a connection between the written and spoken forms of words. In the following lesson, she could then ask questions about the transcript.

Materials

The listening texts were selected to meet Emi's interest in learning about successful female leaders in the world and her interest in authentic videos, a type of material that she had never used before. The five listening texts used in this study were taken from YouTube (https://www.youtube.com/). The videos were interview excerpts taken from the video series *Success with Moira Forbes* hosted by Moira Forbes, the publisher of *ForbesWomen*, a multi-media platform that aims to serve successful women in business and leadership in the world. To control the length of each video excerpt at the same time considering the natural conversation flow, each excerpt was limited to Moira Forbes asking the interviewee three questions. These questions vary in each video, but generally, they are related to the interviewee's life, career, her opinions about leadership, and her advice to other women. To facilitate bottom-up processing, the speech rate was adjusted to 75% of the initial speed of all five videos. This was tested with Emi to ensure the speed was not too fast for her. Two videos were used for the pretest (*Accenture CEO's Advice To Women: Stand Out*) and the posttest (*Atlantic Records' Main Woman*), the same three videos were used for the diagnostic lessons and the review lessons (*Christine Lagarde On The Challenges Of Traversing Different Careers, Ina Garten On Her Unconventional Culinary Career, Christy Turlington's Fresh Perspective*). The speech rate was calculated using the word count divided by the adjusted video length. See Table 1 for details of each video. Because each interviewee has her own speech style, the speech rate varies by person even after adjusting the speed of the video. As Table 1 shows, the non-adjusted speech rate ranges from 146 wpm to 206 wpm and the adjusted speech rate ranges from 114 wpm to 154 wpm. This variation of speech rate could not be controlled due to the nature of authentic

listening materials.

Table 1

Listening Materials

Video	Video length	Word count	Non-adjusted speech rate	Adjusted length	Adjusted speech rate (wpm)
Pretest video	3:45 m.	503	146	4:39 m.	114
Diagnostic/review video 1	3:30 m.	556	168	4:36 m.	127
Diagnostic/review video 2	4:04 m.	835	207	5:22 m.	150
Diagnostic/review video 3	4:22 m.	785	186	5:50 m.	142
Posttest video	3:44 m.	633	184	4:11 m.	154

In terms of lexical coverage, transcripts were analyzed using with Cobb's (2020) *Compleat Web VP*, version 2.5, using the BNC/COCA 1-25k list (lextutor.ca/vp/comp/). Table 2 presents the lexical coverage of the video transcripts of the first 3,000-word frequency band. As shown in Table 2, at the 2,000-word frequency band, the lexical coverage exceeds 95%, a threshold for adequate comprehension as found in Van Zeeland and Schmidt's (2013) study. Judging by Emi's LVLT test scores (119/144), she should know most of the words in the first 3,000-word frequency band. Emi made no mistakes in Part 1 and 2 (1,000 and 2,000- word frequency band), five mistakes in Part 3 (3,000-word frequency band), three mistakes in Part 4 (4,000-word frequency band), seven mistakes in Part 5 (5,000-word frequency band), and 11 mistakes in Part 6 (academic word list). Her test results were used to prepare the instructional materials in the lessons.

Table 2

Video Lexical Coverage

Video	Frequency Level	Cumulated Token (%)
Pretest video	K1 – K2	95.6%
	K1 – K3	98.5%
Diagnostic/review video 1	K1 – K2	96.4%
	K1 – K3	97.1%
Diagnostic/review video 2	K1 – K2	96%
	K1 – K3	97.6%
Diagnostic/review video 3	K1 – K2	96.4%
	K1 – K3	98.9%
Posttest video	K1 – K2	96.4%
	K1 – K3	97.9%

Analysis

Summaries from the pretest, second and third diagnostic lessons, and the posttest were graded based on the correct number of main ideas and minor details given. Because there are three interview questions and answers in each video, there are three main ideas in each video. For each correct main idea identified, three points were given and for each correct minor detail identified, one point was given. For main ideas that were partially correct, one point was deducted. As to minor details, no partial points were given for incorrect answers.

For the qualitative data, the recordings were reviewed, and a data log was created to record any mention or description of problems with listening. When hearing such reports, a short summary was written in the data log. I then categorized these reports into types of problems Emi reported during the lessons such as phonemic encoding of words, unknown words or multiword

sequences, and grammatical processing. After creating the data log, the total number of each problem was tallied. Finally, only the parts of the data concerning listening problems were transcribed and reported in this paper.

Results

Summary Scores

The summaries were scored based on the number of main ideas and minor details correctly identified. Table 3 presents the number of ideas correctly and partially identified and the total score for each video. Based on the results, one can see that Emi understood the general ideas of each video to a certain extent. However, when looking at the minor details, we can see that she did not have a full understanding of the videos except for the pretest video. According to Emi, this was because the interviewee in the pretest video talked about her childhood experience which does not require much background knowledge to understand. The main ideas that Emi either did not understand or completely skipped in her summary writing were about the interviewee's leadership role, her decision to change career trajectory, the non-government organization (NGO) she started, and business strategies she adopted, topics about which Emi reported not having much knowledge. Emi got the lowest summary scores for the diagnostic video 3, which was an interview with the super model Christy Turlington who started an NGO that advocates for maternal health. Emi reported not understanding most of the interview because of a lack of background knowledge and the vocabulary being too difficult.

Table 3

Summary Scores

Lesson	# of main ideas correctly identified	# of minor details correctly identified	Total score
Pretest	1.5	4	5+4=9
Diagnostic 2	1.5	2	5+4=7
Diagnostic 3	1	1	3+1=4
Posttest	2	2	6+2=8

Judging by the scores obtained in the pretest and the posttest, we can see that listening comprehension did not improve over six lessons. This was expected as six lessons are very short. In addition, all the interviewees and topics were unfamiliar to Emi as she gave a rating of 1 for topic familiarity (i.e., I do not know this person and I know nothing about her career) for all the videos. To understand each interview, a certain level of topical knowledge about the industry and the interviewee is required. This type of knowledge is highly specific, therefore, is not transferrable to a new context.

The summary scores reveal Emi's overall understanding of each video. In the next section, specific problems Emi had with listening comprehension will be reported.

Problems with Listening Comprehension

As expected, the participant had problems with both top-down and bottom-up processing. Given that they often interacted each other, it was challenging to categorize the problems. Therefore, problems that involved more than one category were categorized into one category based on what the participant reported as the main problem. For example, the phrase "since its inception" caused problems because of the /s/ phoneme, which is difficult for Japanese

learners to perceive as it does not exist in Japanese phonological system and is caused by the modification of connected speech in English. It was categorized under linked speech instead of as a word recognition problem because the word inception was pre-taught, and Emi reported it as a problem with connected sounds. The other challenge concerns the word recognition category, which contains unfamiliar or unknown words, partial word knowledge, inaccurate phonemic encoding, and unknown or unfamiliar multiword sequences. This is because these problems all impede word recognition. In terms of background knowledge, it also includes a few subcomponents such as a lack of background knowledge, interference of preexisting knowledge, and cultural differences. Because there were few occurrences for each category, for the convenience of categorization, they were combined into one category coded as background knowledge.

Table 4 shows that the top problem that impeded listening comprehension was word recognition. There is a total of 17 occurrences, followed by grammatical processing, background knowledge, and linked speech.

Table 4

Listening Comprehension Problems

Listening problem	# of occurrences
Word recognition	18
Grammatical processing	7
Background knowledge	5
Linked speech	2

Word Recognition

Regarding word recognition, six of the occurrences concerned inaccurate phonemic encoding of orthographically known words caused by L1

phonological influence. For example, the word "fresh" was difficult for the participant to perceive after at least five listening because of the difference between katakana pronunciation フレッシュ (fureshu) and English. Other inaccurate phoneme encoding of combinations that do not have equivalents in Japanese such as /pr/ as in "prepare" was misheard as "repair", /p/ and /b/ sounds caused mishearing between "business" and "place", and "saw" was heard as "sold" were also found to impede listening comprehension. The following occurrence demonstrates how L1 phonological influence affected Emi's listening comprehension:

Emi: For me, I hear "fash", "fash" or something, so "fash" ? not "fresh" ……I don't know why, but I hear "fash", "fash" ……

R: Because "fre" are connected together. It's very common.

E: Ah, and also "fresh" is Japanese. So, in English, *fresh*?

R: Yeah, r and e, we always connect together, in Japanese, you don't, right?

E: In Japanese fresh is フレッシュ.

R: It's "fureshu", "fureshu", right?

E: Yes. So, fresh is not フレッシュ for me.

R: So, it's katakana? Katakana sound?

E: Yeah, katakana sound is フレッシュ, but real English sound is fresh. *fresh* is not フレッシュ for me. Ha ha…

R: So, it was the katakana……

E: Yeah, yeah, yeah. Katakana sound is really difficult.

Due to L1 phonological influence, it took a lot of effort to make connection between the sound and the script of "fresh" during listening. This type of inaccurate phonemic encoding significantly slowed down word recognition as suggested by Broersma (2012).

In addition to inaccurate phonemic encoding, there were also six

occurrences caused by unknown or unfamiliar multiword sequences. Multiword sequences in the three videos include colloquialisms such as "hit the wall" and "wear different hats" and specific word combinations used by the interviewer and the interviewee such as "primary delight" and "entrepreneurial endeavor". Even though some of the multiword sequences or individual words in the sequences were pre-taught, when listening to them in a stream of words, they could be very difficult to comprehend, especially when one of the words is an unknown word and the other word is partially known such as "practicing attorney". After the participant knew the meaning of attorney, she confirmed her understanding by asking "….and practicing attorney means to become a lawyer?". As can be inferred from this question that Emi knows only one meaning of *practice* as "to do something again and again to become better at it" (Meriam-Webster Learner's Dictionary n.d.), not the meaning of "to have a professional medical or legal business" (Meriam-Webster Learner's Dictionary n.d.). Another excerpt shows how even known multiword sequences could cause comprehension problems:

> E: I am sorry, I thought I heard like wear? ……So instead of the r (caterer) or something, wear is also the word you use for career?
>
> R: Oh, "you wear different hats", you mean? That's in the beginning, "you wear so many different hats".
>
> E: Ah~ ……I heard wear, so.. wear? Cook? Wear, cook? Ah~ OK, OK.
>
> R: You wear different hats……cook, caterer, business owner, cookbook writer…
>
> E: ……I missed this……I didn't know why she said wear. Ah~ wear different hats. …It's OK, OK.

It again took a lot of effort for Emi to hear this multiword sequence and make sense of it in the context of this interview.

Relating to the excerpt mentioned earlier, partial knowledge of a word and

unknown or unfamiliar words were another source of problems with listening comprehension, which account for three occurrences each. Emi's listening comprehension was often impeded due to not knowing different parts of speech of known words such as "bored" and "boredom" and "engagement" and "engage".

> E: ….and I thought specialty and special are no difference. So, I thought why specialty? It must be special food store, so, this specialty means t-e-a tea……So I thought special tea store. So, I am confused…

In this excerpt, we can see how not knowing the noun of "special" could lead to misinterpretation of the listening text. We can also see the steps of Emi's semantic processing.

Grammatical Processing

There were seven occurrences of problems concerning grammatical processing. The problems were related to inaccurate interpretation of relative clauses, the subject pronoun "they", and not fully understanding the function of imperative verbs. Emi encountered problems with grammatical processing when the sentences were longer and the grammar was more complex as in this sentence "…the one that I was best prepared for was to be a practicing attorney, to be, to be a lawyer, to operate with clients, and to manage team later on". She had trouble identifying the subject in the relative clause and what the relative clause is supposed to modify in this long sentence. In addition, this long sentence also confused her interpretation of the interviewee's job. She said "…so the one that she was best prepared for was to be a practicing attorney, to be a lawyer, to operate with clients, and to manage team later on. Three? It's one, two, three (Emi was pointing to the transcript while speaking) ?". It seems longer sentences slowed down her processing speed and caused problems with comprehension. Another example

is sentences with "would" and "could" used to explain hypothetical situations as in "The deal I made with the woman who sold me the store was that she would stay with me for a month and teach me what I needed to know. Of course, thinking that a month would do it and immediately it was over my head". Emi was not sure if the interviewee meant one month would be enough or not. Based on her confirmation question "…for one month she could do it?...from her face, it was the opposite (the interviewee was laughing when she said this)", we can see that she could not interpret the meaning and the subtle nuance of 'would' in this context. However, the interviewee's facial expressions gave her a clue that the meaning might be different from what she had thought. Therefore, she wanted to confirm her understanding with me.

Interpreting subject pronouns was challenging for her as in this video excerpt: "It's been my experience in almost every forum in other group settings where I might be with policy makers or other advocates. It's an interesting, like they also see, they recognize me that…". The question Emi asked after listening was "they" means different country's women?" as the conversation involved the NGO which advocates for maternal health. In addition, some women's photos appeared in the video. Therefore, she associated "they" with women from other countries instead of processing the grammar cues in these sentences. The visual in the video obviously had caused confusion in this context.

Background Knowledge

Background knowledge appeared to be problematic despite a brief introduction of the interviewee and her career being given in the beginning of the lesson. The information might have been too new and too unfamiliar for Emi. Often, she was not sure what the interviewee was talking about, for

example, after listening to the part about Christy Turlington's NGO, she reported that she did not understand NGO even though it was explained: "NGO? Does it cover uni..... does it cover all over the world?" as some pictures of women who look like they are from different countries were shown in the video. Hearing about NGO from me is one thing, seeing the images and listening to it is in a wider context is another. The other problem Emi had was caused by preexisting background knowledge about hitchhiking, which she associates with cars only. As a result, when the interviewee was talking about hitchhiking on a motorcycle, she had trouble understanding that. It appears a lack of background knowledge as well as preexisting background knowledge could both interfere with listening comprehension. However, they can also motivate her to learn new knowledge and information. In the interview with Emi at the end of the study, she said she enjoyed learning about these interviewees' career and life because she could learn some new ideas that help her educate her children. For example, she learned from Ina Garten's interview that having fun is important to have a successful career and she learned that encouraging children to take risks can help them become successful leaders from Christine Lagarde's interview.

Linked Speech

There were two occurrences of linked speech causing comprehension problems. One example is "on the list since its inception" as in:

> E: …Even I listened to this part "on the list since its inception" several times, about 10 times, but I couldn't catch up "list since its inception". I hear, I heard different ……
>
> R: What did you hear?
>
> E: ……" since its", hmm, "since its", I couldn't "inception" ……maybe listin..listing or something I heard ……

R: Yeah, yeah, because these two words are connected.

E: Yeah, yeah, yeah, connected.

As can be seen, processing connected sounds in addition to processing the phoneme /s/ in such a short phrase was challenging for the participant. Even after 10 listenings, she was still not able to understand the meaning of this phrase. The other example is "I saw an ad" as in "And I saw an ad for business for sale in New York Times and it was a specialty food store in a place I've never been before". Emi reported not being able to hear "an ad" because both words are short, when connected, it was difficult for her to catch the sounds. I advised her to keep doing reading while listening as we did in the review lessons to help her get used to listening to linked speech with the help of the transcript which shows the written forms of these words. Emi said this activity was very helpful for her when she watched the video by herself at home. She was happy that with the help of the transcript, she was finally able to understand everything in the video. Before the lesson, she had never thought that she would be able to learn English using authentic materials. After these lessons, she started to think that she can keep learning English using authentic videos by herself.

Discussion

The aim of this study is to gain a deeper understanding of problems the participant encountered with listening to authentic YouTube videos. The results showed that the participant encountered many listening problems that involve both bottom-up and top-down processing. In addition, the participant did not make improvement in the posttest. One of the findings that the participant was not able to recognize known lexis during listening is in line with Carney's (2020) and Goh's (2000) study findings. In addition, linked

speech and L1 phonological influence also appeared to cause comprehension problems as found in Carney's (2020) study with high school students. Phonemic encoding seemed to be the main problem that caused word recognition during listening supports Broersma's (2012) finding about inaccurate phonemic encoding affects word recognition speed negatively. The finding that bottom-up processing problems can lead to misinterpretation of the entire sentence supports Hui and Godfroid's (2020) study findings, which state higher-level processing builds on lower-level processing and lexical processing speed predicts L2 listening proficiency. In this study, the participant's problems with word recognition might have affected her listening comprehension. In terms of the role of background knowledge, the participant was found to rely on her schemata when comprehending the listening text instead of using her linguistic knowledge to process the text, which is in line with Long's (1999) study findings that when learners have relevant schemata, linguistic knowledge plays a less prominent role in listening comprehension. Moreover, Long's (1999) finding that existing schemata can both harm and benefit learners' listening comprehension was also found in this study. The participant's preexisting knowledge about hitchhiking was found to impede her comprehension of the video. With regard to the effect of using authentic material in promoting learners' motivation, as in Meyer's (1999) study, the participant in this study also reported using YouTube videos being motivating, and her confidence increased as she could understand and learn new knowledge from them using English. It seems reasonable to conclude that a higher level of linguistic knowledge, topic related schemata, and the ability to efficiently process information in L2 are necessary for comprehending authentic listening materials.

Conclusion

The findings revealed problems learners of intermediate proficiency level could encounter when listening to authentic videos, including word recognition, grammatical processing, background knowledge, and linked speech. However, this study has limitations regarding the study design. First, there is only one participant in this study. The results might not be generalizable to other learners. A second limitation concerns the testing instrument. Summary writing reflects the test-taker's general comprehension of the listening texts, but it does not indicate specific comprehension problems the participant had. In addition, the scoring can be subjective in terms of identifying main ideas versus minor details. A test with multiple choice questions or short answers might allow more objective assessment on listening comprehension. Furthermore, the speech rate was not properly controlled due to technical limitations. Some interviewees spoke faster than others, which caused comprehension difficulty as reported by the participant in the post-study interview. Finally, videos provide extra visual aid for the participant's listening comprehension. As reported by the participant, some interviewees' facial expressions gave her clues when interpreting the meanings of their interviews while the pictures shown in one video interfered with her listening comprehension. Effects of the visual aid on listening comprehension should have been taken into account. For future research, this case study can be replicated with a larger sample size, better designed instruments, and better controlled conditions.

References

Broersma, M. (2012). Increased lexical activation and reduced competition in second-language listening. *Language and Cognitive Processes, 27*(7/8), 1205–1224.

https://doi.org/10.1080/01690965.2012.660170

Carney, N. (2020). Diagnosing L2 listeners' difficulty comprehending known lexis. *TESOL Quarterly*, 1–32. https://doi.org/10.1002/tesq3000

Cobb, T. (2019). Compleat Web VP vol 2.1 [Computer program]. https://www.lextutor.ca/vp/comp/

Goh, C. (2000). A cognitive perspective on language learners' listening comprehension problems. *System, 28*(1), 55-75. http://dx.doi.org/10.1016/S0346-251X(99)00060-3

[ForbesWomen], (2021, May 8). Accenture CEO Julie Sweet's Advice To Women: Stand Out | Success With Moira Forbes [Video]. YouTube. https://youtu.be/IWo15V2BrfM

[ForbesWomen], (2020, Sept. 5). Christine Lagarde On The Challenges Of Traversing Different Careers | Forbes [Video]. YouTube. https://youtu.be/1IPjE2SPpXQ

[ForbesWomen], (2014, Oct. 31). Ina Garten On Her Unconventional Culinary Career | Success With Moira Forbes [Video]. YouTube. https://youtu.be/1AfNwlXe5tI

[ForbesWomen], (2013, Nov. 11). Christy Turlington's Fresh Perspective | Success With Moira Forbes [Video]. YouTube. https://youtu.be/jAlGo7vJfCo

[ForbeWomen], (2021, May 6). Atlantic Records' Main Woman | Success With Moira Forbes [Video]. YouTube. https://youtu.be/dSph8SUbOjo

Griffith, R. (1990). Facilitating listening comprehension through rate-control. *RELC Journal, 21*(1), 55-65. https://doi.org/10.1177/003368829002100104

Hayes, B. (2004). Phonological acquisition in Optimality Theory: The early stages. *Constraints in Phonological Acquisition*, 158-203. https://doi.org/10.1017/CBO9780511486418.005

Hui, B., & Godfroid, A. (2020). Testing the role of processing speed and automaticity in second language listening. *Applied Psycholinguistics*, 1–27. https://doi.org/10.1017/S0142716420000193

Long, D. R. (1999). What you don't know can't help you: An exploratory study of background knowledge and second language listening comprehension. *Studies in Second Language Acquisition, 12*(1), 65-80. https://doi.org/10.1017/S0272263100008743

McBride, K. (2011). The effect of rate of speech and distributed practice on the development of listening comprehension. *Computer Assisted Language Learning*, 24, 131–154. https://doi.org/10.1080/09588221.2010.528777

Meyers, J. R. (1999). The effect of authentic video on communicative competence. *The Modern Language Journal*, *83*(3), 339-349. https://doi.org/10.1111/0026-7902.00026

McLean, S. & Kramer, B. & Beglar, D. (2015). The creation and validation of a listening vocabulary levels test. *Shiken*, 19(2), 1-28. https://doi.org/10.1177/1362168814567889

Oberauer, K. (2019). Working memory and attention—A conceptual analysis and review. *Journal of Cognition, 2*(1), 1–23. https://doi.org/10.5334/joc.58

Omaggio Hadley, A. (1993). *Teaching language in context* (2nd ed.). Boston: Heinle.

Rost, M. (2011). *Teaching and researching listening* (2nd ed.). Harlow, England: Pearson Education.

Van Zeeland, H. & Schmidt, N. (2012). Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension? *Applied Linguistics, 34*(4), 457-479. https://doi.org/10.1093/applin/ams074

Merriam-Webster. (n.d.). Practice. In *Merriam-Webster Learner's Dictionary*. Retrieved August 3rd, 2021, from https://learnersdictionary.com/definition/practice