A Dissertation Submitted to the Hiroshima University Graduate Board

Factors Affecting Aural Decoding Ability for Japanese EFL Learners

広島大学大学院　教育学研究科　博士課程後期

教育学習科学専攻　教科教育学分野　英語教育学領域

Kriss Alexander Lange

Acknowledgements

I would like to express my gratitude to several people whose generous instruction, guidance and support made the completion of this dissertation possible. I'd like to sincerely thank Dr. Tsuido my original advisor, who is now retired, for the kind and patient support he has shown me over this five-year project despite his extremely busy schedule and various commitments continuing into retirement. Also, I'd like to sincerely thank my current advisor Dr. Nishihara for his careful attention to my research and the various requirements for the completion of this degree. In addition to my main advisors, I was very fortunate to have been instructed by Dr. Fukazawa, Dr. Tatsukawa and Dr. Yanase as well. Their assistance was of great value to me during my Thursday visits from Matsue to Hiroshima to participate in seminar class. Many thanks to all of my advising professors and the excellent support they provided for me at Hiroshima University.

I would also like to thank my friend and colleague Dr. Matthews who took a chance on trusting me, a stranger at first, when I reached out by email to him for help in 2018. He kindly agreed to meet with me online every two weeks despite his busy schedule. Although well-known and accomplished as a researcher, particularly in the field of L2 listening, he kindly accepted me as a fellow researcher worthy of collaborating with on studies together. His support and influence has fundamentally shaped my approach to research. I'm fortunate to have learned so much from him through the process of writing and publishing research articles together. His positive approach to research and collaboration has certainly helped me through some difficult times. I truly and thankfully owe him a huge debt of gratitude.

I'd also like to warmly thank all of my colleagues at University of Shimane Matsue Campus who have taken on various extra responsibilities for me as well as generally supported me throughout this doctoral research process. I hope I can make it up to them by taking on more responsibilities on campus from now on. At home, I'm ever grateful to my wife for her tireless patience and loving support. Her and our two sons make all the effort and time devoted to this degree worthwhile. Last but certainly not least, a big thank you to my family in the US who have been cheering me on since long before I can remember and still continue to do so now.

2021/07/05

Kriss Lange

Table of Contents

Abstract

Chapter 1: Introduction

1.1 Background and Context

1.2 Research Questions

1.3 Dissertation Structure

1.3.1 Overview

1.3.2 Research Study 1

1.3.3 Research Study 2

1.3.4 Research Study 3

1.3.5 Research Study 4

Chapter 2: Literature Review

2.1 Overview

2.2 Listening comprehension processes

2.3 Models of listening comprehension processes

2.3.1 The decoding and meaning building model

2.4 Listening comprehension and L2 vocabulary knowledge

2.4.1 L2 aural vocabulary knowledge and L2 listening comprehension

2.5 Decoding and L2 listening comprehension

2.5.1 L2 decoding difficulties

2.5.1.1 Impact of phonological modification on decoding and listening comprehension

2.5.1.2 Impact of function words on decoding

2.6 Conclusions and dissertation aims

Chapter 3: Research Study 1

"Analyzing Difficulties in Aural Word Recognition for Japanese English Learners: Identifying Function Words in Connected Speech"

3.1 Introduction

3.2 The current study

3.3 Research questions

3.4 Materials and methods

3.5 Results

3.6 Analysis

3.7 Conclusion

Summary

Chapter 4: Research Study 2

"Exploring the Relationship that Listening Comprehension has with Bottom-Up Skills and Metacognitive Awareness"

4.1 Introduction

4.2 The current study

4.3 Research questions

4.4 Materials and methods

4.4.1 Participants

4.4.2 Instruments

4.4.2.1 Eiken Pre2 level test

4.4.2.2 Clear Listening Diagnostic Test

4.4.2.3 Metacognitive Awareness Listening Questionnaire

4.4.3 Procedure

4.5 Results

4.5.1 Listening comprehension performance

4.5.2 Bottom- up listening skills assessment

4.5.3 Metacognitive awareness assessment

4.6 Conclusions and implications

Summary

Chapter 5: Research Study 3

"Exploring the relationship between L2 vocabulary knowledge, lexical segmentation and L2 listening comprehension"

5.1 Introduction

5.2 The gaps in the literature and purposes of the study

5.3 Method

5.3.1 Participants

5.3.2 Instruments

5.3.2.1 The Listening Vocabulary Levels Test

5.3.2.2 Paused Transcription Tests

5.3.2.3 TOEIC listening test

5.3.2.4 Eiken Pre-2 listening test

5.3.3 Procedure

5.3.3.1 Data and analysis

5.4 Results

5.4.1 Addressing Research Question 1 – What is the strength of association between the variables that were measured?

5.4.2 Addressing Research Question 2 – To what degree do the variables measured predict L2 listening comprehension?

5.5 Discussion

5.5.1 Pedagogical recommendations

5.5.2. Limitations and future research

5.6 Conclusion

Summary

Chapter 6: Research Study 4

"Analyzing Trends in the Aural Decoding Errors of Japanese EFL Learners"

6.1 Introduction

6.2 Research questions

6.3 Materials and methods

6.3.1 Participants

6.3.2 Instruments and procedures

6.3.2.1 Quantitative test instruments

6.3.2.2 Measures of listening comprehension

6.3.2.3 Qualitative Data Elicitation Protocols

6.3.3 Data analysis

6.4 Results

6.5 Discussion and conclusions

6.6 Limitations and future directions

Chapter 7: Discussions and recommendations for future research

7.1 Overview

7.2 Summary of key findings

7.2.1 Study 1

7.2.2 Studies 2 and 3

7.2.3 Study 4

7.3 Pedagogical implication

7.3.1 Encourage explicit instruction for decoding and collocation

7.3.2 Encourage the utilization of top-down processing for meaning

7.3.3 Encourage more exposure to authentic connected speech

7.3.4 Encourage the use of testing to diagnose listening difficulties

7.3.5 Focus on acquiring aural vocabulary knowledge of high-frequency words in connected speech

7.3.6 Encourage the use of intelligible English pronunciation

7.4 Limitations and future research

7.4.1 Study 1

7.4.2 Studies 2 and 3

7.4.3 Study 4

7.5 Concluding remarks

Reference List

Appendices

Abstract

Developing the ability to understand speech in a foreign language is a difficult task for the second language (L2) learner. Central to this process is the ability to identify words in spoken language and understand their meanings, also known as decoding. Decoding is the fundamental process of breaking up the speech stream into recognizable words making listening comprehension possible. It is a process in which the listener identifies word boundaries, recognizes words, and establishes a literal meaning of a spoken utterance (Field, 2008a). Although decoding is an essential skill for listening comprehension, not enough L2 listening research attention has been focused on understanding its relationship to listening comprehension and the listeners' comprehension processes while decoding. The studies included in this dissertation primarily investigate the decoding abilities of Japanese university EFL students and the relationships that bottom-up skills, metacognitive awareness and aural vocabulary knowledge have with listening comprehension. Bottom-up skills refer to the capacity to perceive linguistic information from the speech signal. Metacognitive awareness refers to one's knowledge of cognitive processes and listening strategies while listening. Aural vocabulary knowledge refers to knowledge of the acoustic representations for words. The basic aim of these studies is to understand the particular English decoding difficulties that are representative of native Japanese learners of English and to determine how decoding skills might be effectively developed for these learners. In order to help fill the gaps in the L2 literature in regard to decoding for Japanese learners, this dissertation presents four research studies which aim to address the following three main research objectives:

1) To investigate the decoding ability of Japanese EFL learners.
2) To examine the relative importance that various factors associated with listening ability have on L2 listening.
3) To identify common sources of decoding difficulties for Japanese EFL learners.

The results for Objective 1 indicate that decoding ability may be limited due to phonological modification which normally occurs in connected speech. Decoding ability for function words was better when words were presented in citation form than when presented in their standard, deemphasized weak forms. In regard to Objective 2, findings from correlation and regression analysis show that decoding ability has a stronger relationship with listening ability than with aural vocabulary knowledge. Finally, results for Objective 3 presented quantitative and qualitative assessments to identify sources of decoding difficulty through a mixed-methods approach. The analysis identified four trends in decoding errors for Japanese EFL learners: limited collocation familiarity, syntactic knowledge constraints, difficulties utilizing co-text, and L1 phonological influence.

Chapter 1: Introduction

1.1 Background and Context

Second language (L2) listening, or the ability to comprehend acoustic input from L2 speech, is a fundamental skill for L2 learners. It promotes the development of other language skills (Rost, 2012; Vandergrift, 2003) and is an essential skill in itself for communication. However, many L2 learners consider listening as the most difficult skill in language learning (Graham, 2006). Keeping up with the speed of the input, segmenting and decoding words correctly, effectively using working memory to quickly process the input, parsing and interpreting meaning for prosodic features such as stress and intonation are some of the reasons given for its difficulty (Vandergrift and Baker, 2015). Listening skills are also difficult for the language teacher to teach because of the "internal" and mental nature of the task which limits the teacher's ability to confirm whether the input was interpreted correctly (Vandergrift, 2007). This has contributed to a "product-focused" approach to listening instruction that resembles testing rather than pedagogy (Sheerin, 1987). Just focusing on the product, i.e. extracting the "answer" from the listening input, "does little to help students understand and control the process leading to comprehension." (Vandergrift, 2007, page 191).

More research attention on understanding listening processes and diagnosing deficiencies in listening comprehension ability is needed in order to provide learners with better listening pedagogy (Field, 2003). One such process in need of further investigation is decoding, a skill which is considered to be critical for supporting listening comprehension ability (Rost, 2012; Hulstijn, 2002; Broersma & Culter, 2008). Decoding involves the listener being able to identify the boundaries between words, recognize words, and establish a literal meaning of a spoken utterance (Field, 2008a). Fluent decoding is necessary for rapidly associating spoken language with word meanings (Hulstijn, 2002). Unlike written texts, spoken language demands the rapid processing of transient acoustic input and therefore requires much concentration for the L2 listener to comprehend. Limited automaticity while decoding has been identified as a significant factor limiting listening comprehension (Goh, 2000). Although it is a critical aspect of listening comprehension, decoding processes and their relationships with listening comprehension has been given relatively little research attention (Broersma & Cutler, 2008). More research into this area is needed to gain a deeper understanding of the difficulties that L2 listeners face while decoding so that pedagogical interventions can be made to rectify them.

1.2 Research Questions

The collection of research presented here is primarily motivated by the need to understand the difficulties that Japanese EFL learners face when decoding connected speech and the impact this has on their listening ability.

Three primary research questions guide the research conducted for this dissertation.

1) To what extent can Japanese university students decode function words from connected speech?
2) What is the strength of association that L2 listening comprehension ability has with awareness of

metacognitive skills, ability to decode function words, knowledge of aural vocabulary and decoding ability.

3) What sources of decoding difficulty can be identified through tests of decoding ability and analysis of introspective self-observation protocol responses.

1.3 Dissertation Structure

1.3.1 Overview

This dissertation begins with a literature review (Chapter 2) to provide the context and background for the studies it contains. Next, four research studies are each presented in separate chapters (Chapters 3 to 6). The section below summarizes each of the research studies. In Chapter 7, findings from each of the research studies are discussed with reference to the three primary research objectives along with pedagogical implications, limitations of the studies and future research implications.

1.3.2 Research Study 1

Research Study 1, which is entitled, *"Analyzing Difficulties in Aural Word Recognition for Japanese English Learners: Identifying Function Words in Connected Speech"* is presented in Chapter 3. This study investigated Japanese listeners' difficulties in recognizing function words in connected speech. The results help to demonstrate the degree of difficulty this task presents to listeners as well as some defining characteristics of the words which were the most difficult to decode. A partial transcription test was administered to a cohort of first- and second-year Japanese university students (N = 29). In the 30-item test, each item aurally presented sentences of five to six words which contained a function word. On the test paper, the first and last words of the sentence were written and students were instructed to transcribe the three to four missing words that were aurally presented. For example, the listeners heard, "I need to talk to you." and attempted to transcribe the four missing words in the test item "I _____ you." However, as this study focuses on function words, only decoding ability for the two instances of the function word *to* were measured. The same students took this same test once more with the exception that the aurally presented sentences were spoken in citation form, without the attributes of connected speech (i.e. phonological modification). By comparing the results from both tests (e.g. citation form and connected speech) inferences can be drawn about the degree of difficulty that attributes of connected speech presented to the listeners. Results showed that only 77.8% of the words presented with connected speech were transcribed whereas 96.3% of the words presented in citation form were transcribed. These results suggest that attributes of connected speech, such as weak form function words, were partially responsible for decoding difficulties. The findings also provide an indication of the degree of difficulty learners have with decoding weak form function words in connected speech as compared with un-connected speech, highlighting the importance of this research focus. Furthermore, these findings help in identifying particular characteristics of some function words, such as voiceless initial phonemes (e.g. /h/), that make them particularly difficult for listeners to decode.

1.3.3 Research Study 2

Research Study 2, which is entitled, *"Exploring the Relationship that Listening Performance has with Bottom-Up Skills and Metacognitive Awareness"* is presented in Chapter 4. This study investigated the strength of relationship between listening comprehension and bottom-up skills in addition to the relationship between listening comprehension and metacognitive awareness evident from a cohort of 54 first-year Japanese university students. The results indicated that the bottom-up skill of decoding deemphasized function words in a partial dictation task was the strongest correlate with listening test scores. This study employed the Clear Listening Diagnostic Test to measure bottom-up skills, the Metacognitive Awareness Listening Questionnaire to assess utilization and awareness of listening strategies, and the Eiken Pre-2 grade listening section to measure listening comprehension, in an effort to explore these relationships. The results demonstrated a significant and moderately strong correlation between listening ability and one of the bottom-up skills tested, deemphasized function word transcription (r = .50). In comparison, the relationship between listening ability and metacognitive awareness was weak (r_s = .21), suggesting further evidence for the greater importance of bottom-up processes for listening development. These results help to legitimize the further investigation of bottom-up skills and their relationship to listening ability which is explored in this dissertation.

1.3.4 Research Study 3

Research Study 3, which is entitled, *"Exploring the Relationships between L2 Vocabulary Knowledge, Decoding, and L2 Listening Comprehension"* is presented in Chapter 5. Based on evidence from Chapters 3 and 4, indicating the important roles of decoding in listening ability, this study aimed to ascertain these relationships more rigorously through correlational and multiple regression analysis. This study further investigated the relationships between listening ability, using two measure of listening comprehension, and measures of aural receptive vocabulary knowledge and decoding ability with 130 first-year Japanese university students. Decoding ability was assessed through a paused transcription test. Paused transcription testing is thought to provide a more accurate assessment of decoding errors than dictation because the listener may more fully utilize top-down as well us bottom-up processing skills. The findings from Research Study 3 suggest that decoding words from connected speech had a stronger relationship with two measures of listening ability (r = .39 and .51) than aural vocabulary knowledge had (r = .15 and .12) as measured by the total scores of the Listening Vocabulary Levels Test. However, scores from the first 1,000-word level of the Listening Vocabulary Test were the strongest predictive variable, independently predicting 22% of variance in TOEIC listening scores and 21% of variance in Eiken Pre-2 listening scores. Regression analyses also indicated that in combination, aural knowledge of vocabulary at the first 1,000-word level and decoding ability could predict 34% and 38% of total variance observed in TOEIC listening and Eiken Pre-2 listening scores respectively. These findings provide evidence that decoding ability, or the ability to decode multiple words

from connected speech, is strongly associated with listening ability and worthy of further investigation and analysis.

1.3.5 Research Study 4

Research Study 4, which is entitled, *"Analyzing Trends in the Aural Decoding Errors of Japanese EFL Learners"* is presented in Chapter 6. This study presents an in-depth analysis of sources of decoding errors using both quantitative and qualitative analysis. Decoding errors were elicited with the same paused transcription test utilized in Research Study 3 along with an introspective self-observation protocol which was used to more accurately determine the sources of decoding errors for 63 first-year Japanese university students. Quantitative findings showed that decoding ability was quite limited for this cohort with only 45.1% of the target words correctly transcribed from connected speech. Supported by the quantitative data, the qualitative findings indicated that four main trends in decoding difficulty could be identified from this study: limited collocation familiarity, syntactic knowledge constraints, difficulties utilizing co-text, and L1 phonological influence. These results help to identify trends in decoding errors for Japanese EFL learners and discuss ways in which decoding ability could be developed based on the findings.

1.4 Ethics

The 4 research studies outlined in section 1.3 were carried out with students attending the University of Shimane, Matsue Campus and were conducted with approval from the Ethics Committee of the university.

Chapter 2. Literature Review

2.1 Overview

The literature review which follows is divided into six sections. This first section describes the content of the literature review. The second section focuses on listening comprehension and the cognitive processes that are fundamental to decoding. The third section presents some of the models of L2 listening which are helpful in conceptualizing this construct. These models provide a theoretical context that is helpful for describing the problems that L2 listeners experience as well as characteristics of skilled and less-skilled listeners. The fourth section focuses on the important role that vocabulary knowledge and aural vocabulary knowledge in particular have on L2 listening. The fifth section focuses on decoding and refers to studies demonstrating the importance of decoding for L2 listening as well as difficulties that L2 listeners face when decoding connected speech. The sixth section summarizes the conclusions of the literature review and justifies the studies presented by indicating how they help to fill the gap in research within this area.

2.2 Listening comprehension processes

Listening comprehension refers to the cognitive processes that utilize linguistic and non-linguistic sources of information to derive meaning from aural input (Buck, 2001). These sources of information are accessed, rapidly processed, and interpreted all while continuing to listen to subsequent acoustic input. Various cognitive processes are performed concurrently while listening such as phoneme discrimination, accessing lexical candidates, syntactic and semantic processing, schema activation, and pragmatic considerations all with the purpose of converting speech into meaning. In order to create meaning from spoken input, the listener must be able to identify words and map them to known words in the lexicon (Hulstijn, 2002; Luce & Pisoni, 1998; Rost, 2012). The lexicon refers to the total knowledge of words including their corresponding syntactic, semantic, morphological, orthographic, and phonological aspects that a language user can access (Jiang, 2000; Levelt, 1989). Accurate aural vocabulary knowledge is particularly important because it allows the listener to quickly match recognized words with words in the lexicon, thereby accessing the linguistic knowledge associated with those words. Aural vocabulary knowledge is the implicit knowledge of acoustic representation for words in the lexicon (Jones & Witherstone, 2011). Decoding cannot occur without sufficient aural vocabulary knowledge to initiate the process of matching lexical entries to speech.

As mentioned in the introduction, decoding is the process of matching speech sounds to known words, resulting in a literal interpretation of the speech input; therefore accurate perception of phonemes (i.e. perceptually unique units of speech sounds) is fundamental to decoding. A number of researchers have suggested that recognition patterns of L1 phoneme acquisition are similar to those of the L2 (Best & Tyler, 2007). L2 learners first learn to detect and analyze the prosodic patterns and phonemes of the second language before recognizing vocabulary words and building their emerging lexicons. As those lexicons grow, the learner is challenged to recognize and identify more and more phonetically similar words such as minimal

pairs (a pair of words which differ by only one phoneme). This process eventually leads to better phoneme discrimination ability, word recognition and thus better decoding ability. This view of L2 speech learning, known as the perception-first view, assumes that phoneme perception drives the learning process leading to production ability (Flege, 1995; Kuhl, 2000). The Speech Learning Model (SLM), developed by Flege (1995), also provides useful insights on how new L2 phonemes are acquired. The model claims that the learner initially may not perceive the distinction between two similar phonemes because the sounds have been combined into a single category that is influenced or biased by the learner's L1. Therefore, phonologically similar sounds, rather than differing sounds, pose more of a challenge to accurate perception due to this phenomenon of lumping similar sounds together. For example, the English /l/ - /r/ distinction is often perceptually categorized by Japanese English learners as the nearest L1 counterpart (Japanese tap /ɾ/) for both of these phonemes thereby masking their difference (Guion, Flege, Ahahane-Yamada, & Pruitt, 2000). As in the perception-first view, the Speech Learning Model recognizes the importance of attention to phonetically similar sounds and words while recognizing the effects of L1 influence. Essentially, the model claims that decoding skills develop through accurate and increasingly nuanced phoneme perception.

Thus, decoding depends on accurate phoneme perception which drives the selection process for lexical entries to be associated with the speech input. Lexical entries with similar phonemes may both be activated in the mental lexicon in the initial stages of decoding (McCelland & Elman, 1986). As more phonemes are perceived, the listener eventually has enough acoustic information to select the appropriate lexical entry candidate from those activated (Rost, 2012). However, there is also evidence that decoding may not proceed in a linear manner driven by perception of individual phonemes. Research by Morais *et al.*, (1979) speculates that recognition may occur at the unit of the syllable based on L1 studies with illiterate participants who had difficulty removing phonemes from the beginning of words. The study suggested that word recognition may occur based on units larger than the phoneme. Some degree of phoneme perception, however, must occur in order for a syllable to be recognized and utilized for decoding. Both accurate phoneme perception as well as aural vocabulary knowledge are critical factors which support decoding ability and thus listening comprehension.

Cognitive listening comprehension processes, which enable decoding ability, have been briefly discussed in this section. These processes are a product of frequent exposure to language. Through the perception of phonemes, which lead to the development of lexical entries containing semantic information, gradually the ability to automatically process speech input for meaning emerges (Hulstijn, 2002). This process of language acquisition for L1 listeners as well as L2 listeners depends on extensive exposure and utilization of linguistic input (Trofimovich, 2008). However, unlike the native speaker, the L2 listener typically has far fewer opportunities for engaging with L2 speech and thus has limited stores of knowledge in the lexicon associated with their lexical entries (Jiang, 2000). In particular, insufficient aural vocabulary knowledge appears to be a key limiting factor for L2 listening comprehension (Goh, 2000).

2.3. Models of listening comprehension processes

Cognitive models have been proposed to explain the processes involved in listening, from speech perception to meaning creation, and speculate on how they interact. Anderson's three phase model defines successful listening as the coordination of perception, parsing and utilization skill components (Anderson, 1995). The first stage, perception, involves segmenting speech into meaningful units such as phonemes and words. Next, parsing refers to the identification of pragmatic and grammatical units from the input and their association to linguistic information in the listener's mental lexicon. Then, the utilization stage involves combining non-linguistic knowledge such as background knowledge (i.e. schemata) with linguistic knowledge, the speech input understood by the listener, in order to create meaning. One limitation of this model pointed out by Graham and Macaro (2008) is that it assumes a linear direction from perception to utilization.

A fundamental concept used to describe the listening process is the notion of top-down and bottom-up processing which broadly describes the direction in which information is processed; either from general to specific in the case of top-down or from specific to general in bottom-up processing (Flowerdew and Miller, 2005). Bottom-up processes utilize discrete linguistic information such as phonemes and syllables and joins them to form larger units of meaning such as words and phrases (Field, 2008a). This dissertation conceptualizes bottom-up processing with Field's (2013, p. 97) model of lower-level processes in listening (see Figure 1). This model utilizes Cutler and Clifton's (1999, p.124) blueprint of the listener, a respected model of L1 listening comprehension as well as Field's (2008a) insight and research expertise in listening comprehension. According to the model, acoustic cues in the speech signal are processed via the interactive stages of phonological knowledge, lexical knowledge and syntactic knowledge.

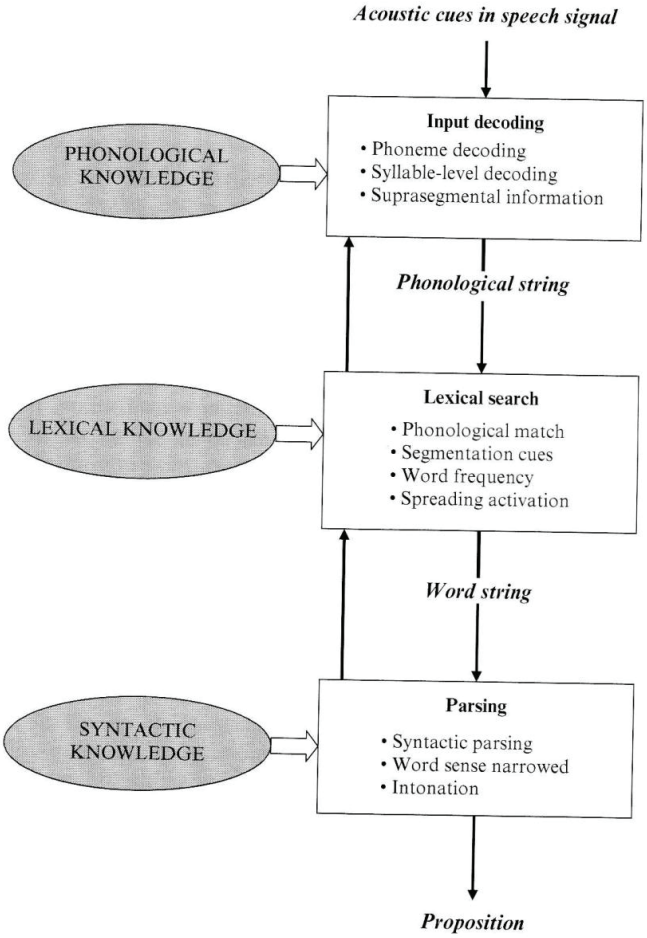

Figure 1

Model of lower-level processes in listening, drawing upon Cutler and Clifton 1999 and Field 2009. *Research and practice in assessing second language listening* (p. 97), by A. Geranpayeh & L. Taylor (Eds.), 2013, Cambridge, England: Cambridge University Press

Although the model appears to operate from top to bottom with the speech signal progressing to a phonological string, then a word string and finally the proposition, bottom-up processes are not always linear in nature. Field (2013) notes that corrected interpretation of information from a higher level (i.e. the word string) can override incorrectly processed information at a lower level (i.e. the phonological string) as reflected by the upward arrows from each of the three stages. The decoded acoustic input is used in the Lexical Search stage to identify words (i.e. decoding). In the Parsing stage, the semantic relationship between words in the phonological string is determined. Word meaning is further refined during parsing using suprasegmental information such as intonation. According to Cutler and Clifton (1999) syntactic and semantic processing is "guided by a language user's knowledge of the structure of his or her language, together with specific structural information made available by the particular words in a sentence" (p. 141).

These bottom-up processes serve to represent a literal interpretation of the speaker's words which is abstract and context independent. Thus, application of top-down processes to create meaning from the decoded input is also required.

Top-down processes utilize non-linguistic information, background knowledge or pre-existing knowledge to help establish meaning from spoken input. It is defined as the "form of language processing that bases inferences on expectations and predictable generalizations cued by the incoming language." (Rost, 2012, p. 284). Top-down processes are essential to creating meaning and are employed when a listener uses knowledge of context and background knowledge to predict words that have not been perceived. For example, those proficient in a language can often guess the next word in an utterance based on the context (Dijkgraaf et al., 2016). If I were to say, "The killer lifted the gun, pointed it at the man and …" they may be able to predict that "fired" or "shot" is likely to be the next word. Both L1 and L2 listeners rely on top-down processing and the local linguistic environment to identify words that could not be decoded or that were partially decoded from speech input with bottom-up processes. The influence of sentence context on decoding ability for L1 speakers was investigated by Norris, Cutler, McQueen and Butterfield (2006) using psycholinguistic priming experiments. The results showed that decoding occurs more accurately and rapidly when there is a semantic priming effect present in the context of the target word thus illustrating the importance of top-down processes for creating meaning.

Listeners' knowledge of strategies that utilize top-down processing, such as using context to guess unknown words, is generally referred to as metacognitive awareness (Vandergrift et al., 2006). Metacognition broadly refers to thinking about thinking and has been defined generally as "knowledge and cognition about cognitive phenomena" (Flavell, 1979, p. 906). It is comprised of the ability to understand and regulate cognitive processes and is thought to play an important role in listening (Vandergrift & Goh, 2012). Research using think-aloud methodology, in which the thought processes of listeners are investigated while listening, has found that skilled listeners use more metacognitive strategies than less-skilled listeners (Vandergrift, 2003). Metacognitive awareness was found to be a significant predictor of listening scores, accounting for 13% of variance, in a study by Vandergrift et al. (2006) with 966 participants. Knowing and utilizing strategies that encourage top-down processing appears to be associated with better listening ability. In the following model of listening proposed by Vandergrift and Goh (2012), perception, parsing and utilization are the three fundamental cognitive processes which interact with top-down, bottom-up and metacognitive processes occurring simultaneously (see Figure 2). Both bottom-up and top-down processes are essential for decoding listening input and usually operate concurrently. Listeners may focus more on one type of process than the other depending on their purposes for listening (e.g. detailed or general focus), extent of background knowledge, working memory capacity and listening proficiency (Vandergrift & Goh, 2012).

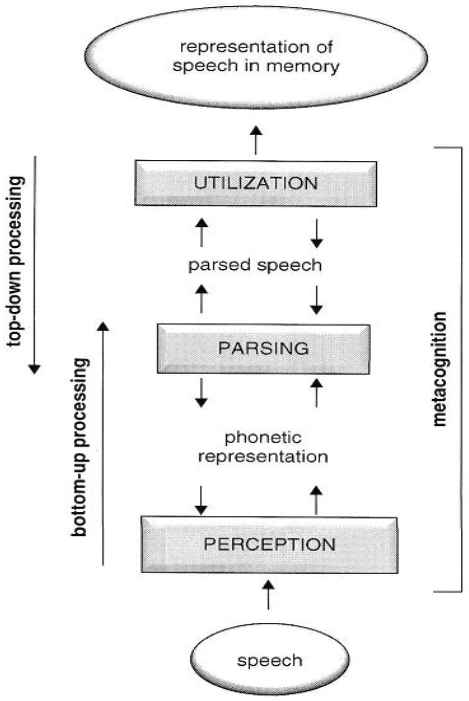

Figure 2

Model of cognitive processes in L2 listening and their interrelationships. *Teaching and Learning Second Language Listening: Metacognition in action* (page 17). By L. Vandergrift & C. C. M. Goh, 2012, New York, NY: Routledge

As shown in Figure 2, the process of transferring speech input to its representation in memory involves interactive processes between perception and parsing and between parsing and utilization which are all supported by metacognition occurring throughout the listening process. Although bottom-up and top-down processing are directionally opposed, it is important to remember that both processes occur interactively and are essential to listening comprehension. Top-down processing is facilitated by contextual knowledge and bottom-up processing depends on the recognition of linguistic items from the spoken input. The interdependent nature of top-down and bottom-up processes in listening comprehension is widely accepted by researchers (Buck, 2001; Cutler, 2012; Rost, 2012).

Although the process of listening and decoding can be explained by models which are similar for L1 and L2 listeners, there are a number of advantages when listening to one's first language. First the ability to discriminate between similar phonemes is generally more developed for L1 listeners. This helps to drive accurate decoding processes while listening as described in section 2.2. The L1 listener is also helped by phonotactic constraints which facilitate parsing the speech stream. Phonotactic constraints refer to the language-specific restrictions in phoneme sequences of syllables (McQueen, 1998). For example, in English

the phoneme /w/ cannot be followed by a labial phoneme such as /b/, /p/ or /v/ and although word onsets may contain the clusters /pr/ and /br/ as in "price" and "break" they are not allowed as word endings (i.e. coda) (adapted from Al-jasser, 2008, p. 95). Native listeners are implicitly aware of these phonotactic constraints and fluently apply them to constrain word candidates while decoding speech (Cutler, 2012). L2 listeners, however, appear to apply their L1 phonotactic knowledge while listening to the L2 as shown by wordspotting tasks (Weber & Cutler, 2006). This is likely to impair decoding and listening ability since the phonotactic constraints of languages differ. Similar to phonotactic knowledge is knowledge of patterns stress or rhythm in words and sentences. In English, roughly 90% of content words have initial stress (Cutler & Carter, 1987). Such knowledge can be helpful to listeners for identifying word boundaries (i.e. segmenting speech). However, as English words contain both strong and weak syllables, the learner needs to be able to differentiate between them. This is especially difficult for Japanese EFL learner's whose first language doesn't utilize strong and weak syllables to the extent that English does. Finally, aural vocabulary knowledge, or the ability to recognize the phonetic forms of words, is another important advantage that the native listener possesses.

2.3.1. The decoding and meaning building model

As we've discussed, listening consists of two broad categories of operation: bottom-up processing which is perceptual and top-down processing which is conceptual (Field, 2008a). The bottom-up perceptual operations decode the linguistic data from the spoken message by perceiving phonemes, syllables, words, and syntax (Field, 2013) and the top-down conceptual operation creates meaning from these raw materials. This process is referred to as the meaning building processes, in which the listener "contextualizes the words that have been decoded" (Field, 2008c, pp. 368-369). Successful comprehension may not always proceed from decoding to meaning building because meaning can be partially established in other ways that aren't dependent on linguistic cues (e.g., interpreting paralinguistic cues, guessing, background knowledge, etc.). However, successful listening comprehension is fundamentally initiated and consistently guided by aural decoding (Field, 2008a, 2008b; Tsui & Fullilove, 1998). Decoding words is of prime importance because they are the linguistic forms that can be reliably matched to meaning. Hulstijn (2002) states that the form-meaning quality of words "makes them more amenable to conscious, metalinguistic reflection than formal units … such as phonemes and grammatical constructions respectively" (p. 204). This means that decoding processes which allow the listener to fluently identify words and the boundaries between them are particularly important for meaning building (Al-Jasser, 2008; Broersma & Cutler, 2008).

Previous studies measuring various constructs of aural decoding strongly suggest these processes are important for successful L2 listening comprehension. For instance, Andringa et al., (2012) evaluated the decoding ability of 113 speakers of Dutch as an L2 and the results showed that decoding accuracy and L2 listening comprehension were correlated strongly (r = .61). In another important study, Tsui and Fullilove (1998) analyzed listening test results from several hundred thousand English language learners in Hong

Kong over a period of seven years. Two versions of a multiple-choice listening test were administered in which one version used 'matching' schema items and the other used 'non-matching' schema items. The sematic relationship between initial and subsequent sections was congruent throughout the matching schema items. Listeners could rely on contextual information learned in the initial section to select the correct answer. However, in the non-matching schema type items the initial sections were not semantically congruent with subsequent sections. In order to correctly answer these questions, participants needed "to be able to continuously process the subsequent input and revise the initial schema accordingly in order to get the correct answer" (Tsui & Fullilove, 1998, p. 439). Results indicated that correct scores for non-matching item types predicted listening test performance better than correct scores on the matching types. Rather than primarily relying on contextual information established from the initial sections of audio, skilled listeners were able to change their meaning building processes based on newly decoded information in subsequent sections. In another study which investigated listeners' ability to revise initial top-down knowledge, Field (2008b) applied a gating procedure to present various ambiguous stems to both expert (native) and non-expert (non-native) listeners. As more of the syllables were heard, a key finding was that in contrast to expert listeners, non-expert listeners were far more reluctant to alter their previous decoding hypotheses even after subsequent input, which contradicted their prior decoding decisions, was presented. These results support the hypothesis that the primary characteristic of a skilled L2 listener is "accurate and automatic decoding [rather than] the ability to make use of context" (Field, 2008a, p. 136), and also aligns strongly with the central findings of Tsui and Fullilove (1998), namely that decoding capacities are of primary (but not sole) importance in skilled L2 listening.

2.4 L2 Listening comprehension and L2 vocabulary knowledge

A key factor affecting decoding ability is the listeners' L2 vocabulary size (e.g. number of words known) and depth (e.g. amount of knowledge about the word) of knowledge. Measures from written vocabulary tests have been shown to have a relatively strong and robust relationship with measures of L2 listening comprehension across a range of learning contexts. For example, Stæhr (2009) investigated the strength of association between advanced Danish EFL students' L2 listening comprehension and vocabulary size (Vocabulary Levels Test; Schmitt, Schmitt & Clapham, 2001) and vocabulary depth (Word Associates Test; Read 1993, 1998) finding that they correlated strongly and significantly ($r = .70$ and $r = .65$ respectively). The generalizability of these results for L2 vocabulary size and L2 listening comprehension was further demonstrated by Andringa et al. (2012). This study investigated factors determining L2 listening comprehension among 113 non-native Dutch speakers with 35 different first language groups, and they found that scores from a 60-item receptive L2 vocabulary test correlated strongly with L2 listening comprehension ($r = .69$). Receptive L2 vocabulary knowledge, measured orthographically in various contexts, appears to have a moderate to strong relationship with L2 listening comprehension.

2.4.1 L2 aural vocabulary knowledge and L2 listening comprehension

The robust relationship between L2 listening comprehension and receptive L2 vocabulary knowledge as measured with written vocabulary tests is relatively well established. However, researchers have tended not to use measures of aural vocabulary knowledge (Stæhr, 2009). This is likely because the available vocabulary tests have only been orthographic tests (Milton, 2013). This tendency is a significant limitation (Stæhr, 2009; Vandergrift & Baker, 2015) as scores from L2 aural vocabulary tests are more strongly associated with L2 listening comprehension than scores from written L2 vocabulary tests (Bonk, 2000). In a study conducted within the Chinese tertiary EFL context with 250 participants, Cheng and Matthews (2018) demonstrated that scores from aural vocabulary tests were more strongly correlated with listening (r = .71) than scores from comparable written vocabulary tests (r = .55).

Other research which has explored the relationships between L2 listening comprehension and other factors associated with it, has also shown a strong link between listening scores and L2 aural vocabulary knowledge. For example, Vandergrift and Baker (2015) investigated which learner variables best predicted L2 listening comprehension among 157 learners of French. They measured a number of variables including receptive aural L2 (French) and L1 (English) vocabulary knowledge, L1 and L2 listening ability, auditory discrimination ability, working memory and metacognition. The strongest correlate with L2 listening proved to be L2 vocabulary knowledge (r = .51). This finding was obtained across three cohorts of learners and was more than double that of all other variables that reached a statistically significant level of correlation (L1 vocabulary, r = .23; metacognition, r = .23; and auditory discrimination, r = .22). Matthews and Cheng (2015) demonstrated that partial dictation test scores which measured knowledge of high-frequency words were correlated strongly with IELTS listening test scores among a cohort of 167 tertiary level Chinese EFL learners (r = .73). McLean, Kramer and Beglar (2015) demonstrated that their Listening Vocabulary Levels Test, which requires test-takers to process aural stimulus material, correlated strongly (r = .54) with parts one and two of the listening component of the Test of English for International Communication (TOEIC). Finally, in the Japanese EFL context, Wallace (2020) examined the relationship between various factors such as aural L2 vocabulary knowledge, metacognitive awareness, memory, attentional control, self-reported topical knowledge, and L2 listening. Results of the structural equation modelling analysis indicated that vocabulary knowledge had the strongest total effect on listening performance. These studies have helped to demonstrate the significant relationship that aural receptive L2 vocabulary knowledge has with L2 listening comprehension ability.

2.5 Decoding and L2 listening comprehension

The research reviewed above demonstrates that there is a relatively strong relationship between aural vocabulary knowledge and L2 listening comprehension across a range of contexts. However, a limitation of previous studies is that they have only measured individual words, and have not measured the capacity to segment multiple words in connected speech. This gap is important to address as spoken language is almost

always delivered in concatenated intonation units (Rost, 2002). Connected words are often acoustically very different from their discrete citation form due to phonological modification (e.g., linking, assimilation, elision, etc.). For this reason, being able to accurately segment strings of connected lexis is an important objective for L2 listeners, and is indicative of high levels of listening proficiency (Field, 2008b). Investigations of L2 learners' capacities to segment and extract meaning from samples of connected speech suggest that phonological modification is strongly associated with listening ability (Field 2008a; Lange, 2018). For example, Sheppard and Butler (2017) used paused transcription tasks to investigate the capacity of 77 L2 learners to segment strings of four or five words in connected speech. Results indicated that only 67% of the words were correctly transcribed.

Other research by Wong et al. (2017) showed that decoding speech containing phonological modification was the strongest correlate with listening (r = .63) from among several others measured such as knowledge of written vocabulary (r = .50) and minimal pairs discrimination (r = .32). These studies suggest that the ability to segment words in connected speech and specifically to mitigate the effects of phonological modification plays an important role in L2 listening. As previously mentioned, Andringa et al. (2012) explicitly addressed the relationship between segmentation ability and L2 listening comprehension demonstrating a strong association (r = .64). However, segmentation was assessed by the test-takers' ability to accurately count the number of words in a string of target speech. Therefore, the method did not directly measure the recognition of specific word forms in connected speech, which is an important factor in L2 listening. In contrast, test formats such as paused transcription can be used to measure a learner's capacity to segment and decode sequences of multiple words presented in connected speech (Field, 2008c). Importantly, such tests can cast light on practical questions, such as which test-takers perceive "… *attracts investment* …" as "… *a tax investment* …" (Matthews & O'Toole, 2015, p. 371) and which recognize "… *don't always notice* …" as "… *don't always know this* …" (Sheppard & Butler, 2017, p. 92). This information is not provided by tests that measure knowledge of single target vocabulary items. For this reason, data gathered from tests that measure knowledge of both single and multiple vocabulary items are likely to offer useful insight into the lexical capabilities of L2 listeners, and how these relate to listening comprehension success.

2.5.1 L2 decoding difficulties

Decoding is especially challenging for L2 learners as authentic spoken language is typically not produced as discrete phonological word forms, but mostly as streams of connected, phonologically modified lexis. Words within fluent speech become co-articulated, with adjacent phonemes influencing each word's phonological form (Field, 2008a). Additionally, the speech signal is transient making it necessary for the listener to segment words rapidly, with an average rate of native speech reaching over six syllables per second in English (Pellegrino, Coupé, Marsico, 2011). Decoding is a complex skill, "which requires a context-sensitive representation of phonemes and phoneme clusters both within and across word boundaries" (Hulstijn, 2003,

p. 420). Considering these challenges, it is unsurprising that decoding of connected speech causes considerable difficulty for L2 listeners (Field, 2008b).

According to Hirai (1999), there are three main problems that less proficient learners face in comprehending auditory input. First, they generally have insufficient syntactic and lexical knowledge to access while listening. Second, lack of familiarity with spoken language results in less automaticity for decoding words and phrases. Third, incorrect phonological knowledge of vocabulary impairs aural recognition even when the learner may be able to recognize the words orthographically. The combination of these factors presents the L2 listener with a formidable challenge when decoding connected speech. We can learn something about the process of listening and its challenges from research that explores the EFL learners' own perceptions of their listening difficulties. Goh (2000) investigated the most frequently occurring English listening comprehension problems for 40 lower-level Chinese students and categorized them based on Anderson's three phase model for L1 listening (Anderson, 1995). This model shares the same three fundamental categories as Vandergrift and Goh's (2012) model; perception, parsing and utilization. Five of the ten problems were connected with perception (phoneme and word recognition), three were problems with parsing and two involved utilization difficulty. The results showed that 22 out of 40 participants reported not recognizing words which they already knew. The most commonly reported perceptual problems were as follows: do not recognize known words, neglect the next part while thinking about meaning, inability to "chunk" streams of speech (decoding), miss the beginning of the spoken input, and concentrate too hard or unable to concentrate. These are mainly problems which seem to indicate difficulties related more to perception and bottom-up processing skills rather than top-down skills.

The most commonly reported listening problem, reported by 26 out of 40 participants, was a tendency to forget what was previously heard while listening so that a global meaning of the section could not be formed. Listeners reported difficulty in connecting decoded input with subsequent input. This could be due to many factors such as limited working memory, the inability to fluently map lexis to acoustic input, syntactic knowledge, limited aural vocabulary knowledge or the cognitive burden of listening. The second most common problem for all of the participants concerned word recognition. Even though students reported knowing the words in orthographic form, they could not recognize them aurally. Goh states, "Many of them said they learnt by memorizing the spelling of words and often neglected to remember how the words sounded." (Goh, 2000, p. 61). Listeners cannot automatically and fluently recognize spoken words unless their mental representations of the words' phonological forms (i.e. aural vocabulary knowledge) are accurate (Segalowitz & Hulstijn, 2005). These findings illustrate the importance of perceptual skills in relation to aural vocabulary knowledge. Orthographic knowledge of a word does not ensure that it can be decoded from spoken English.

In light of the important role that decoding plays in L2 listening, researchers have sought to identify and

categorize specific aural decoding difficulties in L2 listeners. In a study which analyzed paused transcription test results among university students in an intensive English program, Sheppard and Butler (2017) identified four categories of decoding errors: word segmentation, misperceived phonemes leading to incorrect word choices, unknown words/phrases, and top-down fabrications. Although informative, this study determined these categories of decoding error by deduction from transcription error without consulting the learners themselves. Cross (2009) explored common decoding errors for advanced Japanese listeners when listening to short BBC news videos in English. Participants were asked to listen to the videos twice and take notes after each listening. Afterwards, they wrote a summary to demonstrate their understanding of the content and data was collected from these responses for 73 adults. The results showed that 81% of the responses had some form of breakdown in comprehension due to decoding error. Based on Field's (2008a) procedural and textual categories of decoding error, Cross categorized errors in segmenting input or other process-related errors as *procedural* and errors due to unknown vocabulary or insufficient knowledge of the language as *textual*. He also added a third category labelled *intrusion* to account for L1 influence leading to errors in decoding. An example of this type of error is the assimilation of both /b/ and /v/ into the Japanese /b/ leading to aural misperceptions of the word "van" as "bank, bag, bar or back". After analyzing various examples of common decoding errors, Cross conceded that it was very difficult to identify specific reasons for particular errors and that they likely occurred due to a combination of process, text and intrusion problems. For example, in the clause, "you're letting the side down" 7 of the 39 errors decoded "the side" as "decide". Cross reasoned that this could simply be due to a process error in segmenting the reduced form "the", however, he noted that other examples of decoding errors for this clause appeared to be due to intrusion or textural problems. Influence from the loanword collocation "size down", used in Japanese for something decreasing, may explain why 16 of the 39 errors decoded "side down" as "size down". Text error could also be a factor because the idiom in which this clause appeared "let the side down" is very low frequency and decoding error could also be due to insufficient linguistic knowledge. As this example illustrates, the sources of decoding error are varied and difficult to identify. Despite the difficulty in identifying specific reasons for decoding errors, Cross gives pedagogical suggestions for topics to address: awareness of cliticization (i.e. attaching a word or an abbreviated form of a word to the end of a previous one as in contraction) and resyllabification (i.e. linking or liaison), discrimination of consonants /v/ and /b/, /l/ and /r/, discrimination of vowel sounds /ai/, /au/, and /æ/, and confirming or revising word choices using other evidence.

2.5.1.1 Impact of phonological modification on decoding and L2 listening comprehension

Phonological modification, or the characteristic ways in which phonemes are modified in connected speech, imposes another complicating factor on L2 listening comprehension processes. Phonological modification obscures word boundaries by modifying adjacent phonemes as in assimilation and elision as well as deemphasizing function words which convey important grammatical information. These variations in pronunciation occur due to the efficiency principle, whereby speakers tend to use the least amount of energy

for communication, as well as coarticulation which naturally happens when speakers move their articulators (teeth, lips, tongue, jaw) from one position to another. This results in pronunciation that may be significantly different in connected speech than in its isolated, citation form.

Types of phonological modification have been classified in various ways by researchers, however, the three most important phonological modifications as described by Buck (2001) are assimilation, elision and intrusion. Assimilation happens when sounds influence the pronunciation of other adjacent sounds. For example, the /t/ and /tʃ/ are combined so that "won't you" is usually pronounced like "wonchoo". Elision refers to the omission of phonemes to make pronunciation easier. For example, the /t/ phoneme is elided in "next day" so that it is usually pronounced like "nexday". Intrusion describes the introduction of new sounds between other sounds. For example, when saying "please do it" there is a /w/ sound between "do" and "it" like this "please do-w-it" (Adapted from Buck, 2001). He also writes that weak forms are another type of phonological modification that makes recognition of grammatical function words especially difficult in connected speech.

Studies comparing decoding ability in the presence or absence of phonological modification have shown that phonological modification can limit transcription accuracy by approximately 50% (Ernestus, Dikmans, & Giezenaar, 2017; Henrichsen, 1984). Wong et al. (2017) also investigated the role of phonological modification on connected speech comprehension with a cohort of 60 advanced level (mean IELTS band score of 6.5) university students in Hong Kong using a phonological modification dictation test, adapted from Henrichsen (1984). The 33 target sentences used in the dictation test each contained one of nine types of phonological modification (e.g. contraction, vowel weakening, elision, etc.). Results from six variables related to listening ability showed that phonological modification dictation ability had the highest correlation with general listening comprehension, (r = .63), followed by receptive vocabulary (r = .50), non-word repetition (i.e. working memory) (r = .44), speech gating (r = .41), phonemic awareness (r = .38) and minimal pairs discrimination (r = .32). In a speech gating task, listeners hear an utterance that has been divided into segments and are asked to predict how the utterance would continue after each segment (Thorpe & Fernald, 2006).

2.5.1.2 Impact of function words on decoding

Another type of phonological modification which is especially difficult for L2 listeners is the tendency of function words to take a weak form, or be deemphasized, in connected speech. Function words refer to words which signify grammatical relationships such as articles (a, the), auxiliary verbs (have, can, will), personal pronouns (he, you, we), possessive adjectives (her, my, their), demonstrative adjectives (this, that, those), prepositions (on, in, under) and conjunctions (than, and, or) (Celce-Murcia et al., 2016). Function words happen to be very frequent in everyday speech. According to Field (2008a) about 80 of the 100 most frequently spoken words in the British National Corpus are function words. Content words such as nouns and adjectives are given more attention so they are generally stressed in connected speech. Function words usually take a

weak form in connected speech, making them much more difficult to decode. According to Field (2008a), "weak forms may differ from the full forms in four ways: weak vowel quality, loss of phonemes, lack of stress and short duration" (p.146). L2 listeners may have difficulty perceiving known words in connected speech because of these phonological modifications (Goh, 2000; Henrichsen, 1984). Indeed, function words have been shown to be more difficult to transcribe than content words for L2 listeners (Field, 2008c; Lange, 2018; Sheppard & Butler, 2017). Native listeners can use their knowledge of the language to recognize unstressed function words whereas non-native listeners are often unable to (Field, 2008b). We could argue that perceiving function words is not critically important for comprehension as long as the content words (nouns, verbs, adjectives, etc.) are recognized. This may be true in some cases, but function words can drastically alter the meaning as in these two phrases: "I'm looking for the pictures" and "I'm looking at the pictures" (Adapted from Field, 2008b). Function words may seem insignificant but they fulfil the necessary task of conveying important syntactic information. It is reasonable to expect that improving recognition for function words may help L2 listeners improve their listening skills. According to Field (2008a), "Function word recognition is another area where intensive practice will pay dividends" (p.146).

Function words occur very frequently and contain important grammatical information but they are difficult to perceive due to their weak forms in connected speech. Field (2008c) demonstrated that function words were not identified as well as content words by 46 L2 listeners of various L1s. The participants took paused transcription tests which assess the ability to decode authentic, connected speech. In paused transcription tests, the aural text has pauses inserted at irregular intervals and is only played once. When listeners come to a pause, they try to recall and transcribe the phrase that was just heard (usually 4-5 words) from memory. This method is thought to be an effective means of measuring decoding ability because listeners are able to utilize the context through top-down processes as well as rely on their bottom-up perceptual skills to complete the transcriptions (Field, 2008c). Results from this study indicated that, "In most cases, mean percentage recognition for function words was approximately 20% lower than for content words" (p. 424). Field (2008c) also noted that difficulty in recognizing function words seemed to be independent of the listeners L1 or level of proficiency. In another study utilizing paused transcription tests to investigate content and function word decoding ability, Sheppard and Butler (2017) also showed that L2 listeners perceived and transcribed significantly fewer function words than content words. The 77 participants, consisting mostly of Chinese L1 students at a university-based intensive English program, were able to transcribe 76% of the content words and 54% of the function words in the paused transcriptions. These finding align with Field's (2008c) study which showed that recognition for function words was about 20% lower than it was for content words.

A study with Japanese L1 students has also indicated that weak forms of function words in connected speech can be difficult for learners to perceive. Hirose (2007), investigated the ability of 25 junior college students to perceive function words in either weak or strong form in connected speech. The study used a partial dictation

test in which participants listened to 11 sentences ranging from four to ten words and provided one to three missing function words per sentence. Five of the sentences were missing strong-form function words (i.e. citation form) and six sentences were missing weak-form function words. Results showed that when function words appear in their strong forms in connected speech, such as at the end of a sentence, they are much easier to perceive. However, weak forms of the following function words were quite difficult to perceive; *it, from, has, of, her, a, himself*. For example, in the case of *has, her* and *himself*, none of the 25 participants correctly transcribed these missing words in the partial dictation test. This study suggests that function word perception in connected speech is an important factor in listening ability and worthy of further investigation.

2.6 Conclusions and dissertation aims

The studies presented in the literature review attempted to investigate some of the difficulties L2 learners experience when decoding connected speech as well as possible reasons for these difficulties. The studies also suggest that focusing more attention on helping students with developing the ability to decode words from connected speech may be an effective way to improve their listening abilities'. However, previous research has not adequately explored the relationships between L2 decoding and listening ability. There is also a need to investigate factors such as metacognitive strategy utilization, aural vocabulary knowledge and the effects of phonological modification that affect decoding and listening. Furthermore, as listening is an internal process, there is a need for qualitative research to explore L2 listeners' own interpretations of their decoding difficulties to better understand the difficulties they encounter while listening. This dissertation aims to explore these gaps in the research with the following four research studies.

Chapter 3: Research Study 1

Analyzing Difficulties in Aural Word Recognition for Japanese English Learners

3.1 Introduction

The research presented in the literature review demonstrates a need for a better understanding of the decoding process for Japanese EFL learners. Various factors affect decoding and listening ability such as semantic knowledge of vocabulary, phonological knowledge of vocabulary, decoding ability and phonological modification and many others. Research Study 1 begins to explore the strength of influence that phonological modification has on listeners' ability to decode function words (i.e. words that convey syntactic information) from connected speech. Research Study 1 addresses the fundamental question, "Do attributes of connected speech make the decoding of function words more difficult for learners?" This study suggests that attributes of connected speech limit decoding ability for known function words. Furthermore, as function words were identified more correctly when heard in citation form, this suggests they may be limiting decoding ability more than aural vocabulary knowledge for high-frequency words which are likely to be known.

This study examined Japanese listeners' difficulties in recognizing English function words in connected speech. Function words normally take a weak form in connected speech making them particularly difficult to perceive. Japanese first and second year university EFL students (N= 29) listened to short sentences of four to six words containing function words in connected speech and again in un-connected speech (citation form). Students transcribed the sentences and their errors were analyzed. Finding included: 1) students had considerably more difficulty transcribing connected speech than unconnected speech due in part to weak forms 2) function word pronouns beginning with the phoneme /h/ were more difficult to perceive in connected speech. Inadequate knowledge of how function words are modified in connected speech, little attention to sentence syntax, as well as lack of aural vocabulary knowledge are suggested as possible reasons for decoding difficulties[1].

3.2 The current study

Function words are generally thought to be more difficult to perceive than other words due to the deemphasized weak form they assume in connected speech. Hirose (2007) compared strong and weak forms of function words with a partial dictation format in which only recognition of individual function words was assessed. Hirose's study used 11 target sentences with only one, two or three function words missing. Participants listened for the missing function word while reading the target sentence on the test paper. Of the 11 sentences numbers 3, 4, 5, 6 and 10 were identified as strong form function words with the remaining six items labelled weak form. For example, the target item "it" in number 1 "Give (it) a try" is labelled as 'weak form' and the target item "by" in "He came by some money." is labelled as 'strong form'. However, in spoken English both function words would naturally be reduced and pronounced with less stress than the

[1] This study is published in Lange (2018a). The copyright belongs to CASELE academic society.

content words in these sentences. The testing format used in Study 1 presented all of the target sentences in their normal reduced form (i.e. weak form) and then in clearly articulated citation form (i.e. strong form). Transcription ability for the 30 target sentences was carefully examined and compared in each format (i.e. weak and strong). Also, rather than only transcribing the missing function words from the target sentences as in Hirose (2007), all but the first and last words were removed from each target sentence in order to assess the participants' ability to segment the function words in connected speech with little orthographic support from reading the target sentence. Such a testing format allows for stronger construct validity in assessing listening and decoding ability for function words.

3.3 Research Questions

RQ1) How well can students recognize function words in connected speech compared to when the same words are presented in citation form?

RQ2) Which types of function words are the most difficult for students to identify in connected speech?

RQ3) What factors may be contributing to difficulties in aural word recognition?

3.4 Materials and methods

In order to assess the ability to recognize function words in connected speech, 29 first- and second-year students were given a partial dictation test. The test asked students to listen to two repetitions of a short sentence (from 4-6 words) which contained at least one weak form function word. Thirty sentences were used which contained 56 target items (see Appendix A). Sentences were spoken in standard American English, at a normal rate of speech. The first and last words of each sentence were given to help focus test-takers' attention on recognizing and transcribing the three to four missing words containing the function word target items. After hearing the same sentence twice, students were asked to transcribe the missing words on the test paper. Students had approximately 20 seconds to complete each partial dictation question.

The second task of the assessment was to take the test again and transcribe the same sentences from spoken input which did not have phonological modification. Each word in the sentence was pronounced in its isolated, citation form without connected speech. The citation form version of the sentence was aurally presented twice after which the students transcribed the full sentence on the test paper next to the first partial dictation task.

To summarize the test procedure, each sentence was aurally presented twice at normal speed with connected speech and students had 20 seconds to complete each partial dictation. The first and last words of the sentence were written on the test paper. Next, the test was taken again and the same sentences were aurally presented twice in citation form (un-connected speech), after which the whole sentence was transcribed.

The main purpose of this test was to investigate students' ability to perceive function words in connected

speech. The author referred to a list of weak form function words in which there is a change in vowel quality and/or loss of phoneme and from this created 30 very short sentences containing at least one weak form function word (Field, 2008a). For example, the word "him", which takes the weak form /ɪm/ in connected speech, was used in the sentence, "Who made him do that?" The sentence was first presented to students in connected speech /humeɪdɪmˈduðæt/. After listening twice in normal connected speech and completing the partial dictation by transcribing the missing words "made him do" on the test paper, it was presented twice in citation form as in, /hu - meɪd - hɪm - ˈdu - ðæt/ and participants wrote down the full sentence.

The following is a list of the part of speech type and number of instances of all 56 function words included in the partial dictation test (selected with reference to Celce-Murcia, 2016):

articles (e.g. a, the) – an, the, the, a, the (5 instances)

auxiliary verbs (e.g. have, can, will) is, should, have, is, should, must, is, be, do, was (10 instances)

pronouns (e.g. he, you, we) him, her, you, it, his, her, you, her, you, your, I, it, you, you, you, you, me, him, him (19 instances)

prepositions (e.g. on, in, under) to, to, to, to, up, to, in, to, to, for, of, to, to, at, about (15 instances)

conjunctions (e.g. than, and, or, as) – and, and, or, than (4 instances)

copula be (e.g. am, are) are (1 instance)

contractions – don't, shouldn't (2 instances)

3.5 Results

First, the accuracy of the partial dictation responses for function words will be discussed. On average, the twenty-nine students in the study correctly transcribed only 22.6 (*SD* 7.6) of the 56 function words when presented in connected speech (77.8% correct). Results for the citation form speech transcriptions were 53.9 (*SD* 2.7) correct out of 56 function words (96.3% correct). The number of correct responses, common error examples and number of omissions for each of the 56 function words are presented (see Table 1). Transcriptions were checked word by word to determine if each function word was transcribed correctly. As in the marking procedure in Field (2008b), incorrectly spelled words which were accurately segmented and phonetically approximated the target item were accepted. Common error examples are provided in Table 1 to illustrate trends in decoding errors and for analyzing the factors which may contribute to decoding difficulties. The number of instances for the error is listed after the word in parentheses. Omissions were instances when nothing was transcribed for the function word. To illustrate the marking procedure, examples of actual transcriptions are listed for target phrase 1 are listed below (the underlined section is the transcription):

"Who made in do that?" This was marked incorrect because *in* was written instead of *him*.

"Who made do that?" This was marked as an omission because nothing was provided for *him*.

Table 1 *Partial Dictation Test Results for Function Words in Connected Speech*

Target Phrase to Transcribe	*Function Word*	*Function Word Type*	*Number Correct (%)*	*Most Common Errors (number of responses)*	*Omissions (%)*
1. made him do	him	pronoun	8 (27.6)	in (10)	3 (10.3)
2. and see	and	conjunction	8 (27.6)	in (12)	1 (3.4)
3. much is her	is	aux. verb	28 (96.6)	his (1)	0
	her	pronoun	10 (34.5)	the (12)	1 (3.4)
4. you like an	you	pronoun	28 (96.6)		1 (3.4)
	an	article	23 (79.3)	the (1), in (1), a (1)	1 (3.4)
5. got more and more	and	conjunction	21 (72.4)	than (1)	6 (20.7)
6. should have	should	aux. verb	26 (89.7)	show (1)	0
	have	aux. verb	16 (55.2)	on (3)	5 (17.2)
7. had to get some	to	preposition	3 (10.3)	hadn't (1)	23 (79.3)
8. it his or her	it	pronoun	15 (51.7)	the (3)	11 (37.9)
	his	pronoun	11 (37.9)	he's (6), he (3)	5 (17.2)
	or	conjunction	6 (20.7)	of (1), for (1)	16 (55.2)
	her	pronoun	22 (75.9)	have (1), half (1)	5 (17.2)
9. tried talking to	to	preposition	21 (72.4)	with (2), each (1)	4 (13.8)
10. car is bigger than	is	aux. verb	29 (100)		0
	than	conjunction	23 (79.3)	the (2), on (1)	1 (3.4)
11. don't you go	don't	contraction	29 (100)		0
	you	pronoun	28 (96.6)		1 (3.4)
12. wanted something to	to	preposition	19 (65.5)	on (1), it (1)	8 (27.6)
13. should take her	should	aux. verb	29 (100)		0
	her	pronoun	7 (24.1)	a (4), your (4)	3 (10.3)
14. you gone to the	you	pronoun	29 (100)		0
	to	preposition	29 (100)		0
	the	article	27 (93.1)	that (1)	1 (3.4)
15. must get up	must	aux. verb	29 (100)		0
	up	preposition	29 (100)		0
16. is your lucky	is	aux. verb	29 (100)		0
	your	pronoun	28 (96.6)	a (1)	0
17. thought I shouldn't	I	pronoun	14 (48.3)		9 (31.0)
	shouldn't	contraction	20 (69.0)	should (7)	2 (6.9)
18. has to work	to	preposition	22 (75.9)	the (2), a (1), go (1)	1 (3.4)

19. it hot in	it	pronoun	27 (93.1)		1 (3.4)
	in	preposition	23 (79.3)	on (1), than (1)	1 (3.4)
20. like to meet	to	preposition	29 (100)		0
21. are you	are	copula be	29 (100)		0
	you	pronoun	29 (100)		0
22. want you to be	you	pronoun	22 (75.9)		7 (24.1)
	to	preposition	28 (96.6)		1 (3.4)
	be	aux. verb	29 (100)		0
23. do you leave	do	aux. verb	27 (93.1)	did (2)	0
	you	pronoun	28 (96.6)		1 (3.4)
24. met for the first	for	preposition	28 (96.6)		1 (3.4)
	the	article	26 (89.7)	at (1)	2 (6.9)
25. was a lot of	was	aux. verb	29 (100)		0
	a	article	28 (96.6)		0
	of	preposition	28 (96.6)		0
26. need to talk to	to	preposition	21 (72.4)	you (1)	7 (24.1)
	to	preposition	27 (93.1)		2 (6.9)
27. you help me	you	pronouns	29 (100)		0
	me	pronouns	28 (96.6)		0
28. saw him at the	him	pronoun	6 (20.7)	it (1), I'm (1)	15 (51.7)
	at	preposition	23 (79.3)	met (1)	3 (10.3)
	the	article	15 (51.7)	a (4)	7 (24.1)
29. not so sure about	about	preposition	22 (75.9)		3 (10.3)
30. made him so	him	pronoun	16 (55.2)	you (2), me (2), it (2), in (2)	3 (10.3)

3.6 Analysis

Research question 1 asked how well students could recognize function words when they were presented in connected speech and in citation form. It seems clear that transcribing even relatively short sentences of three to four missing words proved to be difficult for most of the participants. Only 77.8% of the function word transcriptions were correct. When the phrases were spoken in citation form, there were very few transcription errors with 96.3% correct. Such a wide gap in transcription ability suggest that phonological modification (i.e. weak form) is a major factor impacting students' ability to decode function words. By extension, this deficiency may be seriously limiting listening comprehension.

It was also interesting to note that when function words were presented in citation form (unconnected speech) they were recognized about 18.5% more than when presented in connected speech (with phonological

modification). Although this study did not compare recognition rates for function words and content words, it did compare recognition rates for connected and unconnected speech. These findings of 18.5% better recognition for unconnected speech are similar to the ratio described in previous research (Sheppard and Butler, 2017; Field, 2008b) which showed L2 listeners can recognize 20% fewer function words than content words. This difference in decoding ability might be attributable to the effects of phonological modification because function words generally take a weak form in connected speech whereas content words are usually more emphasized.

Research question 2 seeks to identify the most difficult type of function words for this group of L2 listeners. Nine function words were especially difficult to recognize, with less than half of the participants correctly transcribing them. They are listed below from most to least difficult with the percentage of words correctly transcribed:

Q. 7. I had to get some more. (to) 10.3%

Q. 8. Was it his or her money? (or) 20.7%

Q. 28. I saw him at the station. (him) 20.7%

Q. 13. You should take her home (her) 24.1%

Q. 1. What made him do that? (him) 27.6%

Q. 2. Come and see me. (and) 27.6%

Q. 3. How much is her car? (her) 34.5%

Q. 8. Was it his or her money? (his) 37.9%

Q. 17. I thought I shouldn't ask. (I) 48.3%

It is interesting to note that six of the nine most difficult function words to recognize were pronouns (him, her, him, her, his, I). However, in the case of the pronoun *you* (and *your*), seven of the eight instances were recognized with over 93% accuracy. Another commonality between these words is that, except for *I*, they all begin with the unvoiced phoneme /h/ which may have contributed to more difficulty with recognition.

An analysis of the number of omissions shows that three function words were omitted by more than half of the participants (*to, or, him*). A brief analysis of reasons for poor decoding ability are discussed below.

Q. 7. I had to get some more. (to) 79.3% omitted

"I had _ get some more" and "I had *to* get some more" sound approximately the same in connected speech. In this case, having the syntactic knowledge that *to* should be placed after *had* is probably a more important factor than phonological perception of *to*. The fact that almost everyone omitted *to* may indicate that students are not effectively utilizing top-down processes, such as their knowledge of English syntax, and had transcribed words based primarily on the phonological form.

Q. 8. Was it his or her money? (or) 55.2% omitted

Recognition difficulty could be attributed here to having four function words occurring consecutively. The function words all normally take weak forms, but in this case *his* and *her* have contrastive stress.

The connector *or* is particularly devoiced and most likely requires the application of top-down syntactic knowledge for most L2 listeners to transcribe.

Q. 28. I saw him at the station. (him) 51.8% omitted

The rates of omission for *him* in test items 1, 28 and 30 offer us an interesting point of comparison. *Him* was omitted in only 10.3% for both items 1 and 30 compared with 51.8% for item 28. This difference could be explained by the difference in the words preceding *him*. The plosive /d/ phoneme at the end of *made* for items 1 (Who made him do that?) and 30 (What made him so angry?) most likely facilitated decoding for the listeners. Whereas, in item 28 (I saw him at the station.), the effects of elision make the /w/ and /h/ phonemes in *saw him* /sɔɪm/ indistinct and difficult to recognize.

Research question 3 asks about the factors that contribute to difficulties in aural word recognition. Through examining some of the most common errors in transcription we may be able to shed some light on the factors contributing to poor decoding ability. For example, 10 students incorrectly transcribed *in* for *him* in Q.1 (What made him do that?). This is a predictable error because the weak form of *him* is /ɪm/ which sounds very much like /ɪn/. Likewise, in Q. 2 (Come and see me.), 12 students transcribed *in* for *and* because /ænd/ sounds similar to /ɪn/ due to assimilation with the /m/ in /kʌm/ such that "come and" approximates /kʌmɪn /. Also, the /d/ in *and* is elided such that it is hardly heard. Furthermore, in this case *in* and *and* could both be grammatically acceptable in the sentence so the perceptual difficulty for the function word *and* is especially great.

We may also learn something from the function words which were most accurately transcribed. Collocations or common phrases that students were most likely taught in school seem to be more familiar and easier to recognize. For example, in Q. 3 (How much is her car?) *is* was almost always transcribed correctly as it was preceded by the familiar collocation "how much". Also, for Q. 4 (Would you like an egg?) and Q. 24 (We met for the first time.) there were hardly any transcription errors presumably due to familiarity with these "chunks" of language (i.e. "would you like…?" and "for the first time").

Listeners sometimes transcribed nonsense words presumably because they could not match the phonetic representation of the word to a matching word in their mental lexicons. For example, in Q. 1 (Who made him do that?), *made* and *him* were combined to make the nonsense word "maden" in four instances. Similarly, in Q. 8 (Was it his or her money?), one student combined *his, or* and *her* into "fiser." These decoding mistakes seem to suggest that lower-level students may rely on interpreting phonological input primarily through bottom-up processes without consideration to meaning. Also, a factor contributing to decoding errors may be that L2 listeners are often unaware that commonly occurring function words, which are usually first learned in their citation forms, are usually pronounced quite differently in connected speech. For example, the word *have* in Q. 6 (I should have gone), sounds like /əv/ in connected speech. If the listener lacks such linguistic

information about the word *have* in their mental lexicon, they will probably be unable to decode it from connected speech.

3.7 Conclusion

The purpose of this study was to investigate the results of a partial dictation test to learn more about the effects of phonological modification on Japanese EFL students' ability to recognize function words. The results showed that when function words were presented in connected speech (with phonological modification), they were about 19% more difficult for the participants to recognize compared to when the same words were presented in citation form (without phonological modification). These results help to illustrate the degree of influence phonological modification has on function word recognition.

An analysis of the most common errors in function word recognition showed that pronouns, with the exception of *you,* seemed to be generally more difficult to recognize. Pronoun beginning with "h" (e.g. her, him, his) were especially difficult to decode in connected speech presumably because the unvoiced initial phoneme /h/ is not pronounced. Other errors demonstrate a tendency to focus mainly on phonological input when decoding speech input without also utilizing top-down knowledge such as syntax such as in the case of the transcription "Who made in do it" which may indicate that little attention was given to the meaning of the sentence. Although, this may also be a limitation of the testing instrument because time constraints and the nature of the partial dictation task did not encourage participants to access top-down information and process the speech for meaning. Phonetically similar words, such as him and in, proved to be especially difficult to perceive. Familiarity with common phrases or collocations also seems to influence how well function words were recognized. More research is needed to investigate the various factors contributing to poor function word recognition ability as well as the strength of relationship these factors have with decoding and listening ability.

A major limitation of this study, however, is the fact that all of the participants listened to the same listening tests a total of four times. Due to the constraints of conducting research within an existing class and limited student numbers, separating students into three randomized groups for the purpose of having a control group, a connected speech test group and a citation form test group were not feasible. However, the marked increase in transcription accuracy from 77.8% percent in the connected speech version to 96.3% in the citation form version (a difference of 18.5%) suggests that weak forms of function words influence decoding difficulty. It is reasonable to assume that the participants' average gain of 18.5% on the same test was not solely because they were able to hear the sentences spoken two more times. However, the author acknowledges that the difference in the number of times the participants heard the target sentences for the two versions of the test is a significant limitation to this study. Further investigation on the effects that phonological modification has on decoding ability for function words and listening comprehension in general is warranted.

Summary

The overarching conclusion of Chapter 3 is that phonological modification of words in connected speech may limit decoding ability. Function words are usually deemphasized in connected speech making them difficult to perceive despite being known and recognizable to the listener in citation form. Also, pronouns beginning with the /h/ phoneme were especially difficult to transcribe and instances of nonsense words suggested that some learners may simply process the input phonetically rather than considering meaning. Engaging top-down processing skills is required for meaning building and these learners may be relying primarily on the bottom-up processing of acoustic input when decoding. This study suggests that attributes of connected speech such as phonological modification limit the decoding ability of function words. In other words, learners may not possess the bottom-up skills necessary to accurately decode function words in connected speech. More investigation is needed to determine the strength of the relationship between bottom-up skills and listening ability. In the next study the relationship that bottom-up skills and strategy use, have on listening comprehension is explored.

Chapter 4: Research Study 2

"Exploring the Relationship that Listening Performance has with Bottom-Up Skills and Metacognitive Awareness"

4.1 Introduction

Research Study 1 demonstrated that attributes of connected speech make decoding more difficult for function words. This chapter attempts to identify whether bottom-up skills or awareness of metacognitive listening strategies has a stronger relationship with listening ability. Instruction in bottom-up skills and top-down skills via metacognitive awareness development are both recommended for improving listening skills (Graham & Macaro, 2008; Yeldham 2016). Investigating the relationship that each of these factors has with listening ability should be prioritized and further researched. Chapter 4 addresses the fundamental question, "Which is more strongly associated with listening ability, bottom-up skills or metacognitive awareness?" Although there is research support for the efficacy of both types of instruction on listening ability, this study aims to demonstrate the factor which is more closely associated with listening comprehension ability.

Listening ability is a fundamental skill for L2 language development but not enough is known about the factors which contribute most to its development. This study attempted to demonstrate the correlation between listening performance and bottom-up skills as well as the correlation between listening performance and metacognitive awareness. A comprehensive English course with 54 Japanese first-year university students participated in the study. Listening performance data was collected using the Eiken Pre2 listening section, bottom-up skill data was collected using the Clear Listening Diagnostic Test and metacognitive awareness data was obtained with the Metacognitive Awareness Listening Questionnaire. Results showed that the ability to correctly transcribe deemphasized function words had a moderately strong correlation (r = .50) with listening performance. Also, a weak correlation (r_s = .21) was found between metacognitive awareness as measured by the MALQ and listening performance as measured by the Eiken Pre-2 listening section. Results from the MALQ questionnaire showed that the Directed Attention subscale had the highest average scores (M = 4.19) and that Person Knowledge had the largest significant relationship with listening ability among the subscales (r_s =.45).

4.2 The current study

This correlational study aims to investigate the degree to which L2 learners' bottom-up listening skills and metacognitive awareness are associated with listening performance. It attempts to develop a more detailed understanding of the challenges these learners face and inform pedagogy by indicating the level of influence each of these factors has on listening performance.

4.3 Research Questions

The current study examined the following three questions:

1) What is the correlational relationship between bottom-up skills and listening performance?

2) What is the correlational relationship between metacognitive awareness and listening performance?

3) Which listening strategies do learners report using more or less?

4.4 Materials and methods

4.4.1 Participants

Fifty-four Japanese first-year university students who were enrolled in the author's comprehensive English course participated in this study. The students had various levels of listening ability and experience with English. The highest EIKEN grades passed by the students are as follows: Grade 2: 13 students, Grade Pre2: 4 students, Grade 3: 5 students, Grade 4: 5 students, Grade 5: 1 student. In total, 27 students had taken an EIKEN test and 27 students reported never having taken one.

4.4.2 Instruments

4.4.2.1 Eiken Pre-2 level test

The listening section of the Pre-2 level Eiken test was used as the measure of listening performance in this study. The Eiken test was developed by the Eiken Foundation of Japan and is widely utilized in Japan. Unlike other popular English proficiency tests, such as the TOEIC, the Eiken is divided into seven levels of difficulty. Grade 1, the highest level, corresponds to C1 proficiency on the CEFR scale and an approximate score of 100 on the TOEFL iBT (600 on the paper-based test). The lowest Eiken level test is Grade 5 indicating a low, A1 level on the CEFR scale. A previous version of the Eiken Pre-2 Grade listening section was used in this study based on the author's assessment of the students' general English ability. This level corresponds to roughly an A2 level on the CEFR and a 20 on the TOEFL iBT (350 on the paper-based test) (Eiken Foundation of Japan, 2016). The listening section, which took approximately 20 minutes to complete, consisted of three parts, each of which had 10 multiple-choice questions. In the first part, the examinee hears a short conversation and chooses the best response from three recorded options. In the second part, the examinee hears a longer conversation and chooses the best answer to written questions about what was said. The third part consists of selecting the best answer to written questions about monologues.

4.4.2.2 Clear Listening Diagnostic Test

The Clear Listening Diagnostic Test (CLDT) is included in the teacher's manual for the pronunciation and listening instruction textbook *Clear Speech* (Gilbert, 2012). The test was chosen because it was designed to measure various bottom-up processing skills. Yeldham (2016) utilized an earlier version of Gilbert's test known as the "Clear Speech Test" for bottom-up skill assessment and suggests Gilbert's diagnostic test as "a standalone resource for researching participants' bottom-up skills" (Yeldham, 2017, p. 15). There are seven sections in the test and each section has 10 items. Five sections were chosen from the test to measure the

following skills: 1) identifying vowel minimal pairs in sentences, 2) identifying consonant minimal pairs in sentences, 3) identifying the number of syllables of words heard in isolation, 4) identifying the most emphasized word in sentences within a conversation, and 5) the ability to decode and transcribe two deemphasized words missing from partially completed sentences (i.e. partial dictation). The two sections of the test not used in this study were Word stress and Thought groups. The test section on word stress was not included because this research is focused mainly on connected speech but the word stress test focused on single words in isolation. Also, the section of the test on thought groups (semantic units of about three to five words that can be identified by prosody) was not used because this was also not considered to be a research focus and efforts were made to limit the duration of the test as it was conducted during class time.

4.4.2.3 Metacognitive Awareness Listening Questionnaire

The Metacognitive Awareness Listening Questionnaire (MALQ) was designed to evaluate learners' perceived use of listening strategies as well as their perceptions of themselves as L2 learners (Vandergrift et al., 2006). The original English MALQ was designed for native English speakers learning French. A Japanese translation, which was adapted for native Japanese speakers learning English, is provided by Watanabe (2008). This Japanese version of the MALQ was used for the current study and is included for reference in Appendix B. The questionnaire contains 21 statements describing various listening comprehension strategies. Respondents were asked to rate their level of agreement with each statement on a scale of 1-6. For example, "I translate key words as I listen." is one of the items included in the questionnaire. The listening strategies are grouped into five subscales: directed attention; mental translation; planning and evaluation; problem solving; and person knowledge. The first four subscales contain items representative of successful listening strategies. For instance, the listening strategies contained in the directed attention subscale assess the ability to maintain focus on the listening task. The respondent's degree of confidence in their listening ability is assessed in the fifth subscale, person knowledge, and is not considered a listening strategy.

Higher responses for items in the mental translation subscale are considered to indicate use of unsuccessful listening strategies (i.e. mental translation while listening) by the questionnaire designers. Vandergrift claims that online mental translation strategies reflect inefficient approaches to listening comprehension and that less-skilled listeners report using these strategies more than skilled listeners (Vandergrift, 2003; Vandergrift et al. 2006). However, other studies have been unable to demonstrate a significant correlation between responses on the MALQ mental translation subscale and listening comprehension (Goh & Hu, 2014; Wang & Treffers-Daller, 2017). Therefore, whether utilizing mental translation strategies while listening has a negative effect on listening comprehension is still unclear.

According to the authors, the MALQ should be administered shortly after the learners have completed an authentic listening task, "so that they would have a specific task on which to base their responses"

(Vandergrift et al., 2006, p. 441). In this study, the MALQ was administered after completing a paused transcription listening activity. It should also be noted that a drawback to the MALQ, as pointed out by Vandergrift and Baker (2015), is that the questionnaire does not assess actual metacognitive awareness, only self-reported awareness.

4.4.3 Procedure

Each of the assessments was administered during normal class periods over three consecutive weeks and took less than two hours in total. No listening instruction, review of the assessments or listening strategy instruction was provided to the participants by the researcher. However, many of the participants took a listening course taught by another instructor which may have included explanations of bottom-up skills, metacognitive strategies and dictation practice.

4.5 Results

4.5.1 Listening performance

Participants' listening ability was assessed by their scores on 30 multiple choice items from a 2009 version of the Eiken Pre-2 listening section. The students (N = 54) had varying levels of English listening ability with scores ranging between 11 and 29 out of 30 points; and a mean of 19.13 (SD = 4.43). The data was normally distributed with acceptable levels of skewness and kurtosis.

4.5.2 Bottom- up listening skills assessment

The results from five sections of the CLDT are presented in Table 2. Each part had 10 questions and was worth ten points.

Table 2. *Correlations between CLDT Parts and EIKEN Pre2*

CLDT Test section (ability measured)	*M*	*SD*	*r*
Part 1 (vowel discrimination)	7.44	1.22	0.11
Part 2 (consonant discrimination)	8.29	1.79	0.19
Part 3 (identify the number of syllables in words)	4.35	2.28	0.02
Part 5 (identify the stressed word in sentences)	6.61	2.31	0.22
Part 6 (partial dictation for deemphasized words)	3.10	2.11	0.50(*)
Total (five sections, 50 points)	28.79	5.06	0.36

Note. *p < .001

The Cronbach's Alpha for the CLDT was 0.29 indicating very low internal consistency in the test. This is to be expected because although each part measures bottom-up skills, the skills and tasks vary greatly, from minimal pair identification tasks to partial dictation. Therefore, the four test sections investigated in this

study were each analyzed independently. Of the 5 sections analyzed, only Part 6 "Deemphasizing with contraction and reductions" produced a moderately strong correlation ($r(52) = 0.50$, $p < .001$) with the EIKEN listening scores (Cortina, 1993).

4.5.3 Metacognitive awareness assessment

The Cronbach's alpha was calculated for all 21 items in the MALQ questionnaire as well as for each of the five subscales to demonstrate internal consistency. They are listed as follows: 0.75 for directed attention; 0.51 for mental translation; 0.68 for planning and evaluation; 0.84 for problem solving; 0.40 for person knowledge; and 0.77 for all of the items indicating overall acceptable internal reliability for the test according to Field (2013). The low internal consistency for the subscale person knowledge may indicate that questions 3, 8 and 15, which assess perceived difficulty or anxiety in regard to English skills, are not strongly related to each other. For instance, high responses to question 8 "English listening is difficult" do not necessarily mean question 15 "I don't feel anxiety when listening" would also receive high responses. Questions designed to measure feelings of difficulty or anxiety would naturally vary by individual and would not be expected to have strong internal reliability. The small number of items included in the computation also produces smaller scores for the Cronbach's alpha.

High responses to questions 3, 8 and 16 meant that the respondents perceived the task in question as difficult (negative response) and low scores meant it was easier (positive response). These items were reverse coded to match the rest of the questions in which high scores indicate positive responses for use of listening strategies and low scores indicate negative responses. In Table 3, the MALQ question items belonging to each subscale are listed to show the relationships between each subscale and listening performance. The average of the responses for all of the items in each subscale is listed along with the standard deviation and correlations calculated with Spearman's Rho (r_s) for ordinal values between listening performance measures and responses to subscale items.

Table 3. *Correlations between MALQ subscales and EIKEN Pre2*

MALQ subscales	MALQ items	*M*	*SD*	r_s
Directed attention	2, 6, 12, 16	4.19	0.57	0.19
Mental translation	4, 11, 18	3.92	0.80	0.14
Planning & evaluation	1, 10, 14, 20, 21	3.20	0.82	0.29(*)
Problem solving	5, 7, 9, 13, 17, 19	3.74	0.85	0.27(*)
Person knowledge	3, 8, 15	4.16	0.69	0.45(**)

Note. $N = 57$, * $p < .05$, ** $p < .01$

In Table 4, all 21 of the items are ranked in order of their average response scores to provide a more detailed

view of the relationship of individual MALQ questions and listening performance. Investigating the individual items in addition to the subscales (in Table 3) may provide more insight into the metacognitive factors which are associated with listening performance. The subscale that each item belongs to, the mean value, standard deviation, and the Spearman's Rho correlation with the EIKEN Pre2 grade scores are included in the table. The mean score for the Eiken Pre-2 listening test was 18.93 (SD = 4.59) out of 30 points possible.

Table 4. *MALQ Items Ranked by Average Score Including Correlations with EIKEN Pre2*

Rank	Item	Subscale	M	SD	r_s
1	2	Directed attention	4.56	1.1	0.03
2	12	Directed attention	4.53	1.1	0.24
3	4	Mental translation	4.40	1.2	0.03
4	6	Directed attention	4.29	1.0	0.10
5	11	Mental translation	4.23	1.1	0.14
6	9	Problem solving	4.04	1.1	0.07
7	17	Problem solving	3.86	1.1	0.28(*)
8	5	Problem solving	3.86	1.2	0.22
9	1	Planning and evaluation	3.67	1.2	0.33(*)
10	16	Directed attention	3.61	1.1	-0.05
11	7	Problem solving	3.63	1.2	0.05
12	13	Problem solving	3.60	1.2	0.17
13	19	Problem solving	3.47	1.1	0.24
14	20	Planning and evaluation	3.30	1.1	0.20
15	15	Person knowledge	3.30	1.4	-0.05
16	14	Planning and evaluation	3.16	1.2	0.15
17	18	Mental translation	3.14	1.1	0.11
18	21	Planning and evaluation	3.09	1.1	-0.07
19	10	Planning and evaluation	2.80	1.4	0.17
20	3	Person knowledge	2.63	1.2	0.26(*)
21	8	Person knowledge	2.19	1.0	0.53(**)

Note. N = 57 * p < .05, ** p < .01

The MALQ items which had significant correlations with listening performance were as follows: Item 8, "I feel that listening comprehension in English is a challenge for me." (r_s = 0.53), Item 1, "Before I start to listen, I have a plan in my head for how I am going to listen." (r_s = 0.33), Item 17, "I use the general idea of

the text to help me guess the meaning of the words I don't understand." (r_s = 0.28), and Item 3, "I find that listening is more difficult than reading, speaking, or writing in English. (r_s = 0.26).

4.6 Conclusions and implications

RQ1) What is the correlational relationship between bottom-up listening skills and listening performance?

In order to answer research question 1, correlations between the CLDT and the EIKEN Pre2 listening section were analyzed. The ability to identify vowel and consonant minimal pairs in sentences measured in Part 1 and 2 of the CLDT had weak correlations with listening performance (r = 0.11 and r = 0.19 respectively). Part 3, which measured the ability to identify the number of syllables in words, also showed almost no correlation with the EIKEN (r = 0.02). Likewise, a weak correlation (r = 0.22) was found for Part 5, in which students identified the most stressed word in each sentence. In addition, each of these correlations did not reach the p = 0.05 significance level. However, a moderately strong and significant correlation (r = 0.50) was found for Part 6, which entailed dictating deemphasized words in connected speech. This part was also the lowest scoring section of the test indicating that it is one of the most challenging bottom-up skills for this cohort.

The deemphasized words in Part 6 are function words which are normally reduced in connected speech but they contain important syntactical meaning. Studies measuring L2 listener ability to decode function words have demonstrated that they are considerably more difficult to perceive in connected speech than content words (Field, 2008b; Lange, 2018). This moderately strong correlation demonstrated a significant relationship between the ability to decode deemphasized words in connected speech and listening performance, whereas the other bottom-up listening skills only had a weak relationship in this study. The ability to perceive and transcribe the missing words for the partial dictation in Part 6 may have been associated with listening performance because it demonstrated perception of function words which convey syntactic information. With inadequate perception of deemphasized function words, the learner must rely mainly on content words (i.e. nouns and verbs), without an adequate understanding of their syntactical relationships, to determine the meaning of the aural input. These results suggest that helping learners develop their function word recognition in connected speech may be an effective bottom-up skill to develop for improved listening performance.

RQ2) What is the correlational relationship between metacognitive awareness and listening performance?

This study attempted to answer research question 2 by measuring the correlation between responses on the MALQ and EIKEN Pre2 listening section scores. A weak correlation (r_s = .21) was found between the two variables. These results show that metacognitive awareness, as assessed by the MALQ, had a weak relationship with actual listening performance for this cohort. The correlation obtained in this study, r_s = .21, is lower but similar to those reported by Vandergrift et al. (2006) r = 0.36, Goh & Hu (2014) r = 0.44, and Vandergrift & Baker (2015) r = 0.23, for listening comprehension and MALQ correlations. Generalizing from

these results, metacognitive awareness and listening performance may be only moderately associated. This suggests that although strategy instruction is helpful, the limited time available for listening instruction may be more effectively spent on developing linguistic skills such as aural word recognition (i.e. partial dictation) which have been shown to correlate more strongly ($r = 0.73$) with listening performance (Matthews & Cheng, 2015).

RQ3) Which listening strategies do learners report using more or less?

Analysis of the subscales in the MALQ indicate that responses for the *directed attention* were higher than for the other categories (see Table 3). However, correlations between these subscale and listening scores were low ($r_s = .19$) and did not reach significance. Responses for *mental translation* were also not significantly correlated with listening scores ($r_s = .14$). The three subscales which correlated significantly with listening scores were *planning and evaluation* ($r_s = .29$), *problem solving* ($r_s = .27$) and *person knowledge*, ($r_s = .45$).

It is interesting to note the weak to moderate but significant correlation that item 3 ($r_s = .26$) and item 8 ($r_s = .53$) of the *person knowledge* subscale had with listening scores. These two items also have the two lowest average scores among all the items because they were reverse coded along with item 16. Item 3, "I find that listening in English is more difficult than reading, speaking, or writing in English" received an average score of 4.37 (reverse coded as 2.63) on a scale of 6 points indicating general agreement with the statement. Item 8, "I feel that listening comprehension in English is a challenge for me." received an average score of 4.81 (reverse coded to 2.19) indicating more agreement with this statement. These views in regard to their self-awareness as listeners demonstrated a significant correlational relationship with their listening scores. In the subscale *mental translation*, students reported often 'translating key words' (item 11, $M = 4.23$) and 'mentally translating while listening' (item 4, $M = 4.40$) but less often for 'translating word by word' as they listen (item 18, $M = 3.14$). Listeners who rely on mental translation are probably diverting some of their cognitive resources to translation while listening which may not allow them to fully focus their attention on the input. More fluent decoding ability could help learners focus more of their cognitive capacities on decoding, meaning building and global comprehension rather than mental translation.

This study explored the relationship that two listener characteristics, bottom-up listening skills and metacognitive awareness, have with listening ability. It found that only one of the bottom-up listening skills tested, partial dictation of deemphasized words, had a moderately strong and significant correlation ($r = .50$) with listening performance as measured by the listening section of the EIKEN Pre2 test. Also, a weak correlation ($r_s = .21$) was found between listening performance and responses to the MALQ, reflecting the moderate relationship metacognitive awareness is thought to have with listening ability. In addition, results from the MALQ suggest that listening performance may be associated with listening strategies in the subscales of problem solving as well as planning and evaluation. However, the correlational nature of this

study can only suggest relationships between listening performance and bottom-up skills or listening strategies. Quantitative research using comparative treatments with control groups are needed to demonstrate a causal relationship between bottom-up skills, metacognitive awareness and listening performance. Qualitative studies which can reveal more about the thought processes of learners while listening would also contribute to developing our knowledge of the listening process and its difficulties. An important pedagogical implication we can draw from this study is that helping students improve their ability to dictate deemphasized function words in connected speech could have positive effects on their listening performance.

Summary

The overarching finding of Chapter 4 is that one assessment of bottom-up skills, namely partial dictation for deemphasized function words, correlated more strongly with listening ability than metacognitive awareness. This suggests that function word decoding as well as the ability to decode connected speech containing phonological modification such as deemphasized function words is a worthwhile focus for additional research. Metacognitive awareness appears to have a weaker correlation to listening ability. However, this study also suggests that not all bottom-up skills are strongly associated with listening ability. Of the five types of assessment in the CLDT, only decoding deemphasized function words correlated strongly with listening scores. Further investigations of the relationship between decoding ability and listening ability are needed to further corroborate and clarify this relationship.

Chapter 5: Research Study 3

Exploring the relationship between L2 vocabulary knowledge, lexical segmentation and L2 listening comprehension

5.1 Introduction

Research Study 2 found a relationship between listening ability and the bottom-up skill of decoding deemphasized function words in connected speech. This chapter attempts to corroborate and clarify the relationship decoding has with listening ability in a more rigorous manner. This study used a much larger sample of learners (N = 130) along with a paused transcription test which was created by the author to specifically assess decoding. This five-part test administered over a semester, aimed to quantify decoding ability for three-word chunks of lexis which included attributes of connected speech. The results were analyzed using two tests of listening ability to more rigorously measure the relationship evident between listening and segmenting connected speech. These findings were also compared with results from a test of aural receptive vocabulary knowledge to explore the relationship between the variables of listening ability, decoding ability and aural receptive vocabulary knowledge. Chapter 5 addresses the fundamental question, "Which is more strongly associated with listening ability, decoding ability or aural receptive vocabulary knowledge?"

The capacity to perceive and meaningfully process foreign or second language (L2) words from the aural modality is a fundamentally important aspect of successful L2 listening. Despite this, the relationships between L2 listening and learners' capacity to process aural input at the lexical level has received relatively little research focus. This study explores the relationships between measures of aural vocabulary, decoding and two measures of L2 listening comprehension (TOEIC & Eiken Pre-2) among a cohort of 130 tertiary level English as a foreign language (EFL) Japanese learners. Multiple regression modelling indicated that in combination, aural knowledge of vocabulary at the first 1,000-word level and decoding ability could predict 34% and 38% of total variance observed in TOEIC listening and Eiken Pre-2 listening scores respectively. The findings are used to provide some preliminary recommendations for building the capacity of EFL learners to process aural input at the lexical level.

5.2 The gaps in the literature and purposes of the study

This study seeks to address some of the many questions that still remain around the relationship between L2 learners' capacity to handle lexical input and L2 listening comprehension. Firstly, it seeks to measure aural receptive L2 vocabulary knowledge and decoding ability among a single cohort of L2 language learners. This will allow us to determine the relative strength of association, as well as the predictive capacities, of these two measures on L2 listening comprehension. Further, unlike previous investigations of the relationship between vocabulary knowledge and a single criterion measure of listening comprehension (e.g., Andringa et

al., 2012; Stæhr, 2009; Vandergrift & Baker, 2015), the current study uses two different measures of L2 listening comprehension. The listening tests that have been chosen for this study, the TOEIC and Eiken Pre-2, are both relevant to the study's context, namely tertiary level EFL in Japan. The Eiken test is not well-known outside of the Japanese EFL context and therefore further information about the test will be provided in section 5.3.2.4. Gathering participant scores on multiple criterion measures of L2 listening comprehension and examining the relationship of these with the lexical capacities mentioned above might provide a more generalizable picture of these relationships. This may then inform testing and teaching practice in the study context. In an effort to do so, the following research questions will be addressed.

1) What is the relative strength of association between aural receptive vocabulary knowledge, decoding ability and the two criterion measures of L2 listening among the study cohort?
2) To what degree does aural receptive vocabulary knowledge and decoding ability predict the two criterion measures of L2 listening?

5.3 Method

5.3.1 Participants

All of the 130 participants (70% females, 30% males) in this study were first-year Japanese university students enrolled in a general English course at a university in western Japan. The focus of the course was mainly to develop reading ability using extensive reading outside of class, along with listening and speaking activities during classes. The participants generally had six years of English education before entering university, which is typical of EFL in Japan; three years in junior high school and three years in high school. An analysis of the participants' average TOEIC listening (229.71, *SD* = 46.14) and reading (151.27, *SD* = 44.11) scores indicated their level of English ability was A2 (Basic user, waystage) on the Common European Framework of Reference for Languages (CEFR) (Educational Testing Service, 2015a).

5.3.2 Instruments

5.3.2.1. The Listening Vocabulary Levels Test

Aural receptive vocabulary knowledge was measured with the Listening Vocabulary Levels Test (McLean, et al., 2015). This test contains 150 items and was designed to measure Japanese learners' lexical knowledge of the first five 1,000-word frequency levels of the British National Corpus/Corpus of Contemporary American English (BNC/COCA) (Nation, n.d.) and the Academic Word List (Coxhead, 2000). Each of the sections from the first 1,000-word frequency level to the fifth 1,000-word frequency level contains 24 items and the final section measuring academic word knowledge contains 30 items. The test uses a multiple-choice format which was based on the Vocabulary Size Test (Nation & Beglar, 2007). Each item consists of the target vocabulary, a non-defining sentence containing the target word and four answer choices (written in Japanese). The target word and non-defining sentence are presented once aurally but are not written on the test paper. Test-takers choose the word, which best represents the meaning of the English target word, from among the four options,

as shown in the example below (English translations added here for clarity).

1. (Test-taker hears: “waited: I waited for a bus.”)

a. 食べた (ate)

b. 待った (waited)

c. 見た (saw)

d. 寝た (slept)

There is a five-second pause between each item and a 15-second pause between test sections (for turning the page). The last section, testing the Academic Word List, contains 30 items and all sections can be completed in about 30 minutes. The audio files were recorded by a native speaker of American English which was appropriate for the cohort of the current study as this is the dialect of English most commonly taught in Japanese EFL. As a demonstration of the validity of the test, a correlation of .54 was reported between the Listening Vocabulary Levels Test and Parts 1 and 2 of the TOEIC listening section (McLean, et al., 2015). See <https://www.iris-database.org/iris/app/home/detail?id=york%3a937862&ref=search> for the test.

5.3.2.2 Paused Transcription Tests

Decoding ability was assessed using a paused transcription test with five sections produced in-house by the author. In paused transcription tests, the test-taker listens to an audio recording and at irregular points in the recording, which correspond to the target items selected for the test, a pause is inserted. During this brief pause, the test-taker tries to transcribe the last phrase of three to five words which immediately preceded the pause. The recording resumes playback after the pause and the test-taker continues listening and transcribing the phrases heard before each pause. One aspect of the paused transcription testing format that is difficult to achieve with other tests of decoding is that it allows for the test-taker to apply their understanding of the aural co-text as well as their own background knowledge to the task of transcribing the target phrase (Field, 2008c). The term co-text refers to a category of context in which, “a group of words which provides syntactic or semantic evidence for the words which occur next.” (Field, 2004, pp. 77). By contrast, standard dictation tests or partial dictation tests usually require the listener to provide the target items without the benefit of hearing a significant amount of the target words’ surrounding co-text.

The audio for each of the five sections of the paused transcription test was recorded in a question-answer format between a Japanese native speaker asking the questions and a North American native speaker answering them. The duration of the audio for each section of the paused transcription test was between 10 to 12 minutes. Each section of the test contained 12 target phrases of three words each for a total of 180 items. A 15-second pause was inserted in the audio text after the intonation unit containing each target phrase. All pauses were located in the speech of the native speaker in an effort to standardize the acoustic features of the target phrases. Each pause in the transcription was preceded with a short beep as a signal to the listener to begin transcribing. The average speech rate for all test audio was 102 words per minute (ranging from 95 to

105 wpm) which is below the 'slower than normal' speech rates for interviews (120 wpm) and comparable to that for lectures (100 wpm) (Tauroza & Allison, 1990). The native English speaker intentionally reduced his speaking rate in order to facilitate aural decoding. The audio texts were only played once.

The content of the dialogues included personalized anecdotes as well as many topics related to Japan that would be familiar to the study cohort. A partial sample of the dialogue used in the first section of the Paused Transcription Test is provided in Appendix C. Note that test-takers were not reading the transcript and filling in blanks while listening to the dialogues; the dialogues were only heard and the test-takers wrote their phrases onto a numbered answer sheet. For example, the listeners heard the following question and answer followed by a beep and a 15-second pause during which they attempted to transcribe the target phrase immediately preceding the beep, 'we could play':

> Speaker 1:"What was it like?"
> Speaker 2: "So growing up in St Louis was fun I lived in a neighborhood with a few kids so we could play" (beep)

When designing the test dialogues, high-frequency vocabulary was almost exclusively used in order to minimize potential errors in decoding due to inadequate vocabulary knowledge. The vocabulary used in the test was analyzed for frequency in the combined COCA/BNC 1-25K corpus using the online computer program Compleat Web VP (Cobb, 2018). Results showed that 94.8% of the 5,278 tokens used in the test were within the first 1,000-word frequency band, 3.30% were in the second, 0.60% in the third, 0.30% in the fourth, 0.50% in the fifth, and 0.10% in the sixth 1,000-word frequency band with the remaining 0.44% of words not included in the corpora (i.e., offlist). A separate frequency analysis of the 60 target phrases showed that 97.2% of the 180 target words were within the first 1,000-word frequency band, 1.70% were in the second and 0.60% in the third. Only five target words were beyond the first 1,000-word frequency band (*castle, seasons, skill, unique* and *Japanese*). All 60 target phrases are listed in Appendix D.

In order to ensure that the target phrases were representative of authentic language in connected speech, each phrase was designed to contain one of three types of phonological modification: reduced function words, transitions between words (i.e., assimilation and elision) or linking (i.e., liaison). These categories of coarticulation are known to be problematic for L2 learners (Sheppard & Butler, 2017; Wong et al., 2017). The target item length was set at three words to reduce the difficulty of the transcription task while adequately representing phonological modification occurring between words. The audio files and materials for the paused transcription test developed for this study are available online at < https://data.mendeley.com/datasets/g278w62zpg/1> in the Mendeley Data repository (Lange & Matthews, 2020b).

5.3.2.3 TOEIC listening

The Test of English for International Communication (TOEIC) Listening and Reading Test is used widely in Japan with approximately 3,400 organizations and educational institutions administering the test in 2017 (Institute for International Business Communication, 2018). The TOEIC listening section takes about 45 minutes to complete and contains four parts with 100 multiple-choice items. Part 1 contains 10 items in which the test-taker selects the most accurate description of a photograph. Part 2 contains 30 items which assess the listener's ability to select the best response to a question. Part 3 contains 10 dialogues with three questions each and Part 4 consists of 10 monologues with 3 questions each to assess listening comprehension. There are 495 points possible for the TOEIC Listening section.

5.3.2.4 Eiken Pre-2 listening

The Eiken test is an English proficiency test developed in Japan and widely used in Japanese secondary schools. There are 7 grades of difficulty from Grade 5 (easiest) to Grade 1 (most difficult). This makes it possible, in contrast to TOEIC, for a test level to be selected that aligns with the known proficiency level of a given cohort. The listening section of the Eiken Pre-2 grade, used in the current study, is ranked between Grade 3 and Grade 2 and adequate achievement on the test positions a test-taker at roughly an A2 level on the CEFR (Eiken Foundation of Japan, 2016), which was the estimated proficiency level of most of the participants in this study. The listening section consists of three parts each containing 10 multiple-choice questions. In Part 1, the test-taker listens to short conversations and chooses the best response from three options. In Part 2, the test-taker hears longer conversations and selects the correct answer to questions. Finally in Part 3, the test-taker hears a monologue and selects the best answers to questions about it. The listening section takes approximately 20 minutes to complete.

5.3.3 Procedures

This study involved the administration of four test instruments: two listening comprehension tests and two lexical measures. For the purposes of analysis, measures of listening comprehension (TOEIC & Eiken Pre-2) were identified as outcome variables, and the two lexical measures were identified as predictor variables. The Listening Vocabulary Levels Test was used to measure aural vocabulary knowledge, and the Paused Transcription Test was used to measure decoding ability.

Tests were administered in the order of Eiken Pre-2, TOEIC and Listening Vocabulary Levels Test. The five sections of the Paused Transcription Test were administered approximately once every two weeks over the course of the 15-week semester. Each section of the PTT took about 40 minutes of class time so it was not feasible to conduct more than one section of the test per class period. Also, two weeks between PTT administrations were needed for conducting individual interviews with the 10 subgroup members. All tests except for the TOEIC were administered during class and necessarily spaced to reduce the cognitive burden

on students and allow time for other teaching activities. Based on previous experience teaching similar students, the slight increases in the listening proficiency of the participants over the duration of the study was not considered necessary to investigate for the purposes of this correlational study. Table 5 lists the instruments, their purposes and time of administration. Formal approval from the university ethics committee was obtained for this study.

Table 5

Procedure Summary

Test	Construct	Administration timing
Eiken Pre-2	L2 listening comprehension	Week 2
TOEIC Listening	L2 listening comprehension	Week 12 (outside of class)
Listening Vocab Levels Test	Aural Vocabulary Knowledge	Week 13
Paused Transcription Test	Decoding Ability	Weeks 3, 5, 7, 9, 11

The directions for all tests, besides the TOEIC, were provided in Japanese with clear examples to illustrate the listening task as well as time to ask any questions about the test. The TOEIC was administered following the standardized rule booklet provided by the testing company and only English instructions for each part of the listening section were provided in the test booklet and spoken aloud on the test CD. The audio for all tests was administered by audio file or CD to the whole class through high-quality speakers in a quiet classroom environment.

The criterion listening tests and two vocabulary tests used multiple-choice formats so scoring was unambiguous. However, the three-word target item transcriptions for the Paused Transcription Test required the development of a scoring protocol to ensure a standard scoring method. A scoring protocol, based on principles described in Matthews and O'Toole (2015, p. 42), was devised to facilitate consistent scoring (See Appendix E). This was not a test of spelling, and so correctly spelled target words and words with minor spelling errors which clearly reflected the phonological form of the target word (e.g., *uniek* for *unique)* received one point each. A score of 0.50 was given to recognizable but more ambiguous representations of the target word (e.g., *unik* for *unique*). A deduction of 0.25 was applied if one of the three target words was transcribed out of order or if additional words were added within the target phrase. Other incorrect words or blanks received zero points. The first author scored the Paused Transcription Test and the second author scored a subset of 10%. The correlation between the author and a co-researcher's scores was very high (r = .997), demonstrating strong levels of inter-rater agreement.

The final scores provided by the TOEIC testing institution, rather than raw scores, were used in this study with a possible score range of 5 to 495. The other three assessments utilized raw scores and their possible range of scores are listed as follows: Eiken Pre-2 listening section 0 to 30, Listening Vocabulary Levels Test 0

to 150 and Paused Transcription Test 0 to 180.

5.3.3.1. Data and analysis

Correlation and multiple regression were the two statistical techniques applied in the current study. Pearson's *r* was used to determine the strength of the relationships between the variables. Further analysis was done using hierarchical multiple regression to demonstrate the capacity that the independent variables (i.e. Listening Vocabulary Levels Test scores and Paused Transcription Test scores) have in predicting variance in the dependent variable (i.e. TOEIC listening section scores and Eiken Pre-2 listening section scores). The advantage of this analysis is that it can demonstrate the relative predictive capacities of multiple independent variables on the dependent variables (Field, 2013). The necessary assumptions associated with linearity, multivariate normality, multicollinearity, and homoscedasticity for regression analysis were confirmed to be unviolated for this data (Tabachnick & Fidell, 2007). The sample size of 130 exceeds the rule of thumb for regression analysis stated by Green (1991) in which *N* should be greater than 104 + *m* (where *m* is the number of predictors) and thus satisfies recommendations for the number of cases-to-independent variables.

5.4 Results

Table 6 shows the minimum, maximum, mean and standard deviation (*SD*) of scores obtained from each test used in the analyses. All instruments had an adequate Cronbach's Alpha (*α*) level of 0.70 or above (Cortina, 1993).

Table 6

Descriptive Statistics for Test Variables

Test	Construct	*N*	Min	Max	Mean	*SD*	Mean %	*α*
TOEIC Listening	L2 listening comprehension	122	115	350	229.71	46.14	45.12	.72
Eiken Pre-2	L2 listening comprehension	130	7	29	18.18	4.88	60.60	.70
Listening Vocab Levels Test	Aural Vocabulary Knowledge	123	68	126	101.84	11.24	67.90	.75
Paused Transcription Test	Decoding Ability	113	1	139	82.66	25.32	46.00	.86

Table 7

Skewness and Kurtosis Statistics for Test Variables

Test	*N*	Skewness	SE Skewness	z-skewness	Kurtosis	SE Kurtosis	z-kurtosis

TOEIC Listening	122	-.20	.22	-0.91	.04	.44	0.10
Eiken Pre-2	130	-.07	.21	-0.33	-.57	.42	-1.36
Listening Vocab Levels Test	123	-.62	.22	-2.84	.22	.43	0.50
Paused Transcription Test	113	-.54	.23	-2.35	.39	.45	0.87

Table 7 shows that z-skewness values for each test fall below 3.29, which indicates normal distribution for medium-sized samples (50 < n < 300) and therefore suitable for further statistical analysis (Kim, 2013).

5.4.1 Addressing Research Question 1 – What is the strength of association between the variables that were measured?

The correlations between all four measures are presented in Table 8. To standardize descriptions of the magnitude of these correlations, Cohen's (1992, p. 157) interpretation of small (r = .10), medium (r = .30) and large (r = .50) effects was used. Firstly, the two measures of listening comprehension were strongly correlated (r = .52). Despite aural vocabulary knowledge and decoding ability each being measures dependent upon processing stimulus through the aural modality, a small (r = .18) but significant correlation was observed between them.

Correlations between aural vocabulary knowledge and measures of L2 listening were small and significant (r = .15 and r = .12). Correlations between decoding ability and L2 listening were medium to strong and significant (r = .39 and r = .51). The trend in the magnitude of the correlation coefficients between the two lexical measures and both measures of L2 listening was the same: decoding ability (stronger) and then aural vocabulary knowledge (weaker). The weak but significant correlation between the Listening Vocabulary Levels Test and the Paused Transcription Test suggests that although both consist predominately of words in the first 1,000-word frequency level, the tests measure different constructs.

Table 8

Summary of intercorrelations between measures from each test instrument used in analyses

Test	1	2	3	4
1. TOEIC Listening Test	—			
2. Eiken Pre-2 Listening Test	.52**	—		
3. Listening Vocab. Levels Test	.15*	.12*	—	
4. Paused Transcription Test	.39**	.51**	.18**	—

*p < .05, ** p < .01

Correlations between the Listening Vocabulary Levels Test and the tests of listening comprehension were too small to warrant further investigation with regression analysis. However, as previous research has shown

that high-frequency aural vocabulary test scores correlate strongly with scores from standardized L2 listening tests (Matthews, 2018), the strength of correlation between each level of the Listening Vocabulary Levels Test and listening test scores was investigated. Table 9 shows that scores from the first 1,000, second 1,000 and third 1,000-word frequency levels of the Listening Vocabulary Levels Test correlated significantly at a medium level with scores from the TOEIC listening test and the Eiken Pre-2 listening test. For both listening tests, smaller non-significant correlations were found for the fourth 1,000, fifth 1,000 and Academic levels of the test.

Table 9

Summary of correlations between L2 listening tests and word frequency level sections (1K – 5K and Academic) of the Listening Vocabulary Levels Test (measuring aural vocabulary knowledge)

Listening Vocab Levels Test frequency level section	TOEIC L	Eiken Pre-2
1K	.48**	.42**
2K	.47**	.44**
3K	.33*	.30*
4K	.11	.25
5K	.03	.20
Academic	.08	.21

Note. 1K to 5K refers to sections of the Listening vocabulary Levels Test which assess knowledge of the first 1,000-word frequency level up to the fifth 1,000-word frequency. The section labelled *Academic* assesses knowledge of vocabulary included in the Academic Word List.

5.4.2 Addressing Research Question 2 – To what degree do the variables measured predict L2 listening?

As presented in Table 9, a medium to strong relationship was found between aural vocabulary knowledge of the first 1,000, second 1,000 and third 1,000-word levels (as measured by the Listening Vocabulary Levels Test) and L2 listening ability (as measured by TOEIC listening section and Eiken Pre-2 listening section). To provide a clearer picture of the relationships and relative predictive capacities these variables have on listening, hierarchical multiple regression analysis was used. The regression modelling used Listening Vocabulary Levels Test scores (1K, 2K and 3K) and Paused Transcription Test scores as predictor variables, to predict the outcome variables, TOEIC Listening and Eiken Pre-2 scores. All analyses entailed entering the Listening Vocabulary Levels Test scores before the Paused Transcription Test scores. The underlying logic of this order entry was that knowledge of single words (as measured by the Listening Vocabulary Levels Test) is fundamental to decoding ability for multi-word chunks (as measured by the Paused Transcription Test). In essence, the Listening Vocabulary Levels Test assesses both knowledge of the target words' phonology as well as their semantics, while the Paused Transcription Test is focused on phonological (i.e. segmental and

suprasegmental) issues and arguably does not directly measure semantic knowledge. When constructing each of the regression models the entry order of the Listening Vocabulary Levels Test scores was as follows: first 1,000-word level, second 1,000-word level, and then the third 1,000-word level. The underlying logic for this decision was that knowledge of higher frequency vocabulary is likely to be more fundamental to L2 listening than knowledge of lower frequency words (Adolphs & Schmitt, 2003).

The first model (See Table 10) sought to determine the degree to which aural vocabulary knowledge of the first 1,000, second 1,000 and third 1,000-word levels and decoding ability predicted variance in TOEIC listening scores. Aural vocabulary knowledge of the first 1,000-word level and decoding ability were the only two statistically significant variables in the model. The first 1,000-word level could account for 22% and decoding ability accounted for an additional 12% of variance in the TOEIC.

Table 10

Hierarchical Multiple Regression Model 1 - Aural vocabulary knowledge for first, second, and third 1,000-word level (AVK) and decoding ability and as predictors of TOEIC listening

Predictor	*R*	R^2	R^2 *Change*
1: First 1,000-word level AVK	.47**	.22**	.22**
2: Second 1,000-word level AVK	.53	.28	.06
3: Third 1,000-word level AVK	.54	.29	.004
4: Decoding Ability	.63*	.40*	.12*

* $p < .01$. ** $p < .001$.

In the second model, (see Table 11) again aural vocabulary knowledge of the first, second and the third 1,000-word levels and decoding ability, were used to predict the outcome variable Eiken Pre-2 listening scores. Similar to Model 1, the first 1,000-word level of the Listening Vocabulary Levels Test accounted for 21% and decoding ability accounted for an additional 17% of the variance in the Eiken Pre-2 scores with both predictive contributions being statistically significant.

Table 11

Hierarchical Multiple Regression Model 2 - Aural vocabulary knowledge for first, second, and third 1,000-word level (AVK) and decoding ability and as predictors of Eiken Pre-2 listening

Predictor	*R*	R^2	R^2 *Change*
1: 1st 1,000-word level AVK	.46*	.21*	.21*
2: 2nd 1,000-word level AVK	.48	.23	.02
3: 3rd 1,000-word level AVK	.48	.23	.001
4: Decoding Ability	.63*	.40*	.17*

* $p < .01$. ** $p < .001$.

Results from Model 1 (See Table 10) indicated that the first 1,000-word level aural vocabulary knowledge scores, and not the second or third, achieved statistical significance in the model and could predict 22% of the variance in TOEIC scores. In addition, Paused Transcription Test scores could predict an additional 12% of the variance in TOEIC scores with the two lexical measures offering a combined predictive capacity of 34% to the model. Results from Model 2 (See Table 11) also revealed similar results in that the first 1,000-word level aural vocabulary knowledge scores and decoding ability could predict 38% of variance observed within Eiken Pre-2 scores. In summary, aural vocabulary knowledge of 2K, 3K, 4K, 5K and Academic word levels added no predictive capacity in regression models for predicting the variance in TOEIC and Eiken Pre-2 listening scores. However, a combination of the first 1,000-word level of the Listening Vocabulary Levels Test and the Paused Transcription Test significantly predicted variance observed in TOEIC listening scores and Eiken Pre-2 listening scores.

5.5 Discussion

Perhaps the most notable finding from this study was the significant predictive capacity that high-frequency aural vocabulary knowledge at the first 1,000-word level contributed to regression models for two tests of listening. Scores from the first 1,000-word level of the Listening Vocabulary Test could independently predict 22% of variance in TOEIC listening scores and 21% of variance in Eiken Pre-2 listening scores. Aural vocabulary knowledge at the 1,000-word level had more predictive power than any other predictor variable used in the models. This finding is surprising because the Listening Vocabulary Levels Test is not a test of listening comprehension and was designed to assess phonological recognition and semantic knowledge of individual words. Correlations between total scores for the Listening Vocabulary Levels Test and the two tests of listening used in this study were weak in magnitude (i.e., $r = .15$ and $r = .12$). However, when correlations were investigated separately by 1,000-word frequency level the first 1,000, second 1,000 and third 1,000-word levels of the Listening Vocabulary Levels Test had medium to large correlations with the listening tests (see Table 9). Upon further investigation with hierarchical multiple regression analysis, it was determined that only scores from the first 1,000-word level of the test contributed significant predictive capacity to both models. This finding highlights the important association that aural knowledge of high-frequency vocabulary has with listening ability. In addition, the consistency in the predictive capacity for the two different standardized tests of listening used in the regression models supports the validity of the claim that aural vocabulary knowledge of the first 1,000-word level is associated with listening ability. Furthermore, these results corroborate previous research demonstrating that knowledge of high-frequency vocabulary is an important foundation for comprehending authentic listening texts and performance on L2 listening tests (Matthews, 2018; Matthews & Cheng, 2015; Webb & Rodgers, 2009).

Another notable finding of the current study was the strength of association between decoding ability and L2 listening. Firstly, this association was evident from correlations between Paused Transcription Tests and the two listening tests ($r = .39$ and $r = .51$ respectively). Secondly, and potentially more importantly, this strength

of association was also observed in the regression analyses. In each instance, decoding ability added a significant predictive capacity beyond that offered by aural vocabulary knowledge at the first 1,000-word level (i.e., an additional 12% and 17%. See Table 10 and 11). This is important as although it is clear that knowledge of the 1,000 most frequent words in the aural modality provides a foundation for L2 listening, the capacity to segment clusters of words in the aural modality adds something extra. The current study also speaks to the relative additional importance of the learners' decoding ability in the prediction of their L2 listening scores.

Stronger correlations were found between decoding ability and L2 listening scores as compared to those found between L2 listening and aural vocabulary knowledge. This result is likely due to the format of the Paused Transcription Test which measures decoding ability and more closely resembles listening processes by utilizing both bottom-up and top-down processing. It is also important to recall that the target items and contextual language used for the Paused Transcription Test consisted of very high-frequency words (0-1K). This in turn emphasizes the importance of the capacity to segment words in the first 1,000-word frequency range, which cover approximately 89% of spoken discourse (Adolphs & Schmitt, 2003). This suggests that a learner's capacity to fluently process the highest frequency words in connected speech is likely to be strongly facilitative of L2 listening comprehension.

This investigation demonstrates that better listeners had a stronger capacity to recognize the phonological form of high-frequency words and could associate these forms with an appropriate semantic representation. Further, better listeners could also more effectively segment clusters of three very high-frequency words that were presented in connected speech.

5.6 Conclusion

Overall, our findings suggest that greater learner familiarity with high-frequency vocabulary, at the first 1,000-word level in particular, may contribute more to overall listening proficiency than aural knowledge of lower frequency words. Further, it seems clear that decoding ability is significantly associated with L2 listening ability. Measurements of decoding ability derived through paused transcription testing provide the opportunity to assess aural recognition of chunks of lexis within connected speech. The listener's ability to establish form-meaning links between high frequency aural vocabulary, and the capacity to recognize phonologically modified chunks of lexis are very useful indicators of general listening comprehension.

Summary

In addition to the important role that aural knowledge of the first 1,000-word level of the Listening Vocabulary Test demonstrated, this study also demonstrates the strong association that decoding ability has with listening ability. Correlations between Paused Transcription Tests and the two listening tests (r = .39 and r = .51 respectively) were stronger than total scores on the Listening Vocabulary Levels Test. Although

quantitative analysis was critical in revealing the relationships that exist between listening ability and the paused transcription test, qualitative analysis is also needed to better understand listeners' perspectives on the difficulties associated in regard to the important role of decoding connected speech. Chapter 6 describes a research study which utilizes student introspections of their listening processes in conjunction with quantitative data to illuminate the difficulties experienced when listening to connected speech.

Chapter 6: Research Study 4

Analyzing Trends in the Aural Decoding Errors of Japanese EFL Learners

6.1 Introduction

The study described in Chapter 5 found a strong predictive relationship between listening ability and the ability to segment words in connected speech. This chapter attempts to investigate and categorize the specific reasons given by learners for their difficulty recognizing words in connected speech. Chapter 6 addresses the fundamental question, "What are the main sources of difficulty when decoding words from connected speech?" This study aims to identify and categorize sources of decoding difficulty through a mixed-methods approach based on paused transcription test results and learner L1 introspections while reflecting on listening difficulties.

Japanese EFL learners' difficulty with accurately decoding connected English speech motivated this mixed methods study. Across one semester, the aural decoding capacities of 63 first-year Japanese university students, with low to intermediate level English proficiency, were first measured with a battery of paused transcription tests (PTT). The transcriptions were clusters of three to five words that each possessed attributes typical of co-articulated speech. In addition, after each test, a subgroup of 10 participants individually listened to the same PTT and recounted introspective self-observations of their perceived difficulties with the aural decoding tasks in their L1. These quantitative and qualitative data were used to identify four trends in decoding errors which were categorized as follows: *limited collocation familiarity, syntactic knowledge constraints, difficulties utilizing co-text, and L1 phonological influence*. This study investigates some of the difficulties associated with aural decoding, highlights the challenges of identifying the origins of decoding errors and suggests that more focus is needed on developing decoding skills as well as knowledge of formulaic language in L2 listening education.

6.2 Research questions

The noted difficulty EFL students have had with aural decoding in the Japanese EFL context and elsewhere (Cross, 2009; Sheppard & Butler, 2017) motivates the current study which seeks to provide a new synthesis of the most common sources of decoding problems evident among a cohort of Japanese EFL listeners. The current study represents a novel contribution to the field as analysis entails the triangulation of both the products and processes associated with L2 aural decoding errors. Studies to date have focused mainly on making assumptions about the nature of decoding problems based on written test results such as dictation or paused transcription (Cross, 2009; Sheppard & Butler, 2017; Wong et al., 2021). An analysis of the quantitative products of erroneous decoding is informative, and is similarly applied in the current study, but it is also important to tap into the dynamic processes that underlie decoding errors. These internal processes are difficult to access; however, introspective self-observation protocols (ISOP) provide one avenue to do so.

As part of an ISOP, learners articulate their perceptions of the processes that occurred as they attempted to decode L2 aural stimulus. The current study not only analyses data drawn from a battery of paused transcription tests (PTT) administered among 63 EFL learners, but also supplements this with qualitative data drawn from a sub-group of 10 of these learners by way of introspective protocols. These quantitative and qualitative data are used in conjunction to identify the most common sources of decoding problems evident among the cohort. In order to achieve this research objective, the following research questions were addressed:

1. What are the lowest-scoring target phrases in the PTT for this cohort?
2. What categories of decoding error can be identified through analysis of learners' introspection?

6.3 Materials and Methods

6.3.1 Participants

Sixty-three Japanese first-year university students (48 females, 15 males) enrolled in a comprehensive English course participated in this study. All participants were 18 to 19 years old and had received approximately six years of formal English instruction. Average scores on the TOEIC listening and reading sections were 224 and 148 points respectively; placing students at an A2 level (Basic user, waystage) on the Common European Framework of Reference for Languages (CEFR) (Educational Testing Service, 2015). This study was approved by this university's ethics committee and written informed consent to participate in the study was obtained from all participants.

6.3.2 Instruments and procedures

6.3.2.1 Quantitative test instruments

Paused Transcription Test. Please see section 5.3.3.2 for the description of the paused transcription test (PTT).

Measures of listening comprehension. To establish the baseline L2 listening proficiency of the participants, two L2 listening assessments were administered. These were the Eiken Pre-2nd grade listening test and the listening section from the Test of English for International Communication (TOEIC). Each test is used broadly in Japan, but the Eiken was developed for Japanese English language learners and is therefore more specific to the research context. The TOEIC is a useful reference point for comparison both within and beyond the research context.

6.3.2.2 Qualitative Data Elicitation Protocols

Introspective Self-Observation Protocol (ISOP). A key objective of this study was to use the participants' introspections about listening to better understand the sources of their decoding problems. To achieve this

goal, an ISOP was facilitated (Gass & Mackey, 2017) by selecting 10 members from the larger cohort of 63. As baseline proficiency among the whole cohort was relatively homogenous, a selection of five relatively low and five relatively high proficiency participants was deemed to be adequately representative of the whole group. The ISOP was conducted individually with each of the 10 participants in the week following an administration of a PTT section. As there were five sections of the PTT, the ISOPs were held approximately once every two weeks over the course of three months with each session lasting approximately 90 minutes. The sessions were recorded with the participant's permission and conducted in Japanese.

During the ISOP, the participant took the same section of the PTT as the one that had been administered during the previous week's class. Unlike the in-class test, however, after transcribing each target phrase the recording was stopped and the participant was encouraged to respond to nine questions regarding the listening task. These questions assessed perception and comprehension of the co-text, target phrase, particular words and sounds that participants found difficult along with any identifiable reasons for their decoding difficulties (see Appendix F for the ISOP questions). The first six questions were asked prior to revealing the correct transcription to the participant. After seeing the correct target phrase, the last three questions were presented in order to help the participant compare and evaluate their initial decoding decisions against the correct answers. This was done to encourage contrastive thinking that might elicit reasons for their decoding errors. After the nine questions were discussed, the researcher continued the audio recording of the test until the next pause at which time the recording was stopped again. The participant transcribed the target phrase, and the same nine questions were discussed in regard to the target phrase. This procedure continued for each of the 12 target phrases in one section of the PTT. In order to clarify the ISOP interview procedure, a translation of an interview transcript for the first six target items of PTT Section 1 is included in Appendix G of the supplementary materials.

6.3.3 Data Analysis

In order to identify trends in decoding errors, an explanatory sequential mixed methods design was applied (Creswell & Plano Clark, 2011). This entailed identifying the most difficult target phrases in the five-part PTTs (quantitative data) followed by an analysis of the participants' reflections on these errors in the ISOPs (qualitative data). These data were analyzed as part of an iterative process which spanned approximately two months. This inductive approach yielded common trends describing the characteristics of decoding errors based on specific observations. The qualitative to quantitative focus aimed to corroborate, complement and strengthen the validity of the trends observed (quan → QUAL) (Riazi & Candlin, 2014). This mixed-methods design attempted to thoroughly investigate both the product and process of decoding.

The analysis procedure started with taking the average scores for each of the 60 target phrases in the paused transcription and ranking them to determine the ten lowest-scoring target phrases. Next, the ISOP audio

recordings from the 10 sub-group members were examined to analyse the reasons given for decoding errors on each of these ten lowest-scoring target phrases. If multiple similar explanations were given in regards to the same type of decoding error, the researchers analysed transcriptions from the entire cohort to see how frequently that particular type of decoding error occurred in the whole cohort's paused transcription test data. Similar examples of a particular type of decoding error evident within the PPT transcriptions were assumed to support the salience of the explanation provided by the subgroup members in the ISOP interviews. This recursive method of analysis continued until all identifiable trends in decoding error evident in the data had been described (Berg, 2007). Salient vignettes from the ISOP have been selected and presented to aid representation of the trends in decoding errors that emerged from the data. Pseudonyms for the participants have been used. The transcriptions of the ISOP were translated from Japanese to English by the first author, who has passed the highest level of the Japanese-Language Proficiency Test (JLPT).

6.4 Results

The descriptive statistics for the listening tests are listed in Table 12. The mean score for the PTT was below 45.1% reflecting the difficulty of the test for this cohort. All listening assessments had acceptable levels of internal consistency, as shown by Cronbach's alpha (α). Table 13 lists test scores for the ten subgroup members on the two listening tests and the PTT. Participants are ranked from higher to lower according to Eiken Pre-2nd grade listening test scores.

Table 12

Listening test descriptive statistics

Test	*N*	Min	Max	Mean	Mean %	*SD*	*α*
TOEIC Listening	58	130	340	223.62	45.20	47.12	.72
Eiken Pre-2	62	9	29	19.02	63.60	4.63	.70
Paused Transcription Test	63	12.76	139	81.21	45.10	25.31	.88

Table 13

Test results for sub-group members

Pseudonym	Eiken Pre-2 listening score (%)	TOEIC listening score (%)	Paused Transcription Test (%)
Kana	29 (96.7%)	300 (60.6%)	111.25 (61.8%)
Sachiko	29 (96.7%)	340 (68.7%)	139 (77.2%)
Maki	27 (90.0%)	255 (51.5%)	110.25 (61.3%)
Chiori	26 (86.7%)	260 (52.5%)	106.5 (59.2%)
Aska	23 (76.7%)	310 (62.6%)	106.25 (59.0%)
Tomoe	15 (50.0%)	(none)	74.01 (41.1%)

Kazuko	14	(46.7%)	150	(30.3%)	38	(21.1%)
Ruka	14	(46.7%)	165	(33.3%)	76.75	(42.6%)
Wakana	13	(43.3%)	215	(43.4%)	73	(40.6%)
Momoe	13	(43.3%)	130	(26.3%)	99.5	(55.3%)

The correlations between all three listening test scores are presented in Table 14.

Table 14

Summary of correlations between measures

Test	1	2	3
1. TOEIC Listening (N = 58)	—		
2. Eiken Pre-2 (N = 62)	.54**	—	
3. Paused Transcription Test (N = 63)	.47**	.48**	—

*p < .05, ** p < .01

Table 15 lists the ten target phrases that had the lowest averages on the paused transcription test, thus representing those that were most difficult for the entire cohort to transcribe (shown in parentheses). The average of these mean scores was .42 out of three points indicating that less than half a point, on average, was awarded per three-word target phrase.

Table 15

The lowest scoring target phrases

Lowest scoring target phrases (ranked low to high)	*M*	*SD*
1. …from (when we were) very young	0.10	0.39
2. …work at (first you're) like a child	0.18	0.51
3. …or just (do a lot) more reading and writing	0.22	0.46
4. …people (aren't allowed) in the group	0.33	0.52
5. …outside is (what I like) to do	0.44	0.71
6. …to see (what it was) like	0.53	0.86
7. …people usually (don't steal) things	0.53	0.54
8. …so that (sort of thing) was	0.56	0.54
9. …pressure (me to do) homework	0.66	0.91
10. …like to (visit those natural) places	0.70	0.69

The recursive analytical approach described above was applied to each of the ten most challenging target phrases shown in Table 15 and gave rise to the following four trends identified in participants' decoding errors. Explanations for these trends which were provided by the subgroup members during the ISOP interviews are

included as vignettes and were translated from Japanese. See Table 16 for examples of target phrases whose transcription errors illustrate the decoding error trend. Percentages are shown in parentheses to indicate the frequency of transcription for the listed words in the PTT data.

Table 6

Decoding error trend categories and frequency of transcriptions

Category	Target phrase*	Transcription frequency percentages*
1. limited collocation familiarity	*(that) sort of thing*	*sort of* (0%) / *all* (29.3%)
	kinds of unique	*kinds of* (80%)
	she made sure	*sure* (25.6%) / *show* (25.6%)
	be able to	*be able to* (89%)
	found out later	*found* (71%), *out* (12.7%) / *a* (32.7%)
2. syntactic knowledge constraints	*(from) when we were*	*when* (5.7%) / *one* (20.8%)
	what I like	*what* (0%), *I* (24.5%)
	comes to mind	*mind* (12.1%) / *mine* (37.9%)
3. difficulties utilizing co-text	*aren't allowed*	*allowed* (3.9%) / *around* (37.3%)
	(see) what it was	*see* (8.5%) / *she* (40.7%)
4. L1 phonological influence	*do a lot (more)*	*love more* (7.6%)
	what it was	*what* (0%), *it* (0%), *was* (50.8%)

*Note: Words in parentheses in the "Target Phrase" column are not part of the three-word target phrase. The percentages following the word or words listed in the "Transcription Frequency Percentages" column indicate their frequency of transcription.

1) *Limited Collocation Familiarity*

One trend that was identified through analysis was that participants' decoding errors resulted from difficulties in perceiving chunks of lexis which they claimed were unfamiliar. The target phrase *sort of thing* in the sentence *…so that (sort of thing) was,* was the eighth lowest-scoring target phrase in Table 15. Transcriptions contained common errors in which *that sort of thing* was mistakenly transcribed as *that's all thing* in 29.3% of the responses. Explanations from subgroup members in the ISOP interviews revealed a lack of familiarity with the collocation *sort of*. Chiori (transcription: *that all things*) indicated that lack of familiarity with *sort of* had contributed to her decoding error:

> "I don't think I've ever heard the expression *that sort of thing* so even if I had heard it correctly I probably would have decoded it (incorrectly) into words that I know."

Although *sort* is within the 1000-word level in the BNC-COCA corpus, none of the 63 participants had transcribed *sort of* on the paused transcription test. In contrast, the semantically similar collocation, *kinds of*

had an 80% transcription rate. Such a stark contrast in scores suggests that limited familiarity with the collocation *sort of* can explain some of the difficulty in decoding the target phrase.

Another interesting contrast in decoding ability was seen between the collocations *made sure* and *be able to*. While attempting to transcribe *she made sure*, participants mistakenly transcribed *show* instead of *sure* in 25.6% of the responses. This transcription error was probably largely due to the phonetic similarity between /ʃoʊ/ and /ʃʊr/. However, lack of familiarity with the collocation *make sure* may have also been a factor in decoding ability. In contrast, the collocation *be able to* was correctly transcribed in 89% of the responses. The individual collocates in these examples are assumed to be known to the participants, however, the disparity in the transcription rates presented here suggests widely differing levels of familiarity with them. Two subgroup members, Momoe and Chiori both transcribed *she made sure* as *she made show we*. When asked to translate the phrase *she made sure* into Japanese, they were both unable to do so. The participant's degree of familiarity with collocation appears to be an influential factor in their decoding ability.

2) Syntactic Knowledge Constraints

Another trend that was identified was the application of erroneous syntactic knowledge when decoding. The target phrase *when we were,* in the sentence …*from (when we were) very young,* was the lowest-scoring target phrase of the whole test. Similar errors, in which *from when* was mistakenly transcribed as *from one*, were found in the subgroup members' transcriptions. Participant explanations revealed that notions of acceptable word order had affected their decoding choices. Tomoe (transcription: *when we were from*) placed *from* at the end of the phrase rather than the beginning due to her ideas of acceptable word order:

> "Personally, I'm not used to the word order in which *when* comes after *from*. I feel like I've heard *when we were from* before but I didn't think *from* comes here (i.e. before when) and I just heard *from* so this may have been a reason – maybe I'm not used to this word order."

In total, there were 18 transcriptions which included *from,* making it the most frequently transcribed word for this test item. Despite the participants' ability to decode *from* relatively well, the complete inability to transcribe the next word, *when,* is noteworthy. Participants seem to have avoided transcribing *from when* due to the limits imposed by perceptions of acceptable word order. Maki (transcription: *from one way out*) commented:

> "I didn't think you could put *when* after *from*. Even if I had heard *from when* I would have thought that it wasn't the word *when*. Grammatically, I didn't think *when* could come after *from* because I assumed both words serve the same function. Even if I had thought it was *when*, I don't think I would have written *when*."

Maki's last comment suggests that notion of correct word order may be prioritized and could override decoding decisions based primarily on phonological input. The words *from one* were transcribed in 17.5% of the transcriptions from the entire cohort whereas the correct answer, *from when*, only appears in 3.2% of

transcriptions, suggesting a clear preference for *from one*.

Further evidence of word order difficulty is provided by transcriptions for the target phrase, *what I like* contained within the broader co-text *traveling and being outside is what I like*. The basic word order of *I like* is undoubtedly familiar and was the most common transcription (14%) followed by *traveling outside* (9.4%). Strangely, *what,* the target word preceding *I like*, occurred in 0% of the responses. Participants commented that unfamiliarity with the word order in which *what* precedes *I like* predisposed them to omit *what.* Chiori (transcription: *traveling outside*) commented:

> "I don't think I can easily hear *is what I like* because I'm not used to the word order."

Similarly, Tomoe (transcription: *I like traveling outside*) explained:

> "Maybe I just connected *I like traveling* without much thought because it's natural to me, instead of connecting *what I like* and I just didn't hear *is*."

This suggests that familiar word order patterns, such as *I like traveling*, were preferred and affected transcription decisions. In addition, as Maki explained in her reasons for not transcribing *when*, syntactic knowledge, such as word order, may be prioritized by L2 listeners such that it outweighs perceptual information and negatively affects decoding hypotheses. The acoustic input is certainly the primary source of information for decoding but in some cases perceived syntactic structures tend to distort interpretation of the input and make it conform to these perceived structures in the mind of the listener. Such trends are indicative of Field's (2008a) description of compensatory processing in that the listener compensates for gaps in bottom-up skills by applying top-down knowledge of syntax.

In addition to word order, misperceived syntactic structures also seemed to have limited participants' decoding ability. The target phrase *first you're*, was the second lowest-scoring target phrase and occurred in the following co-text, *…or how things work, at (first you're) like a child*. Chiori (transcription: *as fast as like a child*) explained her selection of *fast* instead of *first*:

> "I wondered whether the word was *fast* or *first* and I completely assumed that the first word was *as* instead of *at* so I was looking for 'as (something) as' and I thought the word *first* would be strange there (after *as*), so I thought maybe it's *fast*."

Although she initially considered the accuracy of the word *first*, mistaking *at* for *as* prompted her to select the adjective *fast* in order to comply with the "as (adjective) as" syntactic pattern she assumed would follow *as*. Chiori's conviction that she heard *as,* served to restrict her subsequent decoding choices for the next word to adjectives. In this case, as in the previous examples of word order, it appears that the listener's syntactic knowledge influenced and constrained subsequent word candidates while decoding.

Finally, there were also decoding errors which suggest that syntactic knowledge was not adequately applied to the decoding task. The word *mind* within the target phrase *comes to mind* was correctly transcribed in only

12.1% of the responses. The word *mine* was much more frequently transcribed (37.9%) suggesting that the meaning as well as the syntactic structure of the phrase was not adequately considered. If participants had applied their syntactic knowledge to the decoding task, they should have noticed that *comes to mine* is not grammatically correct in this context. It appears that phonological knowledge rather than semantic or syntactic knowledge was mainly utilized while decoding.

3) Difficulties Utilizing Co-text

Another trend that was identified from the analysis was that decoding errors could be attributed to difficulties utilizing the aural text preceding the target phrases, otherwise known as the co-text. Poor comprehension of the co-text made it difficult for some participants to decode *allowed* in the target phrase *aren't allowed*. The co-text referred to opinions on Japanese society stating, *…some people aren't allowed in the group*. Chiori (transcription: *some people aroud*), acknowledged that her lack of co-text comprehension had affected her decoding ability.

> "I had a difficult time understanding the whole section, so this part (i.e. target phrase) is probably based on guessing."

Maki (transcription: *some people are loud*) also commented that lack of general understanding of the section as well as an unknown term had made decoding difficult.

> "I couldn't understand the earlier parts well so I couldn't use the context to help me understand and *allowed* has many possible (phonetic) alternatives such as *around* and I thought a verb might follow, and I just didn't understand. (After the meaning of the section was explained to her) "I also didn't understand 'group identity' so I think that made it more difficult to hear *allow*."

Analysis of the whole cohort's results revealed only 5.4% of the transcriptions contained *allowed*. Phonetically similar words, such as *around* (35.7%) and the homophone *aloud* (12.5%), were more frequently transcribed.

In other cases, correctly comprehended co-text that was misapplied to the decoding task led to errors. The co-text preceding the sixth lowest-scoring target phrase, *what it was* within the sentence *…to see (what it was) like,* also influenced decoding decisions as Maki explained:

> I think I probably heard *she* because you (i.e. the speaker) were talking about your mother.

Decoded co-text referring to the speaker's mother, seems to have predisposed participants to transcribe *she* (40.7%) rather than *see* (8.5%) on this test item. In this case, the strategy of utilizing co-text for meaning building led to an incorrect interpretation of the input due to a small decoding error.

4) L1 Phonological Influence

Another type of error observed from the data may be attributable to L1 phonological influence, or what Cross (2009) referred to as "L1 intrusion", in which the attributes of the listeners' native language lead to decoding difficulties. In the target phrase, *do a lot*, a unique type of decoding error was observed in which the final word *lot* was perceived as *love* in four of the 43 transcriptions provided. Elision of the /t/ in *lot* followed by the

word *more* results in co-articulation resembling /lɑmɔr/ for *lot more*. In addition, the initial phoneme /l/ in *lot* could have been reinterpreted as /ɾ/ (Japanese "r") due to L1 influence to become /ɾɑt/. Since the lips must come together in preparation to pronounce /m/ in the next word *more*, /ɾɑt/ morphs to resemble /ɾɑm/. This series of phoneme shifts ends in the bilabial phoneme /m/ which if mistaken for the similar bilabial phoneme /b/ results in /ɾɑb/. This approximates the Japanese-accented pronunciation of *love*, /ɾɑbu/, a familiar loanword in Japanese. This is all to say that L1 phonological influence may have directed listeners to process L2 input in a way that redefines phonemes to more closely match the L1 sound system, such that the word *lot* (when followed by /m/) can transform into *love* in the mind of the listener.

Another aspect of L1 phonological influence that is important to consider in the analysis of decoding errors is the influence that Japanese-accented English has on decoding ability. Some Japanese students and teachers habitually pronounce English with a Japanese accent presumably to facilitate comprehension for other Japanese native speakers. Tomoe commented that learning English words with attributes of Japanese-accented English limited her ability to decode.

> "Ah, yes I think I couldn't hear it because the native speaker's pronunciation is completely different. When I was studying English, I said the word *eikyousuru* as "in fu ru en su" (i.e. influence). The difference between Japanese pronunciation (and English) is too great so that's why I think I couldn't understand this at all."

Japanese-accented English tends to place vowel phonemes such as /u/ at the ends of English words ending in consonants. This is probably due to the structure of the Japanese mora which usually consists of a consonant followed by a vowel. For example, Tōkyō (とうきょう) has four mora (To-u-kyo-u) that each end in a vowel as in the first mora "To" /to/. This pattern of adding vowels to the end of English words changes them such that *like* would resemble /laiku/. Acclimation to unnaturally emphasized English word endings may also be a factor limiting decoding ability. The words *what* and *it* in the target phrase *what it was*, were very difficult to decode for participants as shown by the percentage of correct transcription for each target word: *what* (0%), *it* (0%) and *was* (50.8%). It is interesting to note that transcriptions included instances of *she, see, was* and *like* but none had transcribed *what it*. Aska explained that the difficulty in decoding *what* and *it* may have been due to the missing final phoneme /t/ which is often over-emphasized with the addition of a vowel in heavily Japanese-accented English (i.e. /wɑ'toʊ/). This addition of vowels to English words ending in consonants might represent a defining characteristic in recognizing them for low-proficiency Japanese listeners. It may be that decoding ability for *what* is diminished if an additional vowel doesn't follow /t/ to become /toʊ/. This problem is magnified if the final /t/ is elided. When decoding the aural input *see what it was like*, Aska, (transcription: *see I like*) interpreted the words *what it* as *I*, due to the elided final /t/ in *what* and *it*:

> This may sound a little strange but I think the /t/ in *what* was dropped and the /t/ in *it* was also dropped so it (i.e. *what it)* became /wai/ and then I think somehow it switched to *I* …perhaps.

Influence from the L1 affects decoding ability and Japanese-accented English appears to also influence

mental perceptions of aural vocabulary and thus decoding ability.

6.5 Discussion

This chapter's study has attempted to identify trends in decoding errors for Japanese EFL listeners through quantitative analysis, followed by an investigation of qualitative data from interviews in a multi-phase iterative design. Four main trends in decoding errors were identified which could account for particular decoding errors: limited collocation familiarity, syntactic knowledge constraint, difficulties utilizing co-text, and L1 phonological influence. However, despite analysis by quantitative and qualitative methods primary explanations for decoding error trends must remain speculative as listeners' reasons for breakdowns in decoding are varied. Also, the implicit nature of listening makes it difficult for listeners to analyze their own sources of decoding difficulty and relate them to researchers as in the ISOP interviews. Although informative, these trends in decoding error are therefore not considered primary explanations but one of many possible explanations for the decoding errors observed. Implications for research and EFL pedagogy based on the trends identified in this study will be discussed next.

1) Limited Collocation Familiarity

Although limited vocabulary knowledge impedes decoding ability, the vocabulary used in this study was concentrated in the 1000-word frequency level and was therefore generally familiar to the participants. Despite this, recognition of the three-word phrases was quite limited. For example, subgroup members expressed a lack of familiarity with the collocation *sort of*. This lack of familiarity was supported by transcription results showing *sort* was not transcribed by any of the participants. In contrast, *kinds of unique* was the fifth highest-scoring target phrase ($M = 2.5$) revealing a large discrepancy in decoding ability for *sort of* and *kind of*. This difference in decoding ability is not likely to be due to differences in frequency for *sort* and *kind.* Both of these words are within the first 500 most frequently used words in the spoken COCA and the collocation *sort of* was ranked 42 out of 100 of the most frequently occurring collocations in the spoken BNC. However, *kind of* was not frequent enough to be included in the list (Shin & Nation, 2008). Frequency of collocation or frequency of collocates were not predictive of decoding ability for *sort of*. This suggests that Japanese EFL education may need to focus more on familiarizing learners with commonly used collocations in English.

Collocations often contain function words which are normally de-emphasized in connected speech and may be extremely difficult to accurately perceive based on acoustic input alone. For example, the homophonous function words *are, a,* and *of* may all be pronounced as /ə/ in connected speech, so differentiating between them can only be achieved through knowledge of collocation or context. Research shows that function words are identified significantly better when contained within formulaic language, such as collocation, than when encountered independently (Yeldham, 2020). Therefore, L2 listeners should be encouraged to rely less on

individual word perception and focus more on decoding multiple-word units.

2) Syntactic Knowledge Constraint

Perceived syntactic constraints, such as acceptable word order, associated with decoded words limited subsequent word candidates and resulted in decoding error. Some participants seemed to evaluate their notions of syntactic acceptability as more reliable sources of information when forming decoding hypotheses rather than their phonological perceptions of the input which they seemed less confident about.

Lack of familiarity with a variety of syntax may have also caused decoding breakdowns. Given the differences in grammatical usage for spoken and written language (Leech, 2000), more focus on the grammatical forms likely to be encountered in spoken language seems warranted. Acclimating Japanese students to authentic spoken texts in which various types of word order and grammatical features are used would likely foster better decoding skills.

3) Difficulties Utilizing Co-text

Some participants in the study claimed their decoding ability was limited because of poor co-text comprehension. Errors due to the misapplication of correctly comprehended information were also evident. Lower-ability listeners need to compensate for poor decoding skills by relying more on the co-text for meaning building (Field, 2008a). Co-text utilization can be improved by developing better listening strategies and strengthening top-down processing (Yeldham, 2020). However, better decoding skills could help listeners avoid misunderstandings such as mistaking *see* for *she*, as shown earlier, and thereby improve co-text utilization to facilitate more accurate decoding and listening comprehension.

4) L1 Phonological Influence

Characteristics of the Japanese phonological system influence the perception of English phonemes and may cause decoding errors (Cross, 2009). This study identified trends in decoding error that appear to be the result of L1 phonological influence on English word perception. More explicit focus on how Japanese phonology affects decoding ability as well as the effects that Japanese-accented English usage may have on decoding ability is also in need of further investigation. Japanese students who become accustomed to hearing Japanese-accented English may develop inaccurate mental representations of English words such that they cannot readily match known words to utterances spoken fluently by expert or native speakers. Repeated exposure to Japanese-accented English is likely to limit decoding ability by establishing two competing phonological representations for a word in the listeners' lexicon; one in Japanese-accented English and one for English as typically spoken by the native or expert speaker. Therefore, it is recommended that English instructors limit the use of Japanese-accented English and provide learners with consistently accurate acoustic models of spoken English.

L2 listening is a process of decoding words from aural input and rapidly interpreting those words to construct meaning. In Research Study 4 participants had difficulty decoding connected speech and often relied heavily on context and guessing to piece together an interpretation from the limited number of words they could recognize. Given that even high-level L2 learners experience pervasive and prolonged difficulty with aural decoding (Cross, 2009; Field, 2008b), establishing a firm grounding in decoding is essential. More attention and instruction to improve this fundamental skill and mitigate the factors limiting successful decoding would allow Japanese EFL learners to systematically improve their listening ability and avoid deficits that could limit listening ability and further skill development.

The four trends presented here provide a useful framework for identifying common patterns of decoding difficulty. Each decoding error, however, is likely to have multiple causes rather than one primary source. Although the common trends in errors suggest a common primary source for the error, multiple influences from both task and listener characteristics led to each listener's particular transcription error (Brunfaut & Révész, 2015). For example, there is a strong likelihood that the presence of reduced forms was an underlying factor affecting each decoding error to some extent (Wong et al., 2021). Due to the difficulty in categorizing sources of decoding error we suggest a data-driven approach which seeks to identify the most commonly occurring decoding errors elicited by a particular aural stimulus for a particular L1 population. Written and spoken English language learner corpora have been valuable in identifying common errors in the output of L1 population (Ishikawa, 2014). Similarly, a large corpus of the decoding errors made for a specified three-word string of connected speech by L2 listeners of a common L1 population could be useful in revealing common trends in their decoding errors. If the data collected were extensive enough and classified by learner proficiency level, it could allow for predicting the likelihood that a listener would make a particular type of decoding error for a given three-word aural stimulus. Such data may allow researchers to identify underlying factors (e.g. L1 phonological influence) which may be primarily responsible for the observed error trends. Although, the primary reason for the decoding error may remain elusive, predictable trends in decoding errors could help teachers anticipate the errors and provide explicit instruction to help learners' remedy them. Systematically addressing these trends in decoding errors with a sustained pedagogical focus can most likely help English learners overcome their unique obstacles to comprehending connected speech and foster continued listening skill development.

Summary

Chapter 6 presents four trends of decoding error that were identified from this mixed-methods study. This extensive investigation provided many insights into Japanese EFL listeners' decoding difficulties through investigations of the product of decoding as well as the listening process. It indicates that decoding spoken English is a very difficult task for lower-level listeners even when vocabulary is very high frequency and

speech rates are minimized. It also suggests that much can be learned about the reasons behind particular decoding errors through paused transcription testing and introspective self-observations with listeners. This study yielded four trends in decoding error that merit further research to gain a better understanding of decoding and L2 listening processes.

This research study was the final one associated with this dissertation. The following chapter considers the key findings of these four research studies, their pedagogical implications, limitations and indications for future research.

Chapter 7: Discussions and recommendations for future research

7.1 Overview

In this chapter, a summary of the four research studies will be discussed in relation to the three research objectives of this dissertation. The key findings of the studies, the pedagogical implications that can be drawn from the results, the limitations and future research directions will be discussed in relation to the three research objectives:

- Research Objective 1: To investigate the decoding ability of Japanese EFL learners. The fundamental question investigated in relation to this objective is as follows: 1) "Do attributes of connected speech make decoding function words more difficult for learners?"
- Research Objective 2: To examine the relative importance that factors associated with listening ability have on L2 listening. The factors examined are bottom-up skills, metacognitive awareness, aural vocabulary knowledge and decoding ability. The fundamental questions investigated in relation to this objective are as follows: 2) "Which is more strongly associated with listening ability, bottom-up skills or metacognitive awareness?" and 3) "Which is more strongly associated with listening ability, knowledge of aural vocabulary or the ability to decode words from connected speech?"
- Research Objective 3: To identify common sources of decoding difficulties for Japanese EFL learners. The fundamental question investigated in relation to this objective is as follows: 4) "What are the main sources of difficulty when decoding words from connected speech?"

7.2 Summary of key findings

7.2.1 Study 1

Research Study 1 investigated the difficulty of decoding connected speech. Research to date has demonstrated that decoding functions words from connected speech is more difficult than decoding content words (Sheppard & Butler, 2017; Field, 2008c). It was also shown that the ability to decode function words improved when function words were presented in strong form rather than weak form (Hirose, 2007). Building on these studies, Research Study 1 attempts to measure decoding accuracy for function words in multiple-word units of two to four words. It compared decoding ability for function words when sentences were presented in connected speech and when they were presented in citation form. This comparison was done to demonstrate the degree of difficulty that an attribute of connected speech (i.e. weak form) exerts on decoding ability.

The main finding is that attributes of connected speech (i.e. weak form function words) appear to limit decoding ability for function words. When the function words were aurally presented in citation form, decoding ability was 96.3%. However, when these same words were spoken normally in connected speech only 77.8% were decoded. This gap in decoding ability suggests that deemphasized function words may have imposed a limitation on decoding ability. Although learners were familiar with the function words as separate

words in citation form, the same word were largely unfamiliar when heard in connected speech. It may be that these learners did not have adequate experience listening to connected speech and noticing the phonological modification of function words.

In response to the fundamental question addressed by this study, "Do attributes of connected speech make decoding function words more difficult for learners?" the answer is tentatively affirmative. Although, a firm conclusion cannot be drawn from the results due to a flaw in the design of this study, when function words were presented in their normal weak forms in connected speech, decoding ability was reduced by approximately 20% as compared to when the same function words were presented in citation form. These findings suggest that attributes of connected speech such as weak forms and phonological modification may contribute to the difficulty learners experience while listening. In addition, results showed that pronouns beginning with /h/ (e.g. him, her, his) were found to be the most difficult type of function word to decode. This trend is likely due to the unvoiced consonant pronunciation of /h/ which does not employ articulation and lacks strong defining characteristics.

7.2.2 Studies 2 and 3

Research Studies 2 and 3 each aimed to investigate the relative importance that factors associated with listening ability have on L2 listening comprehension. These studies helped to clarify the variables that are most strongly associated with listening through correlational and multiple regression analyses. The main finding from Research Study 2 (or ch. 4), is that the bottom-up skill of recognizing deemphasized function words in partial dictations demonstrated the strongest correlation with listening performance from among the five other bottom-up skills measured. Although metacognitive awareness, or the knowledge of listening strategies, was weakly associated (r_s = .21) with listening ability, a bottom-up skill, recognition of deemphasized function words, had a moderately strong and significant association (r = .50) with listening. Such findings encourage the reexamination of the importance of decoding function words in connected speech for listening skill development.

The main finding from Research Study 3, is that decoding ability was more strongly correlated with listening ability and contributed more predictive capacity in regression models for listening ability than aural L2 vocabulary knowledge. Although, L2 aural vocabulary knowledge is fundamentally important for L2 listening, the ability to decode multiple-word chunks of lexis from connected speech is arguably just as important. The association that decoding ability has with listening ability has been underrepresented in the research field (Vandergrift, 2007). This study helps to demonstrate the important role of decoding in listening and highlight the difficulties that attributes of connected speech present to the listener. As shown in research studies 1 and 2, function words in particular appear to be difficult to decode and are likely to play an important role in L2 listening. The importance of aural vocabulary knowledge for L2 listening has been well-documented (Vafaee

& Suzuki, 2020; Vandergrift & Baker, 2015; Wallace, 2021), however the ability to segment multiple words in connected speech is also likely to contribute significantly to listening ability and warrants greater research focus. Hierarchical regression analysis showed that the combination of aural knowledge of vocabulary at the first 1,000-word level and decoding ability could predict 34% and 38% of total variance observed in TOEIC listening and Eiken Pre-2 listening scores respectively. This additional finding highlights the important relationship that aural knowledge of high-frequency vocabulary has with listening ability.

The findings from Research Study 2 inform a response to the fundamental question, "Which is more strongly associated with listening ability, bottom-up skills or metacognitive awareness?" Self-reported metacognitive awareness had a weak correlation with listening scores whereas the bottom-up skill of function word recognition correlated more strongly with listening scores. This is supported by previous correlational studies which have also shown that assessments utilizing bottom-up skills, such as aural vocabulary knowledge and partial dictation, have stronger associations with L2 listening comprehension than measures of metacognitive awareness (Vandergrift & Baker, 2015; Wang & Treffers-Daller, 2017; Wallace, 2021). In response to fundamental question 3, "Which is more strongly associated with listening ability, knowledge of aural vocabulary or the ability to decode words from connected speech?" the findings from Research Study 3 suggest that decoding words from connected speech had a stronger relationship with two measures of listening ability (r = .39 and .51) than aural vocabulary knowledge had (r = .15 and .12) as measured by the total scores of the Listening Vocabulary Levels Test. However, scores from the first 1,000-word level of the Listening Vocabulary Test were the strongest predictive variable, independently predicting 22% of variance in TOEIC listening scores and 21% of variance in Eiken Pre-2 listening scores. Regression analyses also indicated that in combination, aural knowledge of vocabulary at the first 1,000-word level and decoding ability could predict 34% and 38% of total variance observed in TOEIC listening and Eiken Pre-2 listening scores respectively.

7.2.3 Study 4

The aim of Research Study 4, *"Analyzing Trends in the Aural Decoding Errors of Japanese EFL Learners"* was to identify underlying reasons Japanese learners experience difficulty when decoding. The findings determined four trends that help explain sources of decoding difficulty. The findings were determined by first testing learners on their decoding ability using a paused transcription test developed by the author (used in Research Study 4). Then, the average scores for each of the 60 target phrases in the paused transcription test were ranked to determine the ten lowest-scoring target phrases. Next, the audio recordings from ISOP interviews with 10 sub-group members were examined to analyze explanations for decoding errors on the ten lowest-scoring target phrases. Similar explanations that were found in regard to the same type of decoding error in the subgroup were documented and the researcher analyzed transcriptions from the entire cohort to determine how frequently that particular type of decoding error occurred in the whole paused transcription test data. If similar examples of a particular type of decoding error were evident within the PPT transcriptions,

it was assumed to support the legitimacy of the explanation provided by the subgroup members in the ISOP interviews. This recursive method of analysis continued until all categories of decoding error evident in the data had been described. In response to the fundamental question "What are the main sources of difficulty when decoding words from connected speech?" this mixed-method analysis yielded five factors identified as sources of decoding difficulty: limited collocation familiarity, syntactic knowledge constraint, difficulties utilizing co-text, over-reliance on prosodic cues, and L1 intrusion.

7.3 Pedagogical implications

7.3.1. Encourage explicit instruction for decoding and collocation

As mentioned in Research Study 1, function words occur very frequently in English with about 80 of the 100 most frequently spoken words in the British National Corpus being function words (Field, 2008a). These words contribute to L2 learners' decoding difficulties as shown by research studies 1 and 2. Therefore, it is important to help English learners cope with decoding function words in authentic connected speech. In order to do this, explicit instruction in decoding weak form function words is advised. More practice doing partial dictation of multiple-word units containing function words, as conducted in this study, is also recommended for encouraging the learner to notice deemphasized function words when decoding. Additionally, regular practice in simply identifying the number of words heard in a phrase or sentence may be a quick but useful method of helping students to focus on segmentation and developing awareness of decoding skills.

In addition, collocation instruction should receive more emphasis in listening instruction due to the difficulty of decoding the deemphasized function words they contain. Adequate knowledge of commonly used collocations could help learners begin to decode collocations of multiple words as single units of meaning and rely less on the acoustic input for accurate decoding and understanding. Better knowledge of collocation might also allow learners to decode more fluently and free up more attentional resources for meaning building. Also, as indicated by the wide difference in decoding ability in the case of 'kind of' (familiar and thus decoded well) and 'sort of' (unfamiliar and thus decoded poorly) in Study 4, learners may need to be instructed in a wider range of commonly used collocations in Japanese EFL education. Explicit instruction in frequently used collocation would help listeners to identify these chunks of lexis in spoken English. Focusing initially on the 100 most frequently used collocations is recommended (Shin & Nation, 2008). Familiarity with common collocations may help the listener avoid the ineffective strategy of trying to perceive each individual word using phonetic input alone. Also, the presence of formulaic language (e.g. collocation) has been shown to facilitate decoding by presumably reducing the need to analyze words individually (Yeldham, 2020).

7.3.2 Encourage the utilization of top-down processing for meaning

Errors of omission in the partial dictation test used in Study 1 were the most frequent type of error and suggests learners might have difficulty noticing the presence of weak form function words. Consequently, it

may also be helpful to encourage learners to use top-down processes to help them identify missing words based on their knowledge of English syntax. For example, after transcribing the target words on a paused transcription test, learners should be given time to reexamine their transcriptions and judge the grammatical acceptability of their answers for each item. By doing so, they may be able to notice whether a function word is missing from their transcription or not. In subsequent listening practice with partial dictation the learners should attempt to consider not only phonological input while listening but also remember to activate top-down skills to evaluate their transcription attempts for grammatical correctness.

7.3.3 Encourage more exposure to authentic connected speech

The findings from research studies 2 and 3 showed that listening ability has a moderately strong and significant relationship with the ability to decode deemphasized function words (r = .50) and decode three-word units from connected speech (r = .51) on the Eiken Pre-2 listening section. These findings help to support the claim that developing decoding ability for function words and multiple-word units in connected speech would be beneficial for listening development. More exposure to spoken English, such as listening to a story (i.e. Extensive Listening) or watching a video (i.e. Extensive Viewing), is likely to facilitate the development of decoding ability for words in connected speech. Extended periods of listening to connected speech for meaning can help learners to utilize bottom-up processes, such as decoding, as well as the conceptual top-down processes associated with meaning building. Japanese students learning English in an EFL environment are not likely to have enough opportunities to process spoken input for meaning. Consequently, poor listening skills may simply be a result of not getting enough compelling comprehensible input; an essential factor for L2 acquisition according to second language acquisition theory (Krashen, 1982). Practicing the skill of listening for meaning is therefore fundamental to decoding development and fluent listening comprehension (Hulstijn, 2002). Extensive Reading while listening has been shown to improve listening comprehension better than listening alone for low-intermediate EFL learners (Chang & Millet, 2014). This may be an effective way to acquire the patterns of connected speech implicitly while focusing on meaning

7.3.4 Encourage the use of testing to diagnose listening difficulties

The regular use of test formats that require the learner to process lexis through the aural modality is suggested, especially those that target the highest frequency words (e.g., Matthews, 2018; McLean, et al., 2015, etc.). As shown by the results of this study, combining semantic assessment of high-frequency vocabulary via the Listening Vocabulary Levels Test with assessment of form recognition via the Paused Transcription Test may be more predictive of actual listening ability. Such testing is likely to be useful in enabling teachers to stay abreast of the aural vocabulary knowledge status of their students and their ability to comprehend and segment that vocabulary in connected speech. If used as a diagnostic tool as recommended by Field (2003), such testing will provide data that can be used to inform pedagogical decisions aimed at developing learners' capacity to better handle lexis mediated through the aural modality. Keeping records on

the types of decoding errors that occur amongst learners is also suggested, and the regular use of paused transcription tests such as those used in this study is likely to be a valuable way of doing this. Such data may be used to assess and facilitate the development of aural vocabulary knowledge and decoding skills necessary for listening development.

An increased focus on evaluating aural vocabulary knowledge and decoding ability through formats such as the Paused Transcription Test and Listening Vocabulary Levels Tests could help to emphasize the importance of these skills as well as to diagnose listening difficulties for learners. Such a focus on the development of skills for listening proficiency is needed to promote more balanced aural/oral English skills for Japanese learners. Further, this focus on diagnosing listening difficulties through testing may encourage a cultural change within EFL pedagogy in Japan towards assessment for learning (Davison & Leung, 2009), namely increased use of assessment modes that inform ongoing teaching and learning decisions (i.e. formative assessments). Listening difficulties that were identified through testing could then be used to improve students' listening skills. This could also lead to similar approaches to developing other English skills, such as reading or speaking, through ongoing diagnostic assessments. Additionally, finding time to facilitate verbalized introspection, especially in the student's L1, immediately after individual learners engage with paused transcription tests can provide an even deeper insight into the origins of decoding errors. Such information could help to inform other bespoke classroom-based interventions aimed at promoting decoding (e.g., Field, 2003; Siegel & Siegel, 2015).

7.3.5 Focus on acquiring aural vocabulary knowledge of high-frequency words in connected speech

The important role that aural knowledge of high-frequency vocabulary plays in listening was evident from the strong predictive capacity of scores from the Listening Vocabulary Levels Test at the first 1,000 word-level on two tests of listening, 22% for TOEIC and 21% for Eiken Pre-2 (Lange & Matthews, 2020a). These findings support the assertion that helping learners build knowledge of words as they occur in speech is an important strand of vocabulary knowledge development, and that such endeavors are likely to result in positive language learning outcomes (Siegel, 2016). Here we hypothesize that such interventions are likely to be especially impactful in learning contexts within which vocabulary knowledge development has been traditionally addressed through reading and writing largely without also presenting the target words aurally in contextualized speech. Rather than only judging vocabulary to be 'known' when a learner can establish a form-meaning link for written words, educators are also encouraged to reexamine vocabulary learning in terms of learners' aural recognition and comprehension of words in connected speech. Limited development of aural vocabulary knowledge and decoding ability could result in poorer listening ability for even high-frequency words. In addition, instructional approaches should be developed that improve learners' familiarity with the phonological form of words as they occur in connected speech and which also enhance learners' ability to comprehend chunks of lexis under time constraints.

In terms of recommendations for classroom practice, pedagogical activities that build the capacity to aurally recognize and understand words from the high-frequency range should be prioritized. Although there is a need to be somewhat speculative due to the limitations of the correlational research paradigm used here, a general rule of thumb based on the evidence at hand would be to ensure learners have a solid grounding in high-frequency aural vocabulary before explicitly addressing vocabulary beyond the second 1,000-word range. As almost 90% of the vocabulary used in typical spoken discourse is from the first 1,000-word level (Adolphs & Schmitt, 2003), it seems very important that L2 listeners develop fluent recognition of these most frequently occurring words. The vocabulary used in the paused transcription test in Studies 3 and 4 consists of 94.8% in the first 1,000-word level of frequency. These words are generally familiar to the learners when presented in their orthographic form, however learners' aural knowledge of these word appear to be limited judging from the low average scores (45.1%) on the paused transcription test. This suggests that not enough emphasis was placed on aural vocabulary knowledge and decoding in their EFL education up to now. Due to the pressures of factors such as performing well on English examinations for university admission, learners may be pushed to learn increasingly lower-frequency vocabulary (i.e. lexical knowledge breadth) before achieving adequate aural vocabulary knowledge (i.e. lexical knowledge depth) for high-frequency vocabulary. These pedagogical recommendation are particularly important in the Japanese EFL educational context (and others like it) as an inordinate amount of effort from students and teachers is focused on learning increasingly lower frequency vocabulary in preparation for university entrance exams (Kobayashi, 2001). However, this can result in a substantial difference between aural and written vocabulary sizes for learners (Mizumoto & Shimamoto, 2008).

7.3.6 Encourage the use of intelligible English pronunciation

In connection with aural vocabulary knowledge development, learners reported that Japanese-accented English might also be a factor contributing to poor aural decoding ability. Some Japanese teachers of English as well as students speak English with a strong Japanese accent presumably to facilitate comprehension.. This seems to be a habit of convenience because it is easier to use L1 phonemes rather than the phonemes of the L2, however it is likely to have a limiting effect on aural word knowledge. Due to exposure to Japanese-accented English, the learner may store an inaccurate aural form of English words in the lexicon resulting in difficulty when trying to match L2 acoustic input to the L1 influenced form (i.e. Japanese-accented form). So for example, in Study 4 a subgroup member reported remembering the word "influence" in heavily Japanese-accented English which she explained was the reason she could not recognize the word when she heard it in standard spoken English. Based on the author's experience within the Japanese high school EFL context, Japanese-accented English is the "normal" style of spoken English used by students and many teachers and is likely to be the dominant acoustic form for English words stored in the learner's lexicon. Therefore, it is recommended that Japanese EFL learners be consistently provided with accurate acoustic models of English

words and be discouraged from using heavily Japanese-accented English in order to develop accurate lexicons of English aural vocabulary. This suggestion recognizes that native English speakers also pronounce words with an accent, however, the intelligibility of their accented English is arguably greater. Speaking English in an accented manner which utilizes the phonemes of Japanese is likely to result in not only problems with intelligibility when speaking, but also in difficulty with decoding and listening as reported in the ISOP interviews.

7.4 Limitations and future research

7.4.1 Study 1

A major limitation of Study 1 was that learners heard the same target sentences four times in total. This is likely to have produced a testing effect which could have biased the results in favor of the citation form treatment. Each test item was first heard twice for the treatment which measured decoding ability from normal connected speech. The same students then listened to the same target sentences presented twice in citation form. Consequently, hearing the sentences repeatedly could have led to higher scores on the citation form treatment of the test and inflated the difference in transcription ability between the two treatments. In future research, separate groups of students with similar proficiency levels could be used to measure the transcription accuracy resulting from each type of treatment. Testing effects may have resulted in higher scores for the citation treatment, however, large difference in transcription accuracy between the initial two attempts in connected speech and the subsequent two attempts in citation form suggest that transcribing function words from connected speech was considerably more difficult than transcribing them when presented in citation form.

7.4.2 Studies 2 and 3

A main limitation of both Research Study 2 and 3 is that they use correlational analysis (also regression analysis in Study 3) to investigate the relative importance of factors such as bottom-up skill utilization, metacognitive awareness, aural vocabulary knowledge and decoding ability on L2 listening ability. Although the association between these factors and listening ability was established, it does not imply that those factors cause listening ability to improve. Intervention studies between groups of similar students are needed to accurately compare the potential of these factors to increase listening ability. For example, in Research Study 2, a stronger correlation was found between the bottom-up skill of recognizing deemphasized function words in partial dictations and listening ability than what was found between estimates of metacognitive awareness and listening ability. In order to fully demonstrate that partial dictation of function words contributes more to listening development than metacognitive awareness, a pre-test, intervention, post-test research design should be utilized comparing a group of learners who practice partial dictation for function words, a group who receives metacognitive awareness instruction and a control group. Such a study design is more qualified to make stronger claims about the impact each of these factors have on listening ability.

Although correlational analysis may not be able to adequately determine the contribution that these factors have on listening development, in Research Study 3 a more robust means of analysis was incorporated to investigate the importance of aural vocabulary knowledge and decoding ability for listening ability. This was done by increasing the number of participants (N = 130) and employing hierarchical multiple regression analysis. Including more participants in the study satisfied the rule of thumb for regression analysis stated by Green (1991) in which N should be greater than 104 + m (where m is the number of predictors). Hierarchical multiple regression specifically refers to a form of regression modelling in which the order of the predictor variables are selected and entered into the model in order based on their predictive capacity. Each step of the model involves adding a predictor variable and in Research Study 3 this was aural vocabulary knowledge followed by decoding ability. With each step, it can be established whether, and to what degree, the addition of the predictor variable contributes to the predictive capacity of the model overall (see Field, 2013). Building such models presents a standardized approach to explore and communicate the statistical relationship between factors that are assumed to contribute to L2 listening ability. In essence, the degree to which a set of predictor variables can explain variance within an outcome variable provides an estimate of their relative importance. In the case of Research Study 3, aural vocabulary knowledge of the first three 1000-word levels and decoding ability were the predictive variables that were used in the models to predict variance in the outcome variables TOEIC and Eiken pre-2 listening. Results showed that only the first 1,000-word level could account for 22% and decoding ability accounted for an additional 12% of variance in the TOEIC listening section. A similar model was made to explain the variance in the Eiken Pre-2 listening section. The results showed that the first 1,000-word level of the Listening Vocabulary Levels Test accounted for 21% and decoding ability accounted for an additional 17% of the variance in the Eiken Pre-2 scores. In this way, it can be determined that aural vocabulary knowledge of the first 1,000-word level was the most important predictor for listening followed by decoding ability. It also revealed that the 2,000-word and 3,000-word levels were not significant predictors but decoding ability was. Therefore, although Research Study 3 is not an intervention study it was still able to provide detailed and legitimate claims about the relationships and importance of key predictor variables on listening ability. Future research is needed to investigate the efficacy of interventions aimed at enhancing learners' capacity to process words aurally. Longitudinal studies that involve tracking the development of aural vocabulary knowledge and decoding ability as targeted pedagogical interventions are of particular interest. Further, determining the validity of the assertion that improvements in decoding ability and aural vocabulary knowledge can directly improve L2 listening comprehension is key. Confirming or refuting such assertions will require subsequent research studies.

Another possible limitation is that the relatively low proficiency level of the participants indicate they may have been unfamiliar with much of the low-frequency vocabulary from the 1,000 word-level and above. Possible floor effects for sections of the Listening Vocabulary Levels Test containing low-frequency vocabulary

may have diluted the value of the aural vocabulary knowledge data. However, the participants' mean score for the test overall was roughly 67.8% and therefore did not indicate excessively low scores.

A central objective of this study was to provide a preliminary snapshot of the relationships between scores from test instruments measuring lexical capacities and L2 listening among a cohort of Japanese EFL learners. An important area for future research will be to expand the scope of similar studies both within larger cohorts of Japanese EFL students, as well as with learners with different L1 backgrounds and linguistic proficiency levels. Of interest in this regard is to determine the degree to which the generalized trends observed as part of the current study are mirrored or contrasted among other cohorts of learners.

The development of a broader array of tests that measure lexical capacities mediated through the aural modality is another important future research direction. In particular, the development of tests that measure the capacity to handle multiple sequential words is warranted. This seems especially important in light of the specific and robust relationship between L2 listening comprehension and the capacity to segment, recognize and understand lexis mediated through the aural modality.

7.4.3 Study 4

One fundamental limitation of Study 4 is that although the study invited learners to reflect on their difficulties decoding short sections of connected speech in the ISOP interviews, many could not identify reasons for their particular decoding difficulty. Often, the students were unable to provide specific information to explain their errors in decoding presumably because they did not understand the reasons themselves even after the correct answer was provided to them. This is an expected limitation when investigating covert, cognitive processes, however. It is a difficult task to reflect on one's mental processes to identify causes of decoding errors. The implicit nature of the listening process does not lend itself well to metacognitive analysis and is understandably difficult for the learner to conceptualize. Consequently, many subgroup members simply explained that the target items were "too connected" or "too fast". In many cases learners also stated that the target words were "unknown" to them or simply "not heard". Although, such general explanation provided by the learners were interesting, they lacked the specificity required to adequately examine decoding problem. However, studies designed to explore these general claims more specifically would be helpful in the future. For example, a study comparing the influence of different types of phonological modification on decoding ability could provide a way to investigate the claim that these target items are too connected or just poorly perceived. It would be informative to investigate which types of phonological modification in connected speech are perceived as the most connected or most difficult to hear. Also, a study comparing the aural knowledge of particular words with decoding ability for those same words in connected speech may be useful in evaluating the claim that unknown words are primarily responsible for decoding error. In response to the explanation provided by learners that the aural input was "too fast", a study could be conducted in which

decoding ability is assessed for target items spoken at differing speech rates. In this way, the effects of speech rate on decoding ability might be investigated.

Although the PTT format encourages listening for meaning, transcriptions containing nonsense words or familiar words that were uncommonly misspelled (showing a clear lack of recognition) indicated that participants may have relied predominately on acoustic input without adequately attending to meaning. Therefore, future research may benefit from employing an assessment of comprehension for the target phrases in the PTT. More studies demonstrating the relationship between decoding ability and L2 listening as well as intervention studies exploring effective ways to develop decoding ability are needed. Also, studies aiming to clarify the influence of collocation knowledge on decoding and listening ability could help determine the value of interventions to develop formulaic language acquisition.

7.3 Concluding remarks

The studies included in this dissertation demonstrate that although decoding connected speech is a challenging task for L2 learners, it is an important one that is closely associated with listening comprehension. It is clear that aural vocabulary knowledge is strongly tied to listening ability and the studies presented here support the position that decoding ability also has a strong relationship with listening. Decoding ability was shown to be a greater factor in listening performance than aural vocabulary knowledge for all but words at the first 1,000-word level. Knowledge of high-frequency vocabulary along with decoding ability appear to be two of the strongest predictors of L2 listening ability. However, the ability to decode even high-frequency vocabulary was found to be quite challenging for learners. Through systematic analysis of paused transcription data and introspective interviews five sources of decoding difficulty were identified for Japanese EFL learners in this cohort. Future research is needed to expand our understanding of these sources of decoding difficulty as well as to create and test interventions to help learners overcome them. Doing so will hopefully enable them to achieve higher levels of skill in the difficult task of listening comprehension.

REFERENCES

Adolphs, S., & Schmitt, N. (2003). Lexical coverage of spoken discourse. *Applied Linguistics*, *24*(4), 425–438. https://doi.org/10.1093/applin/24.4.425

Al-jasser, F. (2008). The effect of teaching English phonotactics on the lexical segmentation of English as a foreign language. *System*, *36*(1), 94–106.

Anderson, J. (1995). *Cognitive Psychology and its implications* (4th ed.). New York, NY: Worth Publishers.

Andringa, S., Olsthoorn, N., Van Beuningen, C., Schoonen, R., & Hulstijn, J. (2012). Determinants of success in native and non-native listening comprehension: An individual differences approach. *Language Learning, 62,* 49–78. https://doi.org/10.1111/j.1467-9922.2012.00706.x

Berg, B. (2007). An introduction to content analysis. In Berg, B.L. Editor, *Qualitative research methods for the social sciences* (pp. 238–267). Boston, MA: Allyn and Bacon.

Best, C.T. , & Tyler, M. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O. S. Bohn & M. Munro (Eds.), Second-language Speech Learning: The Role of Language Experience in Speech Perception and Production. A Festschrift in Honour of James E. Flege (pp.13-34). Amsterdam: John Benjamins.

Bonk, W. (2000). Second language lexical knowledge and listening comprehension. *International Journal of Listening, 14*, 14–31.

Broersma, M., & Cutler, A. (2008). Phantom word activation in L2. *System*, *36*(1), 22–34. http://dx.doi.org/10.1016/j.system.2007.11.003

Brunfaut, T., & Revez, A. (2015). The role of task and listener characteristics in second language listening. *TESOL Quarterly*, *14*(1), 141–168. http://dx.doi.org/10.1002/tesq.168

Buck, G. (2001). *Assessing listening*. New York: Cambridge University Press.

Celce-Murcia, M., Brinton, D. M., & Goodwin, J. M. (2016). *Teaching pronunciation: a course book and reference guide*. (2nd ed.). New York, NY: Cambridge University Press.

Chang, A. C., & Millett, S. (2014). The effect of extensive listening on developing L2 listening fluency: some hard evidence, *ELT Journal*, *68*(1), 31–40. https://doi.org/10.1093/elt/cct052

Cheng, J., & Matthews, J. (2018). The relationship between three measures of L2 vocabulary knowledge and L2 listening and reading. *Language Testing*, *35*, 3–25. https://doi.org/10.1177/0265532216676851

Cobb, T. *Compleat Web VP v.2* [computer program]. Accessed 01 Nov 2018 at https://www.lextutor.ca/vp/comp/

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155. https://doi.org/10.1037//0033-2909.112.1.155

Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, *78*(1), 98–104. http://dx.doi.org/10.1037/0021-9010.78.1.98

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, *34*(2), 213–238. https://doi.org/10.2307/3587951

Cross, J. (2009). Diagnosing the process, text and intrusion problems responsible for L2 listeners' decoding errors. *Asian EFL Journal, 11*(2), 31–53.

Cutler, A. (2012). *Native listening: Language experience and the recognition of spoken words.* MIT Press.

Cutler, A., & Carter, D. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer, Speech and Language*, *2*, 133–142. http://dx.doi.org/10.1016/0885-2308(87)90004-0

Cutler, A., & Clifton, Jr., C. (1999). Comprehending spoken language: A blueprint of the listener. In C. M. Brown & P. Hagort (Eds.), *The neurocognition of language* (pp. 123-166). Oxford, England: Oxford University Press.

Davison, C., & Leung, C. (2009). Current issues in English language teacher-based assessment. *TESOL Quarterly*, *43*(3), 393–415.

Dijkgraaf, A., Hartsuiker, R., & Duyck, W. (2016). Predicting upcoming information in native-language and non-native-language auditory word recognition. Bilingualism: *Language and Cognition*, *20*, 917 - 930.

Educational Testing Service (2015). Mapping the TOEIC Tests on the CEFR. Princeton, NJ: Educational Testing Service. Retrieved from https://www.ets.org/s/toeic/pdf/toeic_cef_mapping_flyer.pdf

Eiken Foundation of Japan. (2016). *Comparison table*. Retrieved June 23, 2018, from http://www.Eiken.or.jp/Eiken/en/research/comparison-table.html.

Ernestus, M., Dikmans, M., & Giezenaar, G. (2017). Advanced second language learners experience difficulties processing reduced word pronunciation variants. *Dutch Journal of Applied Linguistics*, *6*(1), 1–31. http://dx.doi.org/10.1075/dujal.6.1.01ern

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. Los Angeles etc: Sage.

Field, J. (2003). Promoting perception: Lexical segmentation in L2 listening. *ELT Journal*, *57*(4), 325–334. https://doi.org/10.1093/elt/57.4.325

Field, J. (2004). *Psycholinguistics: The Key Concepts*. New York, NY: Routledge.

Field, J. (2008a). *Listening in the language classroom*. Cambridge, England: Cambridge University Press. https://doi.org/10.1017/CBO9780511575945

Field, J. (2008b). Revising segmentation hypotheses in first and second language listening. *System, 36*, 35–51. https://doi.org/10.1016/j.system.2007.10.003

Field, J. (2008c). Bricks or mortar: Which parts of the input does a second language listener rely on? *TESOL Quarterly*, *42*(3), 411–432. https://doi.org/10.1002/j.1545-7249.2008.tb00139.x

Field, J. (2008d). Face to face with the ghost in the machine: psycholinguistics and TESOL. *TESOL Quarterly*, *42*(4), 361–374.

Field, J. (2013). Cognitive validity. In A. Geranpayeh & L. Taylor (Eds.), *Examining listening: Research and practice in assessing second language listening* (pp. 77-151). Cambridge, England: Cambridge University Press.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist*, *34*(10), 906. doi:10.1037/0003-066X.34.10.906

Flege, J. (1995). Second-language speech learning: theory, findings, and problems. In *Speech perception and linguistic experience: Theoretical and methodological issues* (W. Strange, editor), Timonium, MD: York Press.

Flowerdew, J., & Miller, L. (2005). *Second language listening: Theory and practice*. Cambridge, UK: Cambridge University Press.

Gass, S. M., & Mackey, A. (2017). *Stimulated recall methodology in applied linguistics and L2 research*. New York, NY: Routledge. http://dx.doi.org/10.4324/9781315813349

Gilbert, J. (2012). *Clear speech: Pronunciation and listening comprehension in North American English* (4th ed.). New York, NY: Cambridge University Press.

Goh, C. (2000). A cognitive perspective on language learners' listening comprehension problems. *System*, *28*(1), 55–75.

Goh, C.C.M., & Hu, G. (2014). Exploring the relationship between metacognitive awareness and listening performance with questionnaire data. *Language Awareness*, *23*(3), 255–275.

Graham, S. (2006). Listening comprehension: The learners' perspective. *System*, *34*, 165–182.

Graham, S. & Macaro, E. (2008). Strategy instruction in listening for lower-intermediate learners of French. *Language Learning*, *58*(4), 747–783.

Green, S. B. (1991). How many subjects does it take to do a regression analysis? *Multivariate Behavioral Research*, *26*(3), 499–510. https://doi.org/10.1207/s15327906mbr2603_7

Guion, S., Flege, J., Ahahane-Yamada, R., & Pruitt, J. C. (2000). An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants. *Journal of the Acoustical Society of America*, *107*, 2711–2725.

Hirai, A. (1999). The relationship between listening and reading rates of Japanese EFL learners. *The Modern Language Journal*, *83*, 367–384.

Hirose, K. (2007). Identifying the weak forms of function words in a single sentence when sound changes occur in connected speech. *Meirin Journal of Dental Engineering and Oral Health Science*, *10*(1), 11–16.

Henrichsen, L. (1984). Sandhi-variation: a filter of input for learners of ESL. *Language Learning*, *34*(3), 103–123.

Hulstijn, J. H. (2002). Towards a unified account of the representation, processing and acquisition of second language knowledge. *Second Language Research*, *18*, 193–223. http://dx.doi.org/10.1191/0267658302sr207oa

Hulstijn, J. H. (2003). Connectionist models of language processing and the training of listening skills with the aid of multimedia software. *Computer Assisted Language Learning*, *16*(5), 413–425. https://doi.org/10.1076/call.16.5.413.29488

Ishikawa, S. (2014). Design of the ICNALE-spoken: A new database for multimodal interlanguage analysis. In Ishikawa, S. (Ed.), *Learner Corpus Studies in Asia and the world Vol. 2* (pp. 63-76). Kobe, Japan: Kobe University.

Institute for International Business Communication (2018). *TOEIC Program Data & Analysis 2018*. Tokyo: IIBC. Retrieved from http://www.iibc-global.org/library/default/toeic/official_data/pdf/DAA.pdf

Jiang, N. (2000). Lexical representation and development in a second language. *Applied Linguistics*, *21*(1), 47–77.

Jones, G., & Witherstone, H. (2011). Lexical and sublexical knowledge influences the encoding, storage, and articulation of nonwords. *Memory and Cognition*, *39*(4), 588–599.

Kaneko, M. (2017). Vocabulary Size Targets for the TOEIC Test. *JACET Journal*, *61*, 57–67.

Kim, H. (2013). Statistical notes for clinical researchers: Assessing normal distribution (2) using skewness and kurtosis. *Restorative Dentistry and Endodontics*, *38*(1), 52–54. https://doi.org/10.5395/rde.2013.38.1.52

Kobayashi, Y. (2001). The learning of English at academic high schools in Japan: Students caught between exams and internationalization. *The Language Learning Journal*, *23*(1), 67–72. https://doi.org/10.1080/09571730185200111

Krashen, S. D. (1982). *Principles and practice in second language acquisition*. Oxford: Pergamon.

Kuhl, P. K, (2000). A new view of language acquisition. Proceeding of the National Academy of Sciences USA, *97*, 11850–11857.

Lange, K.A. (2018a). Analyzing difficulties in aural word recognition for Japanese English learners: Identifying function words in connected speech. *CASELE Research Bulletin, 48*, 63–74.

Lange, K. A. (2018b). Exploring the Relationship that Listening Performance has with Bottom-Up Skills and Metacognitive Awareness. *Bulletin of the Graduate School of Education, Hiroshima University. Part. II, Arts and Science Education*,(*1*), 67, 125-133.

Lange, K., & Matthews, J. (2020a). Exploring the relationships between L2 vocabulary knowledge, lexical segmentation, and L2 listening comprehension. *Studies in Second Language Learning and Teaching*, *10*(4), 723–749. https://doi.org/10.14746/ssllt.2020.10.4.4

Lange, K., & Matthews, J. (2020b). Paused Transcription Test (Lange & Matthews, 2020), Mendeley Data, V1, doi: 10.17632/g278w62zpg.1

Leech, G. (2000). Grammars of spoken English: New outcomes of corpus-oriented research. *Language Learning*, *50*(4), 675–724. http://dx.doi.org/10.1111/0023-8333.00143

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: Bradford.

Luce, P., & Pisoni, D. (1998). Recognizing spoken words: the neighborhood activation model. *Ear and Hearing*, *19*, 563–574.

Matthews, J., & Cheng, J. (2015). Recognition of high frequency words from speech as a predictor of L2 listening comprehension. *System*, *52*, 1–13. https://doi.org/10.1016/j.system.2015.04.015

Matthews, J., & O'Toole, J. M. (2015). Investigating an innovative computer application to improve L2 word recognition from speech. *Computer Assisted Language Learning*, *28*, 364–382. https://doi.org/10.1080/09588221.2013.864315

Matthews, J. (2018). Vocabulary for listening: Emerging evidence for high and mid-frequency vocabulary knowledge. *System*, *72*, 23–36. https://doi.org/10.1016/j.system.2017.10.005

McCelland, J., & Elman, J. (1986). The trace model of speech perception. *Cognitive Psychology*, *18*(1), 1–86.

McLean, S., Kramer, B., & Beglar, D. (2015). The creation and validation of a listening vocabulary levels test. *Language Teaching Research, 19*(9), 741–760. https://doi.org/10.1177/1362168814567889

McQueen, J. M. (1998). Segmentation of continuous speech using phonotactics. *Journal of Memory and Language*, *39*, 21-46. doi:10.1006/jmla.1998.2568

Milton, J. (2013). Measuring the contribution of vocabulary knowledge to proficiency in the four skills. In C. Bardel, C. Lindqvist, & B. Laufer (Eds.), *L2 vocabulary acquisition, knowledge and use: New perspectives on assessment and corpus analysis* (pp. 57–78). European Second Language Association. Eurosla Monographs Series 2, European Second Language Association. Retrieved from http://www.eurosla.org/monographs/EM02/EM02home.php

Mizumoto, A., & Shimamoto, T. (2008). A comparison of aural and written vocabulary size of Japanese EFL university learners. *Language Education and Technology*, *45*, 35–51. https://doi.org/10.24539/let.45.0_35

Morais, J., Cary, L., Alegria, J., & Bertelson, P. (1979). Does awareness of speech as a sequence of phones arise spontaneously? *Cognition*, *7*, 323-331.

Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher, 31*(7), 9–13.

Nation, I. S. P. (n.d.). The BNC/COCA headwords lists. [PDF files]. Retrieved from www.victoria.ac.nz/lals/about/staff/paul-nation

Norris, D., Cutler, A., McQueen, J. M., & Butterfield, S. (2006). Phonological and conceptual activation in speech comprehension. *Cognitive Psychology*, *53*(2), 146–193. https://doi.org/10.1016/j.cogpsych.2006.03.001

Pellegrino, F., Coupé, C., & Marsico, E. (2011). Across-language perspective on speech information rate. *Language*, *87*, 539–558. doi:10.1353/lan.2011.0057

Read, J. (1993). The development of a new measure of L2 vocabulary knowledge. *Language Testing, 10*(3), 355–371. https://doi.org/10.1177/026553229301000308

Read, J. (1998). *Validating a test to measure depth of vocabulary knowledge*. In A. Kunnan (Ed.), Validation in language assessment (pp. 41–60). Mahwah, NJ: Erlbaum. https://doi.org/10.4324/9780203053768

Riazi, A. M. & Candlin, C. (2014). Mixed-methods research in language teaching and learning: Opportunities, issues and challenges. *Language Teaching*, *47*, 135–173. https://doi.org/10.1017/S0261444813000505

Rost, M. (2012). *Teaching and researching listening*. Essex, England: Longman.

https://doi.org/10.4324/9781315833705

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing, 18*(1), 55–89. https://doi.org/10.1191/026553201668475857

Segalowitz, N., & Hulstijn, J. (2005). Automaticity in second language learning. In J. Kroll & A. De Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 371-388). Oxford, UK: Oxford University Press.

Sheerin, S. (1987). Listening comprehension: Teaching or testing? *ELT Journal*, *41*(2),126-131. doi:10.1093/elt/41.2.126

Sheppard, B., & Butler, B. (2017). Insights into Student Listening from Paused Transcription. *CATESOL Journal*, *29*(2), 81–107.

Shin, D., & Nation, P. (2008). Beyond single words: The most frequent collocations in spoken English. *ELT journal*, *62*(4), 339–348.

Siegel, J., & Siegel, A. (2015). Getting to the bottom of L2 listening instruction: Making a case for bottom-up activities. *Studies in Second Language Learning and Teaching*, *5*(4), 637–662. https://doi.org/10.14746/ssllt.2015.5.4.6

Siegel, J. (2016). Listening vocabulary: Embracing forgotten aural features. *RELC Journal*, *10*(3), 377–386. https://doi.org/10.1177/0033688216645477

Stæhr, L. S. (2009). Vocabulary knowledge and advanced listening comprehension in English as a foreign language. *Studies in Second Language Acquisition, 31*(4), 577–607. https://doi.org/10.1017/s0272263109990039

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Allyn & Bacon/Pearson Education.

Tauroza, S., & Allison, D. (1990). Speech rates in British English. *Applied Linguistics*, *11*(1), 90–105. https://doi.org/10.1093/applin/11.1.90

Thorpe, K., & Fernald, A. (2006). Knowing what a novel word is not: Two-year-olds "listen through" ambiguous adjectives in fluent speech. *Cognition*, *100*, 389–433. doi:10.1016/j.cognition.2005.04.009

Trofimovich, P. (2008). What do second language listeners know about spoken words? Effects of experience and attention in spoken word processing. *Journal of Psycholinguistic Research*, *37*(5), 309-329.

Tsui, A., & Fullilove, J. (1998). Bottom-up or top-down processing as a discriminator of L2 listening performance. *Applied Linguistics*, *19*, 432–451. http://dx.doi.org/10.1093/applin/19.4.432

Vafaee, P., & Suzuki, Y. (2020). The relative significance of syntactic knowledge and vocabulary knowledge in second language listening ability. *Studies in Second Language Acquisition, 42*(2), 383-410.

Vandergrift, L. (2003). Orchestrating strategy use: Toward a model of the skilled second language listener. *Language Learning, 53*(3), 463–496.

Vandergrift, L. (2007). Recent developments in second and foreign language listening comprehension

research. *Language Teaching*, *40*(3), 191–210. http://dx.doi.org/10.1017/S0261444807004338

Vandergrift, L., & Baker, S. (2015). Learner variables in second language listening comprehension: An exploratory path analysis. *Language Learning, 65*(2), 390–416. https://doi.org/10.1111/lang.12105

Vandergrift, L., & Goh, C. C. M. (2012). *Teaching and learning second language listening*. New York, NY: Routledge.

Vandergrift, L., Goh, C.C.M., Mareschal, C., & Tafaghodtari, M.H. (2006). The metacognitive awareness listening questionnaire (MALQ): Development and validation. *Language Learning*, *56*, 431–462.

Vandergrift, L., & Tafaghodtari, M.H. (2010). Teaching L2 learners how to listen does make a difference: An empirical study. *Language Learning*. *60*(2), 470–497.

Wallace, M.P. (2021). Individual differences in second language listening: Examining the role of knowledge, metacognitive awareness, memory, and attention. *Language Learning*. Advance online publication. doi:10.1111/lang.12424

Wang, Y., & Treffers-Daller, J. (2017). Explaining listening comprehension among L2 learners of English: The contribution of general language proficiency, vocabulary knowledge and metacognitive awareness. *System*, *65*, 139–150. https://doi.org/10.1016/j.system.2016.12.013

Watanabe, S. (2008). Kyoushi ga suisoku suru koukousei no risuninngucyu no ishiki/houryaku shiyou to jittai to no hikaku [Comparing teachers' perceptions of high school students' awareness and strategy use while listening with the students' actual responses]. *Step Bulletin, 20*, 224–232.

Webb, S., & Rodgers, P. (2009). The lexical coverage of movies. *Applied Linguistics*, *30*(3), 407–427. https://doi.org/10.1093/applin/amp010

Weber, A., & Cutler, A. (2006). First-language phonotactics in second-language listening. *Journal of the Acoustic Society of America*, *119*, 597-607. doi:10.1121/1.2141003

Wong, S. W. L., Mok, P. P. K., Chung, K. K., Leung, V. W. H., Bishop, D. V. M., & Chow, B. W. (2017). Perception of native English reduced forms in Chinese learners: Its role in listening comprehension and its phonological correlates. *TESOL Quarterly*. *51*(1), 7–31.

Wong, S. W. L., Leung, V. W. H., Tsui, J. K. Y., Dealey, J., & Cheung, A. (2021). Chinese ESL learners' perceptual errors of English connected speech: Insights into listening comprehension. *System*, https://doi.org/10.1016/j.system.2021.102480

Yeldham, M. (2016). Second language listening instruction: Comparing a strategies-based approach with an interactive, strategies/bottom-up skills approach. *TESOL Quarterly*, *50*(2), 394–420. https://doi.org/10.1002/tesq.233

Yeldham, M. (2017). Techniques for researching L2 listeners. *System*, *66*, 13–26.

Yeldham, M. (2020). Does the presence of formulaic language help or hinder second language listeners' lower-level processing?. *Language Teaching Research*, *24*(3), 338-363. http://dx.doi.org/10.1177/1362168818787828

Appendix A

The following is a list of the thirty sentences used for the partial dictation test. The bolded words are the items students were asked to transcribe. The underlined words are the function words which were initially presented in reduced form.

1. Who **made him do** that?
2. Come **and see** me.
3. How **much is her** car?
4. Would **you like an** egg?
5. I **got more and more** sleepy.
6. I **should have** gone.
7. I **had to get some** more.
8. Was **it his or her** money?
9. She **tried talking to** them.
10. His **car is bigger than** mine.
11. Why **don't you go** too?
12. He **wanted something to** eat.
13. You **should take her** home.
14. Have **you gone to the** doctor?
15. We **must get up** early.
16. This **is your lucky** day.
17. I **thought I shouldn't** ask.
18. He **has to work** today.
19. Is **it hot in** Tokyo?
20. I'd **like to meet** them.
21. How **are you** feeling?
22. I **want you to be** happy.
23. When **do you leave** school?
24. We **met for the first** time.
25. This **was a lot of** fun.
26. I **need to talk to** you.
27. Could **you help me** out?
28. I **saw him at the** station.
29. I'm **not so sure about** that.
30. What **made him so** angry?

Appendix B

Metacognitive Awareness Listening Questionnaire (MALQ) – Japanese version (Watanabe, 2008)
自己のリスニング認知度についての調査

以下のそれぞれの記述について，自分の英語リスニングに最も当てはまる番号（1 全く当てはまらない，2 当てはまらない，3 あまり当てはまらない，4 少し当てはまる，5 当てはまる，6 とても当てはまる）に○をつけてください。

1　計画 / 評価　　聞く前に，どんなふうに聞こうかと考える。
2　集中　　わからないときには，よりいっそう音声テキストに集中する。
3　自己認識　　英語リスニングは，英語で「読む」「書く」「話す」ことよりも難しい。
4　翻訳　　聞きながら頭の中で訳していく。
5　問題解決　　わかった単語から，わからない単語を推測する。
6　集中　　ぼんやりしてしまったら，すぐに集中しなおす。
7　問題解決　　聞いていてわかったことを，すでに自分が知っていることを照らし合わせてみる。
8　自己認識　　英語リスニングは自分には難しい。
9　問題解決　　自分の経験や知識を理解に役立てる。
10　計画 / 評価　　聞く前に，それまでに聞いたことがあるよく似た音声テキストを思い浮かべる。
11　翻訳　　聞きながらキーワードを訳す。
12　集中　　集中が途切れたら，再度，集中しなおそうとする。
13　問題解決　　聞いているとき，自分の解釈が合ってないとわかれば，すぐに修正する。
14　計画 / 評価　　聞き終わってから，自分がどんなふうに聞いたかを振り返り，次にやろうと思う聞き方を考える。
15　自己認識　　英語を聞いていても緊張はしない。
16　集中　　内容が理解できないときは，あきらめて聞くのをやめてしまう。
17　問題解決　　わからない単語の意味を推測するために，音声テキストの概要（あらすじ）を利用する。
18　翻訳　　聞きながら一語一語を訳す。
19　問題解決　　単語の意味を推測するとき，自分の推測で合っているかどうかを確かめるために，それまで聞いたことを振り返る。
20　計画 / 評価　　聞きながら，何度か自分の理解度を確認する。
21　計画 / 評価　　聞いているとき，心の中に何らかの目標を持っている。

※実際の質問紙にカテゴリーは載せていない。

Appendix C

A partial sample of the dialogue used for the first two target phrases for Paused Transcription Test 1

Where did you grow up?

I grew up in St Louis Missouri it's in the center of the United States and it's on the Mississippi River it's a fairly big city.

What was it like?

So growing up in St Louis was fun I lived in a neighborhood with a few kids so we could play. We usually just played sports or rode our bicycles, it was… it was a good childhood.

What were your parents like?

My parents were a little strict I guess. I couldn't stay out very late I guess you know I had to come home when the…. when it began to get dark …dinner time but they didn't pressure me to do homework. On the weekends I usually had to do a lot of housework and there was always washing the dishes or vacuuming or cleaning something so my friend said my parents were strict.

(Note: partial sample only)

Appendix D

Paused transcription test target phrases grouped by reduced forms category

PTT section	*Function Words*	*Linking*	*Transitions Between Words*
PTT 1 Target Phrases	me to do /mi tu du/	quite happy for /kwaɪ tˈæpi fər/	we could play /wi kʊ pleɪ/
	when we were /wɛn wi wɜr/	she got out /ʃi gɑ daʊt/	we had together /wi hæ təˈgɛðər/
	like to play /laɪk tə pleɪ/	which are both /wɪ ʧɑr boʊθ/	visit those natural /vɪzə ðoʊz ˈnæʧərəl/
	do a lot /du ə lɑt/	that's all /ðæt səl/	what I like /wʌ daɪ laɪk/
PTT 2 Target Phrases	lots of pictures /lɑts ə ˈpɪkʧərz/	interested in Japanese /ˈɪntrɪstɪ dɪn ˈʤæpəniz/	she made sure /ʃi meɪ ʃʊr/
	interesting for me /ˈɪntrɪstɪŋ fər mi/	for learning another /fər ˈlɜrnɪŋ gəˈnʌðər/	to learn more /tə lɜr mər/
	at the store /ət ðə stər/	what it was /wʌ dɪ wəz/	eat my favorite /i maɪ ˈfeɪvrət/
	wash your hands /wɑʃ jər hændz/	which is exciting /wɪ ʧɪz ɪkˈsaɪtɪŋ/	it looked like /ɪt lʊk laɪk/
PTT 3 Target Phrases	walk to work /wɔk tə wɜrk/	still in a /stɪ lɪn ə/	a few weeks /ə fju wiks/
	I can do /aɪ kən du/	some books about /səm bʊk səˈbaʊt/	why did you /waɪ dɪ ʤʊ/
	things to learn /θɪŋz tə lɜrn/	can learn about /kən lɜr nəˈbaʊt/	first you're /fɜr stʃʊr/
	as a good	seasons are really	aren't allowed

	/əz ə gʊd/	/ˈsizən zər ˈrɪli/	/ɑrən təˈlaʊd/
PTT 4 Target Phrases	sort of thing /sɔrt əv θɪŋ/ in the evenings /ɪn ðə ˈivnɪŋz/ helped him get /hɛlpt ɪm gɛt/ comes to mind /kʌmz tə maɪnd/	know them as /noʊ ðə məz/ I grew up /aɪ gru wʌp/ was turned away /wəz tɜrn dəˈweɪ/ most of the /moʊ stəv ðə/	the hard part /ðə hɑr pɑr/ would be fun /wə bi fʌn/ get off the /gɛ dɔf ðə/ don't steal /doʊn stil/
PTT 5 Target Phrases	skill that helps /skɪl ðət hɛlps/ you can get /jʊ kɪn gɛt/ near the castle /nɪr ðə ˈkæsəl/ when you walk /wɛn jʊ wɔk/	how important listening /haʊ wɪmˈpɔrtənt ˈlɪsnɪŋ/ be able to /bi ˈjeɪbəl tə/ found out later /faʊn daʊt ˈleɪtər/ kinds of unique /kaɪn zəv juˈnik/	you go on /jʊ goʊ wɔn/ not just yourself /nɑ ʤəs jʊrˈsɛlf/ had to be /hæ tə bi/ thousand years old /ˈθaʊzən jɪr zoʊld/

Appendix E

PTT scoring rubric with rationale and examples

General instructions for scoring individual words in target phrases

Score	Principle	Comments	Example target word → example answer
1.0	The word is spelled correctly.	This answer is easy to score because there is no subjectivity involved.	unique → unique evenings → evenings favorite → favorite
1.0	The word is spelled incorrectly but its phonological form is acceptable according to English phonology.	Subjectivity involved. The test construct is aural decoding therefore slight spelling errors are accepted as long as the spelling approximates the phonological form of the target word.	uniqe / uniek / unieque→ unique wosh → wash wark → work natral → natural heands → hands pictuer → picture turnd → turned thousan → thousand cathle / casltel / castl / casle / catsle → castle lestening / lisning→ listening wuld → would allowd → allowed
0.75	A homophone of the target word is decoded instead of the target word	Although the target word was accurately decoded phonetically, the spelling of the word indicates the wrong word was decoded indicating difficulty with understanding the meaning of the input.	steel → steal witch → which aloud → allowed
0.5	The word is spelled incorrectly and has more ambiguity in the interpretation of its	Subjectivity involved. The incorrect spelling results in an incorrect phonological form that does not	unik / unic / unice/ uniece / unecue→ unique larn / laurn → learn laurning → learning

	phonological form.	approximate the target word. However, there is partial evidence of correct phonological recognition depending on the interpretation of the word's spelling.	alaud / aroud → allowed youself → yourself gat → got leastening / listeing→ listening thousant → thousand exaciting → exciting watsh → wash seson → season reary → really rater → later wark → walk
0.5	Incorrect conjugation/ incorrect form of verb but clear evidence that the root word is recognized.	The core element of the target word is recognized correctly but there is an error in inflection or word form, such as tense or plurality.	played → play visiting → visit interesting → interested look → looked was → is make → made will → would can → could a → the can't → can come → comes are → aren't
0	Significant spelling mistakes make interpretation of the phonological form difficult and its association with the target word tenuous.	The phonological form of the spelling represents a clearly different word from the target word. (Two or more incongruencies with phonological form.)	leran → learn leauning → learning sousend / thouthont / sousond → thousand unirk → unique araude / arowd → allowed
0	A different word, which may be phonologically similar is decoded.	The orthographic form represents a clearly different word from the target word. Despite the phonological similarities, the accurate spelling of the transcribed word	quit / quiet → quite way → away national → natural pray → play a → are leaning → learning fan → fun

		demonstrates that a separate word from the target word was decoded.	mine → mind listing → listening waking → walking latter → later
0	No target word provided		

General instructions regarding deducting points for errors in the target phrases

Score	Principle	Comments	Example target word → example answer
- 0.25	0.25 points are deducted for mistakes of word order in the target phrase.	One of the words in the three-word target phrase is transcribed in an incorrect order relative to the other two target words.	to me do → me to do you hushed hands → wash your hands to walk → walk to you my why → why did you can I do → I can do the most → most of the important how listening → how important listening listening is important → how important listening helps skills → skill that helps
- 0.25	0.25 points are deducted for every extra word in the target phrase transcription.	An extra word is contained within the target phrase. It must come between two of the words in the target phrase.	We have time together → we had together a few day on weeks → a few weeks comes to the mind → comes to mind comes to my mind → comes to mind grow them up → I grew up of the thing → sort of thing

			kind of the unique → kinds of unique you are going → you go on

Appendix F

Introspective Self-observation Protocol (ISOP)

The following is an example of what the researcher said during the ISOP.

You have consented to these sessions being recorded. Are you still happy for me to record this session today? I will ask you to take a listening test in the same way as you did during class. However, after each transcription I will pause the audio recording and ask you about what you heard. I will ask you the following questions in turn while referring to your transcription. Please answer in detail in Japanese.

1) "How well could you understand the content of the section you just heard? On a scale of 1 (low) to 5 (high), please rate your understanding."
2) "Please tell me the meaning of this section's content in as much detail as possible."
3) "Did you only transcribe the phrase as you heard it or did you think about the meaning of the whole section?"
4) "Please tell me the meaning of the sentence containing the target phrase."
5) "Please tell me any parts of the phrase that were difficult to perceive."
6) "Please tell me why you think that part was difficult to perceive."

Next, I will show you the correct transcription and may ask the following questions:

7) "Please tell me the meaning of the sentence containing the target phrase."
8) "Which part(s) were difficult to perceive?"
9) "Why do you think those parts were difficult to perceive?"

Japanese ISOP Questions

1:「今聞いた会話文全体の内容を、どのくらい理解できましたか。よく理解できなかったを 1 とし、よく理解できたを 5 とすると、1〜5 のどれになりますか。」
2:「その部分の内容を、できるだけ詳しく日本語で説明してください」
3:「ターゲットフレーズを聞いたときに、その部分の音声だけを頼りに、解答しましたか。それとも、文脈を参考にして解答しましたか？」
4：「ターゲットフレーズを含む文の内容を言ってください」
5：「聞き取りにくかった部分を教えてください」
6：「聞き取りにくかった理由を教えてください。」

これから、正解を見せます。そして、次の質問をします。

7：「このフレーズを含む文の内容を言ってください」
8：「どの部分が聞き取れなかったですか？」
9：「なぜその部分が聞き取れなかったと思いますか？」

Appendix G

Introspective Self-observation Protocol Transcript

This is the translation of an ISOP interview transcription for the first 6 questions in Paused Transcription Test 1 with Maki, a higher-level sub-group member.

R: Researcher

P: Participant

Number 1

(The recording is played until reaching the beep for test item 1, “with a few kids so (we could play)” then paused for 10 seconds. After the pause, the following dialogue took place in Japanese. The participant’s transcription was “so we could play”).

R: For number 1, how would you rate your understanding of the content on a scale from 1 to 5?

P: (This is in response to question 1, “How well could you understand the content of the section you just heard? On a scale of 1 (low) to 5 (high), please rate your understanding.”).

About 2.

R: OK, now please tell me about the content in Japanese in as much detail as you can.

P: (This is in response to question 2, “Please tell me the meaning of this section’s content in as much detail as possible.”)

Well, I focused too much on listening for the words I would need to transcribe at the end and didn’t try to understand what I was listening to.

R: I see. And you said that you didn’t pay much attention to the meaning of this section and were just waiting for the beep so number 3 is you transcribed the phrase based on the sound right?

P: (This is in response to question 3, “Did you only transcribe the phrase as you heard it or did you think about the meaning of the whole section?”)

That’s right.

R: OK. So for number 4 please tell me the meaning of the content of this target phrase.

P: (This is in response to question 4, “Please tell me the meaning of the sentence containing the target phrase.”)

“And so, we could do that” (Her transcription was “so we could play” but she interpreted “play” as “do”)

R: I see. If there were any parts that were hard to hear, please tell me.

P: (This is in response to question 5, “Please tell me any parts of the phrase that were difficult to perceive.”)

I’m not sure if this is correct or not but if this is “could” then the “d” part isn’t pronounced strongly so it sounds like “cou” so I’m not sure if this word is correct. The word before “play” had that pronunciation so I thought it might be “could”. If that part (/d/) is missing then it’s probably this word (could).

(In this the participant already explained her response to question 6, “Please tell me why you think that part was difficult to perceive.”, so the researcher continued to question 7.)

R: So, next we'll look at the answer first and discuss the next questions. You transcribed it correctly and you told me the meaning of the phrase.

P: (This is in response to question 7, "Please tell me the meaning of the sentence containing the target phrase.")

I know the meaning. (the student wrote "so we could play")

R: Right. Were there any parts that were difficult to hear?

P: (This is in response to question 8, "Which part(s) were difficult to perceive?")

It was OK.

(Based on this response, number 9, "Why do you think those parts were difficult to perceive?" wasn't asked.)

R: OK, let's go on to the next item and I'll ask you the same questions about it.

P: *OK.*

Number 2

(The recording is played until reaching the beep for test item 2, "they didn't pressure (me to do) homework" then paused for 10 seconds. After the pause, the following dialogue took place in Japanese. The participant's transcription was "pressure me homework").

R: What was your level of comprehension for this section?

P: (This is in response to question 1, "How well could you understand the content of the section you just heard? On a scale of 1 (low) to 5 (high), please rate your understanding.").

Three.

R: OK so could you tell me the basic meaning of this passage and the context?

P: (This is in response to question 2, "Please tell me the meaning of this section's content in as much detail as possible.")

The parents come to the house, come to the house. When the parents came to the house, they said something like, "Do your homework instead of playing sports." I think.

R: OK. So, as you listened did you mainly use pronunciation to understand as in the first item or did you focus more on the whole meaning?

P: (This is in response to question 3, "Did you only transcribe the phrase as you heard it or did you think about the meaning of the whole section?")

I tried to focus this time and listen to the whole section.

R: Oh, I see. So you understood more this time.

P: (This is in response to question 4, "Please tell me the meaning of the sentence containing the target phrase.") *Yes.*

R: What parts were difficult to understand?

P: *In this phrase?*

R: Yeah.

P: *Let's see, this verb part* (i.e. pressure), in my English classes we didn't use this word so much and in listening I don't think I've heard it before so…(inaudible).

R: I see, the verb "pressure" is a little unfamiliar isn't it. Are the other words besides "pressure" alright?

P: *Maybe.*

R: OK here is the answer.

P: *Oh, my answer was totally wrong* (the student wrote "pressure me homework").

R: But before this part there was the word "pressure".

P: *I wonder why I missed them* (i.e. "to do")

R: Can you tell me the meaning of this phrase?

P: (This is in response to question 7, "Please tell me the meaning of the sentence containing the target phrase.")

I know the meaning – in order for me to do homework they pressured me

R: Right. So "to" and "do" were difficult to hear?

P: (This is in response to question 8, "Which part(s) were difficult to perceive?")

Yes.

R: Why do you think those words were difficult to hear? Just that part.

P: (This is in response to question number 9, "Why do you think those parts were difficult to perceive?")

Words like "pressure" and "homework" were spoken clearly but "to do" is really fast so that's why it was left out. Only the words with a stronger impression remained (in my mind).

R: OK, let's continue. Ask me questions any time and just tell me if you'd like to stop.

P: *OK.*

Number 3

(The recording is played until reaching the beep for test item 3, "the time we had together" then paused for 10 seconds. After the pause, the following dialogue took place in Japanese. The participant's transcription was "time hard we together")

R: Please tell me you level of understanding for this section.

P: (This is in response to question 1, "How well could you understand the content of the section you just heard? On a scale of 1 (low) to 5 (high), please rate your understanding.").

Four

R: Do you understand what was said before the target phrase?

P: (This is in response to question 2, "Please tell me the meaning of this section's content in as much detail as possible.")*You have an older brother and older sister who are about 10 years older than you. You are the youngest. The question was, what was most memorable to you when you were a child, and you said that's a difficult question.*

R: You remember well but the part after that?

P: *The next part went by quickly and I didn't understand.*

R: OK. Did you focus mainly on pronunciation or did you use the context for the whole section to try and understand?

P: (This is in response to question 3, "Did you only transcribe the phrase as you heard it or did you think about the meaning of the whole section?")

I used the whole section.

R: Next, please tell me the meaning of the target phrase.

P: (This is in response to question 4, “Please tell me the meaning of the sentence containing the target phrase.”). *When we are together time feels fast. But “hard” doesn’t have that meaning ! But that is my impression or feeling.*

R: I often rely on feeling too when trying to understand Japanese. Please tell me the parts that were hard for you to understand.

P: (This is in response to question 5, “Please tell me any parts of the phrase that were difficult to perceive.”)*The end part – the part I was supposed to listen to was very difficult for me to hear and I didn’t understand the meaning so well either.*

R: By the “end part” do you mean this part ending in “together” (indicating the word on the student’s answer sheet)?

P: *I felt like it was too fast. I understood the section well, I marked it as four for comprehension, but the last two sentences or so, I totally couldn’t understand.*

R: Number three is “time we had together” (showing on the paper).

P: *Oh, I mistook “had” for “hard”. I also put the “we” in the wrong word order.*

R: Is the meaning of this phrase OK?

P: (This is in response to question 6, “Please tell me why you think that part was difficult to perceive.”). *we spent time together.*

R: Which parts of the phrase were hard to understand and why?

P: (This is in response to question 8, “Which part(s) were difficult to perceive?” and question number 9, “Why do you think those parts were difficult to perceive?”). *I thought ‘had’ was ‘hard’ and since it was fast I confused the word order a little.*

R: Yes, with the *we* coming before *hard*. Ok let’s go on to the next one.

Number 4

(The recording is played until reaching the beep for test item 4, “was quite happy for me” then paused for 10 seconds. After the pause, the following dialogue took place in Japanese. The participant’s transcription was “quiet happy for me”)

R: Please tell me your level of understanding for this section.

P: (This is in response to question 1, “How well could you understand the content of the section you just heard? On a scale of 1 (low) to 5 (high), please rate your understanding.”).

Three

R: Next, please tell me the meaning of the section before the target phrase.

P: (This is in response to question 2, “Please tell me the meaning of this section’s content in as much detail as possible.”).

It was a talk about a time at camp. You got up in the morning, watched birds, ate a campfire breakfast and so on.

R: Did you just focus on the sound of the target phrase or did you try to use the context to help you understand?

P: (This is in response to question 3, “Did you only transcribe the phrase as you heard it or did you think about the meaning of the whole section?”).

I used the context.

R: Please tell me the meaning of the phrase.

P: (This is in response to question 4, “Please tell me the meaning of the sentence containing the target phrase.”).

For me it was a little happiness? A small happiness.

R: Were there any parts that were hard to understand?

P: (This is in response to question 5, “Please tell me any parts of the phrase that were difficult to perceive.”).

I heard "quiet" but I don't think I've ever seen a sentence like this before (looking at her test paper) so I'm not sure about this word.

R: Next please tell me why you think this was difficult to hear – especially "quiet"

P: (This is in response to question 6, "Please tell me why you think that part was difficult to perceive.").

I know that "quiet" just means quiet so this sentence just doesn't make much sense to me and I think maybe it wasn't this word (i.e. quiet).

R: So this is the target phrase (showing the student) – "quite happy for me"

P: *Oh, I see. That's what it was.*

R: These two words *quite* and *quiet* are very similar. I think you understand this but please tell me the meaning in Japanese.

P: (This is in response to question 7 after the participant has seen the correct answer, "Please tell me the meaning of the sentence containing the target phrase.")

As for me, I was very happy

R: (Question 8 was skipped because it was clear that "quite" had been the part that was difficult to perceive.)That's right and why was this part difficult to catch?

P: (This is in response to question number 9, "Why do you think those parts were difficult to perceive?")

I can understand this phrase "quite happy" when I read it but I've never used this combination of words in speech or writing so I wasn't familiar with it. So, I think I had almost forgotten this word "quite".

R: What do you think about the connection between the words "quite happy", "quite happy".

P: *When you say them together, it sounds like "quiet" to me.*

R: That's interesting. Let's go on to the next one.

Number 5

(The recording is played until reaching the beep for test item 5, "she got out of" then paused for 10 seconds. After the pause, the following dialogue took place in Japanese. The participant's transcription was "she got out off")

R: how would you rate your comprehension?

P: (This is in response to question 1, "How well could you understand the content of the section you just heard? On a scale of 1 (low) to 5 (high), please rate your understanding.").

Four.

R: Next please tell me what you could understand from this section.

P: (This is in response to question 2, "Please tell me the meaning of this section's content in as much detail as possible.").

This was about your worst childhood memory, there weren't so many but maybe, when your dog hit a car was the worst memory. This "She" refers to the dog so maybe the phrase is about when you parted with your dog?

R: Which parts were difficult for you to understand?

P: (This is in response to question 5, "Please tell me any parts of the phrase that were difficult to perceive.").

Everything after "she" was difficult for me to understand (i.e. the target phrase).

R: Why do you think this part was hard to understand "got out off"?

P: (This is in response to question 6, “Please tell me why you think that part was difficult to perceive.”).

It is fairly connected so I didn’t know where the word endings were.

R: Now, please tell me the meaning of the target phrase.

P: (This is in response to question 4, “Please tell me the meaning of the sentence containing the target phrase.”).

Earlier there was the word “when” – when I parted with her

R: Now let’s look at the answer – “she got out of” is correct.

P: *Oh, I see, of* is correct.

R: What do you think this means? It is sort of difficult to imagine without hearing the rest of the sentence.

P: (This is in response to question 7 after the participant has seen the correct answer, “Please tell me the meaning of the sentence containing the target phrase.”)

I don’t know. I think I’ve learned this before (referring to “got out”) *but …*

R: Got out means to escape from something. OK, next one.

Number 6

(The recording is played until reaching the beep for test item 6, “together from when we were” then paused for 10 seconds. After the pause, the following dialogue took place in Japanese. The participant’s transcription was “from one way out”).

R: Please rate your level of understanding for this section.

P: (This is in response to question 1, “How well could you understand the content of the section you just heard? On a scale of 1 (low) to 5 (high), please rate your understanding.”).

Four.

R: So you understood the context fairly well.

P: *Yes.*

R: Could you tell me the meaning of the section?

P: (This is in response to question 2, “Please tell me the meaning of this section’s content in as much detail as possible.”).

This is a continuation of the story about the dog that jumped the fence and hit the car. The dog hit the car quite hard and died – you cried and were scared of the experience. After that something about your best friend. You had one close friend named Mikee and you called him Mike and you played after school and I think that’s all.

R: That’s great, you really understood this section well. Next, could you tell me the meaning of the target phrase?

P: (This is in response to question 4, “Please tell me the meaning of the sentence containing the target phrase.”).

The last part of this section I really couldn’t understand. I could catch “from”, maybe “from one” but after this it is so connected that I completely couldn’t understand.

R: Yes.

P: *If I write down what I heard it would be this (referring to her transcription) but I have no idea what it means, so I’m sure it’s wrong.*

R: I see. Could you tell me the parts that were difficult to hear for you?

P: (This is in response to question 5, “Please tell me any parts of the phrase that were difficult to perceive.”).

The part after "from".

R: And why do you think that part was difficult to hear?

P: (This is in response to question 6, "Please tell me why you think that part was difficult to perceive.").

It was very connected and all I could hear was just sounds like "bararara"

R: That's right. If I say it again it sounds like "from when we were", "from when we were very young".

P: *So after "when" comes "from"?*

R: Yes. Now please tell me the meaning of the target phrase.

P: (This is in response to question 7 after the participant has seen the correct answer, "Please tell me the meaning of the sentence containing the target phrase.")

I don't really understand. I just heard it like this. It's something like "we (something) from that time"

R: OK. So which part was the most difficult to understand?

P: (This is in response to question 8, "Which part(s) were difficult to perceive?").

"When we were"

P: (This is in response to question number 9, "Why do you think those parts were difficult to perceive?")

After "from" was "when" and the phrase "from when" and I didn't know you could put "when" after "from" – I didn't think you could do that so even if I thought I had heard "when" after "from" I don't think I would have written "when". I thought this word order was not grammatically acceptable.

R: I see. That's interesting. "From when" and "from one" sound very similar too. OK let's continue.