

Spoken Word Form Recognition with a Mobile Application: Comparing Azerbaijani and Japanese Learners



Joshua Matthews , Kriss Lange , and Gunther M. Wiest 

Abstract To develop adequate levels of automaticity in second language (L2) word recognition learners need to engage with a large amount of meaningful spoken target language input. Unfortunately, there is often not enough in-class time available for this. Thus, using technology to facilitate out-of-class development of L2 word recognition holds strong potential. This study explores and compares the use of a mobile language learning application (app) by Azerbaijani and Japanese learners of English. The app was designed to improve second language (L2) learners' recognition of the spoken form of high-frequency vocabulary by giving learners repeated and self-paced opportunities to listen to and transcribe target words presented in fluently articulated, contextual sentences. First, a quasi-experimental approach was applied in both contexts to compare improvements in word recognition among those who used the app and those in a control group that did not. Next, learner interaction data collected in the app's database were used to investigate links between learner engagement and improvements in word recognition. Further, the most challenging words for learners to recognize and transcribe were identified and compared between treatment group members from each context. Finally, stimulated recall protocols in the learners' respective first languages (L1) were conducted among a subgroup from each context to investigate the origins of their difficulty with aurally recognizing their most challenging words. Suggestions are provided for the development of word recognition from speech that may be useful both within and beyond the classroom.

Keywords L2 listening · L2 word form recognition from speech (WFRS) · Mobile-assisted language learning · Out-of-class learning

J. Matthews (✉)

University of New England, Armidale, Australia
e-mail: jmatth28@une.edu.au

K. Lange

University of Shimane Matsue Campus, Matsue, Japan
e-mail: k-lange@u-shimane.ac.jp

G. M. Wiest

ADA University, Baku, Azerbaijan
e-mail: gwiest@ada.edu.az

1 Introduction

Recognizing words in speech is a fundamental starting point in the second language (L2) listening comprehension process. Without first recognizing the phonological form of words in speech, a listener is unlikely to be able to access the associated meaning. Rapid and accurate recognition of words from speech is a key attribute of skilled L2 listeners (Field, 2008a). As well as being a fundamental element of skilled listening, the recognition of words in L2 speech can present significant challenges for L2 learners (Lange & Matthews, 2020, 2021). These challenges can be understood by considering the intrinsic nature of spoken words. Unlike static words on a page which can be revisited by the reader's eye as required, words in speech are available to the listener for only a very brief duration. In short, words in speech are temporal. Further, words in speech are often blended (i.e., coarticulated) such that boundaries between consecutive words in an utterance (intonation unit) are not explicit. This can make the phonological form of spoken words variable and dependent on the acoustic context within which they occur. The specific challenges associated with L2 word recognition from speech highlight the merit of interventions that help learners effectively deal with this fundamental aspect of L2 learning.

Efforts to enhance L2 learners' spoken word recognition from speech should focus on developing automatized phonological knowledge of words. Firstly, to account for the temporal nature of speech learners must be able to quickly recognize spoken words without drawing excessively from their finite cognitive resources, which are required for higher level L2 listening comprehension processes. Automaticity in language processing is underpinned by implicit knowledge which develops in response to the frequency effects of exposure to input (Ellis, 2002). In other words, a learner's automaticity in word recognition develops in step with the number of opportunities to successfully recognize words in the spoken target language. Furthermore, to account for the blended and variable nature of spoken words, learners must also possess adequate levels of phonological knowledge. Simply put, learners need to know what words actually *sound like* when articulated in speech. Importantly, an L2 learner's knowledge of words in written form is typically not equivalent to knowledge of words in spoken form (Cheng & Matthews, 2018). For instance, it is not unusual for a language learner to know a word in the written form but be unable to recognize that same word when it is encountered in speech (Carney, 2021; Goh, 2000; Lange & Matthews, 2021). As word knowledge is modality specific, it is important to provide learners with extensive and meaningful opportunities to engage with *spoken* target language input.

Unlike most L1 listeners who develop automatic word recognition effortlessly throughout their lifetime, L2 listeners need to devote time to actively develop word recognition from speech. However, the outlay of time needed to facilitate the automatization of L2 word recognition from speech is a significant problem for language learning. Firstly, there is inadequate time within in-class settings to adequately engage with this involved task. Additionally, in English as a foreign language (EFL) contexts, the extent of a learner's English language usage and exposure may typically not

extend beyond in-class learning. For example, even at the tertiary level and after years of formal instruction, learners may still not have the phonological knowledge needed to recognize the 3,000 most frequently occurring words of the English language (Matthews & Cheng, 2015). In such circumstances, it is especially important to assist learners in the development of their L2 word recognition from speech as part of targeted out-of-class learning. Mobile-assisted language learning holds particularly strong promise in facilitating language learning that extends beyond the classroom, especially in the current era, where most language learners almost always have powerful smartphones (mini-computers) close at hand. The current study explores the potential of mobile-assisted language learning for the development of L2 word recognition in out-of-class learning conducted within two different formal EFL university contexts in Azerbaijan and Japan.

2 Literature Review

2.1 *The Importance of Word Knowledge in L2 Listening Comprehension*

Recent research has made clear the strong connection between L2 word knowledge and L2 listening comprehension (Cheng et al., 2022; Matthews, et al., 2023; Vafae & Suzuki, 2019; Wallace, 2022). The special importance of word knowledge in language learning can be attributed to the strong form-meaning unity that occurs at the lexical level (Hulstijn, 2002). In terms of listening, this means that if the phonological form of a word (or string of words) can be recognized, the listener has a chance to access the appropriate corresponding meaning. However, in relation to L2 listening, the temporal, blended, and variable nature of the form of words in speech makes their recognition a considerable challenge. Meeting this challenge entails the listener skillfully utilizing both linguistic information derived from bottom-up processing (i.e., aural decoding) and contextual information from top-down processing (Flowerdew & Miller, 2005). Therefore, difficulty recognizing L2 words from speech may generally be attributable to a combination of inadequate linguistic knowledge needed for bottom-up processing and inadequate utilization of contextual information (e.g., background knowledge, comprehension of the preceding aural text, pragmatic knowledge).

Listening comprehension problems often stem from a difficulty in recognizing words in speech despite those words being known by the L2 listener in the written form (Goh, 2000; Masrai, 2020). For example, Carney (2021) conducted interviews and analysis of 15 Japanese EFL learners' difficulties in comprehending English speech consisting of high-frequency vocabulary. The most common reasons for comprehension breakdown were difficulties with L1 phonological influence, word segmentation, and word recognition. Lange and Matthews (2021) also used a mixed methods approach among Japanese EFL learners to determine that a significant cause

of misunderstanding in L2 listening was an inability to recognize the phonological form of L2 words. Through the application of L1 interviews, Lange and Matthews (2021) showed that phonological representations of high-frequency words stored in learners' mental lexicons were often strongly influenced by L1 intrusion. This saw learners struggle to map words from speech onto their corresponding semantic representations despite knowing the word's meaning and written form. Learners reported that their mental representations of English words had been altered by their Japanese phonological system making even known words difficult to recognize from speech. These issues suggest the importance of helping learners more accurately recognize the spoken form of words including those that the learners are likely to have encountered many times in their previous language learning experiences (i.e., high-frequency words).

In terms of pedagogical recommendations, research makes clear that it is important for learners to apply non-linguistic knowledge such as background knowledge and strategies to assist their L2 listening comprehension (Yeldham, 2016). However, early-stage learners are unlikely to have developed sufficient levels of automaticity in L2 language processing necessary to effectively integrate both linguistic and non-linguistic knowledge sources. When trying to comprehend L2 speech, such learners are likely to experience a heavy cognitive burden simply trying to catch words here and there from largely unrecognizable sequences of spoken words. Indeed, this inevitable circumstance speaks to the importance of applying listening strategies to try to accommodate for limitations of automaticity in linguistic processing. However, as Graham et al., (2010) assert "... a minimum level of vocabulary recognition is required before nonlinguistic knowledge ... can be brought into play effectively" and that "... without accurate word recognition, applications of such knowledge are little more than guesses imposed on the text" (p. 14). A central objective of the current study is to investigate approaches to help learners develop better L2 word recognition from speech.

2.2 Approaches to Develop L2 Word Form Recognition from Speech

In its fullest sense, the construct of word recognition entails the capacity to both recognize the phonological form of a word and to map this form onto an appropriate meaning. In the current research, however, we have focused on the learner's capacity to recognize the phonological form of a word and map it against its corresponding written form, so-called *word form recognition from speech* (WFRS). A clear limitation of this construct is that it does not directly measure knowledge of word meaning. However, the practical advantage is that this construct facilitates convenient provision of automated computer-mediated feedback on learner performance. Further,

as learners often have a more complete knowledge of words in the written form, providing learners with systematic opportunities to map the phonological form onto the corresponding written form has pedagogical value (Field, 2008b; Hulstijn, 2003).

Although a number of researchers have suggested the potential of technology in improving WFRS (Hulstijn, 2003; Jia & Hew, 2021a), few have empirically investigated the efficacy of such approaches in language classrooms (Matthews & O'Toole, 2015; Matthews et al., 2015, 2017). Furthermore, none to our knowledge have done so specifically in out-of-class contexts by way of the affordances of mobile devices. To our knowledge, the only study that has addressed the computer-mediated development of L2 WFRS by way of a quasi-experimental design is Matthews et al. (2015), which investigated the effectiveness of a prototype online app used in an in-class context among 96 Chinese EFL tertiary level learners. Results indicated that learners in a treatment group who used the app across a five-week period had significantly greater improvements in L2 WFRS when compared to a control group that did not use the app. The app played short sections of simple speech to learners, thereby giving them repeated opportunities to transcribe the text and receive subsequent feedback on performance. These results provide preliminary empirical support for the general recommendations put forward by scholars concerning how to develop L2 WFRS (Field, 2008b; Hulstijn, 2003) and demonstrate the capacity of computers to facilitate these recommendations in authentic learning contexts. However, many questions remain, for example, little is known about the extent of mobile technology's usefulness in the development of L2 WFRS in out-of-class locations. Moreover, related research has only been undertaken in a few research contexts (Matthews & O'Toole, 2015); little is known about how generic suggestions for the development of L2 WFRS may be differentially effective in different language learning contexts.

2.3 *The Current Study*

The current research can be positioned within Benson's (2011) model of *language learning beyond the classroom* (i.e., location, formality, pedagogy, and locus of control) in the following way. In terms of the location, the use of the application was undertaken out-of-class. The portability and omnipresence of mobile devices is a key advantage in this regard. Learners can engage in learning at almost any time and anywhere. The learning associated with the app in both Azerbaijan and Japan was formal in the sense that it was linked with tertiary level courses, albeit undertaken in locations beyond the formal classroom itself. In terms of the dimension of pedagogy, the app was used by the learners in a self-instruction mode. As Benson (2011) describes, "in self-instruction specially designed ... [affordances] ... take on the role of the classroom instructor and there is a strong intention to learn on the part of the learner" (p. 11). In relation to locus of control, as the use of the app was initiated as part of formal learning in both Azerbaijan and Japan, its use can best be described as other-directed. However, as the app was used in out-of-class settings and learners needed to make decisions about their use of the app (e.g., when, where

and for how long), this can also be described as self-regulated learning (Lai et al., 2022).

This study seeks to help fill the ongoing gap in the literature by exploring the efficacy of a free mobile-assisted language learning app designed to improve the L2 WFRS of early-stage L2 learners. A key feature of the current research is the out-of-class implementation of the mobile app across two EFL contexts—Azerbaijan and Japan. This will not only enable us to critically interrogate the overall potential usefulness of the app but will also cast light on how learners interact with the app in different language learning environments. A key objective of the current study is to not only investigate the use of a mobile app to enhance word recognition from speech but to also use data from the app to draw insight about how to enhance in-class learning.

The following research questions will be addressed:

1. Is out-of-class usage of the app associated with significant improvements in WFRS and if so, does the magnitude of improvement vary between L1 groups?
2. What relationship is evident between the number of times learners listen to the app and improvements in WFRS?
3. From the 1,000 target words presented in the app, which are most challenging for learners to recognize and transcribe, and what could be learned about the origin of learner difficulty with these words through stimulated recall protocols?

3 Method

3.1 *Participants*

3.1.1 *Azerbaijani Participants*

The Azerbaijani treatment group ($n = 16$) and control group ($n = 16$) consisted of first-year students (17 to 18 years old) enrolled in a year-long English foundation program in which L1 instruction is minimal. Years and consistency of English education varied from one individual to another before entering university. Both groups were involved in this study via their respective course in listening and speaking. Foundation program students receive approximately 22 hours of English instruction per week before moving on to general education courses conducted in English from their second year (out of five). According to mean scores on a locally developed, university-led English proficiency exam taken prior to the study, all participants were within the Common European Framework of Reference for Languages (CEFR) A1 level (basic user, beginner).

3.1.2 Japanese Participants.

The Japanese treatment group ($n = 17$) consisted of second-year students (19 to 20 years old) enrolled in an English writing course conducted mainly in English. The control group ($n = 16$) also consisted of second-year students in a general English course. All Japanese participants had approximately three to six hours of English courses per week during this study. Most of these English courses were conducted predominantly in the learners' L1 by Japanese instructors. Prior to entering university, learners generally had received six years of English education. Scores from the Test of English for International Communication (TOEIC) for the treatment group ($M = 572.3$, $SD = 137.7$, $n = 15$) and the control group ($M = 537.7$, $SD = 101.9$, $n = 13$) indicated that their level of English proficiency was CEFR A2 (basic user, elementary) (Educational Testing Service, 2019).

3.1.3 Stimulated Recall Protocol Treatment Subgroup

From each of the L1 treatment groups, seven participants (14 in total) were selected to participate in stimulated recall protocols. The Japanese and Azerbaijani subgroup members were matched based on their pretest scores which varied by less than 10 points between the paired learners. Pairing subgroup members from Japan and Azerbaijan was intended to enable comparison between learners of similar levels of English proficiency in the two contexts.

3.2 The Mobile App

The *C-levels Vocab* app was designed by the first and second author to develop L2 WFRS by providing learners with multiple opportunities to listen to and transcribe high-frequency words. To be clear, the app is a free resource from which the designers of the app gain no financial benefit. The app presents the first 1,000 words from the Corpus of Contemporary American English (COCA) in blocks of 100 words (i.e., 10 *c-levels*). The words are presented in descending frequency of occurrence, thus generally progressing from relatively easy to relatively difficult. Figure 1 presents selected screenshots of the app's user interface. In Fig. 1A, an example of a contextual sentence is shown. At this stage the full spoken sentence is played and is heard through the learner's device, as in "I like him because he's good." All sentences were spoken by a North American English speaker.

After listening, the learner types the target word into the corresponding text box; there are four attempts to listen and do so. If not transcribed on the first attempt, the word is presented to the learner again after other words of the *c-level* have been engaged with. It is only when a word is transcribed correctly on the first attempt (of four) that the app's logic categorizes it as *known* and it is withdrawn from the cycling target word list of that *c-level*. Instant feedback makes clear to the learner that they

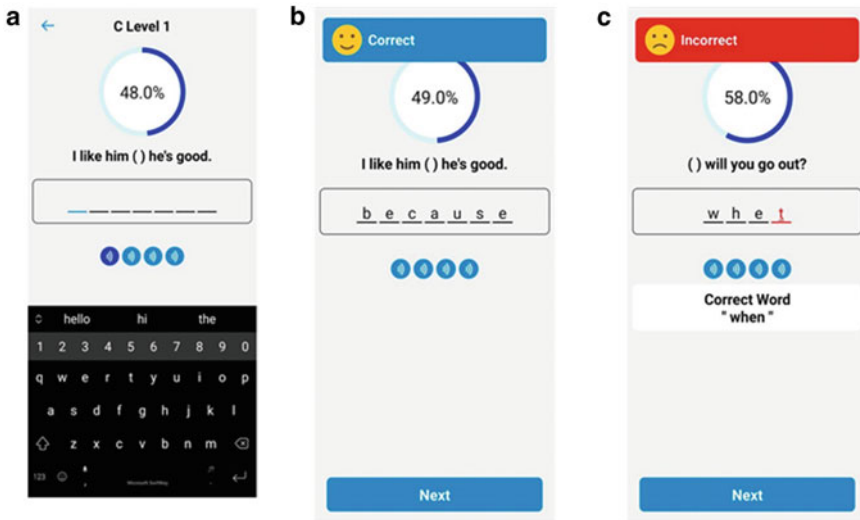


Fig. 1 Selected Screenshots of the App's User Interface

have correctly recognized and transcribed the word (Fig. 1B). It is only when all of the words of one c-level are *known* that the next c-level is made available to the learner.

Sentences and accompanying feedback were designed to provide as much contextual assistance to the learner as possible. The contextual sentences were piloted with native English language speakers to ensure the target words could be guessed without the benefit of the spoken sentence. This made sure the sentence itself afforded sufficient contextual support. Further, based on learner performance, incorrect letters were indicated in red and the correct target word was provided after four incorrect attempts (Fig. 1C). Learners were also provided with an overview of their progress through each c-level; this was shown as a percentage of words currently *known* for each c-level displayed on the screen. A range of data, such as the number of listens to each target word, was stored in the app's database for each learner.

3.3 Word Form Recognition from Speech (WFRS) Test

Before and after the intervention, the same 100-word pretest and posttest was administered to all participants from each context. Ten target words were randomly selected from each of the 10 c-levels of the app and these words were presented in both the pre- and posttest, albeit in a different order. As with the cloze format of the app, the test items consisted of a contextual sentence with the target word missing. Half of the target words were presented using contextual sentences which were identical to those used in the app and half had contextual sentences which were different from

the app. This was done to confirm that learning effects from target words presented in the same contextual sentences were negligible. During administration, the test audio was played once to the respective treatment group and learners used their smartphones to enter the missing target words into an online cloze template. Words were automatically scored as either correct or incorrect.

3.4 Stimulated Recall Protocol

A structured stimulated recall protocol (Gass & Mackey, 2017) was applied in each context with the 14 subgroup members (seven from each context). Prior to the protocol, the 20 target words which had been listened to most by each subgroup member while using the app (i.e., the most challenging words) were identified from stored app data. Each subgroup member's list of words and the corresponding cloze phrases were printed onto a reference paper for use during the stimulated recall protocol. The target word's audio (the same as that of the app) was played and the subgroup member was asked to transcribe the missing target word. Afterward, the participant was asked to articulate the missing target word, provide its L1 meaning, and translate the contextual sentence with the target word into their L1. The researchers evaluated the subgroup members on each task binomially (correct as 1 or incorrect as 0). Finally, the participant was asked to explain why the target word had been especially difficult to recognize while using the app. This was done with the aim to identify and categorize the primary source of difficulty in transcribing each of the 20 target words. All of the subgroup member interviews were conducted in the participants' respective L1. The interviews lasted for approximately one hour and were audio recorded with the informed consent of each subgroup member.

Sources of WFRS error for each word were categorized based on previous research in the field (Lange & Matthews, 2020, 2021). The error categories are listed as follows:

1. Semantically unknown: The listener could not provide a L1 definition for the target word even after seeing its orthographic form.
2. Semantically known but phonologically unfamiliar: The meaning of the target word is known in the L1 but insufficient knowledge of its phonological form was primarily responsible for difficulty transcribing the target word.
3. Semantically known but phonologically unfamiliar due to the influence of connected speech: The meaning of the target word is known but attributes of connected speech, such as coarticulation, were primarily responsible for difficulty transcribing the target word.
4. Semantically known but spelled incorrectly: The meaning of the target word is known but spelling the target word incorrectly was primarily responsible for multiple failed transcription attempts while using the app.
5. Other reasons.

3.5 Procedure

Data for this study was collected independently by researchers in Azerbaijan and Japan. First, all of the participants in each context took the pretest via mobile devices. Next, the treatment group members downloaded the app and began to use it for WFRS development outside of class as homework over a period of approximately six weeks with the loose goal of completing one to two c-levels each week. After an initial practice session in class, the participants were asked to use the app outside of class at their own pace. This inevitably resulted in individual variation in how quickly participants completed the app tasks. All of the treatment group members completed the 1,000 words assessed with the app before taking the posttest. No control group participants used the app and they took the posttest during the same week as the treatment group. Finally, the subgroup members individually undertook the stimulated recall protocol with the respective researchers in each of the two contexts.

3.6 Data Analysis

Quantitative analyses included comparison of mean WFRS difference scores between groups (research question 1). Mean difference scores were determined by subtracting pretest scores from posttest scores. To test our hypothesis that those who used the mobile app (treatment groups) achieved greater mean WFRS difference scores than those that did not (control groups), independent samples *t*-tests were performed. Pearson correlation was also used to examine links between learner engagement with the app (e.g., the total number of times the participants listened to the words) and mean difference scores (research question 2).

The quantitative data collected through stimulated recall protocol were assessments of participants' ability to articulate the target word, provide an L1 definition for it, and translate the contextual sentence. Percentages of correct answers for each task were calculated to identify the most challenging target words and investigate factors which may explain suboptimal WFRS (research question 3). The qualitative data collected during stimulated recall protocols were audio recorded. Participant L1 responses to the question of why the target word had been difficult to transcribe were analyzed for explanatory themes. The themes were identified via a thorough examination of the responses from participants based on aspects of thematic analysis (Braun & Clarke, 2006). Researchers noted recurring explanations of comprehension difficulty such as the latter part of the target word being difficult to hear and influence from the L1 affecting comprehension. The majority of these responses were aligned with the error categories presented to participants during the protocol. These are explored in more detail in the discussion section.

4 Results

Research Question One —Is out-of-class usage of the app associated with significant improvements in WFRS and if so, does the magnitude of improvement vary between L1 groups?

Measures of internal consistency for both the pretest (Cronbach's $\alpha = 0.96$) and the posttest (Cronbach's $\alpha = 0.94$) were very good. The distribution of mean difference scores was sufficiently normal for these analyses (i.e., skewness & kurtosis each < 2).

For the Azerbaijani group, the assumption of homogeneity of variance was satisfied with Levene's test, $F(30) = 0.002, p = 0.966$. Independent samples t -test results showed a significant effect, $t(30) = 3.17, p = 0.004$ with Azerbaijani treatment group members achieving significantly greater WFRS difference scores ($M = 24.56, SD = 9.70$) than those in the Azerbaijani control group ($M = 13.63, SD = 9.82$) (Table 1). A Cohen's d of 1.12 suggested a large effect size (Plonsky & Oswald, 2014).

For the Japanese group, Levene's test results, $F(31) = 0.273, p = 0.605$, again verified the assumption of homogeneity of variance. Independent samples t -test results showed a significant effect, $t(31) = 3.17, p = 0.002$. As with the treatment group members from Azerbaijan, Japanese treatment group members achieved significantly greater WFRS difference scores ($M = 9.47, SD = 7.25$) than those in the Japanese control group ($M = 1.31, SD = 6.20$). As before, a large effect size was indicated (Cohen's $d = 1.12$).

The mean WFRS difference scores for the Azerbaijani treatment group ($M = 24.56$) were greater than those of the Japanese treatment group ($M = 9.47$). An independent samples t -test result showed that this difference reached a level of statistical significance ($t(31) = 1.731, p = 0.000$), with a large effect size (Cohen's $d = 1.8$). Of note is that the mean pretest score for the Azerbaijani treatment group ($M = 56.13, SD = 17.90$) was lower than that of the Japanese treatment group ($M = 77.17, SD = 10.25$). An independent samples t -test demonstrated that the difference in mean

Table 1 A comparison of test scores for treatment and control group members

	<i>n</i>	Mean pretest score	<i>SD</i>	Mean posttest score	<i>SD</i>	<i>MD</i>	<i>SD</i>
Azerbaijani treatment group	16	56.12	17.90	80.69	11.98	24.56	9.70
Azerbaijani control group	16	53.75	19.47	67.38	14.01	13.63	9.82
Japanese treatment group	17	77.18	10.25	86.65	7.11	9.47	7.25
Japanese control group	16	72.88	7.72	74.19	10.82	1.31	6.20

Note *MD* = mean difference between the pre- and posttest scores

Table 2 Descriptive statistics and correlations for total number of listens, difference score, and pretest score

Variable	<i>M</i>	<i>SD</i>	1	2	3
1. Total number of listens	1915.97	626.17	-		
2. Difference score	16.79	11.36	0.66**	-	
3. Pretest score	66.97	17.80	-0.87**	-0.85**	-

Note $N = 33$, ** $p < 0.001$

pretest scores between the treatment groups was statistically significant ($t(31) = -4.179$, $p = 0.000$).

Research Question Two —What relationship is evident between the number of times learners listen to the app and improvements in WFRS?

Stored data from the app provided a raw score of the number of times learners listened to each target sentence. As a learner's existing word recognition capabilities are likely to influence the patterns of interaction with computer-mediated language learning interventions (Matthews et al., 2017), another variable of interest in this analysis was pretest score. Table 2 shows the Pearson correlation matrix of all treatment group learners ($n = 33$).

A significant positive correlation was observed between total number of times listened and difference score (Table 2). The magnitude of this relationship was large (e.g., $r > 0.6$). The general trend evident is that more repeated listening to the app was strongly associated with greater WFRS improvement. There was a strong, negative correlation between pretest scores and number of times listened ($r = -0.872$, $p < 0.001$) and pretest scores and difference scores ($r = -0.851$, $p < 0.001$). The trend evident here is that learners with a lower pretest score listened more and were associated with greater difference scores than those with a higher pretest score.

Table 3 presents a breakdown of the number of times learners from each language group listened to the app across the duration of the intervention. Figure 2 visualizes the relationship between engagement with the app and WFRS difference scores.

Table 3 Descriptive statistics on mean number of times listened by L1 Group

	<i>n</i>	<i>Min</i>	<i>Max</i>	Mean number of times listened	<i>SD</i>
Azerbaijani treatment group	16	1,257	3,799	2,209	668.69
Japanese treatment group	17	399	2,160	1,640	445.31

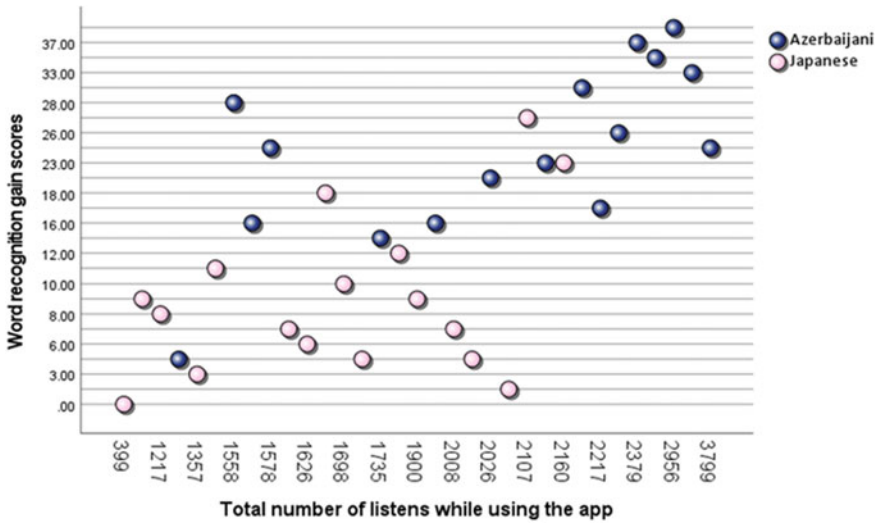


Fig. 2 Scatter Plot of WFRS Difference Score by Total Number of Listens

Research Question Three —From the 1,000 target words presented in the app, which are most challenging for learners to recognize and transcribe, and what could be learned about the origin of learner difficulty with these words through stimulated recall protocols?

Back-end data were used to rank each participant's target words according to the number of times listened when using the app. From this list, the 20 target words which were listened to most frequently were selected as a unit of analysis and are referred to here as the 20 most challenging target words. The mean number of times these words were listened to for the Azerbaijani cohort ($n = 16$) was 10.15 and 8.24 for the Japanese cohort ($n = 17$). Next, instances of the same target words in the 20 most challenging target words for each cohort were tallied. The target words which were shared five or more times within each cohort are listed in Table 4. For example, the first target word in the Azerbaijani cohort, *unidentified*, is the 274th most frequent word in the spoken section of the COCA and is shared on 10 of the 16 learners' lists of their 20 most challenging target words (indicated in the *Shared* column). Also, *unidentified*, *attorney*, and *correspondent* are followed by an asterisk indicating these three words were shared five or more times in both cohorts.

Table 5 presents the selection percentages for each of the five error categories for the Azerbaijani and Japanese subgroup members. These categories were selected by the participant and researcher during the stimulated recall protocol for each member's most challenging target words.

Commonalities within the subgroups as per their most challenging words were investigated by identifying the target words which repeat most often within each cohort member's list of most challenging words. The scope of investigation was

Table 4 Target words which were shared 5 or more times in lists of learners' 20 most challenging words

Japanese Cohort (<i>n</i> = 17)			Azerbaijani Cohort (<i>n</i> = 16)		
Frequency Ranking	Shared	Target word	Frequency ranking	Shared	Target word
475	9	attorney*	274	10	unidentified*
310	9	certainly	512	7	committee
264	9	senator	114	6	any
410	7	governor	256	6	court
855	7	career	425	6	sort
822	7	Iraqi	475	6	attorney*
261	6	democrat	594	6	correspondent*
274	5	unidentified*	263	5	political
594	5	correspondent*	81	5	him
73	5	our	236	5	whether
68	5	their	23	5	(<i>n</i> 't
357	5	administration	433	5	threat
513	5	heart	705	5	able
829	5	soldier			

Note. *Frequency Ranking* refers to the frequency ranking for the target word in the spoken section of the COCA. *Shared* refers to the number of times the target word is shared in the lists of the 20 most difficult target words for each participant. Words followed by an asterisk (*) appear in both L1 cohorts' lists. (*n*'t represents the contracted form of *not*)

Table 5 Percentage of transcription error category selections

Transcription error category	Azerbaijani subgroup (<i>n</i> = 7)	Japanese subgroup (<i>n</i> = 7)
(1) Semantically Unknown	14%	26%
(2) Semantically known but phonologically unfamiliar	29%	31%
(3) Semantically known but phonologically unfamiliar due to the influence of connected speech	13%	19%
(4) Semantically known but spelled incorrectly	8%	21%
(5) Other (Could not explain why the item was difficult)	36%	4%

reduced from 20 target words to 10 target words, thus allowing a focus on only the most challenging items. The target words which repeated three or more times within the subgroups' 10 most challenging target words are listed in Table 5. To elucidate why these particular words were the most challenging, the frequency with which error categories 1 to 5 were selected are indicated in the far-right column of

the table. For example, in Table 6 the first target word *attorney* is the 475th most frequently occurring word in the spoken section of the COCA. It appears on five of the seven subgroup members' lists of their 10 most challenging target words and was categorized as error category 1 by three learners and as category 2 by two learners. Table 7 presents data for the Japanese subgroup members in the same format.

Table 6 Difficult target words shared among the Azerbaijani subgroup

Freq	Shared	Target word	Target sentence	Categories				
				1	2	3	4	5
475	5	attorney*	If you have broken the law, you need to get an ()	3	2	0	0	0
274	5	unidentified*	We do not know all of their names, so some () people have been killed	1	1	1	2	0
512	5	committee	Maybe I can help. I am on the () that runs this event	2	1	0	1	1
594	3	correspondent*	The foreign () will send news back to America	0	1	0	1	1
397	3	candidate	He is the democratic () for President	2	1	0	0	0

Note. $n = 7$, *Freq.* = frequency ranking for the target word. *Shared* = times shared in lists of 10 most difficult target words. (*) shows words that appeared in both L1 cohorts' lists. *Categories* are defined as: (1) Semantically unknown, (2) Semantically known but phonologically unfamiliar, (3) Semantically known but phonologically unfamiliar due to the influence of connected speech, (4) Semantically known but spelled incorrectly, (5) Other

Table 7 Difficult target words shared among the Japanese subgroup

Freq	Shared	Target word	Target sentence	Categories				
				1	2	3	4	5
310	6	certainly	If there is an answer I () don't know it	1	4	1	0	0
822	4	Iraqi	The war started over American interest in () oil. They wanted more control of the oil from that foreign country	4	0	0	0	0
248	3	police	Stop now or I'll call the ()!	0	3	0	0	0
425	3	sort	You can't do that () of thing here	2	0	0	1	0
594	3	correspondent*	The foreign () will send news back to America	3	0	0	0	0
507	3	commercial	It made a lot of money so the movie did well in a () sense	1	0	0	2	0

Note $n = 7$, *Freq.* = frequency ranking for the target word. *Shared* = times shared in lists of 10 most difficult target words. (*) shows words that appeared in both L1 cohorts' lists. (See note in Table 5 for details of the *Categories*)

5 Discussion

The findings from research question one show that the treatment groups in both L1 cohorts made significant improvements in WFRS. The Azerbaijani treatment group gained an average of approximately 10.93 words more than the Azerbaijani control group with a large effect size (i.e., $d = 1.12$). The Japanese treatment group gained an average of 8.16 words more than the Japanese control group ($d = 1.12$). This provides evidence that WFRS was enhanced through use of the app for both L1 groups, however, the Azerbaijani learners' gains were larger. This difference may be attributable in part to the initially lower proficiency levels of the Azerbaijani learners since instruction in aural word recognition tends to produce better results for lower proficiency learners (Jia & Hew, 2021b). Another contributing factor may be the Azerbaijani group's greater use of the app with a mean difference of 569 more listening times than the Japanese cohort. Moreover, their gains may have been larger from having up to 19 more hours of English instruction per week than their Japanese counterparts.

This overall positive result for both groups provides preliminary evidence that the affordances of the app may be beneficial in enhancing WFRS among learners of diverse language backgrounds and proficiency levels. Due to the various possible contributing factors, however, more empirical research is required to determine the sole contribution of the app on WFRS development. In terms of a comparison with the one previous study that we are aware of that was specifically directed towards computer-mediated development of L2 WFRS, results from the current research appear to be positive. The Matthews et al. (2015) in-class study noted an improvement in WFRS between treatment and control group members, but the magnitude of this improvement (approximately 1.5 words difference between control and treatment group) was less than that noted in the current study and had a smaller effect size (i.e., $d = 0.47$). Although more research is needed before more assertive conclusions can be made, this comparison is at least suggestive of the feasibility of developing WFRS capacities in out-of-class contexts via the use of mobile apps.

Research question two explored the relationship between the number of times learners listened to the app and WFRS difference scores. There was a strong significant correlation between the number of times learners listened and their difference scores ($r = 0.66$). Pretest scores had a strong inverse relationship with both number of listens and WFRS score gains, which shows that learners with lower proficiency listened more and improved more than higher proficiency learners. This result is not entirely unexpected based on previous research. For example, Matthews et al. (2017) demonstrated that learners of different proficiency levels experienced significantly different WFRS gains after computer-mediated intervention, with mid-level proficiency learners achieving greater difference scores than both lower and higher proficiency learners. In sum, these findings reinforce the importance of computer-mediated approaches that are adaptive. A future area for improvement with the C-levels app (and others like it) would be the addition of an algorithm that modulated the difficulty of the target words depending on the learner's preceding performance

with the task. For example, if the learner was recognizing all words correctly, the app could increase the difficulty (decrease the frequency) of the target words (e.g., skip a c-level). Also, the Azerbaijani treatment group listened to the app on average 569 times more than the Japanese treatment group. This increased level of engagement may have been influenced by several factors, including proficiency and affective factors. For example, the Azerbaijani groups' generally lower proficiency may have meant the app content was more immediately relevant to their learning needs and this may have motivated higher levels of engagement.

Suggestions for developing WFRS based on our findings are provided next. Because participants who used the app more frequently made greater improvements, it is recommended that learners focus on developing WFRS through frequent listening and cloze transcription practice. In addition to practice via the app (or similar affordances), explicit instruction focusing on patterns of phonological modification in connected speech, word stress patterns, and utilization of contextual meaning to support WFRS development is also recommended. As the analysis of participants' most challenging target words demonstrated, individual learners have unique difficulties with WFRS due to a variety of factors. Screening for difficult words via a tool such as the C-Levels Vocab app allows researchers and learners to focus on addressing the unique challenges presented by each word for each learner. Developing WFRS for words that are especially difficult may require explicit instruction tailored to the individual learner and their L1 group rather than simply more generalized practice.

The first part of research question three investigated the most challenging target words for learners to recognize and transcribe when using the app. In both groups, there were at least 13 target words which five or more learners found particularly difficult. Many of these were high-frequency words, which is consistent with previous research demonstrating difficulty with WFRS for known words (Carney, 2021; Lange & Matthews, 2021). For the Japanese participants, seven of the 14 words in Table 3 (*certainly*, *senator*, *democrat*, *unidentified*, *our*, *their*, *administration*) had frequency rankings which were higher than the 400th ranked word in the spoken section of the COCA. Among the Azerbaijani participants' lists as well, seven of the 13 words in Table 3 (*unidentified*, *any*, *court*, *sort*, *attorney*, *political*, *him*, *whether*, *(n't)*) rank higher in frequency than the 400th ranked word. To illustrate, *their* (ranked 68th in frequency) was shared on the lists of five Japanese learners' most challenging words and was a noted challenge for Azerbaijani learners also. It is likely that this difficulty is at least partially attributable to Japanese and Azerbaijani not having a voiced dental fricative /ð/. The finding that very high-frequency words can cause difficulty for language learners is of strong interest. Teachers and researchers should be aware that the phonological form of some high-frequency words may not be well known by some learners and that these problematic high-frequency words may vary depending on the learner's L1. It is important for teachers to identify these problematic high-frequency words and offer them explicit attention in an effort to raise learners' awareness of the potential challenges of recognizing their phonological forms. For example, providing learners with repeated opportunities to hear authentic samples of connected speech containing these problematic words and then again

while listening and reading an accurate corresponding written transcript is an important first step. The provision of metalinguistic explanations of how the phonological form of these challenging high-frequency words may vary when articulated in fluent speech (e.g., variable acoustic contexts) is also warranted.

It is also interesting to note that scores for the Azerbaijani subgroup members on the target word articulation and transcription tasks were over 10 average percentage points greater than those of the Japanese learners. Although speculative, this slight advantage for the Azerbaijani learners may be attributable to a stronger similarity between the orthographic systems of Azerbaijani and English, when compared to that of Japanese and English. As an example, consider the words *Azerbaijan* (i.e., *Azərbaycan*) and *Japan* (i.e., 日本) written in Azerbaijani and Japanese respectively.

The second part of research question three addressed why learners in both contexts had experienced difficulty in WFRS for certain target words. Five error categories were used to clarify the primary reasons for transcription difficulty. The largest error category in the Japanese subgroup was (2) *Semantically known but phonologically unfamiliar* which, as confirmed via the stimulated recall protocols, represents a pervasive limitation on WFRS. Learners often recounted incongruences between their mental representation of a word's phonological form and the phonological form they perceived from the audio recording. Data suggested that these differences stem from extensive prior exposure to Japanese-accented English that seemed to have created Japanese-accented phonological forms of English words in the mental lexicon. For example, one subgroup member explained their inability to recognize *hand* (/hænd/) in the spoken utterance "It's more work than I thought. Could you give me a () with this?" was due to an inaccurate phonological representation of the target word. Japanese requires mora to have a vowel in the syllable-final position so *hand* was modified to /hando/ in the learner's lexicon to conform to Japanese phonotactics. Another learner had difficulty recognizing *school* in the spoken utterance "Our child said she didn't want to go to () today." The learner explained that "The end of the word is hard to hear" due to the consonant in the syllable-final position and further stated that "[they] have a habit of hearing in *katakana*" (i.e., the Japanese phonetic syllabary). By this, we assume the learner was describing the process of mapping English words onto Japanese-accented representations in the mental lexicon. Thus, when *hand* is recognized it is associated with /hando/ and if *school* is recognized, it is mapped to /sukuuru/ in the mind of the listener. In another example, the learner explained that her difficulty perceiving *police* in "Stop now or I'll call the ()!" was due to influence from Japanese. She repeatedly entered *please* for *police* despite having semantic knowledge of the target word. Every subgroup member described at least one similar difficulty related to influence from Japanese-accented English and previous studies have also documented this trend (Lange & Matthews, 2020, 2021). Influence from Japanese-accented English input can be more generally understood as L1 phonotactic constraints whereby the listener tries to apply phonological conventions of the L1 (e.g., placing vowels in the syllable-final position) erroneously to the L2 (Cutler, 2012). The largest error category for the Azerbaijani subgroup was (5) *Other* which mainly reflected the learners' difficulty in explaining why WFRS had been difficult. Speculatively, this may be attributable to learners in this context

having limited experience with critical analysis of their own English language performance. This also underscores the general difficulty in self-assessing and describing limitations in one's implicit L2 linguistic knowledge, especially among learners with relatively low proficiency levels. This speaks to the particular importance of teachers working closely with lower proficiency learners to help them identify and resolve the specific challenges they have in recognizing the phonological form of words.

Although informative, this study had several limitations. First, WFRS development for the Azerbaijani learners was likely to be disproportionately affected by the 20 or more hours of English instruction they received each week during the study compared to the Japanese learners' three to six hours. Another limitation was the inclusion of culturally-bound words within the COCA (e.g., *senator*, *attorney*, *Iraqi*) which may have affected WFRS improvement differently in the two contexts. The use of knowledge-based vocabulary lists specific to L1 language groups (when they become available) to guide the order and selection of the content of similar interventions is advised for future studies (see, Schmitt et al., 2022). An additional limitation was that due to feasibility constraints only one rater in each context scored the stimulated recall protocol assessments and analyzed the qualitative data from the learners' L1 responses. Scores from multiple raters are recommended to accurately assess inter-rater reliability.

6 Conclusion

Recognizing words from speech is a fundamental skill in L2 learning and is particularly important for listening development. Unlike information processed through the orthographic modality, WFRS must be executed fluently to keep up with the transient nature of spoken input. Regular engagement with out-of-class mobile-assisted language learning like that at the center of this study is likely to help develop this capacity across different language learning contexts. Although all learners regardless of L1 shared some of the same challenging words in the current study, an important takeaway was that there was sizable variation in the specific words that learners found challenging. Another key finding was that the degree of difficulty encountered with any given word did not necessarily correspond to the frequency of occurrence of that word (i.e., as indicated by the COCA). Therefore, an individualized approach to WFRS development is suggested in which each learner's *problem* words are identified and the underlying factors responsible for suboptimal WFRS are addressed. The approach applied in the current study has enabled us to cover new ground, but there is more work to be done. Moving forward, we call for more mixed-methods research that triangulates out-of-class language learning app usage data with one-on-one interview data, both within and across learning contexts. Such research will inform us on how to strategically apply technological affordances to facilitate effective out-of-class learning.

References

- Benson, P. (2011). Language learning and teaching beyond the classroom: An introduction to the field. In P. Benson & H. Reinders (Eds.), *Beyond the language classroom* (pp. 7–16). Palgrave Macmillan.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101.
- Carney, N. (2021). Diagnosing L2 listeners' difficulty comprehending known lexis. *TESOL Quarterly*, 55(2), 536–567. <https://doi.org/10.1002/tesq.3000>
- Cheng, J., & Matthews, J. (2018). The relationship between three measures of L2 vocabulary knowledge and L2 listening and reading. *Language Testing*, 35(1), 3–25. <https://doi.org/10.1177/0265532216676851>
- Cheng, J., Matthews, J., Lange, K., & McLean, S. (2022). Aural single-word and aural phrasal verb knowledge and their relationships to L2 listening comprehension. *TESOL Quarterly*. <https://doi.org/10.1002/tesq.3137>
- Cutler, A. (2012). *Native listening: Language experience and the recognition of spoken words*. MIT Press.
- Educational Testing Service (2019). *Mapping the TOEIC test on the CEFR*. <https://www.ets.org/s/toeic/pdf/toeic-cefr-flyer.pdf>
- Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, 24(2), 143–188. <https://doi.org/10.1017/S0272263102002024>
- Field, J. (2008a). Revising segmentation hypotheses in first and second language listening. *System*, 36(1), 35–51. <https://doi.org/10.1016/j.system.2007.10.003>
- Field, J. (2008b). *Listening in the language classroom*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511575945>
- Flowerdew, J., & Miller, L. (2005). *Second language listening: Theory and practice*. Cambridge University Press.
- Gass, S. M., & Mackey, A. (2017). *Stimulated recall methodology in applied linguistics and L2 research*. Routledge. <https://doi.org/10.4324/9781315813349>
- Goh, C. C. (2000). A cognitive perspective on language learners' listening comprehension problems. *System*, 28(1), 55–75. [https://doi.org/10.1016/S0346-251X\(99\)00060-3](https://doi.org/10.1016/S0346-251X(99)00060-3)
- Graham, S., Santos, D., & Vanderplank, R. (2010). Strategy clusters and sources of knowledge in French L2 listening comprehension. *Innovation in Language Learning and Teaching*, 4(1), 1–20. <https://doi.org/10.1080/17501220802385866>
- Hulstijn, J. (2002). Towards a unified account of the representation, processing and acquisition of second language knowledge. *Second Language Research*, 18(3), 193–223. <https://doi.org/10.1191/0267658302sr2070a>
- 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>
- Jia, C., & Hew, K. F. (2021a). Meeting the challenges of decoding training in English as a foreign/second language listening education: Current status and opportunities for technology-assisted decoding training. *Computer assisted language learning*, 1–30. <https://doi.org/10.1080/09588221.2021.1974051>
- Jia, C., & Hew, K. F. (2021b). Toward a set of design principles for decoding training: A systematic review of studies of English as a foreign/second language listening education. *Educational Research Review*, 33, 100392. <https://doi.org/10.1016/j.edurev.2021.100392>
- Lai, Y., Saab, N., & Admiraal, W. (2022). University students' use of mobile technology in self-directed language learning: Using the integrative model of behavior prediction. *Computers & Education*, 179, 104413. <https://doi.org/10.1016/j.compedu.2021.104413>

- Lange, K., & Matthews, J. (2020). 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/ssl.2020.10.4.4>
- Lange, K., & Matthews, J. (2021). Analyzing trends in the aural decoding errors of Japanese EFL learners. *International Review of Applied Linguistics in Language Teaching*. <https://doi.org/10.1515/iral-2020-0050>
- Masrai, A. (2020). Exploring the impact of individual differences in aural vocabulary knowledge, written vocabulary knowledge and working memory capacity on explaining L2 learners' listening comprehension. *Applied Linguistics Review*, 11(3), 423–447. <https://doi.org/10.1515/applirev-2018-0106>
- 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., Cheng, J., & O'Toole, J. M. (2015). Computer-mediated input, output and feedback in the development of L2 word recognition from speech. *ReCALL*, 27(3), 321–339. <http://doi:https://doi.org/10.1017/S0958344014000421>
- Matthews, J., Masrai, A., Lange, K., McLean, S., Alghamdi, E. A., Kim, Y. A., Shinhara, Y., Tada, S. (2023). Exploring Links Between Aural Lexical Knowledge and L2 Listening in Arabic and Japanese Speakers: A Close Replication of Cheng, Matthews, Lange and McLean (2022). *TESOL Quarterly*, <https://doi.org/10.1002/tesq.3212>
- Matthews, J., & O'Toole, J. M. (2015). Investigating an innovative computer application to improve L2 word recognition from speech. *Computer Assisted Language Learning*, 28(4), 364–382. <https://doi.org/10.1080/09588221.2013.864315>
- Matthews, J., O'Toole, J. M., & Chen, S. (2017). The impact of word recognition from speech (WRS) proficiency level on interaction, task success and word learning: Design implications for CALL to develop L2 WRS. *Computer Assisted Language Learning*, 30(1–2), 22–43. <https://doi.org/10.1080/09588221.2015.1129348>
- Plonsky, L., & Oswald, F. L. (2014). How big is “big”? Interpreting effect sizes in L2 research. *Language Learning*, 64(4), 878–912. <https://doi.org/10.1111/lang.12079>
- Schmitt, N., Dunn, K., O'Sullivan, B., Anthony, & L., Kremmel, B. (2022). *Knowledge-based vocabulary lists user manual: A supplementary resource*. British Council. https://www.britishecouncil.org/sites/default/files/kvl_user_manual_final.pdf
- Vafaei, P., & Suzuki, Y. (2019). The relative significance of syntactic knowledge and vocabulary knowledge in second language listening ability. *Studies in Second Language Acquisition*, 42(2), 383–410. <https://doi.org/10.1017/S0272263119000676>
- Wallace, M. P. (2022). Individual differences in second language listening: Examining the role of knowledge, metacognitive awareness, memory, and attention. *Language Learning*, 72, 5–44. <https://doi.org/10.1111/lang.12424>
- 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>

Joshua Matthews is a Senior Lecturer and Chair of the School Education Committee at the School of Education, University of New England, Australia. His major research interests include computer-assisted language learning, L2 listening, L2 vocabulary, L2 teaching, and language testing. His previous publications have appeared in journals including *Computer-Assisted Language Learning*, *ReCALL*, *Language Learning and Technology*, *TESOL Quarterly* and *Language Testing*. His current research projects include those addressing L2 aural vocabulary knowledge, the role of noticing in L2 development, multi-modality in language testing, and the design of MALL.

Kriss Lange is an Associate Professor at the University of Shimane Matsue Campus, Japan. His research interests include L2 listening, vocabulary acquisition and mobile-assisted language learning. Based in Japan, his research has been mainly focused on understanding English listening difficulties for Japanese university students and on L2 listening education. His previous research has been published in journals such as *TESOL Quarterly* and *International Review of Applied Linguistics in Language Teaching*.

Gunther M. Wiest has taught English in Japan, Tunisia, Azerbaijan, and his native USA. The foundation level courses in listening, speaking, and academic vocabulary that he facilitates as Senior Instructor at English for Academic and Professional Purposes (EAPP), ADA University, inspire him to research contemporary methods of language acquisition. These include student telecollaboration or COIL with faraway peers, and as in the current volume, self-study with educational technology.