EMPIRICAL STUDY

Frequency of Exposure Influences Accentedness and Comprehensibility in Learners' Pronunciation of Second Language Words

Takumi Uchihara ,[a] Stuart Webb ,[b] Kazuya Saito ,[c] and Pavel Trofimovich [d]

[a]Waseda University, [b]University of Western Ontario, [c]University College London, and [d]Concordia University

Abstract: The current study investigated the effects of repetition on the learning of second language (L2) spoken word forms. Japanese university students learning L2 English were randomly assigned to one of three treatment conditions (one, three, and six exposures) and learned 40 words while hearing them and viewing their corresponding pictures. A picture-naming test was administered before, immediately after, and approximately one week after the treatment. The elicited speech samples were evaluated for two aspects of spoken vocabulary knowledge: pronunciation (accentedness and comprehensibility) and form–meaning connection (spoken form recall). Results showed that (a) the number of exposures positively affected measures of form–meaning

connection and pronunciation immediately after the treatment, and (b) cognateness moderated how strongly repetition impacted the pronunciation of L2 words. Moderate learning gains occurred for comprehensibility after six exposures to new words. However, with six exposures, only small effects of repetition were observed for accentedness.

Introduction

Frequency of exposure is a key determinant of first language (L1) and second language (L2) acquisition and processing (Ellis, 2002). One of the most extensively researched areas exploring frequency effects is incidental vocabulary acquisition (Uchihara, Webb, & Yanagisawa, 2019). Earlier studies suggested varying numbers of exposures to be necessary for significant vocabulary learning to occur, spanning six (Rott, 1999), eight (Horst, Cobb, & Meara, 1998), 10 (Webb, 2007), and more than 20 exposures (Waring & Takaki, 2003). This line of research has advanced our understanding of frequency effects in vocabulary acquisition by measuring not only knowledge of form–meaning connection but also other aspects of word knowledge, including knowledge of collocation (Webb, Newton, & Chang, 2013), grammar (van Zeeland & Schmitt, 2013), association (Horst et al., 1998), and spelling (Webb, 2007). However, findings are predominantly based on testing word knowledge in written form, and knowledge of pronunciation remains underexplored. The lack of attention to pronunciation in vocabulary research is surprising in view of the prominence placed on it as one of the fundamental aspects of word knowledge (Nation, 2013) and speaking proficiency (de Jong et al., 2012).

Building on psycholinguistic models of the bilingual lexicon (Bordag, Gor, & Opitz, 2021; Jiang, 2000) and frameworks of sound and word learning (K. Saito, 2018; Werker, 2018), the current study conceptualizes L2 word pronunciation learning as an advanced stage of L2 spoken vocabulary acquisition subsequent to the development of form–meaning connection. Learners are first assumed to encode novel forms of words and map acquired L2 forms to existing L1 meanings (Jiang, 2000). At this stage of learning (form–meaning mapping), articulated forms of words cued by meanings or messages may be partially specified but not sufficiently accurate for the spoken words to be fully and easily understood by the listener. In real-life situations, for example, it is not rare to find learners who can pronounce all the phonemes of a word in a correct sequence yet whose word pronunciations are heavily accented and difficult

to understand. At the subsequent stage of learning (refinement of phonological form), increased exposure to spoken input is expected to further strengthen the form–meaning connection for a word (Jiang, 2000) and to facilitate the development of its phonological representation (Flege, 1995). As they develop fully specified representations for L2 words, learners are expected to use their spoken vocabulary knowledge not only in semantically appropriate but also in phonologically intelligible and comprehensible ways.

To provide a nuanced understanding of how L2 learners acquire spoken vocabulary knowledge, it is important to distinguish the processes of establishing form–meaning connections for novel words (i.e., mapping L2 forms to L1 meanings) from those of further developing knowledge of spoken forms (i.e., refining the phonological forms of words whose form–meaning connections are already established). On the basis of this two-step model of spoken vocabulary acquisition, the current study aimed to explore the effects of repeated exposure to spoken forms of words on form–meaning mapping and word pronunciation learning (phonological refinement). This research sheds further light on our understanding of frequency effects in L2 acquisition and provides important implications for L2 vocabulary and pronunciation learning.

Background Literature

Defining Pronunciation Knowledge: Accentedness and Comprehensibility

Since Munro and Derwing's (1995a) seminal study, several global constructs, including accentedness and comprehensibility, have been widely researched in L2 pronunciation studies (Huensch & Nagle, 2021; Munro & Derwing, 2020; K. Saito, 2021). Accentedness (or linguistic nativelikeness) is defined as listeners' judgments of how different L2 speech sounds from the expected language variety. Comprehensibility refers to listeners' perceived ease or difficulty in understanding L2 speech. These two constructs are measured through listeners' ratings of speakers using numerical point scales (e.g., 1 = *no accent*, 9 = *heavily accented*; 1 = *easy to understand*, 9 = *hard to understand*). Comprehensibility is often distinguished from intelligibility, which captures listeners' actual understanding of L2 speech (Derwing & Munro, 2009), measured through a variety of methods including listener transcription of words or utterances, responses to true/false statements, and perception of nonsense sentences (Kang, Thomson, & Moran, 2018). However, conceptualized broadly, comprehensibility is an intuitive and easy-to-use measure used frequently as an alternative metric of listener understanding of words and utterances (Martin, 2020).

Accentedness and comprehensibility are partially independent (see K. Saito, 2021, for a review). For example, L2 speakers with a stronger foreign

accent are not necessarily less comprehensible or intelligible (Munro & Derwing, 1995a). Similarly, when listeners rate L2 utterances for comprehensibility and accentedness, the processing cost indicated by response latency significantly predicts raters' comprehensibility but not accentedness judgments (Munro & Derwing, 1995b), implying that the two constructs can be distinguished through a reaction-time measure. The smaller amount of effort to decode the intended message (i.e., meanings), indexed by faster response, might thus be closely associated with higher comprehensibility rather than decreased accentedness. According to cross-sectional and longitudinal investigations (Derwing & Munro, 2013; K. Saito, 2015), L2 learners appear to continue to improve various dimensions of language competence relevant to comprehensibility (temporal, lexical, grammatical, and prosodic features) when they use the target language daily. In contrast, although the degree of foreign accent tends to diminish within early phases of L2 immersion (Derwing & Munro, 2013), this is likely to be followed by a plateau, and further development may be limited to learners with higher phonetic aptitude, memory, and motivation (Suzukida, 2021). In light of prior work that considers accentedness and comprehensibility as separate constructs, the current study targets these two constructs—through scalar ratings of accentedness and comprehensibility—to measure L2 word pronunciation learning.

Repetition and Vocabulary Learning
Repetition is an important condition contributing to L2 vocabulary learning (Webb & Nation, 2017). Although the positive effect of repetition has been demonstrated in deliberate vocabulary learning (Nakata, 2017), the prominence given to the effect has mostly stemmed from research on incidental vocabulary learning (Uchihara et al., 2019). This line of work involves looking for the optimal number of encounters with words necessary for significant learning to occur while learners engage in a meaning-focused activity such as reading a short story (Horst et al., 1998), listening to songs (Pavia, Webb, & Faez, 2019), viewing television (Peters, 2019), and listening to academic lectures (Dang, Lu, & Webb, 2021).

Webb (2007) conducted an experimental study with Japanese learners of English as a foreign language who read sets of sentences including 10 target words. Participants were randomly assigned to one of four treatment groups that encountered the target words one, three, seven, and 10 times. After the treatment, learning was assessed in tests measuring receptive and productive knowledge of five aspects of vocabulary knowledge (orthography, association, syntax, form–meaning connection, and grammar). Webb found that repeated

encounters promoted vocabulary learning, and also reported considerable variation of the repetition effect across different aspects of word knowledge. At one encounter, sizable gains in both receptive and productive knowledge of orthography were found, such that six out of 10 target words were learned receptively and five out of 10 productively. However, participants were less successful at learning form–meaning connections, as measured through a meaning recall test, where they demonstrated learning of only three out of 10 words after encountering target words 10 times.

Building on Webb's (2007) study, Chen and Truscott (2010) conducted a replication study in which participants encountered target words one, three, and seven times, and their learning was measured for receptive and productive knowledge of four aspects of vocabulary (orthography, form–meaning connection, grammar, and association). According to Plonsky and Oswald's (2014) effect-size benchmarks, the results showed large effects of repetition involving between one and three encounters for receptive and productive knowledge of orthography ($d = 1.02$ and 1.13, respectively), compared to medium effects for other aspects of word knowledge such as receptive knowledge of form–meaning connections ($d = 0.66$). These findings suggest that formal aspects of word knowledge may be sensitive to repetition effects and that measurable learning gains might arise after a small number of encounters (e.g., one to three).

Although findings of repetition effects are mostly based on written input (e.g., Chen & Truscott, 2010; Rott, 1999; Waring & Takaki, 2003; Webb, 2007), studies have started exploring spoken input, such as listening to academic lectures (Dang et al., 2021; Vidal, 2011), teacher talk (Jin & Webb, 2020), songs (Pavia et al., 2019), and TV interviews (van Zeeland & Schmitt, 2013), and viewing full-length TV programs (Peters & Webb, 2018). It appears that repetition effects are diminished in spoken input compared to written input. For example, Jin and Webb (2020) did not find a frequency effect for words encountered between three and 10 times in teacher speech. Van Zeeland and Schmitt (2013) found that 15 encounters with spoken words did not lead to larger gains than seven or 11 encounters in knowledge of form, grammar, or meaning at either immediate or delayed posttests. Vidal (2011) reported a larger correlation between frequency of encounters and learning gains in reading ($r = .69$) than in listening ($r = .49$). One explanation for these results is that, during listening, learners experience difficulty segmenting continuous speech, which makes it harder for them to identify target words and notice them appearing multiple times (Vidal, 2011). Accordingly, the relationship between exposure frequency and the learning of spoken forms may not be linear.

However, this research area is still in its infancy; few studies have been conducted, only limited aspects of word knowledge have been tested, and most results have focused on receptive knowledge (e.g., form or meaning recognition). To our knowledge, no studies have examined the effects of repetition on productive knowledge of spoken word forms (i.e., spoken form recall), let alone knowledge of pronunciation.

Repetition and Pronunciation Learning
The lack of research investigating the effects of repetition on pronunciation learning may be due to the discrepancy in modality between learning (i.e., listening) and testing (i.e., speaking). However, there are theoretical perspectives that account for a close interface between L2 perception and production, supporting the hypothesis that repeated exposures to L2 words would first help establish new phonological representations, which will in turn support improved perception and production accuracy. According to Flege's (1995) speech learning model, difficulties in perception are responsible for difficulties in production. Once an adequate perceptual specification of a L2 sound is established, such that it is not confused with a L1 sound, production will become more accurate with continued exposure over time. This view that perception is closely related to production has been empirically tested via a considerable body of perceptual training studies (for a review, see Sakai & Moorman, 2018). In the seminal work conducted by Bradlow et al. (1997), Japanese learners who completed three to four weeks of input-only perception training (i.e., focusing on identification of English /r/ and /l/) showed improvement not only in their perception but also in their production accuracy for these target sounds.

Prior work on auditory word priming also provides support for the view that repetition promotes pronunciation learning. Auditory word priming refers to the phenomenon in which prior exposure to spoken words leads to more rapid or more accurate processing of the same words at subsequent exposures (Church & Fisher, 1998). This processing advantage that repeated words have over unrepeated words is characterized by unconscious and unintentional facilitation, supporting the learning of spoken word forms. Such a repetition-driven processing advantage for words, observed in L1 speakers, appears also to be available to L2 learners (Trofimovich, 2005; Trofimovich & Gatbonton, 2006). If L2 learners are indeed more sensitive to spoken words they have recently encountered than words that they have not, then manipulating the frequency of exposures to spoken word forms might have great pedagogical value for improving L2 learners' pronunciation through classroom instruction.

Cognateness and Vocabulary Learning

Studies have examined how vocabulary learning is influenced by multiple word-related variables, including corpus-based frequency, cognateness, word length, part of speech, and concreteness (Peters, 2020). Among these variables, cognateness has received particular attention in L2 vocabulary research (e.g., Lotto & de Groot, 1998; Peters, 2019; Peters & Webb, 2018; Rogers et al., 2015). Cognates are typically defined as words that are phonologically or orthographically, semantically, and etymologically related across languages (Peters, 2020). However, this definition has been extended to word pairs that are shared across languages in form and meaning regardless of the presence or absence of an etymological relationship (Rogers et al., 2015). An example falling under this extended definition includes loanwords in Japanese such as *cable*/ケーブル (keeburu) and *cup*/カップ (kappu). Research has consistently indicated that cognates are easier to learn than noncognates regardless of the learning condition. For instance, in paired-associate learning, learners were more accurate and faster at recalling the forms of cognates than noncognates with fewer exposures (Lotto & de Groot, 1998). Similarly, in incidental learning research, participants were likely to acquire cognates before noncognates (Peters, 2019; Peters & Webb, 2018), and the positive effect of cognateness might be larger for learning through spoken than written input (Vidal, 2011).

However, our understanding of how cognateness affects L2 vocabulary learning is primarily based on the acquisition of limited aspects of word knowledge, particularly form–meaning connection, measured through translation tests to elicit the recall of L1 meanings (Peters, 2019) or L2 written forms (Lotto & de Groot, 1998) and through multiple-choice tests where learners choose L1 meanings cued by L2 written forms (Peters & Webb, 2018). To the best of our knowledge, research has yet to compare the learning of word pronunciation for cognates versus noncognates.

The Present Study

There are several reasons why research on pronunciation learning as a function of frequency of exposure is needed. First, it will advance our understanding of an underexplored dimension of vocabulary learning beyond form–meaning connections. Second, determining the number of encounters necessary to learn the pronunciation of L2 words might provide a useful guide to introducing new words in the classroom. Such guidance may also help to indicate the importance of systematically providing spoken input when teaching new words. Given the relevance of comprehensible speech to international com-

munication (Levis, 2020), it is crucial for the pronunciation of newly learned words to be available for immediate use in oral communication.

Third, the present study may help to reveal the extent to which pronunciation of L2 words can be learned as a by-product of exposure alone without explicit attention being drawn to specific phonological properties of these words.[1] Given that the time available for pronunciation instruction is often limited in L2 curricula (Martin, 2020), it is important to optimize in-class time by prioritizing L2 words whose pronunciation is difficult to learn and using out-of-class time to target other words that are easy enough to be learned incidentally. Exploration of how pronunciation learning is influenced by cognateness might help determine the learnability of words, informing teachers of specific word characteristics that are associated with a lower learning burden.

Finally, this research can help bridge the gap between vocabulary and pronunciation research. Pronunciation studies tend to focus on specific sound features extracted from words whose form–meaning connections are already established (such as high-frequency items), in order to control for the effects of word familiarity (e.g., Field, 2005; Lee & Lyster, 2016; Y. Saito & Saito, 2017; see Munro & Derwing, 2008, for evidence suggesting that word frequency affects L2 vowel acquisition). Exploring pronunciation learning through exposure to novel, previously unknown words could thus inform extant L2 pronunciation research and extend the L2 vocabulary literature by focusing more attention on input modality.

The present study was guided by the following research questions:

1. How does frequency of exposure (one, three, or six exposures) influence learners' recall of the spoken forms of previously unknown L2 words?
2. How does frequency of exposure (one, three, or six exposures) influence two aspects of learners' pronunciation (accentedness and comprehensibility) of previously unknown L2 words?
3. To what extent does cognateness moderate the relationship between repetition and learners' performance on measures of pronunciation learning?

Method
Overview of the Study
The study adopted a pre–post intervention design with three experimental groups (one, three, and six exposures) and three testing trials (pretest, immediate posttest, and delayed posttest). Participants were randomly assigned to the three experimental groups and received different frequencies of exposure to target words: one exposure, three exposures, or six exposures. The range

in exposure was based on earlier vocabulary studies that suggest that three exposures trigger an initial boost in learning and seven exposures lead to substantial learning gains (Chen & Truscott, 2010; Nakata, 2017; Webb, 2007). With respect to the learning of various aspects of pronunciation, detectable effects in the processing of L2 speech emerge after a single exposure to the target token (Church & Fisher, 1998; Trofimovich, 2005) and increase through two, four, and up to eight exposures (Gullberg, Roberts, & Dimroth, 2012; Vroomen et al., 2007), with long-lasting effects detected after listening to two to four sentences (e.g., Clarke & Garrett, 2004) or to 20 words (e.g., Norris, McQueen, & Cutler, 2003), or after several minutes of experience with novel phonetic material (e.g., Escudero & Williams, 2014). In an attempt to arrive at a frequency manipulation that would be compatible with both vocabulary and pronunciation research, frequency was compared between three and six repetitions, with one exposure designated as a baseline condition. Six repetitions (rather than seven or eight) were chosen as the highest exposure level based on the results of a pilot study, to mitigate a potential fatigue effect given the length of an extended learning sequence.

During the treatment, participants learned 40 English words through listening to the words and viewing their corresponding pictures. A picture-naming test was administered at the three testing times, and the elicited samples were evaluated for measures of form–meaning connection and pronunciation. We considered the listener's perspective to operationalize and measure the accuracy of word pronunciation: accentedness (degree of foreign accent) and comprehensibility (ease of understanding). The rationale for the use of listener-based measures was motivated from the standpoint of ecological validity (Derwing & Munro, 2009), given that what matters in real-life communication is how listeners perceive spoken L2 words and that successful recognition of spoken words is expected to result in successful oral communication (Field, 2005).

Participants
The participants were 79 Japanese university students in Japan who were learning English as a foreign language. Four participants were excluded from the subsequent analysis because they had lived abroad for an extended period of time (2–12 years). The remaining 75 participants had studied English for a minimum of 6 years in instructional settings. They had scored 90% or higher on the 1,000-word level of the Vocabulary Levels Test (Webb, Sasao, & Ballance, 2017), and all except two had scored 80% or higher on the 2,000-word level of the test. Their mean score at the 2,000-word level was 28.44, indicating that most participants had a considerable knowledge of the 2,000 most

frequent words. The 75 participants were randomly assigned to three experimental groups with 25 participants per group: one exposure (E1), three exposures (E3), and six exposures (E6). There was no significant between-group difference in vocabulary test scores, $F(2, 72) = 1.70, p = .191, \eta_p^2 = 0.05$. All participants reported normal hearing.

Target Items

Forty target words were quasi-randomly selected from a pool of candidate words collected according to the following three criteria (Table 1). First, because the purpose of this study was to examine the learning of "unknown" or "new" words rather than already known words, a pool of low-frequency words was created by collecting English word items that were beyond the most frequent 5,000 word families in Nation's BNC/COCA word lists (Nation, 2012). Second, because the treatment involved learning spoken forms attached to meanings conveyed in visual images (pictures), only concrete nouns were selected as candidate target items. Third, words that could be replaced with high-frequency synonyms were avoided to reduce the possibility that high-frequency synonyms of the target items would be produced in the picture-naming test.

Cognateness of target items was determined by having five L1 Japanese-speaking raters judge whether the target word was a loanword (Rogers et al., 2015). If an item was considered a cognate by the majority of raters (at least three out of five), it was coded as a cognate item (Peters, 2019). There was 90% agreement among the five raters. As a result of this procedure, 13 words were coded as cognates out of 40 target words.

Each of the 40 target words was recorded twice by a female native speaker of English using a TASCAM DR-05 audio recorder and digitized into a WAV format (44.1 kHz sampling rate with 16-bit quantization). The better of the two productions was selected in terms of clarity, naturalness, and lack of background noise and then stored as an individual sound file. To minimize the influence of between-speaker variations in loudness, the peak intensity for all speech samples was normalized using Praat (Boersma & Weenink, 2019). The stimuli were clear and comprehensible based on the judgment of another native English speaker.

Treatment and Testing

Paired-associate vocabulary learning was implemented as the learning intervention. The learning and testing schedule was programmed with PsychoPy (Peirce, 2007). Before the treatment began, participants put on headphones equipped with a microphone (AT810 Cardioid Headset Microphone) and familiarized themselves with the vocabulary learning task by working through

Table 1 Target words (n = 40) with basic item information and cognate status

No.	Target word	Number of phonemes	Number of syllables	Cognate
1	abalone	7	4	No
2	acorn	5	2	No
3	armadillo	8	4	Yes
4	binoculars	10	4	No
5	caramel	6	3	Yes
6	carousel	7	3	No
7	catapult	8	3	No
8	celery	6	3	Yes
9	chameleon	8	4	Yes
10	chandelier	9	4	Yes
11	chisel	4	2	No
12	cicada	6	3	No
13	clover	6	2	Yes
14	crayon	5	2	Yes
15	croissant	7	2	Yes
16	escalator	9	4	Yes
17	ladle	4	2	No
18	loquat	6	2	No
19	lotus	5	2	No
20	maracas	7	3	Yes
21	marshmallow	8	3	Yes
22	mermaid	6	2	Yes
23	pacifier	8	4	No
24	parakeet	8	3	No
25	persimmon	8	3	No
26	podium	6	3	No
27	porcupine	9	3	No
28	protractor	10	3	No
29	raccoon	5	2	No
30	raisin	4	2	Yes
31	razor	5	2	No
32	spatula	7	3	No
33	strainer	7	2	No
34	syringe	6	2	No
35	tadpole	6	2	No
36	toboggan	7	3	No
37	toupee	4	2	No
38	treadmill	7	2	No
39	walrus	6	2	No
40	xylophone	7	3	No

three practice examples. During the treatment, participants saw the meanings of the target words conveyed in visual images (i.e., copyright-free pictures retrieved from the Internet, standardized to a size of 400 × 400 pixels) while hearing the spoken forms of the words. For each target item, the picture was displayed on the computer screen for 4 s, with the auditory presentation of the target word beginning 750 ms after the picture appeared. The picture remained visible for the entire 4 s. A 2-s blank interval was inserted between trials.

During the treatment, the 40 target items were presented in a sequence of eight blocks of five items. The different experimental groups (E1, E3, and E6) received different numbers of exposures to the 40 target items. Thus, the total number of exposures to target items varied between the groups: 40 exposures (40 × 1 exposure) in E1, 120 (40 × 3 exposures) in E3, and 240 (40 × 6 exposures) in E6. For all groups, the order of item presentation within and between blocks was randomized across participants. For E3 and E6, five items in each block were presented in a fixed order so that the interval between the first exposure and the next exposure to the same target word remained constant, in order to control for spacing effects (e.g., Item A → B → C → D → E → A → B → C …).

Immediately after the final exposure to each block of five items, a picture-naming test was administered. In this test, participants were presented with the same pictures that were presented during the learning trial and asked to twice orally produce the words corresponding to the pictures shown on the computer screen. If participants did not remember a word, they were instructed to move to the next item. Their speech was recorded with a TASCAM DR-05 audio recorder and digitized into a WAV format (44.1 kHz sampling rate with 16-bit quantization). One out of two productions per word (i.e., a speech sample without fillers or self-corrections during articulation) was selected and stored in an individual sound file, with peak intensity normalized using Praat. Prior to data collection, issues with clarity of visual stimuli, trial procedures, and testing procedures were resolved through a pilot study with 20 university students with a similar learning background. Data for pilot study participants were not included in the main data analysis. Visual stimuli (Uchihara, Webb, Saito, & Trofimovich, 2022b) have been made publicly available via IRIS (https://www.iris-database.org) and the Open Science Framework (https://osf.io/zersy).

Procedure

The experiment was conducted over two sessions on two different days. On Day 1, participants who consented to participate in this study were informed

that they would learn 40 English words and that their oral production of these words would be elicited and recorded for the purpose of investigating the number of words they could remember. After this, participants took a pretest and then completed the treatment, an immediate posttest, and the Vocabulary Levels Test. For participants listening to words multiple times (those in groups E3 and E6), a 5-min break was provided halfway through the treatment to reduce participant fatigue. Participants were told to learn the English words but were not forewarned that their pronunciation would be assessed. When participants were observed reciting words during practice trials, they were encouraged to focus on listening rather than repeating words. Participants were not allowed to take notes of words they heard. After completing the treatment and immediate posttest, all participants agreed to meet with the researcher for the second session. Participants were not informed about the administration of a delayed posttest nor allowed to take home any learning materials, including a list of target words and sound files. On Day 2, approximately one week ($M =$ 6.05 days, $SD = 3.53$)[2] after the first session, participants completed a delayed posttest and filled out language background questionnaires. The treatment and tests were conducted individually with the researcher or a research assistant, and all speech samples were recorded in a sound-attenuated booth at the university. A total of 4,443 speech samples were elicited from 75 speakers on three test trials and evaluated for form–meaning connection and pronunciation measures. After all experiment sessions were completed, participants were debriefed about the purpose of the study and the fact that their production of words would be rated by English-speaking listeners in terms of pronunciation accuracy.

Form–Meaning Connection and Pronunciation Measures

To assess knowledge of form–meaning connection, spoken form recall (i.e., production of accurate forms of words in a picture-naming test) was measured. Form recall is considered the most difficult aspect of form–meaning knowledge for learners to master compared to form recognition, meaning recognition, and meaning recall (Laufer & Goldstein, 2004). For pronunciation measures, following Derwing and Munro (2015), two global constructs were measured: accentedness (i.e., listener rating of the extent to which learners' word productions deviated from a native variety of the target language) and comprehensibility (i.e., listener rating of the degree of effort needed to comprehend learners' word productions). The measure of spoken form recall was derived from participants' word productions, as transcribed orthographically

by two English language teachers. The accentedness and comprehensibility ratings were provided by an additional group of 24 raters. The test format (i.e., picture naming) across the three time points was the same, except that 10 high-frequency items were added to the pretest to boost motivation; these items were not counted for any measures.

Spoken Form Recall

To measure participants' productive knowledge of form–meaning connection (i.e., spoken form recall), two native English-speaking teachers, both speakers of North American English (one female, one male), were recruited to complete a timed dictation task programmed using PsychoPy (Peirce, 2007). In this task, raters listened to each of the 4,443 speech samples and typed the spelling of the word they heard as fast as possible. Raters were presented with 44 blocks of 100 samples and a block of 43 samples that contained a random selection of pretest, immediate posttest, and delayed posttest items. Recordings were played only once. Form recall scores per rater were derived from transcription accuracy, with minor misspellings considered accurate (e.g., *chisle*, *camelieon*, *ladel*). Coded dichotomously (1 = accurate, 0 = inaccurate), the form recall scores captured the ability to productively retrieve the spoken form of target words (cued by picture prompts) with sufficient accuracy, as judged by listeners. Before completing the rating task, raters completed a practice set of 15 samples representing varying pronunciation qualities (not included in the main dataset). Due to the large sample size and task demand, the two raters completed the listening task in multiple sessions (i.e., 14 to 16 sessions of 1 hr each). All listening sessions were implemented individually in the researcher's office.

Accentedness and Comprehensibility

Twenty-four native English speakers of North American English (13 females, 11 males) were recruited to participate in rating sessions. Fifteen speakers had never taught English, whereas the remaining nine speakers had some experience with English language teaching, such as language tutoring, teaching conversational English, and/or teaching academic English. Their familiarity with Japanese-accented speech (1 = *not familiar at all*, 6 = *very familiar*) was relatively high ($M = 4.58$, $SD = 1.65$), for various reasons such as having Japanese friends, taking Japanese language classes, and/or teaching English to Japanese students. They had no hearing problems. Although L2 speech ratings by listeners having L2 teaching experience and L1 familiarity might not perfectly

reflect how novice listeners perceive L2 spoken words, the potential effects of listeners' backgrounds on their ratings were considered minimal based on previous studies reporting no significant difference between the ratings assigned by experienced and inexperienced raters (Isaacs & Thomson, 2013; Kennedy & Trofimovich, 2008).

Because the goal of this study was to explore the learning of pronunciation of unknown words (defined as productive knowledge of form–meaning connection through a measure of spoken form recall), cases in which participants knew the words in the pretest were removed from the data for the immediate and delayed posttests. The resulting samples for the 75 speakers (2,051 words) were subsequently divided into four sets, and raters were randomly assigned to one of four sets. The allocation of speech samples was made such that raters listened to seven speakers from each of the three experimental groups (seven from E1, seven from E3, and seven from E6). Fifty-two samples from three out of seven speakers (one from E1, one from E3, and one from E6) occurred across all four sets, for a total of 2,207 samples across the four sets (Set 1: 580, Set 2: 528, Set 3: 565, Set 4: 534). This rating scheme allowed us to examine the consistency in rating behavior across and within sets while reducing the burden of rating tasks for each rater.

Individual rating sessions were scheduled with the researcher in a virtual environment. Five blocks of 100 words plus another block of the remaining words (28 to 80 words) were played once, and rating responses were recorded using Gorilla, an online experiment builder (Anwyl-Irvine et al., 2020). Following existing literature (e.g., Derwing & Munro, 1997), raters were first asked to familiarize themselves with 40 target words and received a brief description of the two pronunciation criteria: accentedness (1 = *no accent*, 9 = *extremely strong accent*) and comprehensibility (1 = *easy to understand*, 9 = *extremely difficult to understand*) (see Uchihara, Webb, Saito, & Trofimovich, 2022a, for instructions and rating scales provided to raters). Raters went through a practice set of 12 items, three of which were produced by native speakers of English. The researcher confirmed that all 24 raters understood the rating procedure and provided ratings of 1 (*no accent* and *easy to understand*) for native-speaker samples. The practice trial was followed by a main rating session in which raters evaluated speech samples with interim breaks provided between blocks. Raters completed three or four blocks in the presence of the researcher (approximately 1 hr, including background survey, instructions, practice trial, and breaks), and they were asked to complete the remaining blocks in their free time within a week. The presentation of speech stimuli in each set was randomized within and between blocks per rater.

Table 2 Rater consistency (Cronbach's α) for accentedness and comprehensibility ratings

Rating	Set 1	Set 2	Set 3	Set 4	Overlap
Accentedness	.83	.83	.87	.87	.96
	[.81, .85]	[.81, .85]	[.85, .89]	[.85, .89]	[.94, .97]
Comprehensibility	.92	.88	.90	.90	.98
	[.91, .93]	[.86, .89]	[.88, .91]	[.88, .91]	[.97, .99]

Note. Square brackets indicate 95% confidence intervals.

Preliminary Analysis

Before addressing the research questions, a preliminary analysis was conducted targeting the form–meaning connection and pronunciation measures. First, interrater agreement in the dictation task was checked (98% agreement). Second, interrater reliability for accentedness and comprehensibility ratings was examined. Due to technical problems, some rating scores for comprehensibility were not properly recorded and so were treated as missing data: Set 1 (0.2%), Set 2 (0.4%), Set 3 (1.1%), and Set 4 (4.3%). The resulting numbers of observations in total were 13,242 for accentedness and 13,211 for comprehensibility ratings. As presented in Table 2, the interrater consistency values (Cronbach's α) for all sets and within sets exceeded an acceptable benchmark of .70 (Larson-Hall, 2010); these results corresponded to those of earlier studies focusing on the production of sentences or paragraphs ($\alpha = .87$–$.92$; see K. Saito, 2021, for a review) and those of one previous study known to us (Martin, 2020) measuring the accentedness ($\alpha = .85$–$.92$) and comprehensibility ($\alpha = .89$–$.95$) of individual words.

In order to further inspect whether the 9-point scale for measuring accentedness and comprehensibility functioned properly, a many-facet Rasch analysis was conducted using Facets (Linacre, 2020). Based on the guidelines for the functionality of rating scales proposed by Eckes (2015), the preliminary results of average measures (i.e., monotonic increase with scale category 1 to 9) and data-model fit statistics (i.e., outfit mean square < 2.0) supported the functionality of the current rating scales for our raters (for detailed results, see Appendix S1 in the Supporting Information online). The distributional patterns for accentedness and comprehensibility ratings (Figure 1) show that ratings appeared to be harsher for accentedness, tending to be clumped around the strongly accented end of the scale, whereas they were more lenient for comprehensibility, tending to cluster closer to the easy-to-understand end. Correlation
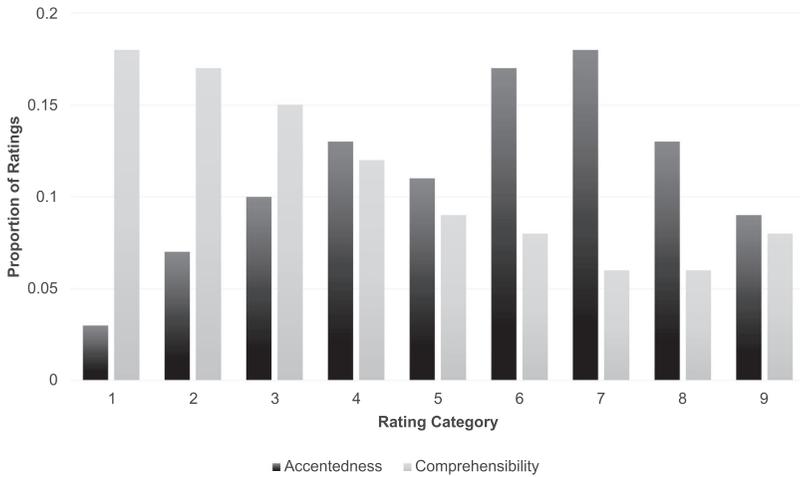
Figure 1 Distribution of ratings provided by 24 raters for accentedness and comprehensibility.

analysis showed that accentedness and comprehensibility were associated, $r =$ .72, 95% CI [.71, .73], $p < .001$, yet the disproportional patterns of the two rating sets suggest that these are distinct constructs at the word level, with at least 48% of variance being distinct between the two constructs, in line with earlier studies measuring two constructs at the sentence or passage level (Derwing & Munro, 2009, p. 480). These preliminary findings together confirm that the 9-point rating scale used for measuring accentedness and comprehensibility functioned properly for our raters, and that the two global constructs, measured at the word level, were correlated but partially independent of each other, supporting the construct validity of the two pronunciation measures in line with earlier L2 pronunciation studies.

Data Analysis

We addressed the three research questions through analyses of spoken form recall and of accentedness and comprehensibility using statistical analysis software, jamovi (Version 1.1; The jamovi project, 2019). Prior to conducting data analyses, we confirmed statistical assumptions: normality (through inspection of skewness and kurtosis statistics and residual distributions), homogeneity of variance, and collinearity.

To answer the first research question, regarding the effect of repetition on spoken form recall, a generalized mixed-effects model analysis was conducted

with exposure as a between-participants variable (E1, E3, and E6) and time as a within-participants variable (immediate and delayed posttest). Only two levels of test times were available because cases in which the participants knew the words in the pretest (indicated by the measure of spoken form recall) were removed. All independent variables were dummy coded (reference categories = E1 and immediate posttest), and an interaction term between time and exposure was tested in the model.

To answer the second and third research questions, regarding the effect of repetition on pronunciation learning and the influence of cognateness, we analyzed accentedness and comprehensibility ratings in a mixed-effects model. In this model, the fixed variables included exposure (E1, E3, E6) as a between-participants variable, time (immediate and delayed posttests) and cognateness (cognate and noncognate) as two within-participants variables, and all interaction terms among them. All independent variables were dummy coded (reference categories = E1, immediate posttest, and noncognate). We included random intercepts for participant (75 levels), word (40 levels), and rater (24 levels). Mixed-effects modeling was conducted for accentedness and comprehensibility ratings separately. For the three models of spoken form recall, accentedness, and comprehensibility, we ran each model twice with a different baseline each time, allowing us to examine the contrasts between E1 versus E3, E1 versus E6, and E3 versus E6 in order to interpret the main effect for exposure (Sinkeviciute et al., 2019). All models were fitted using a maximum likelihood technique. The magnitude of effect size (Cohen's d) with 95% confidence intervals was calculated and interpreted according to Plonsky and Oswald's (2014) effect-size benchmarks for between-groups contrasts: small ($d = 0.40$), medium ($d = 0.70$), and large ($d = 1.00$); and for within-group contrasts: small ($d = 0.60$), medium ($d = 1.00$), and large ($d = 1.40$). The raw data and the model code (Uchihara, Webb, Saito, & Trofimovich, 2022c and 2022d, respectively) are publicly available via IRIS (https://www.iris-database.org) and the Open Science Framework (https://osf.io/zersy).

Results
Spoken Form Recall
The descriptive statistics for spoken form recall and the two pronunciation measures (accentedness and comprehensibility ratings) are presented in Table 3. In order to address the first research question, regarding the effect of repetition on spoken form recall, a generalized mixed-effects model was fitted to the binary data, the results of which are summarized in Table 4 and

Table 3 Means, standard deviations, and 95% confidence intervals for spoken form recall and the two pronunciation ratings by group

Participant subgroups	Spoken form recall			Accentedness		Comprehensibility	
	Pretest	Immediate posttest	Delayed posttest	Immediate posttest	Delayed posttest	Immediate posttest	Delayed posttest
E1	9.84 (2.32)	22.00 (3.83)	13.68 (2.95)	3.65 (0.23)	3.43 (0.49)	2.74 (0.38)	2.31 (0.74)
	[8.88, 10.80]	[20.42, 23.58]	[12.46, 14.90]	[3.55, 3.74]	[3.23, 3.63]	[2.58, 2.89]	[2.01, 2.62]
E3	10.28 (2.13)	31.72 (4.76)	14.24 (3.15)	3.52 (0.34)	3.60 (0.35)	2.54 (0.48)	2.47 (0.53)
	[9.40, 11.16]	[29.76, 33.68]	[12.94, 15.54]	[3.38, 3.66]	[3.45, 3.75]	[2.34, 2.73]	[2.25, 2.70]
E6	10.16 (2.04)	35.12 (3.64)	15.12 (3.52)	3.41 (0.23)	3.40 (0.35)	2.35 (0.24)	2.38 (0.50)
	[9.32, 11.00]	[33.62, 36.62]	[13.67, 16.57]	[3.31, 3.50]	[3.25, 3.54]	[2.25, 2.45]	[2.17, 2.59]

Note. Maximum score for spoken form recall was 40. Accentedness ratings ranged from 1 (*no accent*) to 9 (*extremely strong accent*); comprehensibility ratings ranged from 1 (*easy to understand*) to 9 (*extremely difficult to understand*). Standard deviations are in parentheses, and upper and lower 95% confidence intervals are in square brackets. E1 = one-exposure subgroup; E3 = three-exposures subgroup; E6 = six-exposures subgroup.

Table 4 Results of generalized mixed-effects regression model for spoken form recall

Fixed effects	B	SE	95% CI		z	p
			Lower	Upper		
Intercept	0.49	0.35	−0.19	1.18	1.41	.159
Time	−2.09	0.16	−2.40	−1.77	−13.07	< .001
Exposure (1): E3–E1	1.79	0.22	1.35	2.23	7.97	< .001
Exposure (2): E6–E1	2.71	0.24	2.24	3.17	11.47	< .001
Time × Exposure (1)	−1.54	0.22	−1.97	−1.11	−7.02	< .001
Time × Exposure (2)	−2.29	0.23	−2.74	−1.84	−9.93	< .001

Random effects	SD	Variance	ICC
By-participant intercepts	0.65	0.42	0.08
By-word intercepts	1.94	3.77	0.73

Note. Marginal R^2 = .32; conditional R^2 = .70. E1 = one-exposure subgroup; E3 = three-exposures subgroup; E6 = six-exposures subgroup; ICC = intraclass correlation coefficient.
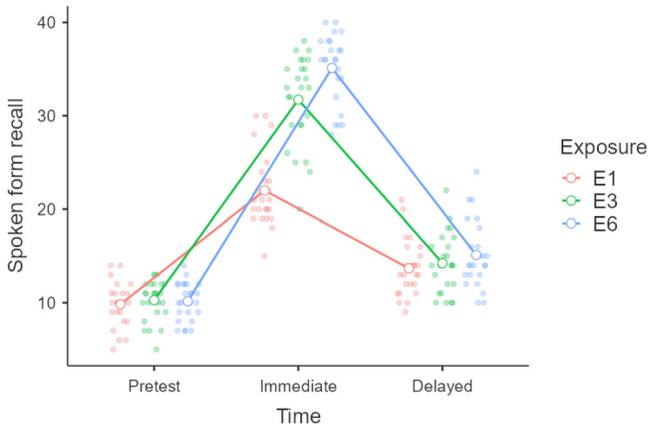
Figure 2 Estimated marginal means for spoken form recall by exposure with observed scores. Immediate = immediate posttest; delayed = delayed posttest; E1 = one-exposure subgroup; E3 = three-exposures subgroup; E6 = six-exposures subgroup.

Figure 2. Two main effects (time and exposure) were found statistically significant. As time and exposure variables were dummy coded (reference categories = immediate posttest and E1 group), the result of a significant exposure effect indicated that at the immediate posttest the E6 group recalled a greater number of spoken word forms than did the E1 group, $z = 11.47$, $p < .001$, $d = 3.51$, 95% CI [2.63, 4.40], and the E3 group outperformed the E1 group, $z = 7.97$, $p < .001$, $d = 2.25$, 95% CI [1.54, 2.96]. With the E3 group coded as a baseline group, the E6 group outperformed the E3 group, $z = 3.90$, $p = .001$, $d = 0.80$, 95% CI [0.23, 1.38]. However, a significant interaction of time and exposure showed that the repetition effect was significantly diminished at the delayed posttest: E6 versus E1, $d = 0.44$, 95% CI [–0.12, 1.01]; E6 versus E3, $d = 0.26$, 95% CI [–0.29, 0.82]; and E3 versus E1, $d = 0.18$, 95% CI [–0.37, 0.74]. Post hoc comparison tests with Bonferroni correction ($\alpha = .017$) showed no significant differences between the three exposure groups, whereas significant decreases in spoken form recall with large effects were found between immediate and delayed posttests for the E1 group, $d = –2.04$, 95% CI [–2.74, –1.37], the E3 group, $d = –4.07$, 95% CI [–5.01, –3.07], and the E6 group, $d = –6.29$, 95% CI [–6.46, –3.76]; for detailed results, see Appendix S1 in the Supporting Information online.

Accentedness and Comprehensibility

Mixed-effects modeling was conducted with the two pronunciation measures (accentedness and comprehensibility ratings) as dependent variables (see Tables 5 and 6). The results of the random components showed that 26% and 30% of the variance in accentedness and comprehensibility ratings respectively were explained by three random effects for accentedness (participant = 10%, word = 8%, and rater = 8%) and for comprehensibility (participant = 10%, word = 9%, and rater = 11%). In the accentedness model, three main effects (time, cognateness, exposure: E6–E1 contrast) were statistically significant.

In response to the second research question, regarding the effect of repetition on measures of L2 word pronunciation, the significant effect of exposure was further examined. As time, cognateness, and exposure variables were dummy coded (reference categories = immediate posttest, noncognate, and E1 group), the results for the significant exposure effect showed that at the immediate posttest the E6 group pronounced noncognates in a more nativelike manner than did the E1 group, $t = -2.78$, $p = .007$, $d = 0.39$, 95% CI [0.03, 0.76], but the difference between the E3 and E1 groups was not significant, $t = -1.76$, $p = .081$, $d = 0.25$, 95% CI [–0.12, 0.61]. With the E3 group coded as a baseline category, no significant difference in accentedness was observed between the E6 and E3 groups, $t = -1.02$, $p = .309$, $d = 0.14$, 95% CI [–0.21, 0.50]. In the comprehensibility model, two main effects (cognateness and exposure: E6–E3 and E3–E1 contrasts) were statistically significant. The results for the significant exposure effect showed that at the immediate posttest the pronunciation of noncognates by the E1 group was significantly less comprehensible than the pronunciation by the E6 group, $t = -3.96$, $p < .001$, $d = 0.56$, 95% CI [0.15, 0.98], and by the E3 group, $t = -2.05$, $p = .044$, $d = 0.29$, 95% CI [–0.13, 0.71]. With the E3 group coded as a baseline category, the analysis showed a marginally significant difference between the E6 and E3 groups, $t = -1.94$, $p = .057$, $d = 0.27$, 95% CI [–0.14, 0.68], indicating that the E6 group tended to pronounce noncognates more comprehensibly than did the E3 group.

In short, the results of immediate posttests showed that repeated exposure had significant effects on the accentedness and comprehensibility of noncognates. However, Figures 3–6 reveal that the repetition effect appears to be influenced by the timing of the posttest (immediate vs. delayed) and the cognateness of the word (cognate vs. noncognate). These observations are supported by a number of significant two-way interactions between the three predictor variables (time, exposure, and cognateness), although three-way interactions were

Table 5 Results of mixed-effects regression model for accentedness rating

Fixed effects	B	SE	95% CI Lower	95% CI Upper	t	p
Intercept	5.96	0.21	5.56	6.37	28.85	< .001
Time	−0.28	0.12	−0.52	−0.04	−2.32	.021
Cognateness	−0.44	0.22	−0.86	−0.02	−2.04	.045
Exposure (1): E3–E1	−0.33	0.19	−0.69	0.04	−1.76	.081
Exposure (2): E6–E1	−0.52	0.19	−0.88	−0.15	−2.78	.007
Time × Cognateness	0.45	0.18	0.11	0.80	2.56	.010
Time × Exposure (1)	0.50	0.16	0.19	0.80	3.20	.001
Time × Exposure (2)	0.46	0.15	0.16	0.76	3.03	.002
Cognateness × Exposure (1)	−0.01	0.14	−0.30	0.27	−0.10	.917
Cognateness × Exposure (2)	0.53	0.14	0.24	0.81	3.68	< .001
Time × Cognateness × Exposure (1)	−0.01	0.25	−0.50	0.48	−0.03	.978
Time × Cognateness × Exposure (2)	−0.35	0.24	−0.83	0.13	−1.43	.151

Random effects	SD	Variance	ICC
By-participant intercepts	0.63	0.39	0.10
By-word intercepts	0.55	0.31	0.08
By-rater intercepts	0.57	0.33	0.08

Note. Marginal $R^2 = .01$; conditional $R^2 = .23$. E1 = one-exposure subgroup; E3 = three-exposures subgroup; E6 = six-exposures subgroup; ICC = intraclass correlation coefficient.

Table 6 Results of mixed-effects regression model for comprehensibility rating

Fixed effects	B	SE	95% CI Lower	95% CI Upper	t	p
Intercept	4.65	0.26	4.14	5.16	18.02	< .001
Time	−0.20	0.14	−0.47	0.08	−1.41	0.160
Cognateness	−0.98	0.27	−1.51	−0.46	−3.68	< .001
Exposure (1): E3–E1	−0.44	0.21	−0.85	−0.02	−2.05	.044
Exposure: (2): E6–E1	−0.84	0.21	−1.26	−0.42	−3.96	< .001
Time × Cognateness	0.48	0.20	0.08	0.88	2.35	.019
Time × Exposure (1)	0.46	0.18	0.10	0.81	2.54	.011
Time × Exposure (2)	0.55	0.18	0.20	0.89	3.12	.002
Cognateness × Exposure (1)	0.06	0.17	−0.27	0.38	0.34	.736
Cognateness × Exposure (2)	0.86	0.17	0.54	1.19	5.23	< .001
Time × Cognateness × Exposure (1)	−0.20	0.29	−0.76	0.37	−0.68	.494
Time × Cognateness × Exposure (2)	−0.44	0.28	−0.99	0.11	−1.56	.118

Random effects	SD	Variance	ICC
By-participant intercepts	0.72	0.51	0.10
By-word intercepts	0.70	0.49	0.09
By-rater intercepts	0.78	0.60	0.11

Note. Marginal $R^2 = .02$; conditional $R^2 = .26$. E1 = one-exposure subgroup; E3 = three-exposures subgroup; E6 = six-exposures subgroup; ICC = intraclass correlation coefficient.
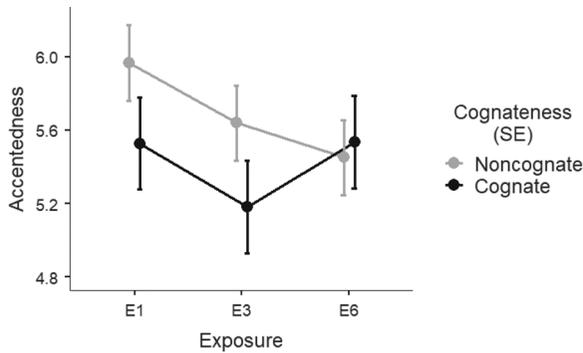
Figure 3 Estimated marginal means for accentedness ratings by exposure and cognateness (immediate posttest). Brackets enclose ± 1 standard error. E1 = one-exposure subgroup; E3 = three-exposures subgroup; E6 = six-exposures subgroup.
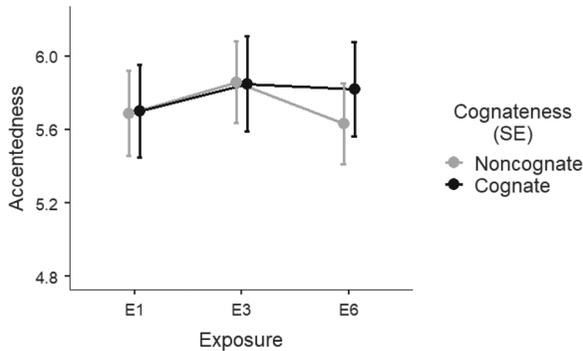

Figure 4 Estimated marginal means for accentedness ratings by exposure and cognateness (delayed posttest). Brackets enclose ± 1 standard error. E1 = one-exposure subgroup; E3 = three-exposures subgroup; E6 = six-exposures subgroup.

not significant. In order to interpret the effect of each variable on pronunciation learning, in what follows, each interaction effect is examined closely.

First, there were significant interactions between time and exposure for the E6–E1 contrast, $t = 3.03$, $p = .002$ (accentedness), $t = 3.12$, $p = .002$ (comprehensibility), and for the E3–E1 contrast, $t = 3.20$, $p = .001$ (accentedness), $t = 2.54$, $p = .011$ (comprehensibility), but not for the E6–E3 contrast, $t = -0.27$, $p = .785$ (accentedness), $t = 0.59$, $p = .554$ (comprehensibility). Figures 3–6 show that the repetition effect observed in the performance at the immediate posttest, particularly for noncognates, appears to be diminished at the delayed posttest. Post hoc tests confirmed no significant differences across
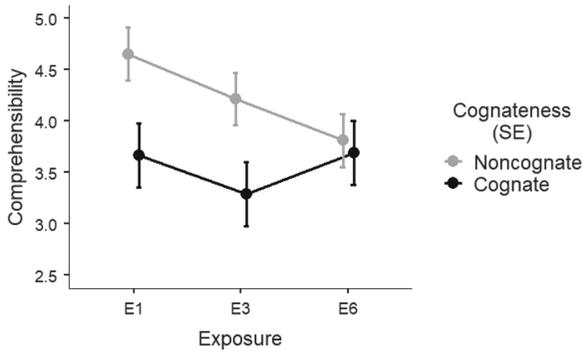
Figure 5 Estimated marginal means for comprehensibility ratings by exposure and cognateness (immediate posttest). Brackets enclose ±1 standard error. E1 = one-exposure subgroup; E3 = three-exposures subgroup; E6 = six-exposures subgroup.
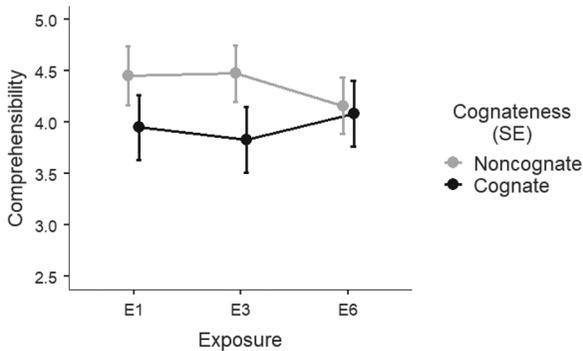

Figure 6 Estimated marginal means for comprehensibility ratings by exposure and cognateness (delayed posttest). Brackets enclose ±1 standard error. E1 = one-exposure subgroup; E3 = three-exposures subgroup; E6 = six-exposures subgroup.

the three exposure groups for delayed posttest performance. Regarding retention of pronunciation learning (immediate vs. delayed posttests), after a delay of approximately one week, the productions of words by the E3 and E6 groups were significantly more heavily accented in the E3 group, $t = -4.94$, $p < .001$, $d = -0.70$, 95% CI [–0.87, –0.52], and in the E6 group, $t = -2.74$, $p = .006$, $d = -0.39$, 95% CI [–0.55, –0.22], and also less comprehensible in the E3 group, $t = -3.89$, $p < .001$, $d = -0.55$, 95% CI [–0.75, –0.35], and in the E6 group, $t = -3.79$, $p < .001$, $d = -0.54$, 95% CI [–0.73, –0.34]. However, the E1 group showed no significant differences between the immediate and delayed posttests for either accentedness, $t = -0.16$, $p = .870$, $d = -0.02$, 95% CI [–0.43, 0.38],

or comprehensibility, $t = -0.42$, $p = .672$, $d = -0.06$, 95% CI [–0.26, 0.14]; for detailed results, see Appendix S1 in the Supporting Information online.

In response to the third research question regarding the influence of lexical cognateness, the interaction of exposure and cognateness was further examined. Figures 3–6 indicate that with increased exposure the pronunciation of noncognates tended to become more comprehensible and nativelike than that of cognates. Such emerging patterns were supported by significant interactions between cognateness and exposure for the E6–E1 contrast, $t = 3.68$, $p < .001$ (accentedness), $t = 5.23$, $p < .001$ (comprehensibility), and the E6–E3 contrast, $t = 3.77$, $p < .001$ (accentedness), $t = 4.87$, $p < .001$ (comprehensibility), although there was no significant interaction for the E3–E1 contrast, $t = -0.10$, $p = .917$ (accentedness), $t = 0.34$, $p = .736$ (comprehensibility). Post hoc comparison tests with Bonferroni correction ($\alpha = .008$) showed no significant differences between cognates and noncognates for accentedness and comprehensibility. However, there was a trend showing that the E6 group produced more comprehensible pronunciation of noncognates than did the E1 group, $t = 2.56$, $p = .012$, $d = 0.36$, 95% CI [–0.07, 0.80]. This result contrasts with the smaller effect of repetition on the pronunciation of cognates between the E6 and E1 groups, $t = -0.33$, $p = .742$, $d = -0.05$, 95% CI [–0.51, 0.41]. Further exploration of the effect size values for the E6 and E1 group difference suggests that the stronger effect of repetition for noncognates can be attributed to the immediate posttest performance, $d = -0.01$, 95% CI [–0.51, 0.48] (cognate), $d = 0.56$, 95% CI [0.15, 0.98] (noncognate), compared to the delayed posttest performance, $d = -0.07$, 95% CI [–0.59, 0.45] (cognate), $d = 0.16$, 95% CI [–0.35, 0.67] (noncognate); for detailed results, see Appendix S1 in the Supporting Information online.

Finally, significant interactions between time and cognateness were observed for accentedness and comprehensibility. Post hoc comparison tests with Bonferroni correction ($\alpha = .013$) showed that for noncognates there was no significant difference for either accentedness or comprehensibility between the immediate and delayed posttests. For cognates, the production of the words became significantly less nativelike and comprehensible, as indicated by the significant differences between the immediate and delayed posttests, $t = -4.64$, $p < .001$, $d = -0.66$, 95% CI [–0.81, –0.50] (accentedness), $t = -4.35$, $p < .001$, $d = -0.62$, 95% CI [–0.59, –0.22] (comprehensibility). Regarding the cognate-versus-noncognate comparison, no significant differences were found for either accentedness or comprehensibility in the delayed posttest. In the immediate posttest, there was a significant difference for comprehensibility, $t = 2.71$, $p = .010$, $d = 0.38$, 95% CI [0.19, 1.17], but not for accentedness, $t = 1.36$, p

$= .182$, $d = 0.19$, 95% CI [–0.20, 0.58], indicating that the pronunciation of cognates was more comprehensible (but not necessarily more nativelike) than that of noncognates immediately after the treatment; for detailed results, see Appendix S1 in the Supporting Information online.

Discussion

Repetition Influences Spoken Form Recall and Pronunciation of Noncognates

In answer to the first research question, the results showed that at immediate posttests learners receiving six exposures successfully recalled a larger number of spoken word forms than learners receiving one or three exposures, and that learners receiving three exposures outperformed those receiving one exposure. These findings support earlier studies highlighting the important role of repetition in developing learners' knowledge of form–meaning connection (Nakata, 2017; Uchihara et al., 2019; Webb, 2007). This result reveals that the positive effects of repetition can be extended to improving form recall in an aural modality. However, the absence of significant repetition effects on the delayed posttest suggests that the effect may not be long-lasting. This might be expected given that form recall is the most difficult aspect of form–meaning connection for learners to master (Laufer & Goldstein, 2004), the paired-associate learning program adopted in this study did not involve retrieval practice (Nakata, 2017), and there was a mismatch between the learning condition (i.e., recognizing spoken word forms) and the testing condition (i.e., producing spoken word forms; Morris, Bransford, & Franks, 1977).

In answer to the research questions regarding the repetition effects on word pronunciation learning, the results showed that at immediate posttests learners receiving six exposures produced noncognate words that were more comprehensible and more nativelike than learners receiving one exposure. Similarly, learners receiving three exposures outperformed those receiving one exposure in comprehensibility. These findings indicate that repetition enhanced the quality of spoken forms for unfamiliar words, while also likely contributing to the development of form–meaning connections. Put differently, learners' production of spoken word forms became more comprehensible and less accented after learners had encountered these spoken forms multiple times while attempting to remember word meanings. Learners might benefit from multiple auditory exposures to novel word forms because repetition might help learners refine the phonetic detail they perceive and subsequently store such phonetic information for these words in their lexicons, at least temporarily. A more refined lexical representation may then guide learners' production, resulting in

listeners perceiving the intended word to be more comprehensible and less accented.

However, the obtained repetition effects were not durable, and they diminished at the delayed posttests, suggesting that exposure to the spoken forms of words in a single learning session is not sufficient for the benefit of repetition to hold in the long run. The finding that no significant differences in pronunciation ratings emerged between immediate and delayed posttests for the E1 group (in contrast to the E3 and E6 groups) was unexpected, given that we predicted that more exposures should lead to greater learning and retention. One possible explanation for this finding is that with a relatively small amount of knowledge gained after one exposure, there was less knowledge to decay, resulting in little difference between the initial and subsequent tests.

Although repetition promoted the initial stage of pronunciation learning, accentedness appeared to be less impacted by repetition effects compared to comprehensibility. The analyses revealed (a) a larger effect of one versus six exposures on comprehensibility ($d = 0.56$, $p < .001$) compared to accentedness ($d = 0.39$, $p = .007$), (b) a small but significant effect of one versus three exposures for comprehensibility ($d = 0.29$, $p = .044$) but not accentedness ($d = 0.27$, $p = .081$), and (c) a larger effect of three versus six exposures for comprehensibility ($d = 0.27$, $p = .057$) than accentedness ($d = 0.14$, $p = .309$). These findings tentatively point to a different learning trajectory as shown through the ratings of accentedness and comprehensibility. Accentedness, which often requires extensive learning experience (Munro & Derwing, 2008), appears to develop more slowly to the extent that six exposures may only bring about moderate improvement compared to that for comprehensibility, which improved to a greater extent after six exposures (Derwing & Munro, 2013; K. Saito, 2015).

The findings of the current study add to an ongoing discussion revolving around how different aspects of word knowledge are developed with increased exposure (Chen & Truscott, 2010; Webb, 2007). Previous research has suggested that knowledge of word form (i.e., spelling) is learned more easily and quickly than other aspects of word knowledge, including word meaning or collocation. However, the present findings of relatively small effects of repetition on accentedness ($d = 0.14$–0.39) and comprehensibility ($d = 0.27$–0.56) contrast with the findings of previous studies measuring productive knowledge of orthography ($d = 0.43$–1.41 in Chen & Truscott, 2010; $d = 0.52$–1.33 in Webb, 2007). This was an unexpected finding, considering the different methodological approaches adopted in comparison with previous studies (Chen & Truscott, 2010; Webb, 2007). Because the target vocabulary items were presented in a

decontextualized manner with the word's meaning illustrated through pictorial information, it was expected that the learners in this study would pay attention to target word forms at every encounter. In contrast, the learners in the studies by Chen and Truscott (2010) and Webb (2007) encountered the target vocabulary in short sentences, requiring them to infer the meanings of unfamiliar words using contextual information. In light of the current findings and various methodological differences across studies, it appears that the impact of repetition on the learning of word forms depends on the mode (written vs. spoken) in which vocabulary learning is measured.

Cognateness Moderates Repetition Effects on Pronunciation Learning
In answer to the third research question, the effect of repetition was significantly moderated by cognateness. The significant interaction of cognateness and exposure indicated that the positive effects of repetition were predominantly attributable to improved pronunciation of noncognates, whereas little improvement was observed for cognates irrespective of exposure. A possible reason for this finding is that cognates could be pronounced with sufficient accuracy after a single exposure, so that little room was left for further improvement. The L1–L2 form and meaning overlap for cognates may provide learners with sufficient learning benefit immediately after the initial exposure, enabling them to produce cognates in a nativelike and comprehensible way. In contrast, learning noncognates involves encoding new information ranging from individual phonemes (e.g., vowels and consonants) to sound sequences (e.g., syllable structure) as well as mapping novel forms to meanings, requiring greater amounts of input and practice before such form-related knowledge is fully specified and acquired (Elgort et al., 2018).

The possible ceiling effect is especially true for comprehensibility, because the magnitude of cognateness effects does not appear to be consistent across the ratings of accentedness and comprehensibility. The results of immediate posttests showed that on average cognate status tends to have a larger impact on comprehensibility ($d = 0.38, p = .010$) than on accentedness ($d = 0.19, p = .182$). This finding is not surprising given that the two pronunciation measures reflect different constructs (Derwing & Munro, 2009). Despite some overlap between words in English and loanwords in Japanese, large differences still remain at the phonological and phonetic levels, providing cues to foreign accent that listeners easily detect. Consequently, the positive effect of cognateness was not as salient for accentedness as for comprehensibility.

In addition, the significant interaction of time and cognateness indicates that the extent to which pronunciation accuracy backslides from immediate

to delayed posttesting was greater for cognates than noncognates. Learners' pronunciation of cognates became less comprehensible and more strongly accented approximately one week after exposure compared to that of noncognates, shown through the ratings of both accentedness ($d_{\text{cognate}} = -0.66$ vs. $d_{\text{noncognate}} = -0.09$) and comprehensibility ($d_{\text{cognate}} = -0.41$ vs. $d_{\text{noncognate}} = -0.14$). The immediate learning gains for cognates thus appear less durable over time. For cognates, learning gains might be available to learners after their initial exposure to words; however, this knowledge might degrade rapidly and may be difficult to access one week after learning. During posttesting, learners might have relied on existing L1 knowledge about cognates, consequently pronouncing them in a more heavily accented and less comprehensible manner. Although cognates are generally pronounced more accurately than noncognates, further improvement in the pronunciation of cognates is likely to pose a challenge for L2 learners. At least in the short term, cognates appear to be easy to produce, yet learners might require extended exposure and practice opportunities to maintain this initial level of performance, given that cognates demonstrate close similarity to L1 word forms.

Limitations, Implications, and Future Directions

The current study provides initial evidence indicating that the frequency hypothesis (i.e., more exposure leads to greater learning) applies to the learning of word pronunciation and form–meaning mapping. However, the finding that pronunciation gains were not consistently retained in this study suggests that a single-time, on–off exposure session does not help learners retain improved pronunciation even after six exposures to target words. This finding implies that the process of learning spoken word forms is incremental and might require a greater number of exposures (more than six) over an extended period of time (more than one session).

Whereas exposure to the target items was carefully controlled in this study, an aim of most teachers and pedagogical resources is to provide repeated exposure to items over time. Thus, the treatment conditions in this study reflect learning without any subsequent exposure to target items, and this led to initial learning but not retention. In the classroom, we would hope that any initial exposure to spoken words forms would be supplemented with later exposure to expand on the early learning gains. In order to improve the effectiveness of repetition, future research should investigate whether long-term and spaced exposures to L2 words in classroom settings consolidate pronunciation gains. A recent meta-analysis by Kim and Webb (2022) suggests that a shorter spacing interval during learning is particularly beneficial, in light of the high degree

of complexity involved in the learning of pronunciation. When the spacing is longer, learners may have difficulty accessing phonological information during subsequent exposures to auditory input. It may therefore be important to provide the opportunity for learners to listen to spoken words intensively, for example, with a few days apart between multiple encounters.

The findings from treatments where learners encounter L2 words in isolation should not be generalized to situations of contextualized word learning where learners encounter L2 words in short sentences (e.g., Chen & Truscott, 2010; Webb, 2007) or longer passages (e.g., Dang et al., 2021; Waring & Takaki, 2003). Encountering words in connected speech might complicate the learning of pronunciation. Because the phonetic quality of words is influenced by their immediate environment, such as the preceding and following sounds (Field, 2014), resulting in variability in spoken forms, accurate recognition of spoken words encountered in varying phonetic contexts might become more challenging yet might eventually lead to more robust learning.

Although cognates generally enjoy a learning advantage over noncognates (Lotto & de Groot, 1998; Nation, 2013; Peters & Webb, 2018; Vidal, 2011), cognates may need to be taught explicitly, as their pronunciation is less likely to be improved through up to six encounters. Assuming that the goal for the majority of learners is to first achieve comprehensible pronunciation of L2 words (Levis, 2020), teaching cognates may not need to be prioritized at least for L2 beginners, given that sizable gains can be expected for cognates with a few exposures to their spoken forms. Also, as shown through informal interviews with raters, when learners pronounce Japanese loanwords in a way that makes them sound like English words, these words are sometimes harder to understand. It would therefore be important to draw intermediate and advanced learners' attention to the pronunciation of cognates, for example, through raising awareness of the differences between the spoken forms of Japanese loanwords and their English counterparts. Such focused practice also needs to be provided repeatedly over time, as initial improvement for cognates is more likely to decay rapidly than that for noncognates.

Several elements of the current design might be modified in future research to provide further insight into how repetition impacts L2 pronunciation. First, although the current study defined prior knowledge of target words in terms of form recall, participants might have had partial knowledge of some words such as form and meaning recognition. To further probe into the influence of form–meaning knowledge on pronunciation learning, other test formats such as multiple choice and L2-to-L1 translation tasks could be used to capture partial knowledge of form–meaning connections. Researchers, however, need to

carefully control the effects of exposure to test prompts, for example, by including a test-only group to determine the extent to which taking multiple tests might lead to improvement, independent of the repetition effects. Second, the learning approach adopted in the current study did not offer the best conditions for developing form–meaning connection: It did not involve, for example, retrieval opportunities or productive practice, because our goal was to examine the role of exposure. It would be more practically and pedagogically valuable to explore the degree to which learners' knowledge of spoken forms develops under more favorable learning conditions, considering depth of processing (Yanagisawa & Webb, 2021) and retrieval practice with corrective feedback (Nakata, 2017).

Third, more research is needed to further investigate the effect of cognateness on L2 word pronunciation learning. The current study was exploratory in that spacing for exposure to cognates and noncognates and different characteristics of cognates (e.g., the degree of similarity in phonological features between Japanese loanwords and English words) were not controlled. Controlling the spacing effect and exploring various word-related features of cognates, including phonological similarity, word length, and familiarity with loanwords, would provide a more nuanced understanding of how cognateness impacts L2 word pronunciation learning. Finally, different aspects of pronunciation could also be measured (K. Saito & Plonsky, 2019), given that, for example, prosodic features (e.g., word stress) and segmental accuracy differentially contribute to listeners' global judgments of pronunciation proficiency (Suzukida & Saito, 2022). Future studies should explore the extent to which repetition affects different aspects of pronunciation, including the accuracy of individual sounds and the placement of word stress (Field, 2005). Using a diverse toolkit of pronunciation measures would provide further insight into the role of repetition in L2 pronunciation development.

Conclusion

Through the current study, we have provided further insights into how repetition affects L2 vocabulary learning by adopting a listener's perspective to operationalize and examine knowledge of word pronunciation. We also found that cognates were not subject to repetition effects, which implied that cognateness is an important word-related moderator in L2 word pronunciation learning. The current findings need to be interpreted carefully, considering the lack of durable repetition effects, which invites future work to optimize the effectiveness of repeated exposures to spoken input for pronunciation learning. The main takeaway from the present findings is that repetition impacts

L2 word pronunciation learning without explicit attention being drawn to the phonetic features of individual words. In order to advance our understanding of how L2 input and instruction promote (or preclude) the acquisition of L2 vocabulary (e.g., in terms of form–meaning mapping) and pronunciation (e.g., operationalized as form specification) in tandem, researchers must engage in work that bridges the domains of vocabulary and pronunciation research. This study suggests the possibility that in future vocabulary research, researchers can track—at the level of individual words—the development of comprehensibility and accentedness as distinct pronunciation constructs (see Uchihara, 2022, for more evidence supporting the distinctiveness of the two constructs). Given the importance of testing word knowledge in a way that reflects learners' ability to use words in real-life communication (Kremmel & Schmitt, 2016), measuring L2 word pronunciation through global, listener-based constructs, such as comprehensibility and accentedness, offers a useful way to capture learners' ability to use words in spontaneous oral communication.

Final revised version accepted 15 April 2022

Open Research Badges

This article has earned an Open Data badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data are available at https://osf.io/zersy/.

Notes

1 The treatment task adopted in this study was a focused vocabulary learning activity (paired-associate learning) but not a deliberate pronunciation learning activity (at least from a methodological perspective), given that no guidance as to how to hear or articulate specific L2 sounds (for a definition of explicit instruction, see K. Saito & Plonsky, 2019) or information about pronunciation assessment was provided to participants. However, it is possible that some participants deliberately tried to improve their pronunciation of words as a consequence of the way in which target items were presented (in isolation rather than in context).

2 The interval between the immediate and delayed posttesting was not significantly different across the three exposure groups, $F(2, 72) = 0.56$, $p = .571$, $\eta_p^2 = 0.02$.

References

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*, 388–407. https://doi.org/10.3758/s13428-019-01237-x

Boersma, P., & Weenink, D. (2019). Praat: Doing phonetics by computer (Version 6.1.07) [Computer program]. Praat. http://www.praat.org

Bordag, D., Gor, K., & Opitz, A. (2021). Ontogenesis model of the L2 lexical representation. *Bilingualism: Language and Cognition*, *25*, 185–201. https://doi.org/10.1017/S1366728921000250

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/and /l/: IV. Some effects of perceptual learning on speech production. *The Journal of the Acoustical Society of America*, *101*, 2299–2310. https://doi.org/10.1121/1.418276

Chen, C., & Truscott, J. (2010). The effects of repetition and L1 lexicalization on incidental vocabulary acquisition. *Applied Linguistics*, *31*, 693–713. https://doi.org/10.1093/applin/amq031

Church, B. A., & Fisher, C. (1998). Long-term auditory word priming in preschoolers: Implicit memory support for language acquisition. *Journal of Memory and Language*, *39*, 523–542. https://doi.org/10.1006/jmla.1998.2601

Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign accented English. *Journal of the Acoustical Society of America*, *116*, 3647–3658. https://doi.org/10.1121/1.1815131

Dang, T. N. Y., Lu, C., & Webb, S. (2021). Incidental learning of single words and collocations through viewing an academic lecture. *Studies in Second Language Acquisition*. Advance online publication. https://doi.org/10.1017/S0272263121000474

de Jong, N. H., Steinel, M. P., Florijn, A. F., Schoonen, R., & Hulstijn, J. H. (2012). Facets of speaking proficiency. *Studies in Second Language Acquisition*, *34*, 5–34. https://doi.org/10.1017/S0272263111000489

Derwing, T. M., & Munro, M. J. (1997). Accent, intelligibility, and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition*, *19*, 1–16. https://doi.org/10.1017/S0272263197001010

Derwing, T. M., & Munro, M. J. (2009). Putting accent in its place: Rethinking obstacles to communication. *Language Teaching*, *42*, 476–490. https://doi.org/10.1017/S026144480800551X

Derwing, T. M., & Munro, M. J. (2013). The development of L2 oral language skills in two L1 groups: A 7-year study. *Language Learning*, *63*, 163–185. https://doi.org/10.1111/lang.12000

Derwing, T. M., & Munro, M. J. (2015). *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research*. Amsterdam: John Benjamins.

Eckes, T. (2015). *Introduction to many-facet Rasch measurement* (2nd ed.). Frankfurt am Main: Peter Lang.

Elgort, I., Candry, S., Boutorwick, T. J., Eyckmans, J., & Brysbaert, M. (2018). Contextual word learning with form-focused and meaning-focused elaboration. *Applied Linguistics*, *39*, 646–667. https://doi.org/10.1093/applin/amw029

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*, 143–188. https://doi.org/10.1017/S0272263102002024

Escudero, P., & Williams, D. (2014). Distributional learning has immediate and long-lasting effects. *Cognition*, *133*, 408–413. https://doi.org/10.1016/j.cognition.2014.07.002

Field, J. (2005). Intelligibility and the listener: The role of lexical stress. *TESOL Quarterly*, *39*, 399–423. https://doi.org/10.2307/3588487

Field, J. (2014). Pronunciation teaching has to establish in the minds of language learners a set of distinct consonant and vowel sounds. In L. Grant (Ed.), *Pronunciation myths: Applying second language research to classroom teaching* (pp. 80–106). Ann Arbor: University of Michigan Press.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 229–273). Timonium, MD: York Press.

Gullberg, M., Roberts, L., & Dimroth, C. (2012). What word-level knowledge can adult learners acquire after minimal exposure to a new language? *International Review of Applied Linguistics in Language Teaching*, *50*, 239–276. https://doi.org/10.1515/iral-2012-0010

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

Huensch, A., & Nagle, C. (2021). The effect of speaker proficiency on intelligibility, comprehensibility, and accentedness in L2 Spanish: A conceptual replication and extension of Munro and Derwing (1995a). *Language Learning*, *71*, 626–668. https://doi.org/10.1111/lang.12451

Isaacs, T., & Thomson, R. I. (2013). Rater experience, rating scale length, and judgments of L2 pronunciation: Revisiting research conventions. *Language Assessment Quarterly*, *10*, 135–159. https://doi.org/10.1080/15434303.2013.769545

Jiang, N. (2000). Lexical representation and development in a second language. *Applied Linguistics*, *21*, 47–77. https://doi.org/10.1093/applin/21.1.47

Jin, Z., & Webb, S. (2020). Incidental vocabulary learning from listening to teacher talk. *The Modern Language Journal*, *104*, 550–566. https://doi.org/10.1111/modl.12661

Kang, O., Thomson, R. I., & Moran, M. (2018). Empirical approaches to measuring the intelligibility of different varieties of English in predicting listener comprehension. *Language Learning*, *68*, 115–146. https://doi.org/10.1111/lang.12270

Kennedy, S., & Trofimovich, P. (2008). Intelligibility, comprehensibility, and accentedness of L2 speech: The role of listener experience and semantic context.

The Canadian Modern Language Review, *64*, 459–489.
https://doi.org/10.3138/cmlr.64.3.459

Kim, S. K., & Webb, S. (2022). The effects of spaced practice on second language
learning: A meta-analysis. *Language Learning*, *72*, 269–319.
https://doi.org/10.1111/lang.12479

Kremmel, B., & Schmitt, N. (2016). Interpreting vocabulary test scores: What do
various item formats tell us about learners' ability to employ words? *Language
Assessment Quarterly*, *13*, 377–392.
https://doi.org/10.1080/15434303.2016.1237516

Larson-Hall, J. (2010). *A guide to doing statistics in second language research using
SPSS*. New York: Routledge.

Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and
computer adaptiveness. *Language Learning*, *54*, 399–436.
https://doi.org/10.1111/j.0023-8333.2004.00260.x

Lee, A. H., & Lyster, R. (2016). Effects of different types of corrective feedback on
receptive skills in a second language: A speech perception training study. *Language
Learning*, *66*, 809–833. https://doi.org/10.1111/lang.12167

Levis, J. (2020). Revisiting the intelligibility and nativeness principles. *Journal of
Second Language Pronunciation*, *6*, 310–328. https://doi.org/10.1075/jslp.20050.lev

Linacre, J. M. (2020). Facets computer program for many-facet Rasch measurement
(Version 3.83.4). [Computer program].

Lotto, L., & de Groot, A. (1998). Effects of learning method and word type on
acquiring vocabulary in an unfamiliar language. *Language Learning*, *48*, 31–69.
https://doi.org/10.1111/1467-9922.00032

Martin, I. A. (2020). Pronunciation can be acquired outside the classroom: Design and
assessment of homework-based training. *The Modern Language Journal*, *104*,
457–479. https://doi.org/10.1111/modl.12638

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus
transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior*,
16, 519–533. https://doi.org/10.1016/S0022-5371(77)80016-9

Munro, M. J., & Derwing, T. M. (1995a). Foreign accent, comprehensibility, and
intelligibility in the speech of second language learners. *Language Learning*, *45*,
73–97. https://doi.org/10.1111/j.1467-1770.1995.tb00963.x

Munro, M. J., & Derwing, T. M. (1995b). Processing time, accent, and
comprehensibility in the perception of native and foreign-accented speech.
Language and Speech, *38*, 289–306. https://doi.org/10.1177/002383099503800305

Munro, M. J., & Derwing, T. M. (2008). Segmental acquisition in adult ESL learners:
A longitudinal study of vowel production. *Language Learning*, *58*, 479–502.
https://doi.org/10.1111/j.1467-9922.2008.00448.x

Munro, M. J., & Derwing, T. M. (2020). Foreign accent, comprehensibility and
intelligibility, redux. *Journal of Second Language Pronunciation*, *6*, 283–309.
https://doi.org/10.1075/jslp.20038.mun

Nakata, T. (2017). Does repeated practice make perfect? The effects of within-session repeated retrieval on second language vocabulary learning. *Studies in Second Language Acquisition*, *39*, 653–679. https://doi.org/10.1017/S0272263116000280

Nation, I. S. P. (2012). *The BNC/COCA word family lists*. Victoria University of Wellington. https://www.wgtn.ac.nz/lals/resources/paul-nations-resources/vocabulary-lists

Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge, UK: Cambridge University Press.

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, *47*, 204–238. https://doi.org/10.1016/S0010-0285(03)00006-9

Pavia, N., Webb, S., & Faez, F. (2019). Incidental vocabulary learning through listening to songs. *Studies in Second Language Acquisition*, *41*, 745–768. https://doi.org/10.1017/S0272263119000020

Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*, 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Peters, E. (2019). The effect of imagery and on-screen text on foreign language vocabulary learning from audiovisual input. *TESOL Quarterly*, *53*, 1008–1032. https://doi.org/10.1002/tesq.531

Peters, E. (2020). Factors affecting the learning of single-word items. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 125–142). London: Routledge.

Peters, E., & Webb, S. (2018). Incidental vocabulary acquisition through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition*, *40*, 551–577. https://doi.org/10.1017/S0272263117000407

Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning*, *64*, 878–912. https://doi.org/10.1111/lang.12079

Rogers, J., Webb, S., & Nakata, T. (2015). Do the cognacy characteristics of loanwords make them more easily learned than noncognates? *Language Teaching Research*, *19*, 9–27. https://doi.org/10.1177/1362168814541752

Rott, S. (1999). The effect of exposure frequency on intermediate language learners' incidental vocabulary acquisition and retention through reading. *Studies in Second Language Acquisition*, *21*, 589–619. https://doi.org/10.1017/S0272263199004039

Saito, K. (2015). Experience effects on the development of late second language learners' oral proficiency. *Language Learning*, *65*, 563–595. https://doi.org/10.1111/lang.12120

Saito, K. (2018). Advanced segmental and suprasegmental acquisition. In P. Malovrh & A. Benati (Eds.), *The handbook of advanced proficiency in second language acquisition* (pp. 282–303). Hoboken, NJ: Wiley Blackwell.

Saito, K. (2021). What characterizes comprehensible and native-like pronunciation among English-as-a-second-language speakers? Meta-analyses of phonological, rater, and instructional factors. *TESOL Quarterly*, *55*, 866–900. https://doi.org/10.1002/tesq.3027

Saito, K., & Plonsky, L. (2019). Effects of second language pronunciation teaching revisited: A proposed measurement framework and meta-analysis. *Language Learning*, *69*, 652–708. https://doi.org/10.1111/lang.12345

Saito, Y., & Saito, K. (2017). Differential effects of instruction on the development of second language comprehensibility, word stress, rhythm, and intonation: The case of inexperienced Japanese EFL learners. *Language Teaching Research*, *21*, 589–608. http://doi.org/10.1177/1362168816643111

Sakai, M., & Moorman, C. (2018). Can perception training improve the production of second language phonemes? A meta-analytic review of 25 years of perception training research. *Applied Psycholinguistics*, *39*, 187–224. https://doi.org/10.1017/S0142716417000418

Sinkeviciute, R., Brown, H., Brekelmans, G., & Wonnacott, E. (2019). The role of input variability and learner age in second language vocabulary learning. *Studies in Second Language Acquisition*, *41*, 795–820. https://doi.org/10.1017/S0272263119000263

Suzukida, Y. (2021). The contribution of individual differences to L2 pronunciation learning: Insights from research and pedagogical implications. *RELC Journal*, *52*, 48–61. https://doi.org/10.1177/0033688220987655

Suzukida, Y., & Saito, K. (2022). What is second language pronunciation proficiency? An empirical study. *System*, *106*, 102–754. https://doi.org/10.1016/j.system.2022.102754

The jamovi project. (2019). jamovi (Version 1.1) [Computer software]. https://www.jamovi.org

Trofimovich, P. (2005). Spoken-word processing in native and second languages: An investigation of auditory word priming. *Applied Psycholinguistics*, *26*, 479–504. https://doi.org/10.1017/S0142716405050265

Trofimovich, P., & Gatbonton, E. (2006). Repetition and focus on form in processing L2 Spanish words: Implications for pronunciation instruction. *The Modern Language Journal*, *90*, 519–535. https://doi.org/10.1111/j.1540-4781.2006.00464.x

Uchihara, T. (2022). Is it possible to measure word-level comprehensibility and accentedness as independent constructs of pronunciation knowledge? *Research Methods in Applied Linguistics*. Advance online publication. https://doi.org/10.1016/j.rmal.2022.100011

Uchihara, T., Webb, S., Saito, K., & Trofimovich, P. (2022a). *Speech perception task. Materials from "Frequency of exposure influences accentedness and comprehensibility in learners' pronunciation of second language words"* [Language test]. IRIS Database, University of York, UK. https://doi.org/10.48316/7zcz-5257

Uchihara, T., Webb, S., Saito, K., & Trofimovich, P. (2022b). *Word list and visual stimuli. Materials from "Frequency of exposure influences accentedness and comprehensibility in learners' pronunciation of second language words"* [Collection: stimuli]. IRIS Database, University of York, UK. https://doi.org/10.48316/7jzv-0h72

Uchihara, T., Webb, S., Saito, K., & Trofimovich, P. (2022c). *Data. Datasets from "Frequency of exposure influences accentedness and comprehensibility in learners' pronunciation of second language words."* IRIS Database, University of York, UK. https://doi.org/10.48316/16hw-zh44

Uchihara, T., Webb, S., Saito, K., & Trofimovich, P. (2022d). *Analysis code. Datasets from "Frequency of exposure influences accentedness and comprehensibility in learners' pronunciation of second language words"* [Analysis code]. IRIS Database, University of York, UK. https://doi.org/10.48316/68kf-k642

Uchihara, T., Webb, S., & Yanagisawa, A. (2019). The effects of repetition on incidental vocabulary learning: A meta-analysis of correlational studies. *Language Learning*, *69*, 559–599. https://doi.org/10.1111/lang.12343

van Zeeland, H., & Schmitt, N. (2013). Incidental vocabulary acquisition through L2 listening: A dimensions approach. *System*, *41*, 609–624. https://doi.org/10.1016/j.system.2013.07.012

Vidal, K. (2011). A comparison of the effects of reading and listening on incidental vocabulary acquisition. *Language Learning*, *61*, 219–258. https://doi.org/10.1111/j.1467-9922.2010.00593.x

Vroomen, J., van Linden, S., de Gelder, B., & Bertelson, P. (2007). Visual recalibration and selective adaptation in auditory-visual speech perception: Contrasting build-up courses. *Neuropsychologia*, *45*, 572–577. https://doi.org/10.1016/j.neuropsychologia.2006.01.031

Waring, R., & Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader? *Reading in a Foreign Language*, *15*, 130–163. https://doi.org/10.1177/003368828501600214

Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics*, *28*, 46–65. https://doi.org/10.1093/applin/aml048

Webb, S., & Nation, I. S. P. (2017). *How vocabulary is learned*. Oxford, UK: Oxford University Press.

Webb, S., Newton, J., & Chang, A. (2013). Incidental learning of collocation. *Language Learning*, *63*, 91–120. https://doi.org/10.1111/j.1467-9922.2012.00729.x

Webb, S., Sasao, Y., & Ballance, O. (2017). The updated Vocabulary Levels Test: Developing and validating two new forms of the VLT. *ITL - International Journal of Applied Linguistics*, *168*, 33–69. https://doi.org/10.1075/itl.168.1.02web

Werker, J. F. (2018). Speech perception, word learning, and language acquisition in infancy: The voyage continues. *Applied Psycholinguistics*, *39*, 769–777. https://doi.org/10.1017/S0142716418000243

Yanagisawa, A., & Webb, S. (2021). To what extent does the involvement load hypothesis predict incidental L2 vocabulary learning? A meta-analysis. *Language Learning*, *71*, 487–536. https://doi.org/10.1111/lang.12444

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Appendix S1. Summary of Analyses.

Appendix: Accessible Summary (also publicly available at https://oasis-database.org)

Does hearing a word one, three, or six times help learners pronounce second language words?

What This Research Was About and Why It Is Important

Seeing or hearing previously unknown words repeatedly enhances vocabulary learning in a learner's second language. Thus far, the role of repetition in second language vocabulary learning has been examined in relation to various aspects of vocabulary, including how well learners remember the meanings of new words or their spelling. However, this work has generally focused on the learning of written words, and there is little knowledge about how learners pronounce previously unknown spoken words. Understanding how learners pronounce words is important because being able to say words is a core aspect of vocabulary knowledge and a key contributor to speaking. Therefore, the researchers in this study studied whether presenting previously unknown words one, three, or six times during learning helps learners not only connect the word to its meaning but also to pronounce the word in such a way that it does not sound heavily accented and is easy to understand.

What the Researchers Did

- The researchers tested 75 Japanese university students in one of three groups (one, three, and six exposures).
- The students completed a word learning task, where they heard 40 unfamiliar English words with pictures illustrating their meaning. The words were presented one, three, or six times, depending on the group.
- The students' word knowledge was tested by showing them the same pictures and asking them to say the word. This test was given three times: before the learning task, immediately after, and about 1 week later.
- The researchers measured if the students said the correct word for each picture and also asked 24 external raters to evaluate how accented (on a scale between not accented and accented) and how comprehensible (on a scale

between easy to understand and hard to understand) the words sounded to them.

What the Researchers Found
- Hearing the words more frequently during the learning resulted in better recall and also in the words being perceived by raters as less accented and more comprehensible.
- Hearing the words six times during the learning had a more pronounced effect on how comprehensible than how accented the words sounded to external raters.
- Positive effects of repeated exposure to spoken words during learning dissipated after 1 week.
- Hearing words repeatedly was particularly useful for noncognate words (words which have no similar-sounding Japanese words) than for cognates.

Things to Consider
- Frequency of exposure plays an important role in how learners pronounce previously unknown words.
- Vocabulary learning is incremental, so it might require more than six exposures over a longer time period for the learning benefits to arise and stay, especially for words such as cognates whose pronunciations might be particularly confusable across a learner's two languages.

Materials, data, open access article: Materials and data are publicly available on IRIS (https://www.iris-database.org/) and OSF (https://osf.io/zersy).
How to cite this summary: Uchihara, T., Webb, S., Saito, K., & Trofimovich, P. (2022). Does hearing a word one, three, or six times help learners pronounce second language words? *OASIS summary* of Uchihara et al. (2023) in *Language Learning*. https://oasis-database.org

This summary has a CC BY-NC-SA license.