Cross-Situational Word Learning in Two Foreign Languages: Effects of Native Language and Perceptual Difficulty

Alba Tuninetti [1,2,3*†], Karen E. Mulak [1,2,4†] and Paola Escudero [1,2]

[1] The MARCS Institute for Brain, Behavior, and Development, Western Sydney University, Penrith, NSW, Australia, [2] Australian Research Council Centre of Excellence for the Dynamics of Language, Canberra, ACT, Australia, [3] Department of Psychology, Bilkent University, Ankara, Turkey, [4] Department of Hearing and Speech Sciences, University of Maryland, College Park, MD, United States

Cross-situational word learning (CSWL) paradigms have gained traction in recent years as a way to examine word learning in ambiguous scenarios in infancy, childhood, and adulthood. However, no study thus far has examined how CSWL paradigms may provide viable learning pathways for second language (L2) word learning. Here, we used a CSWL paradigm to examine how native Australian English (AusE) speakers learned novel Dutch (Experiment 1) and Brazilian Portuguese (Experiment 2) word-object pairings. During each learning phase trial, two words and objects were presented without indication as to which auditory word belonged to which visual referent. The two auditory words formed a non-minimal or vowel minimal pair. Minimal pairs were classified as "perceptually easy" or "perceptually difficult" based on the acoustic-phonetic relationship between AusE and each L2. At test, participants again saw two visual referents but heard one auditory label and were asked to select the corresponding referent. We predicted that accuracy would be highest for non-minimal pair trials (in which the auditory words associated with the target and distractor object formed a non-minimal pair), followed by perceptually easy minimal pairs, with lowest accuracy for perceptually difficult minimal pair trials. Our results support these hypotheses: While accuracy was above chance for all pair types, in both experiments accuracy was highest for non-minimal pair trials, followed by perceptually easy and then perceptually difficult minimal pair trials. These results are the first to demonstrate the effectiveness of CSWL in adult L2 word learning. Furthermore, the difference between perceptually easy and perceptually difficult minimal pairs in both language groups suggests that the acoustic-phonetic relationship between the L1-L2 is an important factor in novel L2 word learning in ambiguous learning scenarios. We discuss the implications of our findings for L2 acquisition, cross-situational learning and encoding of phonetic detail in a foreign language.

Keywords: second language (L2) acquisition, word learning, cross-situational learning, minimal pairs, acoustic-phonetic

INTRODUCTION

For adults, learning a second language (L2) can be difficult and time-consuming. Compared to young children, adults need more time and exposure to achieve native-like proficiency in an L2 (e.g., Johnson and Newport, 1989; deKeyser, 2000), and despite these efforts, are rarely rated as sounding like native speakers of the language (see Piske et al., 2001). While adulthood appears to confer some advantage in learning aspects of a language, for example, syntax and morphology (Krashen et al., 1979; Hartshorne et al., 2018), compared to younger L2 learners, later learners have greater trouble with language skills in the L2 such as pronunciation and production (Seliger et al., 1975; Oyama, 1976; Tahta et al., 1981; Piske et al., 2001), grammar learning (Johnson and Newport, 1989; deKeyser, 2000), and lexical access (Jared and Kroll, 2001; Kroll and Sunderman, 2003).

One contributing factor to this difficulty is the influence of the L2 learner's first (or native) language (L1). L2 word learning is affected by the relation between the phonetics and phonology of the learner's L1 and the L2. Models of L2 speech perception focus on how individual vowels and consonants are perceived based on these relations, with the expectation that difficulty or ease in discriminating a phonetic contrast in the L2 extends to respective difficulty or ease in discriminating minimal word pairs that differ by that contrast (e.g., Perceptual Assimilation Model-L2 [PAM-L2]: Best and Tyler, 2007; Second Language Linguistic Perception [L2LP]: Escudero, 2005, 2009; Speech Learning Model [SLM]: Flege, 1995). For example, words such as *rock* and *lock* in English are difficult for native Japanese speakers to learn, even with specific training (e.g., McCandliss et al., 2002). This is because L1 Japanese speakers perceive the initial sounds in English *rock* and *lock*, [ɹ] and [l], as instances of a single Japanese phoneme, /r/ (e.g., Aoyama et al., 2004), making it difficult to perceive *rock* and *lock* as two separate words. Similarly, L1 Spanish speakers learning novel Dutch minimal pair words that differed in a single Dutch vowel (e.g., [piχ] and [pɪχ]) in an explicit word learning task, in which each word is explicitly paired with its corresponding referent, showed poorer learning of minimal pairs when the Dutch vowel contrast differentiating the word pair ([ɪ] and [i]) was predicted to be perceived as a single Spanish vowel (/i/; Escudero et al., 2013, 2014).

But while such dissimilarities between the L1 and L2 sound inventories impede L2 word learning, these obstacles are not present when L1 and L2 sound contrasts align. The same L1 Spanish participants mentioned above showed stronger learning of Dutch minimal pair contrasts when an analogous contrast existed in their native Spanish (e.g., Dutch [i] and [y] in [piχ] and [pyχ] were predicted to be perceived as Spanish /i/ and /u/, respectively). This was true both for L2 learners who were naïve to Dutch, as well as those who had been learning and using the language in an immersive environment, suggesting this perceptual influence of the L1 on L2 perception and explicit word learning is relatively stable and not readily altered by L2 experience or proficiency (Escudero et al., 2013, 2014; see also Antoniou et al., 2015).

The value of this research in growing our understanding of the factors involved in L2 word learning is clear, specifically regarding our understanding of how the perceptual biases shaped by the L1 that the L2 learner brings with them affects this process. Nevertheless, these findings are limited in scope due to exclusive use of explicit word learning paradigms. In an explicit auditory word learning paradigm, participants undergo a training phase in which each novel word is explicitly, unambiguously paired with its corresponding referent. Typically, in each trial participants are shown a picture of a novel object on a screen in tandem with the auditory label for the object. This is followed with a test phase in which participants typically hear an auditory word and are asked to select the corresponding referent from a set of two or more visual objects (Smith, 2000; Escudero et al., 2013, 2014). This type of learning may more closely mimic classroom learning, in which teaching of L2 words generally occurs explicitly, such that a person is presented with the unambiguous one-to-one association between an unfamiliar L2 word and an existing concept (e.g., Spada, 1997). Such learning can, for instance, take the form of an activity in which students are shown pictures of concepts and their associated L2 word and are asked to repeat words out loud, with explicit instructions and corrective feedback (see Spada, 1997).

While undoubtedly effective in increasing L2 vocabulary (e.g., Ellis, 2015), explicit teaching methods do not reflect all the ways a language can be learned both in the classroom and in more naturalistic and immersive environments. The process of associating new words with objects can happen when the connection between the two is not explicitly taught, and no explicit feedback is provided (see e.g., Kriengwatana et al., 2016, for evidence that providing explicit and corrective feedback in non-native learning environments enhances performance compared to no feedback). Instead, the referent belonging to an auditory word is derived across multiple exposures to the word, narrowed down from an infinite set of possible referents. Determining these novel word-object pairings is supported through statistical tracking of word-referent co-occurrences over time, forming associations between words and referents that co-occur with the greatest probability (e.g., Yu and Smith, 2007) as well as top-down, hypothesis-checking techniques whereby the learner tests a possible word-object association by seeing whether the word and object co-occur in subsequent exposures (e.g., Trueswell et al., 2013; Berens et al., 2018). Indeed, this type of learning, termed cross-situational word learning (CSWL), is likely a primary way in which we learn words in our native language (Yu and Smith, 2007) and subsequent languages in an immersive environment.

In the lab, a CSWL paradigm comprises a learning phase followed by a test phase. In the learning phase, participants are not informed that this is a word learning task, instructed instead to simply view and attend to the stimuli presented to them. Participants see multiple novel images (candidate referents) on a screen, while the spoken label for each object (or in some cases, only one label) is presented in random order so that there is no indication of which spoken label refers to which novel image, resulting in referential ambiguity. In this way, while it is not possible to derive word-object associations in a single trial, participants can draw inferences about the relation between pseudowords and candidate referents across multiple

exposures by tracking pseudoword-object co-occurrences across trials. These associations are then tested in a forced-choice test phase, in which participants hear one word and are asked to select its referent from more than one presented on the screen. No feedback is given at any point throughout the learning or testing phases.

Research using this paradigm supports CSWL as a real-world word-learning strategy (e.g., Yu and Smith, 2007; Fitneva and Christiansen, 2011; Vlach and Sandhofer, 2014; Escudero et al., 2016). In their seminal study, Yu and Smith (2007) showed that university students could associate novel English pseudowords with novel objects with differing degrees of within-trial ambiguity, as defined by the number of pseudowords and novel objects presented in a single trial. Participants saw two to four pictures and heard two to four pseudowords in each trial, with no indication of picture-word mappings. Therefore, participants' degree of certainty about which pseudoword corresponded with which object varied, based on the number of pseudowords and objects presented within a trial (i.e., less ambiguity with two pseudowords and two objects compared with four pseudowords and four objects). At test, they selected the correct referent significantly above chance in all conditions, demonstrating that adults can use cross-situational learning to derive the correct word-object associations for novel words produced in their native language. Research since then has continued to show support for CSWL as a real-world mechanism, demonstrating that adults can retain these novel word-object pairings for at least a week (Vlach and Sandhofer, 2014), and can encode novel words learned via CSWL in fine phonological detail (Escudero et al., 2016; Mulak et al., 2019). Australian English (AusE) speakers learned and subsequently identified pseudoword-object pairings in a non-minimal pair (e.g., *bon-deet*), consonant minimal pair (e.g., *bon-ton*), and vowel minimal pair (e.g., *deet-dit*) context above chance in all conditions, though accuracy was lower in the vowel minimal pair context, suggesting weaker encoding of vowels compared to consonants (Escudero et al., 2016).

While the research on CSWL in adults supports CSWL as a viable word learning mechanism in the L1, adults' ability to learn words in an L2 via CSWL has not been investigated. This is because novel words across experiments to date have conformed to the phonology and phonotactic rules of the learner's L1. Of course, there is no reason to believe that adults cannot use CSWL to learn L2 words at all. While not previously tested in adults, children have been found to learn L2 words in this way. In a direct comparison of L2 CSWL and a more explicit paradigm, 8-year-old Mandarin-speaking students who were studying English were exposed to four real English words that were unknown to the students (*clamp, wedge, snood*, and *dart*) that were paired with novel images (i.e., not with the actual referent in English). Half of the participants were exposed to word-object pairings in a CSWL paradigm, in which a target word was presented with another target word in each trial. The other participants were taught words in a more unambiguous, mutual exclusivity paradigm, in which the novel image paired with the auditory word in a trial was presented alongside an image for which children knew the corresponding English word. In this way, the auditory label could

be inferred as belonging to the novel referent by a process of elimination, since participants already know the label associated with the alternate referent. In both conditions, children learned all four words. While immediate testing revealed a disadvantage for words learned via CSWL, there were no differences across conditions when retention was examined 15 min after the task had ended (Hu, 2017).

Similarly, Junttila and Ylinen (2020) demonstrated that 5- to 8-year-old Finnish children could learn real English word-object pairs in a CSWL paradigm in which they were presented with two spoken words and two images in each trial. The authors found no evidence that CSWL differed in effectiveness compared to an intentional, explicit learning paradigm in which children were asked to memorize the word object pairs (with some children also being asked to produce the words) or an incidental learning paradigm in which children were not asked to memorize the words and were asked to produce the Finnish word for each visual referent, such that any learning of the English labels would have occurred incidentally.

While these studies demonstrate that children can learn L2 words via CSWL, the L2 words used by Hu (2017) and Junttila and Ylinen (2020) were all very phonologically distinct. As discussed above, certain sounds in an L2 can be particularly difficult for a learner to discriminate based on the relation between the sounds in the L2 and the listener's native L1 sound categories. Because CSWL involves tracking the co-occurrences between auditory words and candidate referents, if listeners are unable to reliably distinguish between certain minimal pair words between the L1 and the L2, that could greatly impact the efficacy of CSWL in the L2. A similar situation arises when learners are tasked with learning two words for one referent, as is common in the bilingual CSWL literature. Indeed, in a CSWL task in which participants were taught two auditory labels for each visual referent, participants with experience with more than one language (i.e., they knew English and had knowledge of at least one other language) were better at learning both labels compared to monolingual English participants, but only when the two auditory labels were very phonologically distinct (disyllabic words ending in a vowel vs. monosyllabic words ending in /k/; Benitez et al., 2016). These results highlight the possibility that if monolinguals are unable to distinguish between the auditory minimal pairs, their ability to track the word-object pairs across learning trials may be disrupted and may lead to competition between certain word pairs depending on their phonetic similarity.

To investigate whether adults can learn L2 words via CSWL and in particular whether perceptual difficulties in the L2 obstruct CSWL in the L2, the current study compared L2 learners' ability to learn phonologically distinct L2 pseudowords and vowel minimal pairs in which the vowel contrast is predicted to be perceptually easy or difficult for the listener to discriminate based on the phonological relation between the L1 and to-be-learned language. Specifically, we tested monolingual AusE speakers' ability to learn referents associated with novel words produced by native speakers of Dutch (Experiment 1) or Brazilian Portuguese (Experiment 2), which conformed to Dutch or Brazilian Portuguese phonology and phonotactics.

Comparing learning in two languages allows us to test the idea that the acoustic-phonetic relationship between the L1 and L2 is a deciding factor in how well learners acquire new words.

Our predictions regarding the effects of the L1-L2 acoustic-phonetic relationship were based off models of L2 speech perception, such as the L2LP model (e.g., van Leussen and Escudero, 2015) and the PAM-L2 (Best and Tyler, 2007), which examine how this early or initial perception of L2 sounds can help or hinder discrimination based on acoustic (L2LP) or articulatory (PAM-L2) characteristics. Specifically, when two sounds in the L2 map onto one category in the L1, this can make discrimination of the contrast more difficult, since both L2 sounds are perceived as belonging to a single L1 sound (e.g., Spanish speakers confusing English [i] in *bean* and [ɪ] *bin* because both map onto the sole Spanish /i/ category). This is known as a "new scenario" in the L2LP model. Another difficult scenario is a "subset scenario," in which one non-native vowel can be categorized into two or more native categories. For example, in a categorization study, native AusE listeners categorized the non-native Dutch vowel /ɤ/ across three different native AusE vowels fairly equally: /ɛ/ - 19%, /ʊ/ - 19%, /ʉ/ - 14% (Alispahic et al., 2017). In that same study, Dutch /ɪ/ and /i/ were mapped most frequently to AusE /ɪ/ (40 and 48%, respectively), in an example of a "new scenario." These two scenarios highlight the difficulties in vowel perception and categorization across different languages, such that depending on the acoustic relationship between the L1 and L2, certain vowels may be perceived as belonging to one or more native categories, leading to difficulty in learning to discriminate them in lexical contexts.

Following the L2LP model, we used acoustic measurements of the Dutch, Brazilian Portuguese, and AusE vowels to classify the target Dutch and Brazilian Portuguese minimal pairs as perceptually easy or difficult for native AusE listeners to discriminate. We focused on vowel minimal pairs over consonant minimal pairs because previous work has demonstrated that while native AusE listeners learn both consonant and vowel minimal pairs via CSWL in their own language, performance is weakest for native vowel minimal pairs (Escudero et al., 2016; Mulak et al., 2019). Thus, if L2 perceptual difficulties do affect CSWL in the L2, the ability to learn L2 vowel minimal pairs may be particularly affected.

We predicted that overall, native AusE participants would be able to learn L2 pseudowords in both Dutch and Brazilian Portuguese, given evidence that adults can learn non-minimal and minimal pair L1 pseudowords via CSWL (Escudero et al., 2016) and evidence that children can learn phonologically distinct L2 words in a version of CSWL (Hu, 2017; Junttila and Ylinen, 2020). Given that when tested in the L1 adults show better learning of non-minimal than vowel minimal pairs (Escudero et al., 2016), we predicted performance would be best for non-minimal pairs compared to perceptually easy minimal pairs. We further predicted that accuracy would be worst for perceptually difficult minimal pairs compared to perceptually easy pairs, presumably because the difficulty in discriminating between words would disrupt cross-situational tracking of word-object co-occurrences.

METHODS

Experiment 1: Dutch

Participants

Participants were 20 undergraduate students at Western Sydney University (15 females, 5 males, mean age = 22.87 years, SD = 3.72 years). All students were monolingual AusE speakers as revealed by a language background questionnaire administered at the beginning of the session (i.e., participants self-reported only using English in their daily lives, including schooling and work). Participants received course credit for their participation.

Stimuli

The 12 Dutch auditory words were the same as those used in Escudero et al. (2013). They were produced by a native female Dutch speaker and were recorded at the University of Amsterdam in a soundproof booth. All words adhered to Dutch phonology and phonotactics. Half of the words were in a /p-vowel-χ/ (/pVχ/) context with the Dutch vowels /ɪ, i, ɑ, a, y, ʏ/. **Table 1** reports the first (F1) and second (F2) vowel formants, which approximately correspond to the position of the tongue during vowel production with regards to height and backness, respectively. Overlap between these values can help predict whether a contrast will be perceptually easy or difficult for a non-native listener to perceive. These measurements were originally reported by Elvin et al. (2016) for AusE (Western Sydney area) and Adank et al. (2004) for Dutch.

Pairings of these six /pVχ/ words comprised the minimal pair (MP) set. The remaining six words were disyllabic words adapted from Shatzman and McQueen (2006). These words contained different consonants and vowels from the MP set, arranged in variable, phonologically distinct contexts (/.beːptuː/, /.foːmpəl/, /.jɔmtoː/, /.kɛstə/, /.surkɛt/, /.tœykfɔm/). These formed non-minimal pairs (non-MPs) when paired with one another or with a /pVχ/ word. The 12 Dutch pseudowords were randomly paired with 12 black-and-white line drawings of nonsense objects from Shatzman and McQueen (2006).

As mentioned in the introduction, categorization of vowel contrasts as easy or difficult was based on patterns of expected acoustic-phonetic mapping of L2 vowels to the L1 phonological space, following the L2LP model (see e.g., van Leussen and Escudero, 2015; Alispahic et al., 2017). We used the categorization results from Alispahic et al. (2017) to predict how our participants would categorize the Dutch vowels tested here, as this study also tested native AusE speakers from the Western Sydney area. As can be seen in **Table 2**, difficult minimal pairs were those that contained vowel contrasts that could be classified as belonging to the same L1 vowel category or that could be categorized across more than one L1 category, whereas easy minimal pairs contained vowel contrasts that were expected to be categorized clearly to two separate L1 vowel categories.

Procedure

Participants completed a CSWL task which consisted of a learning phase and a testing phase. They were seated in front of a 17-inch laptop computer and were told they would see images

TABLE 1 | Formant values (Hz) of vowels across languages used in present study.

Australian English			Brazilian Portuguese			Dutch		
Vowel	F1	F2	Vowel	F1	F2	Vowel	F1	F2
/iː/	320	2,339	/i/	307	2,676	/ɑ/	578	1,172
/ɪ/	332	2,336	/e/	425	2,468	/a/	670	1,425
/e/	467	2,085	/ɛ/	646	2,271	/ɪ/	361	2,162
/æ/	695	1,763	/a/	910	1,627	/i/	278	2,162
/ɐː/	757	1,349	/o/	681	1,054	/ɣ/	366	1,595
/ɐ/	743	1,386	/ɔ/	442	893	/y/	259	1,734
/ɔ/	584	1,040	/u/	337	812			
/oː/	439	846						
/ʊ/	378	948						
/ʉː/	341	1,796						
/ɜ/	468	1,637						
/ɪə/	329	2,343						
/eː/	452	2,092						
/ɑe/	660	1,099						
/æɪ/	745	1,613						
/oɪ/	480	956						
/æɔ/	698	1,844						
/əʉ/	636	1,442						

For Dutch and Brazilian Portuguese, only vowels that appear in minimal pair stimulus words are included. Values were originally reported in Elvin et al. (2016) for AusE, Elvin et al. (2014) for Brazilian Portuguese, and Adank et al. (2004) for Dutch.

and hear words. Participants were not told that the words were associated with the images or that they would later be tested on their association between the pictures and words.

Learning Phase

The learning phase consisted of 72 trials (12 words presented 6 times each) in which they were presented auditorily with two pseudowords and visually with two black-and-white non-sense line drawings on the screen on the left and right sides (from Shatzman and McQueen, 2006). All pseudowords were presented with every other pseudoword at least once and trials were counterbalanced such that each novel line drawing was presented an equal number of times on the left or right and each pseudoword was presented an equal number of times as the first or second word in each trial. Trials had a 500 ms delay between picture onset and word onset and a 500 ms inter-stimulus interval.

Testing Phase

The testing phase consisted of 264 trials (12 words presented 22 times each as target, twice with every other word). This created 28 trials that were considered difficult MPs, 32 trials that were considered easy MPs, and the remaining 204 trials were non-minimal pairs. While pairing each word with every other word resulted in a greater number of non-minimal pair trials compared to minimal pair trials, this ensured minimal pairs did not stand out as a primary focus of our investigation (see also Escudero et al., 2014), and is also reflective of naturalistic language exposure in which minimal pairs appear at a much lower frequency.

TABLE 2 | Easy and difficult vowel mappings of Dutch vowels to AusE vowels.

Dutch

Easy		Difficult	
Contrast	Expected AusE mapping	Contrast	Expected AusE mapping
/ɪ/-/ɑ/	/i/-/ɔ/	/ɪ/-/i/	/ɪ/
/ɪ/-/a/	/ɪ/-/ɐː/	/ɑ/-/a/	/ɔ/ or /ɐː/-/ɐː/
/ɪ/-/ɑ/	/ɪ/-/ɔ/	/ɣ/-/y/	/ɛ/ or /ʊ/ or /ʉː/ - /ʊ/ or /ʉː/ or /ɪ/
/ɪ/-/a/	/ɪ/-/ɐ/	/ɪ/-/y/	/ɪ/ - /ʊ/ or /ʉː/ or /ɪ/
/ɑ/-/ɣ/	/ɔ/-/ʉː/	/ɪ/-/ɣ/	/ɪ/ - /ɛ/ or /ʊ/ or /ʉː/
/ɑ/-/y/	/ɔ/-/ʉː/	/i/-/y/	/iː/ or /ɪ/ or /i/ or /ɛ/ - /ʊ/ or /ʉː/ or /ɪ/
/a/-/ɣ/	/ɐː/-/ʉː/	/i/-/ɣ/	/iː/ or /ɪ/ or /i/ or /ɛ/ - /ɛ/ or /ʊ/ or /ʉː/

Participants heard one pseudoword and were presented with two line drawings on the screen and were asked to indicate which line drawing corresponded to the word they heard. They pressed a key on the keyboard to make their response. While it is possible that participants could learn or strengthen pseudoword-object pairings during the training phase, importantly, no feedback was given on test trials. Thus, the test trials were effectively additional cross-situational trials in which one auditory word co-occurred ambiguously with two candidate referents, and therefore still reflect CSWL. See
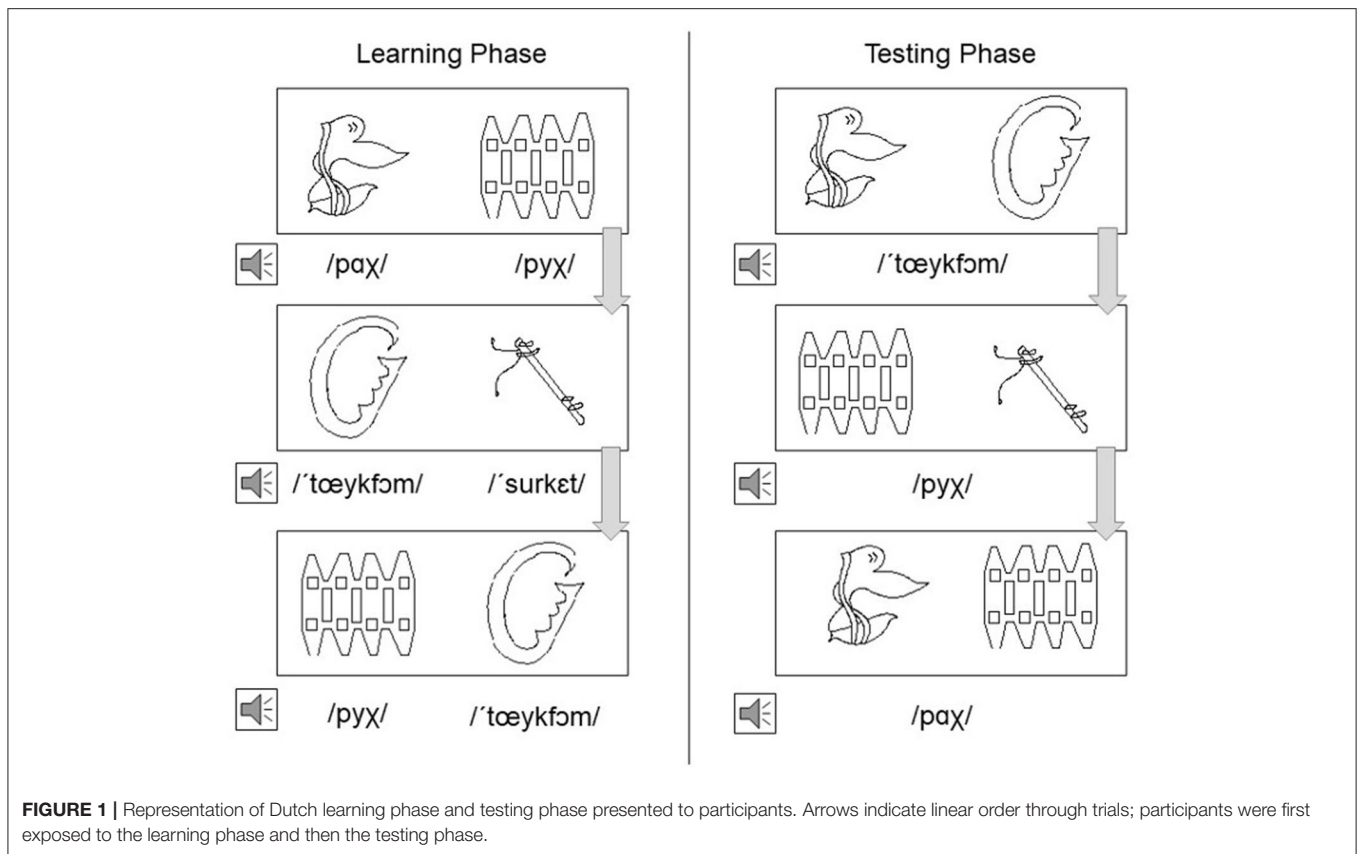
FIGURE 1 | Representation of Dutch learning phase and testing phase presented to participants. Arrows indicate linear order through trials; participants were first exposed to the learning phase and then the testing phase.

Figure 1 for a representation of the learning and testing phase for Dutch stimuli.

Experiment 2: Brazilian Portuguese
Participants
Participants were 20 undergraduate students at Western Sydney University (13 females, 7 males, mean age = 24.10 years, SD = 8.23 years). As with Experiment 1, all students were monolingual AusE speakers as revealed by a language background questionnaire administered at the beginning of the session and they received course credit for their participation.

Stimuli
Brazilian Portuguese words were selected from the Escudero et al. (2009) corpus. The words were produced for Escudero et al. (2009) by a female native speaker of Brazilian Portuguese recorded at the Escola Superior de Propaganda e Marketing in São Paolo, and adhered to Brazilian Portuguese phonology and phonotactics. Fourteen words were selected comprising the seven Brazilian Portuguese vowels /i, e, ɛ, a, ɔ, o, u/ placed within a CVCV context; seven of these contexts were /fVfe/ for each of the vowels, comprising the MP set. The remaining seven words were /kɔko/, /kuke/, /pipe/, /popo/, /sase/, /sɛso/, /teko/, which formed non-MPs when paired with one another or with an /fVfe/ word. See **Table 1** for first (F1) and second (F2) vowel formants for AusE [originally reported by Elvin et al. (2016)] and Brazilian Portuguese [originally reported by Elvin et al. (2014)].

Similar to the Dutch stimuli, categorization of vowel contrasts as easy or difficult was based on patterns of expected acoustic-phonetic mapping of L2 vowels to the L1 phonological space. We used the categorization results from Elvin et al. (2014) to predict how participants would categorize the Brazilian Portuguese vowels. **Table 3** shows the easy and difficult minimal pair vowel contrasts between AusE and Brazilian Portuguese.

Procedure
The procedure was the same as that in Experiment 1.

Learning Phase
The learning phase was identical to the learning phase in Experiment 1 except that there were 84 learning trials (14 words presented 6 times each) rather than 72, due to there being 2 additional words in the Brazilian Portuguese set. Counterbalancing and timing were identical to that in Experiment 1.

Testing Phase
The testing phase consisted of 364 trials (14 words presented 26 times each as target, twice with every other word). In total, there were 24 pairs that were considered difficult minimal pairs, 60 pairs that were considered easy minimal pairs, and 280 pairs that were non-minimal pairs. The procedure was the same as in Experiment 1. See **Figure 2** for a representation of the learning and testing phase for Brazilian Portuguese stimuli.

TABLE 3 | Easy and difficult vowel mappings of Brazilian Portuguese vowels to AusE vowels.

Brazilian Portuguese

Easy		Difficult	
Contrast	Expected AusE mapping	Contrast	Expected AusE mapping
/i/-/ɛ/	/i/-/ɜ/	/i/-/e/	/ɪ/ or /iː/- /e/ or /eː/
/i/-/a/	/i/-/ɐː/	/e/-/ɛ/	/e/ or /eː/ - /ɜ/ or /e/
/i/-/o/	/i/-/ʊ/	/ɛ/-/a/	/ɜ/ or /e/ - /ɐː/
/i/-/ɔ/	/i/-/ɔ/	/a/-/ɔ/	/ɐː/ - /ɔ/ or /oː/
/i/-/u/	/i/-/ʉ/	/o/-/ɔ/	/ʊ/ or /oː/ - /ɔ/ or /oː/
/e/-/a/	/e/-/ɐ/	/o/-/u/	/ʊ/
/e/-/o/	/e/-/ʊ/		
/e/-/u/	/e/-/ʊ/		
/e/-/ɔ/	/e/-/ɔ/		
/ɛ/-/o/	/ɜ/-/ʊ/		
/ɛ/-/u/	/ɜ/-/ʊ/		
/ɛ/-/ɔ/	/ɜ/-/ɔ/		
/a/-/o/	/ɐː/-/ʊ/		
/a/-/u/	/ɐː/-/ʊ/		
/u/-/ɔ/	/ʊ/-/ɔ/		

RESULTS

We were interested in whether AusE participants could learn non-minimal pair and vowel minimal pair words in one of two unfamiliar languages (Brazilian Portuguese and Dutch) in an ambiguous word-learning paradigm, and whether the relationship between the vowel inventories in the L1 and target L2 influenced word learning. All analyses were conducted in R version 4.0.2 (R Core Team, 2020). Because very long reaction times were observed for a small number of trials, suggesting participants had taken breaks during these trials, prior to analysis, we applied a conservative approach to removing outlier test trials. Using the skewness function in the moments package (Komsta and Novomestky, 2015), we found reaction times for both the Brazilian Portuguese and Dutch experiments to be right-skewed, and so reaction times were log-transformed for the outliers analysis. We removed trials in which the reaction time was more than three times the median absolute deviation (Leys et al., 2013), applied to each experiment separately. This removed 1.28% of the trials in the Brazilian Portuguese language experiment, and 0.38% of the trials in the Dutch language experiment. Mean accuracy and reaction time values and standard deviations summarized across participants and pair types are available in S1.

Accuracy Analysis
Dutch
We fit a mixed-effect binomial logistic model to participants' correct and incorrect responses to test trials using the glmer function from the lme4 package (Bates et al., 2015). We first fit a maximal random effects structure that comprised random intercepts for participants and target words, and random

slopes for the within-subject effect of pair type, following recommendations from Meteyard and Davies (2020). The maximal random effects structure converged, at which point we added our fixed effect of pair type and used the anova function in R to calculate an F value for the three-level pair type factor. The effect of pair type was significant ($F_{[2,21]} = 44.02$, $p < 0.001$). Tukey-adjusted follow-up comparisons of the log odds ratio, calculated using the emmeans package (Lenth, 2020), showed that accuracy was higher for non-minimal pair (nonMP) trials than for easy minimal pair (easyMP; $\beta = 1.12$, SE = 0.18, $z = 6.20$, $p < 0.001$) and difficult minimal pair (diffMP; $\beta = 2.27$, SE = 0.26, $z = 8.88$, $p < 0.001$) trials. Further, accuracy was higher for easyMP trials than for diffMP trials ($\beta = 1.16$, SE = 0.24, $z = 4.90$, $p < 0.001$). Thus, performance was best for nonMP trials, followed by easyMP trials, with the worst performance in the diffMP trials. Nonetheless, one-tailed, one-sample t-tests against chance (0.5) conducted using the t.test function in R confirmed that performance was above chance for all pair types (nonMP: $M = 0.90$, $t_{[4,066]} = 87.98$, $p < 0.001$, lower 95% CI [0.90]; easyMP: $M = 0.76$, $t_{[635]} = 15.17$, $p < 0.001$, [0.73]; diffMP: $M = 0.57$, $t_{[556]} = 3.38$, $p < 0.001$, [0.54]). Overall accuracy for the Dutch stimuli was 85%. While we analyzed participants' individual responses to each test trial, for visualization, participants' proportion of correct responses across pair types in each experiment are presented in **Figure 3**.
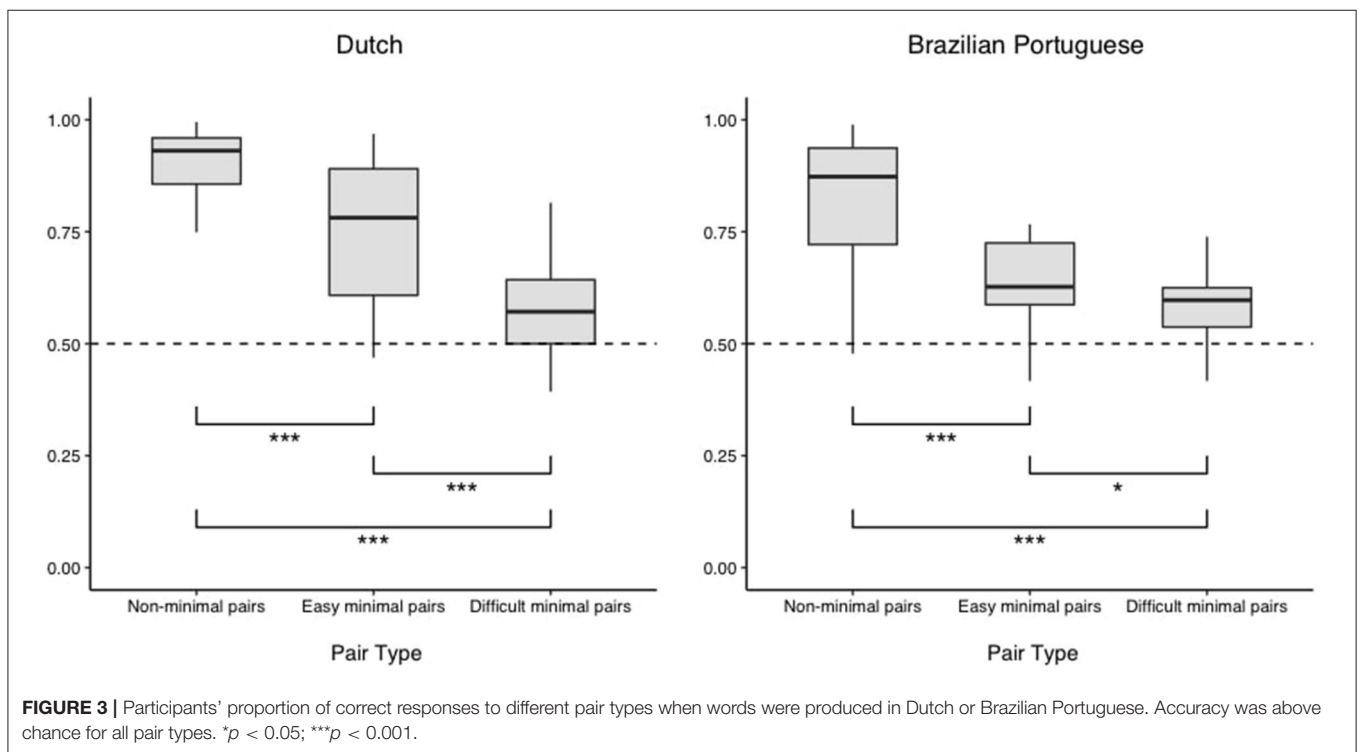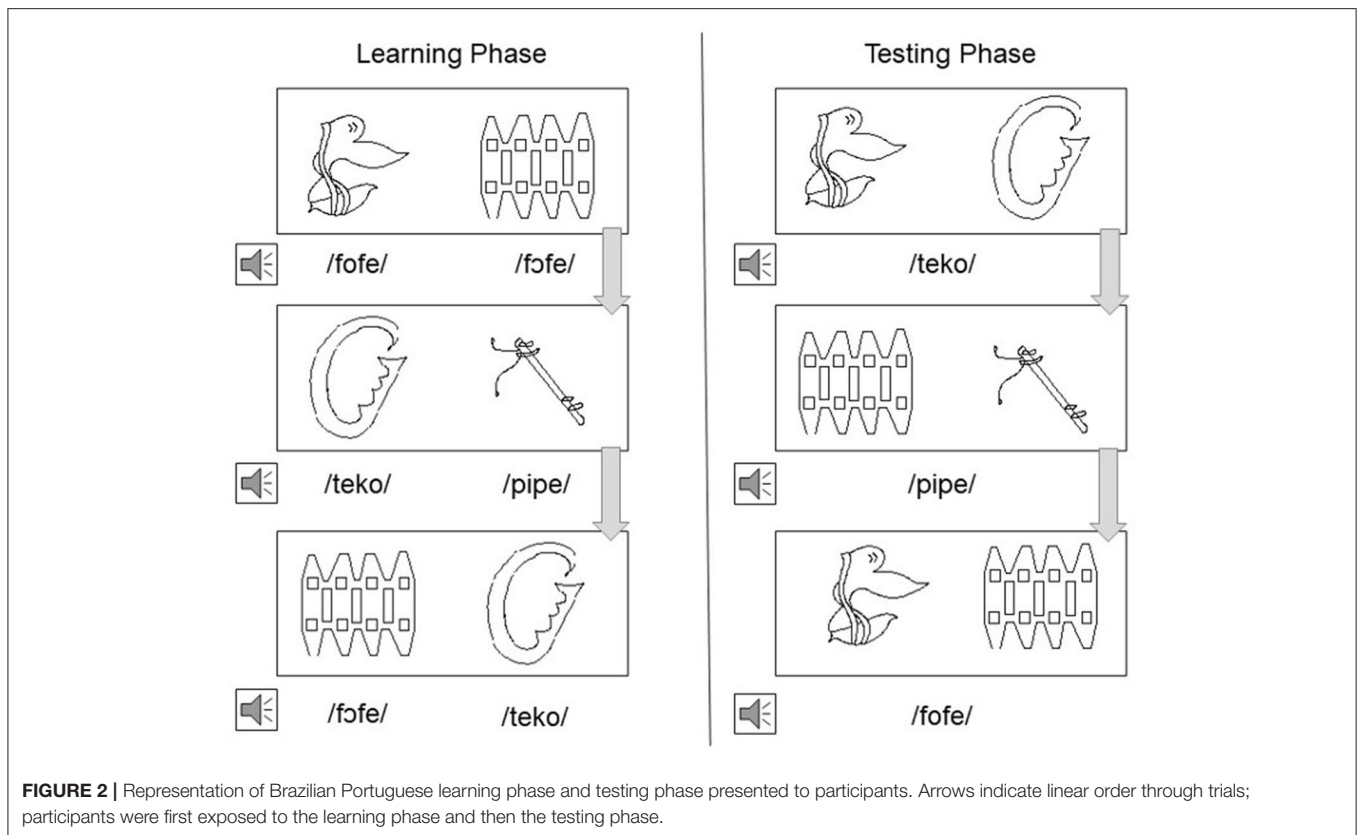
Brazilian Portuguese
We followed the same approach in analyzing the Brazilian Portuguese experiment. The final random effects structure included random intercepts for participants and target words. As in the Dutch experiment, there was a main effect of pair type ($F_{[2,4,244]} = 78.50$, $p < 0.001$), with follow-up comparisons showing the same pattern of better performance for nonMP than easyMP ($\beta = 0.87$, SE = 0.09, $z = 9.87$, $p < 0.001$) and diffMP ($\beta = 1.19$, SE = 0.12, $z = 10.23$, $p < 0.001$) trials, and better performance for easyMP than diffMP trials ($\beta = 0.33$, SE = 0.12, $z = 2.73$, $p = 0.018$). Accuracy for all pair types was again above chance (nonMP: $M = 0.82$, $t_{[5,547]} = 62.61$, $p < 0.001$, lower 95% CI [0.81]; easyMP: $M = 0.65$, $t_{[1,171]} = 10.71$, $p < 0.001$, [0.65]; diffMP: $M = 0.58$, $t_{[466]} = 3.42$, $p < 0.001$, [0.54]). Overall accuracy for the Brazilian Portuguese stimuli was 78%.

Reaction Time Analysis
We next analyzed participants' reaction times for correct responses in a mixed-effects linear model using the lmer function from the lme4 package (Bates et al., 2015), using the same approach as in the accuracy analysis. Because reaction times were right-skewed, they were log-transformed as well as scaled and centered prior to analysis. For visualization, participants' reaction times in ms as a function of minimal pair type are presented in **Figure 4**.

Dutch
We first fit a maximal random effects structure with random intercepts for participants and target words, and a random slope for the effect of pair type by participant. The maximal structure
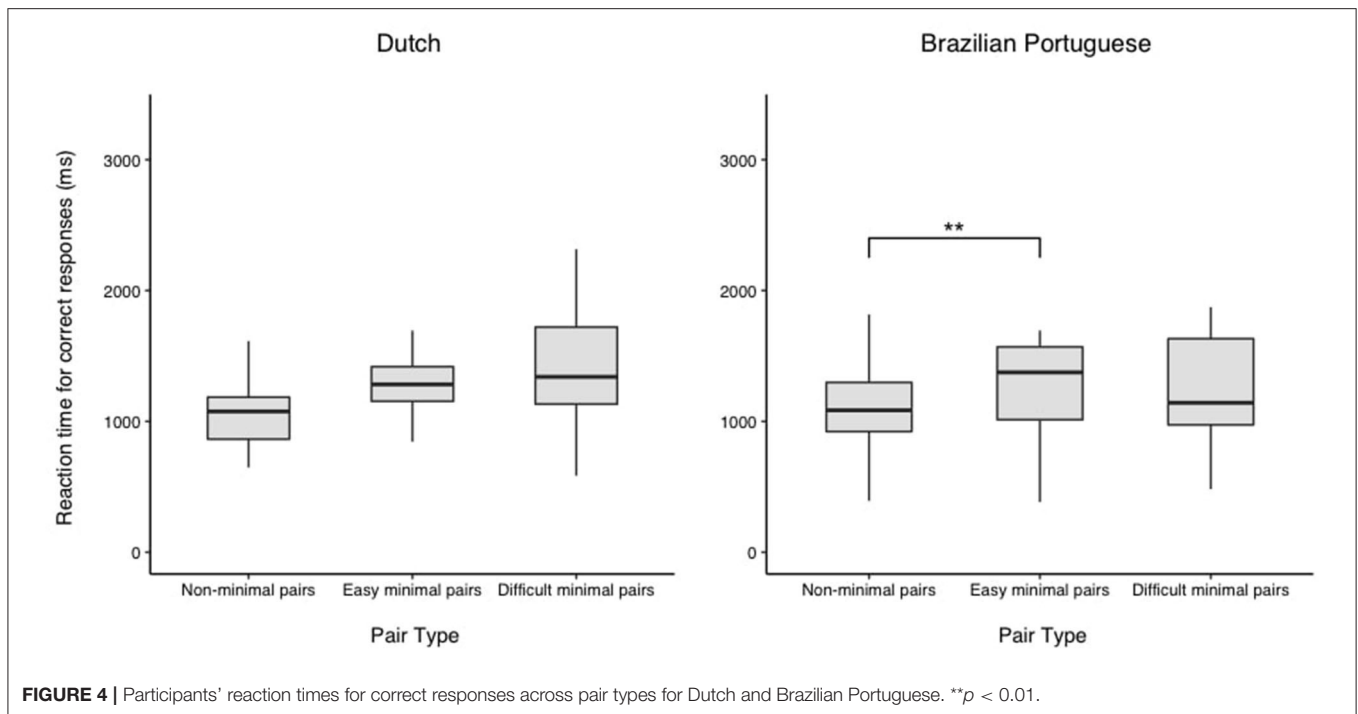
FIGURE 2 | Representation of Brazilian Portuguese learning phase and testing phase presented to participants. Arrows indicate linear order through trials; participants were first exposed to the learning phase and then the testing phase.

FIGURE 3 | Participants' proportion of correct responses to different pair types when words were produced in Dutch or Brazilian Portuguese. Accuracy was above chance for all pair types. $*p < 0.05$; $***p < 0.001$.

FIGURE 4 | Participants' reaction times for correct responses across pair types for Dutch and Brazilian Portuguese. **$p < 0.01$.

converged, and so we added our main effect of pair type. The effect of pair type was marginal ($F_{[2, 21]} = 3.41$, $p = 0.052$).

Brazilian Portuguese

The maximal random effects structure that would converge included random intercepts for participants and target words. The effect of pair type was significant ($F_{[2, 5,542]} = 4.90$, $p = 0.007$), with Tukey-adjusted follow-up pairwise comparisons calculated from emmeans showing faster reaction times for nonMP trials compared to easyMP trials ($\beta = -0.10$, SE = 0.03, $z = -3.05$, $p = 0.007$). There was no evidence that reaction time differed between nonMP and diffMP trials ($z = -1.34$, $p = 0.373$) or between easyMP and diffMP trials ($z = 0.65$, $p = 0.792$).

DISCUSSION

We examined if adults could learn foreign language words in an ambiguous word-learning paradigm. While previous research demonstrates that adults can learn words in their native language via CSWL, we were interested in whether CSWL extends to foreign language word learning, particularly in situations in which the foreign words are predicted to be perceptually difficult for the listener to discriminate, since this would be expected to impact learners' ability to accurately track word-object co-occurences across trials. To that end, we taught native L1 AusE listeners novel word-object pairings in one of two unfamiliar languages (Experiment 1: Dutch; Experiment 2: Brazilian Portuguese) and tested their word learning in non-minimal pair and minimal vowel pair contexts. Minimal pairs were further characterized as being perceptually easy or difficult

for native AusE listeners to discriminate based on expected acoustic mappings between the L1 and each L2.

In line with our predictions, participants learned new foreign language words in both languages, associating the learned pseudowords with the correct image above chance level across both minimal pair types and non-minimal pairs. Participants' performance followed the same pattern in both languages: Accuracy was highest when the label for the distractor image formed a non-minimal pair with the target word, and lowest (though still above chance) when the distractor label formed a minimal pair predicted to be perceptually difficult for the learner. Participants were also faster to respond when tested with non-minimal pairs compared to perceptually easy minimal pairs in the Brazilian Portuguese experiment, likely reflecting that while non-minimal pairs could be differentiated based on the first segment, minimal pairs were not differentiated until the second segment. These results mimic previous work showing that minimal pair words are not learned as accurately as non-minimal pair words via CSWL even when stimuli conformed to the L1 phonotactics of the learners (Escudero et al., 2016; Mulak et al., 2019).

Thus, these results tell us that CSWL is a viable L2 word learning strategy, even in cases where the L2 words contained contrasts that are perceptually confusable for the listener. Previous work has demonstrated that adults can learn new word-image associations in CSWL paradigms when words are produced by a speaker of their native language, including minimal pair words that differ either in a single consonant or vowel (Escudero et al., 2016). Escudero et al. (2016) nonetheless showed accuracy was lowest for vowel minimal pairs, perhaps reflecting the tendency of listeners to perceive vowels less categorically than consonants (Fry et al., 1962; Liberman

et al., 1967; Beddor and Strange, 1982). This baseline difficulty would be expected to be further exacerbated by the perceptual difficulties incurred when foreign vowels do not have a direct correspondence with native vowels. Because CSWL is supported by reliably tracking the co-occurrences of auditory labels in the context of candidate referents, an inability or difficulty in perceptually distinguishing between two or more words would greatly impact the ability to derive one-to-one mappings between those words and their candidate referents, potentially to the point of preventing word learning. However, while we did find that word learning for perceptually difficult vowel minimal pairs was lower than our perceptually easy pairs, learning of these difficult vowel minimal pairs was nonetheless above chance. Thus, although L2 perception comes with additional perceptual challenges compared to L1 perception, L2 word learning can still be achieved via CSWL in the face of these challenges.

Our results also support the tenets of the L2LP model that the perceptual ease or difficulty in perceiving vowel contrasts in the L2 is influenced by the acoustic relationship between L2 sounds and L1 vowel categories and that difficulties in perceiving individual contrasts will translate to difficulties in learning words that differ by those contrasts (Escudero, 2005; Escudero and Chládková, 2010; van Leussen and Escudero, 2015). Following L2LP, we classified vowel minimal pairs as easy or difficult based on the relationship between the F1 and F2 acoustic measurements of the L2 vowels differentiating the minimal pair and the F1 and F2 characteristics of the listeners' native AusE vowels. It was predicted that perceptually easy vowel contrasts in both Dutch and Brazilian Portuguese would be mapped by AusE listeners to two or more native L1 categories that do not overlap, therefore supporting reliable discrimination. For difficult minimal pairs, both vowels in the contrast would be expected to be perceived as either belonging to a single L1 category or each vowel potentially being categorized as two or more native categories, making discrimination more challenging. L2LP classifies the former as a "new scenario" and the latter as a "subset scenario" (Escudero, 2005) and evidence for it has been shown with AusE and Peruvian Spanish listeners discriminating Dutch sounds (Alispahic et al., 2017) and Iberian Spanish and AusE listeners discriminating Brazilian Portuguese (Elvin et al., 2014). Indeed, performance was worse for perceptually difficult than perceptually easy vowel minimal pairs. Because patterns across both languages were similar as well, this extends previous results demonstrating an influence of perceptual difficulty on explicit word learning (Escudero et al., 2013, 2014; Escudero, 2015; Elvin et al., 2020) to L2 word learning in an ambiguous learning scenario via CSWL.

While this study did not directly compare word learning in an explicit, one-to-one mapping scenario to word learning in an ambiguous scenario where one-to-one correspondences are not available, there is evidence that both learning scenarios lead to similar results with these stimuli. AusE listeners learned the same Dutch (Escudero, 2015) and Brazilian Portuguese (Elvin et al., 2020) pseudowords when taught in an explicit mappings word learning task, demonstrating comparable patterns as observed here. When presented with explicit mappings, Escudero (2015) found that AusE learners were more accurate at identifying the

image corresponding to a novel Dutch word when target words were in a non-minimal pair context compared to a minimal pair context, and that learners were more accurate for minimal pairs that were perceptually easy compared to perceptually difficult. Similarly, Elvin et al. (2020) found that AusE participants learning Brazilian Portuguese words in a similar explicit-mappings paradigm were most accurate at identifying non-minimal pairs, followed by perceptually easy minimal pairs, and showed the worst performance with perceptually difficult minimal pairs. In both studies, like in the current study, word learning was nonetheless above chance for each pair type. This suggests that L2 word learning in ambiguous scenarios may be as effective a learning strategy as when explicit word-object mappings are provided, even for perceptually difficult stimuli that present challenges for tracking occurrences of individual words (and the referents that co-occur with them) in CSWL and highlights a direct comparison of the two learning strategies as an area for future research. Furthermore, CSWL is proposed to be supported both by automatic statistical tracking mechanisms, as well as top-down hypothesis-testing strategies (e.g., see Yu and Smith, 2012). While our experiment was not designed with the intention of exploring these different mechanisms, determining the efficacy of each strategy with regard to L2 word learning is an area for future research.

Overall, our results add to the body of work showing viable word learning in a CSWL paradigm by monolinguals, and we extend this to the learning of words in a foreign language. Because monolinguals only have single word-to-concept mappings before they approach learning a new language, it may be easier for them to build new word-to-concept mappings via CSWL. Thus far, some studies using learning paradigms that provide explicit word-object pairings have shown that bilinguals outperform monolinguals when learning novel words that do not conform to the phonotactics of the participants' languages (Kaushanskaya and Marian, 2009) and when learning novel words that share phonemes across monolinguals' and bilinguals' languages (Kaushanskaya, 2012). Some studies using learning paradigms such as CSWL in which word-object pairings are ambiguous and must be derived have also shown bilinguals to outperform monolinguals when the novel language conformed to the phonotactics of the monolinguals' language and one of the bilinguals' languages (e.g., Benitez et al., 2016; Escudero et al., 2016; Poepsel and Weiss, 2016). Therefore, future work should examine how specifically the acoustic-phonetic relationship between the languages of multiple-language speakers (i.e., bilinguals and multilinguals) and the stimulus language affects novel word learning. For example, in a CSWL paradigm, Poepsel and Weiss (2016) showed that Mandarin Chinese-English bilinguals showed better learning of monosyllabic compared to bi- and trisyllabic novel word forms, which may be due to the fact that single syllables in Mandarin are considered the phonological encoding unit of the language, as opposed to segmental information in languages such as English or Dutch (O'Seaghdha et al., 2010). This suggests that the native phonological and lexical structure of the L1 affected learning in an L2. The authors also suggest that examining how native phonology affects novel

word learning would be useful to understand the performance difference between groups. Extending the L1-L2 relationship to, for example, L1-L3/L2-L3 relationships would clarify the impact of the acoustic-phonetic relationship between multiple languages and how it interacts with word-learning paradigms, such as CSWL.

In summary, although it is well-established that adults can learn new words in their native language in an ambiguous learning paradigm such as CSWL (e.g., Yu and Smith, 2007; Escudero et al., 2016), here we show that CSWL also extends to second language word learning. Native AusE speakers successfully learned novel Dutch and Brazilian Portuguese words. These results provide evidence that L2 learning can occur in the absence of presentation of direct one-to-one mappings, but is modulated by the L1-L2 acoustic-phonetic relationship, as has been previously found in studies of L2 word learning in explicit mapping scenarios. To understand the efficacy of L2 word learning in ambiguous learning scenarios by adults and in more naturalistic settings, future work should examine how this relationship affects learning with other stimuli in a CSWL paradigm, such as consonant minimal pairs and with real L2 word-concept pairings.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Human Ethics Committee at Western Sydney University (H9373). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcomm. 2020.602471/full#supplementary-material

REFERENCES

Adank, P., van Hout, R., and Smits, R. (2004). An acoustic description of the vowels of Northern and Southern standard dutch. *J. Acoust. Soc. Am.* 116, 1729–1738. doi: 10.1121/1.1779271

Alispahic, S., Mulak, K. E., and Escudero, P. (2017). Acoustic properties predict perception of unfamiliar dutch vowels by adult Australian English and peruvian Spanish listeners. *Front. Psychol.* 8:52. doi: 10.3389/fpsyg.2017.00052

Antoniou, M., Liang, E., Ettlinger, M., and Wong, P. C. M. (2015). The bilingual advantage in phonetic learning. *Bilingualism* 18, 683–695. doi: 10.1017/S1366728914000777

Aoyama, K., Flege, J. E., Guion, S. G., Akahane-Yamada, R., and Yamada, T. (2004). Perceived phonetic dissimilarity and L2 speech learning: the case of Japanese /r/ and English /l/ and /r/. *J. Phon.* 32, 233–250. doi: 10.1016/S0095-4470(03)00036-6

Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

Beddor, P. S., and Strange, W. (1982). Cross-language study of perception of the oral-nasal distinction. *J. Acoust. Soc. Am.* 71, 1551–1561. doi: 10.1121/1.387809

Benitez, V. L., Yurovsky, D., and Smith, L. (2016). Competition between multiple words for a referent in cross-situational word learning. *J. Mem. Lang.* 90, 31–48. doi: 10.1016/j.jml.2016.03.004

Berens, S., Horst, J., and Bird, C. (2018). Cross-situational learning is supported by propose-but-verify hypothesis testing. *Curr. Biol.* 28, 1132–1136. doi: 10.1016/j.cub.2018.02.042

Best, C. T., and Tyler, M. D. (2007). "Nonnative and second-language speech perception: Commonalities and complementarities," in *Language Learning & Language Teaching*, Vol. 17, eds O.-S. Bohn and M. J. Munro (Amsterdam: John Benjamins Publishing Company). doi: 10.1075/lllt.17.07bes

deKeyser, R. (2000). The robustness of critical period effects in second language acquisition. *Stud. Second Lang. Acquisition* 22, 499–533. doi: 10.1017/S0272263100004022

Ellis, N. C. (2015). "Implicit AND explicit language learning: their dynamic interface and complexity," in *Implicit and Explicit Learning of Languages*, ed P. Rebuschat (Amsterdam: John Benjamins Publishing Company), 1–24. doi: 10.1075/sibil.48.01ell

Elvin, J., Escudero, P., and Vasiliev, P. (2014). Spanish is better than English for discriminating portuguese vowels: acoustic similarity versus vowel inventory size. *Front. Psychol. Lang. Sci.* 5:1188. doi: 10.3389/fpsyg.2014. 01188

Elvin, J., Williams, D., and Escudero, P. (2016). Dynamic acoustic properties of monophthongs and diphthongs in Western Sydney Australian English. *J. Acoust. Soc. Am.* 140, 576–581. doi: 10.1121/1.4952387

Elvin, J., Williams, D., and Escudero, P. (2020). "Learning to perceive, produce and recognise words in a non-native language," in *Linguistic Approaches to Portuguese as an Additional Language*, eds K. V. Molsing, C. B. L. Perna, and A. M. T. Ibaños (Amsterdam: John Benjamins Publishing Company). doi: 10.1075/ihll.24.03elv

Escudero, P. (2005). *Linguistic Perception and Second Language Acquisition.* Utrecht: Utrecht University.

Escudero, P. (2009). "Linguistic perception of "similar" L2 sounds," in *Phonology in Perception*, eds P. Boersma and S. Hamann (Berlin: Mouton de Gruyter), 151–190.

Escudero, P. (2015). Orthography plays a limited role when learning the phonological forms of new words: the case of Spanish and English learners of novel Dutch words. *Appl. Psycholinguist* 36, 7–22. doi: 10.1017/S014271641400040X

Escudero, P., Boersma, P., Rauber, A. S., and Bion, R. A. H. (2009). A cross-dialect acoustic description of vowels: Brazilian and European portuguese. *J. Acoust. Soc. Am.* 126, 1379–1393. doi: 10.1121/1.3180321

Escudero, P., Broersma, M., and Simon, E. (2013). Learning words in a third language: effects of vowel inventory and language proficiency. *Lang. Cogn. Process* 28, 746–761. doi: 10.1080/01690965.2012.662279

Escudero, P., and Chládková, K. (2010). Spanish listeners' perception of American and Southern British English vowels. *J. Acoust. Soc. Am.* 128, EL254–EL260. doi: 10.1121/1.3488794

Escudero, P., Mulak, K. E., and Vlach, H. A. (2016). Cross-situational learning of minimal word pairs. *Cogn. Sci.* 40, 455–465. doi: 10.1111/cogs.12243

Escudero, P., Simon, E., and Mulak, K. E. (2014). Learning words in a new language: orthography doesn't always help. *Bilingualism* 17, 384–395. doi: 10.1017/S1366728913000436

Fitneva, S. A., and Christiansen, M. H. (2011). Looking in the wrong direction correlates with more accurate word learning. *Cogn. Sci.* 35, 367–380. doi: 10.1111/j.1551-6709.2010.01156.x

Flege, J. E. (1995). "Second-language speech learning: theory, findings, and problems," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research,* ed W. Strange (Baltimore, MD: York Press), 229–273.

Fry, D. B., Abramson, A. S., Eimas, P. D., and Liberman, A. M. (1962). The identification and discrimination of synthetic vowels. *Lang. Speech.* 5, 171–189. doi: 10.1177/002383096200500401

Hartshorne, J. K., Tenenbaum, J. B., and Pinker, S. (2018). A critical period for second language acquisition: evidence from 2/3 million English speakers. *Cognition* 177, 263–277. doi: 10.1016/j.cognition.2018.04.007

Hu, C.-F. (2017). Resolving referential ambiguity across ambiguous situations in young foreign language learners. *Appl. Psycholinguist.* 38, 633–656. doi: 10.1017/S0142716416000357

Jared, D., and Kroll, J. F. (2001). Do bilinguals activate phonological representations in one or both of their languages when naming words? *J. Mem. Lang.* 44, 2–31. doi: 10.1006/jmla.2000.2747

Johnson, J. S., and Newport, E. L. (1989). Critical period effects in second language learning: the influence of maturational state on the acquisition of English as a second language. *Cogn. Psychol.* 21, 60–99. doi: 10.1016/0010-0285(89)90003-0

Junttila, K., and Ylinen, S. (2020). Intentional training with speech production supports children's learning the meanings of foreign words: a comparison of four learning tasks. *Front. Psychol.* 11:1108. doi: 10.3389/fpsyg.2020.01108

Kaushanskaya, M. (2012). Cognitive mechanisms of word learning in bilingual and monolingual adults: the role of phonological memory. *Bilingualism Lang. Cogn.* 15, 470–489. doi: 10.1017/S1366728911000472

Kaushanskaya, M., and Marian, V. (2009). The bilingual advantage in novel word learning. *Psychonomic Bull. Rev.* 16, 705–710. doi: 10.3758/PBR.16.4.705

Komsta, L., and Novomestky, F. (2015). *Moments: Moments, Cumulants, Skewness, Kurtosis and Related Tests* (R package version 0.14) [Computer software]. Available online at: https://CRAN.R-project.org/package=moments (accessed October 15, 2020).

Krashen, S. D., Long, M. A., and Scarcella, R. C. (1979). Age, rate and eventual attainment in second language acquisition. *TESOL Q.* 13, 573–582. doi: 10.2307/3586451

Kriengwatana, B., Terry, J., Chládková, K., and Escudero, P. (2016). Speaker and accent variation are handled differently: evidence in native and non-native listeners. *PLoS ONE* 11:e0156870. doi: 10.1371/journal.pone.0156870

Kroll, J. F., and Sunderman, G. (2003). "Cognitive processes in second language learners and bilinguals: the development of lexical and conceptual representations," in *The Handbook of Second Language Acquisition* (Oxford: Blackwell Publishing), 104–129. doi: 10.1002/9780470756492.ch5

Lenth, R. (2020). *Emmeans: Estimated Marginal Means, aka Least-Squares Means* (R package version 1.4.8) [Computer software]. Available online at: https://CRAN.R-project.org/package=emmeans (accessed October 15, 2020).

Leys, C., Ley, C., Bernard, P., and Licata, L. (2013). Detecting outliers: do not use standard deviation around the mean, use absolute deviation around the median. *J. Exp. Soc. Psychol.* 49, 764–766. doi: 10.1016/j.jesp.2013.03.013

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychol. Rev.* 74, 431–461. doi: 10.1037/h0020279

McCandliss, B. D., Fiez, J. A., Protopapas, A., Conway, M., and McClelland, J. L. (2002). Success and failure in teaching the [r]-[l] contrast to Japanese adults: tests of a hebbian model of plasticity and stabilization in spoken language perception. *Cogn. Affect. Behav. Neurosci.* 2, 89–108. doi: 10.3758/CABN.2.2.89

Meteyard, L., and Davies, R. A. I. (2020). Best practice guidance for linear mixed-effects models in psychological science. *J. Mem. Lang.* 112:104092. doi: 10.1016/j.jml.2020.104092

Mulak, K., Vlach, H., and Escudero, P. (2019). Cross-situational learning of phonologically overlapping words across degrees of ambiguity. *Cogn. Sci.* 43:e12731. doi: 10.1111/cogs.12731

O'Seaghdha, P., Chen, J.-Y., and Chen, T.-M. (2010). Proximate units in word production: Phonological encoding begins with syllables in Mandarin Chinese but with segments in English. *Cognition* 115, 282–302. doi: 10.1016/j.cognition.2010.01.001

Oyama, S. (1976). A sensitive period for the acquisition of a nonnative phonological system. *J. Psycholinguist. Res.* 5, 261–283. doi: 10.1007/BF01067377

Piske, T., MacKay, I. R. A., and Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: a review. *J. Phon.* 29, 191–215. doi: 10.1006/jpho.2001.0134

Poepsel, T. J., and Weiss, D. J. (2016). The influence of bilingualism on statistical word learning. *Cognition* 152, 9–19. doi: 10.1016/j.cognition.2016.03.001

R Core Team (2020). *R: A Language and Environment for Statistical Computing.* Vienna: R Foundation for Statistical Computing. Available online at: https://www.R-project.org/

Seliger, H. W., Krashen, S. D., and Ladefoged, P. (1975). Maturational constraints in the acquisition of second language accent. *Lang. Sci.* 36, 20–22.

Shatzman, K. B., and McQueen, J. M. (2006). Prosodic knowledge affects the recognition of newly acquired words. *Psychol. Sci.* 17, 372–377. doi: 10.1111/j.1467-9280.2006.01714.x

Smith, L. B. (2000). "How to learn words: an associative crane," in *Breaking the Word Learning Barrier,* eds R. M. Golinkoff, and K. Hirsh-Pasek (Oxford: Oxford University Press), 51–80.

Spada, N. (1997). Form-focussed instruction and second language acquisition: a review of classroom and laboratory research. *Lang. Teach.* 30, 73–87. doi: 10.1017/S0261444800012799

Tahta, S., Wood, M., and Loewenthal, K. (1981). Foreign accents: factors relating to transfer of accent from the first language to a second language. *Lang. Speech.* 24, 265–272. doi: 10.1177/002383098102400306

Trueswell, J., Medina, T., Hafri, A., and Gleitman, L. (2013). Propose but verify: Fast mapping meets cross-situational word learning. *Cogn. Psychol.* 66, 126–156. doi: 10.1016/j.cogpsych.2012.10.001

van Leussen, J.-W., and Escudero, P. (2015). Learning to perceive and recognize a second language: The L2LP model revised. *Front. Psychol.* 6:1000. doi: 10.3389/fpsyg.2015.01000

Vlach, H. A., and Sandhofer, C. M. (2014). Retrieval dynamics and retention in cross-situational statistical word learning. *Cogn. Sci.* 38, 757–774. doi: 10.1111/cogs.12092

Yu, C., and Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychol. Sci.* 18, 414–420. doi: 10.1111/j.1467-9280.2007.01915.x

Yu, C., and Smith, L. B. (2012). Modeling cross-situational word–referent learning: prior questions. *Psychol. Rev.* 119, 21–39. doi: 10.1037/a0026182