Laboratory Phonology
Journal of the Association for Laboratory Phonology

Zhang, S., Babel, M., Hammerly, C., & Tessier, A.-M. (2026). The online processing of phonological contrasts in an L2: Mapping L1 allophones to L2 phonemes. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 17(1), pp. 1–46. DOI: https://doi.org/10.16995/labphon.18470

Open Library of Humanities

The online processing of phonological contrasts in an L2: Mapping L1 allophones to L2 phonemes

Sijia Zhang*, Department of Linguistics, University of British Columbia, Canada, sijia01@student.ubc.ca

Molly Babel, Department of Linguistics, University of British Columbia, Canada

Christopher Hammerly, Department of Linguistics, University of British Columbia, Canada

Anne-Michelle Tessier, Department of Linguistics, University of British Columbia, Canada

*Corresponding author.

First language (L1) phonological knowledge may influence second language (L2) processing at different levels. The current study disentangles L2 low-level perception and a higher level of phonological encoding in the lexicon. L1-Mandarin L2-English bilinguals listened to English low vowel + nasal (loVN) contrasts in which the vowels pattern as allophones in Mandarin. The AX discrimination task revealed that sequential Mandarin-English bilinguals accurately perceived the sequences. A spoken word recognition task in a Visual World Paradigm, however, showed that in response to words with loVN sequences Mandarin-English bilinguals experienced more lexical competition than English L1 listeners in a relatively later time window. Taken together, this suggests that the influence of Mandarin phonology affects phono-lexical encoding to a greater extent than lower-level phonetic encoding. Moreover, the symmetric L2 competition between the L1-licit and L1-illicit loVN contexts suggests that L2 listeners are able to repurpose L1 allophones to phonemes in the L2. In addition, across language backgrounds, a larger receptive vocabulary size facilitates spoken word recognition. The phonological knowledge generalized across larger lexicons resolves competition more quickly for both bilingual and monolingual listeners. Overall, the study suggests that allophones, rather than phonemes, are mappable units in phono-lexical encoding by sequential bilinguals.

1. Introduction

Second language (L2) speech perception can be influenced by first language (L1) sound systems (for summaries and reviews see Best & Tyler, 2007; Flege & Bohn, 2021; Chang, 2019). The influence from an L1 can emerge at different levels of representation and processing. For instance, the L1 sound system can influence L2 low-level phonetic discrimination (e.g., Werker et al., 1981; Johnson & Babel, 2010), as well as higher-level phonological encoding in the lexicon (e.g., Weber & Cutler, 2004; Cutler et al., 2006; Turner, 2022; Amengual, 2016). For lexical items that have non-L1 sound contrasts, L2 learners make errors that L1 listeners do not (Broersma, 2012; Darcy et al., 2013; Gor, 2018), yet it is unclear whether the errors arise from inaccurate phonetic processing or higher-level encoding, or both (for a recent review, see Darcy et al., 2024). Although some studies suggest a connection between L2 phonetic discrimination and phonological categorization (e.g., Boomershine et al., 2008), others find a discrepancy between the processing of phonetic and lexical information, such that L2 listeners demonstrate accurate low-level processing yet poor lexical decision performance (e.g., Darcy et al., 2013; Amengual, 2016). Several eye-tracking studies on L2 lexical competition have shown that L2 learners can encode distinctions in the lexicon without forming target-like phonological contrasts (Weber & Cutler, 2004; Cutler et al., 2006). For example, in online processing, Turner (2022) revealed that L2 listeners were able to leverage acoustic details in an early time window, and they experienced phono-lexical competition at a later time.

The current study approaches these questions about L1 influences in L2 phono-lexical processing from a novel direction. By querying how Mandarin L1 listeners perceive and process low vowel + nasal sequences in their English L2, we examine how language users repurpose L1 allophones into an L2 phonemic contrast. More specifically, whereas English phonology permits the low vowels /æ/ and /ɑ/ in combination with coda /m/, /n/ and /ŋ/, Mandarin phonology only has [æn] and [ɑŋ]. This phonotactic pattern is analyzed within Mandarin phonology as an allophonic patterning of the low vowels conditioned by the place of articulation of the coda nasals. To preview our results, we observe high accuracy for the Mandarin listeners' perception of the English contrasts in an AX discrimination task, suggesting listeners have accurate acoustic-level perception of the sound sequences. The results of an eye-tracking experiment reveal that compared to English L1 listeners, there is more lexical competition for Mandarin-English bilinguals with low vowel + nasal sequences at a later time window, suggesting L2 differences at a later and more phono-lexical level of encoding. There are no differences for Mandarin-English bilinguals, however, with respect to the particular English low vowel + nasal sequences–that is, whether the English sequences are present or absent in Mandarin, which we interpret as the effective mapping of the L1 allophones [æ] and [ɑ] to the phonemic-level contrast in their L2 (German et al., 2013). We also demonstrate that receptive vocabulary size of L1 and L2 listeners contributes to the quality of phono-lexical processing and lexical competition.

1.1. L1 influences at different levels of L2 processing

L1 sound systems can influence listeners' processing of L2 speech sounds at different levels. At a lower phonetic level of perception, L2 listeners might experience difficulties distinguishing sounds that are not contrastive in their L1. For instance, Japanese listeners were reported to have difficulty discriminating between English /l/ and /ɹ/ (Goto, 1971), for which the latter does not exist in the Japanese phoneme inventory. Similarly, L1 American English listeners find it challenging to discriminate between Hindi dental and retroflex stops, as these contrasts are not phonologically differentiated in their L1 (Werker et al., 1981). In addition to individual phonemes, low-level perceptual difficulties also apply to unfamiliar sound sequences. For instance, listeners experienced perceptual illusions when discriminating between a phonotactically unattested consonant cluster and a licit alternative. For example, Japanese listeners treat sequences like [ebzo] and [ebuzo] as perceptually equivalent (Dupoux et al, 2011; Berent et al., 2007). Darcy and Thomas (2019) demonstrate that cross-language phonotactic repairs are sensitive to exact pronunciations. They observe that Korean learners of English accept insertions of /ʊ/, but not /ɪ/ as an epenthetic vowel, labelling items like [bʊˈluː] as a viable production of "blue," but not [bɪˈluː]. This suggests some degree of precision in the phonolexical representations of second language learners. Thus, speech perception can be warped by the L1 sound system not only at low-level phonetic processing but also at the phonological level.

L2 listeners develop a different perceptual system compared to L1 listeners when they apply the different L1 phonological relationships between sounds (Boomershine et al., 2008; Huang & Johnson, 2010; Johnson & Babel, 2010) to their L2. Listeners tend to perceive sounds in allophonic alternations as more similar to each other than those that are phonemically contrastive. Using a perceptual similarity rating task that is argued to tap phonological-level processing, Boomershine et al. (2008) found that L1 Spanish listeners perceived [ð] and [d] as more similar to each other than did L1 English listeners, as [ð] and [d] are allophones in Spanish but are distinct phonemes in English. Conversely, the pair [d] and [ɾ], which are in an allophonic relation in English but contrastive in Spanish, are perceived as more similar by L1 English listeners compared to L1 Spanish listeners. Huang and Johnson (2010) found a similar L1 effect in the perception of tonal allophony by L1 Mandarin compared to L1 American English listeners in a similarity rating task. Moreover, although Boomershine et al. and Huang and Johnson observed that the L1 sound system (i.e., language-specific effects) warps both low-level discrimination and high-level phonological processing, Johnson and Babel (2010) found no low-level warping by the L1. Babel and Johnson (2010) suggest that the degree of language influence on phonetic tasks may be subject to the speed of the response; faster responses may reflect judgments that are more influenced by the psycho-acoustic similarity between pairs of phones, whereas slower discrimination responses may allow for more language-specific similarity to affect a listener's decision.

In another study that looked at the creation of a new phoneme in L2 acquisition, Darcy et al. (2013) examined the processing of Japanese geminate and singleton contrasts and German front rounded and back rounded vowel contrasts by L1 American English speakers. While the English L1 listeners show sensitivity to the contrasts in the phonetic categorization task, the lexical level of encoding is not target-like. Darcy et al. (2013) attribute this to English L1 listeners making reference to English L1 categories, which do not contrast consonant length or the front/back dimension on rounded vowels, making the Japanese and German sounds poor exemplars of the familiar L1 phoneme (Darcy et al., 2013). The accuracy at the phonetic or phonological level of processing does not necessarily transfer to the target-like encoding of phonological contrast at the word level.

Altogether, it is unclear exactly how and when L1 influences emerge at the phonetic and phonological levels, especially when testing sound patterns that involve different phonological relations (allophones vs. phonemes) across languages. Barrios and Hayes-Harb (2021) attempted to tease apart where difficulties in L2 listening and processing reside using a lexical decision task with English word and nonword materials and L1 English, Mandarin, and Korean participants, where the L1 Mandarin and Korean participants were L2 learners of English. Barrios and Hayes-Harb reasoned that the pattern of errors for lexical decisions will vary for words and nonwords based on whether the dimension of difference was a category that was present in the L1 (what they call a dominant category) or not (a non-dominant category, in their usage) and whether the difficulty is in the perceptual coding or matching to a phonolexical representation. Using English /æ/-/ɛ/ and /l/-/ɹ/ as the contrasts, they demonstrate that when using the same materials, Mandarin learners of English exhibit error patterns that align with difficulties in perceptual coding, and Korean learners of English exhibit behaviors concordant with difficulties in phonolexical processing. Interestingly, they demonstrate that all listener groups–including the L1 English listeners–show an asymmetry in processing /æ/ and /ɛ/, whereby listeners have higher accuracy for [æ] nonwords than [ɛ] ones. This indicates that listeners accept /æ/ pronunciations that are [ɛ]-like more than /ɛ/ pronunciations that are [æ]-like. That the English L1 listeners showed this asymmetry is important because it demonstrates that all asymmetries are not due to a lack of phonological contrast, but may be due to phonetic and phonological variation experienced by the speech community more generally (Nesbitt, 2023).

1.2. L2 phonological representation and mapping: models and hypotheses

Several different models and hypotheses have been proposed concerning the development of L2 representations at the phonological and/or lexical level, with allophones as opposed to phonemes as the level of unit in L1-to-L2 sound mapping (see Chang, 2019, for a review of bilingual phonetic and phonological learning models). The revised Speech Learning Model (SLM-r), for instance, posits that L2 learners start by automatically linking non-native sounds to existing

L1 categories (Flege & Bohn, 2021). According to the SLM-r, L2 learners gradually discern the phonetic differences between L1 and L2 sounds as they receive more L2 input. The likelihood that a new phonetic category is formed for an L2 sound depends on the perceived dissimilarity of the L2 sound from its closest L1 sound, as well as the quality and quantity of the input received by the L2 learners (Flege & Bohn, 2021, p. 33). The sound-level unit in the SLM-r is a position-specific allophone, which means that the mapping of an L1 allophone to an L2 phoneme is part of the organization of the bilinguals' sound inventory mapping across languages.

In the context of second dialect acquisition, German et al. (2013) present evidence that allophones in one's first dialect (D1) can be readily remapped to a different phonemic contrast in a second dialect (D2), suggesting that allophones are an adaptable sound level unit for listeners. Speakers of North American English (NAE) were trained to imitate and produce words containing [ɾ] and [t^h] in Glaswegian English (GE). All speakers produced some flaps [ɾ] for the phoneme /r/ in GE, which was an effective remapping of an allophone for /t/ in NAE. Speakers reassigned [ɾ] to /r/ in both word-medial and word-initial contexts, though the performance was better word medially where [ɾ] typically occurs in their D1. They also produced [t^h] for /t/ reliably in contexts where that phoneme is usually realized by [ɾ] in their D1. The speakers systematically generalized the D2 patterns to new words they did not hear in the sample of Glaswegian pronunciations to which they were exposed. These findings suggest that learners updated their coding system to allow particular allophonic variants of phonemes to be (1) used outside of their usual phonological context and (2) reassigned as the realization of a different phoneme (German et al., 2013). This research suggests allophones are an independent and accessible level in phonological encoding.

It is unclear to what extent this type of systematic modification of phonological encoding and generalizations of novel mappings between phonemes and allophones in D2 learning can be generalized to the phonological learning and adjustments required in L2 learning. In particular, it is a question whether allophones are the level of mapping unit in L2 learning similar to D2 learning, as hypothesized in SLM-r model (Flege & Bohn, 2021).

1.3. L2 lexical competition and phonological encoding

The inaccuracy or the imprecision in the forming of L2 phono-lexical representation can lead to L2 challenges at the lexical level of processing and encoding, such as over-activating lexical competitors compared to L1 listeners. Lexical competition occurs when more than one relevant lexical item in the L1 or L2 is activated; we focus here on the many studies that illustrate the involvement of the L2 in lexical competition processes (Weber & Cutler, 2004; Cutler et al., 2006; Broersma, 2012; Darcy et al., 2012; Desmeules-Trudel & Zamuner, 2023). For instance, the Dutch vowel inventory includes /ɛ/ but not /æ/, whereas English has both vowels contrastively. The absence of /ɛ/-/æ/ phonemic contrast in Dutch and the perceptual similarity between the two sounds has been shown to cause “spurious phoneme matches” for L1 Dutch speakers, thus increasing the lexical competition when L1 Dutch listeners process L2 English words containing these sounds (Weber & Cutler, 2004, p. 3). Hearing the beginning part of a word with the non-native sound [æ] (e.g., *p*[æ]*nda*) will likely activate a competitor that includes the L1 category [ɛ] (e.g., *p*[ɛ]*ncil*). In an eye-tracking study of Weber and Cutler (2004), L1 Dutch-speaking and L1 English-speaking listeners were presented with English words in auditory forms and asked to select the correct visual item on the screen (i.e., the visual world paradigm). It was found that L1 Dutch listeners fixed longer on the competitor item /ɛ/-word (*pencil*) when an /æ/-word (*panda*) was the target item. L1 English listeners did not show such difference in the fixation time. Moreover, the competition was asymmetric, meaning that /ɛ/-words were likely to be activated when L1 Dutch listeners heard an /æ/-word, but not vice versa (Weber & Cutler, 2004). In other words, when hearing words including a non-L1 phoneme, L1 Dutch listeners tend to activate words consisting of a perceptually similar phoneme in their L1; when hearing words including an L1 phoneme, they are unlikely to activate words including a non-L1 phoneme.

Crucially, however, Escudero et al. (2008) demonstrated that the observed asymmetry in Weber and Cutler (2004) hinges on being able to explicitly tie the phonetic form to a particular phonological category. Escudero and colleagues taught proficient Dutch-English bilinguals English nonwords with /ɛ/ and /æ/, varying whether participants received orthographic support for the nonwords or only an auditory signal. At test, listeners who were only presented with the auditory signal showed symmetric looks to /ɛ/ and /æ/ items, whereas those who were presented with written examples of the nonwords, which then mapped the acoustic signal to either an /ɛ/ or /æ/, exhibited the asymmetric pattern observed by Weber and Cutler. This indicates that orthography cued phonological categories.

1.4. The role of lexical knowledge in L2 phonology

As apparent in the literature surveyed above, accurate L2 perception does not necessarily lead to accurate lexical encoding (Curtin, Goad et al., 1998; Hayes-Harb & Masuda, 2008; Hayes-Harb, 2005), and interference from the L1 phonological system may not fully explain the development of L2 phono-lexical representation (Daidone & Darcy, 2021). For instance, the Fuzzy Lexical Representation Hypothesis (FLR) argues for overall imprecision in encoding L2 phonological contrasts at a global level, independent of local ambiguity in relation to the mapping to specific L1 phonemes (Gor et al., 2021; Cook et al., 2016; Gor, 2018; Darcy et al., 2024). The phono-lexical representations of L2 words that involve perceptually difficult phonemes (such as those that are perceptually similar yet not contrastive in their L1) are not fully differentiated and lack details in the mental lexicon (Gor, 2018). Given the global property of fuzzy phono-lexical representations in L2 (and perhaps in L1 as well), *flesh-flash* was observed for L1 Dutch-L2 English listeners but not for L1 English listeners, suggesting that the English /æ/-/ɛ/ contrast was not robustly encoded by L1 Dutch listeners, and these items were treated as identity pairs (Broersma, 2012). Other factors, such as receptive vocabulary size, have been found to be a crucial predictor for L2 phonological development (Bundgaard-Nielsen et al., 2011, 2012) and for lexical encoding accuracy (Daidone & Darcy, 2021; Llompart 2021).

Research on child language acquisition suggests that development of the lexicon encourages phonological development (Metsala & Walley, 1998; Vihman & Croft, 2007; Vihman, 2017; Daidone & Darcy, 2021; Munson et al., 2011). Infants begin with storing a few numbers of familiar word forms or perhaps start out with a rather holistic phonological knowledge (Vihman, 2017) and soon build more detailed phonological representations at segmental and featural levels from the growing number and density of words that they recognize and produce (e.g., the PRIMIR model in Werker & Curtin, 2005; Byun & Tessier, 2016). As children learn more vocabulary and produce different word types as well as words in related forms, their phonological awareness increases and they become more sensitive to the phonological relations between words (Metsala & Walley, 1998; Vihman, 2017), soon showing an emergent capacity to generalize both form patterns and meanings (Vihman, 2017). The processes of learning sounds and words feed each other and proceed in parallel (Vihman, 2017).

In addition to L1 acquisition, the effect of vocabulary knowledge on phonological development has been found in L2. Daidone and Darcy (2021) examined the lexical encoding of Spanish contrasts by English learners of Spanish through a lexical decision task. It was found that a larger L2 vocabulary size predicted a greater accuracy in the lexical encoding. Daidone and Darcy argue that the acquisition of more phonologically similar words may force learners to create more detailed phonological representations to differentiate such words.

Similarly, Llompart (2021) tested the effects of phonetic categorization and vocabulary size on lexical decision performance for German learners of English with different levels of L2 proficiency. The study revealed that categorization predicted nonword rejection accuracy for intermediate learners, whereas vocabulary size predicted nonword rejection accuracy for advanced learners. This suggests that once L2 learners have a robust phonetic categorization scheme in place, the size of the lexicon plays a more important role in creating target-like phono-lexical representations.

1.5. The pattern: processing English low vowel + nasal sequences (loVN) by L1 Mandarin listeners

L2 perception of non-L1 nasal place contrasts is particularly challenging (Harnsberger, 2001; Narayan et al., 2010). The studies in Harnsberger (2000, 2001) on cross-language perception of nasal contrasts reveal that the discrimination and the identification of non-L1 nasal contrasts can be predicted by phonetic cues relevant to L1 nasal contrasts. Narayan (2008) and Narayan et al.

(2010) show that perceptual-acoustic salience can influence nasal perception, while L1 language experience and exposure can facilitate discrimination of acoustically similar contrasts.

The challenge of perceiving L2 nasal place can arise from the influence of L1 nasal contrast, as different phonological restrictions apply to L1 concerning the surrounding segments. In the case of Mandarin and English, Mandarin has multiple phonological restrictions on low vowel + nasal sequences (loVN, e.g., English [hæm] "ham" and [dɑn] "dawn"). North American English has two low vowel phonemes, the low front vowel /æ/ and the low back vowel /ɑ/. In English, the three nasal contrasts /m n ŋ/ are all allowed in coda position, while only /m n/ are such in onset positions. In contrast, Mandarin allows the alveolar nasal /n/ and the velar nasal /ŋ/ in coda position, but not the bilabial nasal /m/, and, like English, only /m n/ are permitted in onsets (Duanmu, 2007). The low vowels [a] and [ɑ] are allophones in Mandarin, and their backness is conditioned by the place of the following nasal coda (Duanmu, 2007; Luo et al., 2020). Namely, the low front vowel must precede the alveolar nasal (i.e., [an] but *[aŋ]), and the low back vowel must precede the velar nasal (i.e., [ɑŋ] but *[ɑn]; Rhyme Harmony as in Duanmu, 2007). While the only contrastive loVN permitted in Mandarin are [an] and [ɑŋ], English permits all the combinations of loVN sequences as contrastive (i.e., /æm/, /æn/, /æŋ/, /ɑm/, /ɑn/, /ɑŋ/, as in **Table 1**). Notably, only English /æn/ and /ɑŋ/ follow Mandarin phonotactic restrictions. Therefore, despite the subphonemic difference in the vowels of English /æn/ and Mandarin /an/, they are assumed as equivalent at the phonological level, as low front vowels.

	Nasal coda	/m/	/n/	/ŋ/
Low vowel	/–back/	*æm	**æn**	*æŋ
	/+back/	*ɑm	*ɑn	**ɑŋ**

Table 1: Inventory of English loVN. Boxes in grey and with "*" represent those sequences that fail to follow Mandarin phonotactic restrictions but are possible in English.

Mandarin learners of English exhibit variations in producing the English vowel + nasal (VN) sequences, such as in the place of articulation of the nasal coda contrast, as well as the place of the vowel (Hansen, 2001; Liu, 2016). It is unclear to what extent these production differences arise; perhaps Mandarin-L1 English-L2 learners do not perceive the English patterns accurately or possibly L1 Mandarin listeners have yet to develop sufficiently detailed phono-lexical representations for their L2.

1.6. This study

This study aims to examine and compare how L1 Mandarin listeners process the L2 English loVN phonological contrast at different levels, specifically with respect to low-level phonetic discrimination (Experiments 1a and 1b) and phono-lexical encoding (Experiment 2). Previous research has found a disconnect between lower-level and higher-level processing (e.g., Amengual, 2016; Darcy et al., 2013; Curtin et al., 1998; Hayes-Harb, 2005; Hayes-Harb & Masuda, 2008; Turner, 2022). For instance, L2 listeners can have difficulty encoding non-L1 phonological contrast despite high accuracy in phonetic categorization (Darcy et al., 2013) or have high acceptance of non-words as words in lexical decision (Amengual, 2016). In the current study, we aim to identify and disentangle the L2 processing difficulties at comparatively lower and higher levels in two experiments. Experiment 1 tests L2 phonetic-level processing through a speeded AX discrimination task, supplemented by a perceptual similarity rating task. Experiment 2 examines L2 lexical-level online processing and competition using an eye-tracking technique with the Visual World Paradigm. Across these studies, we examine whether difficulty in lexical encoding may be attributable to imprecision at the acoustic-level perception or phonological-level encoding. Building upon the findings in Darcy et al. (2013) and Turner (2022), the study examines how late Mandarin-English bilinguals apply their phonological knowledge of the Mandarin loVN sequences to their burgeoning English phonology.

The second goal of this study is to probe the phonological unit that bilinguals map from their L1 to L2. If Mandarin-English bilinguals can map Mandarin allophones to English phonemes (see German et al., 2013, as discussed in Section 1.2), we predict symmetry in how Mandarin-English bilinguals process English low vowel + nasal sequences that are either licit (e.g., English sequences like /æn/ and /ɑŋ/) or illicit (e.g., English sequences like /æŋ / and /ɑn/) in Mandarin. If phonemes are the mappable sound level unit, we then predict asymmetries in lexical competition, with the illicit sequences differing from the licit sequences.[1] In Experiment 2, we also test how listeners' vocabulary size contributes to the precision of phonological encoding.

2. Experiment 1

Experiment 1 consisted of two perception tasks that tested and compared the processing of English loVN contrasts by L1 Mandarin listeners and L1 North American English listeners. A speeded AX discrimination task (Experiment 1a) was adopted to tap a phonetic level of processing (Werker & Logan, 1985), while a perceptual similarity task (Experiment 1b) aimed to tap phonological level of processing (Boomershine et al., 2008; Johnson & Babel, 2010; Huang & Johnson, 2010). Given the potential effects of L2 experience on the two levels of processing (Flege & Bohn, 2021; Darcy

---

[1] If we were dealing with less proficient bilinguals who are at an earlier stage in their L2 English learning, one might predict symmetrical performance with cross-language phonemic mapping because the licit Mandarin sequence in English would be processed with high fidelity and the illicit sequence would be implicitly repaired and fixed to align with Mandarin phonotactics (e.g., the Japanese ebzu case in Dupoux et al., 2011). In short, it is likely that the language learning trajectory is an important aspect of how a listener parses the sound structure of an additional language (e.g., see discussions in Darcy et al., 2013).

et al., 2013), two groups of L1 Mandarin listeners were tested who differed in their physical locations (Canada or China), which we assumed would index the quality and quantity of their English language experience.

We expect that L1 Mandarin listeners will have overall relatively high accuracy in the AX discrimination task, given the relatively high L2 proficiency level of this population (for details, see Section 2.1). Since the L1 sound system can warp low-level perception, as in Huang & Johnson (2010), L1 Mandarin listeners may experience longer reaction time processing the loVN sequences that are not contrastive in their L1 (e.g., [ɑn] vs. [ɑŋ]) compared to L1 English listeners. In the similarity rating task, L1 Mandarin listeners may perceive higher similarity between English loVN sequences that are non-contrastive in Mandarin compared to L1 English listeners, following the findings in Boomershine et al. (2008). Sequences that differ in nasals will be perceived as more similar than those that differ in vowels, as nasal cues are perceptually less robust than vowel cues in general (Harnsberger, 2001).

2.1. Participants

Two groups of Mandarin-speaking participants originally from Northern regions of China[2] were recruited through Chinese social media. Twenty-four of the L1 Mandarin-speaking participants were living in China and learned English as L2 (Chi-M group) (8 male, 16 female, aged 20–55, Mean = 25). Twenty-six of the L1 Mandarin speakers were living in Canada and learned English as L2 (Can-M group) (10 male, 16 female, aged 22–30, Mean = 25). None spoke a Chinese language other than Mandarin, and all reported that they learned a North American variety of English (mean onset age of acquisition of Chi-M = 7.7, Can-M = 4.7). Participants in the Can-M group lived in Canada for 3 to 8 years (mean age of arrival = 19.23, range = 15–26).

Twenty-six L1 English-speaking participants of a North American English variety were recruited from the University of British Columbia's participant pool as the L1 control group (4 male, 22 female, aged 18–25 with as a mean of 20.8). The English control (Eng) group reported no proficiency in any Chinese languages.

The Can-M group was assumed to have more exposure to English than the Chi-M group. Individual L2 experience and proficiency level were quantitively measured through the Lexical Test for Advanced Learners of English (LexTALE; Lemhöfer & Broersma, 2012) and the questionnaire Bilingual Language Profile (BLP; Birdsong et al., 2012). The LexTALE was used to assesses English lexical knowledge, where they made a judgement about whether the sequence of letters displayed on the screen on each trial was an existing English word. The means of the

---

[2] Speakers of Northern Mandarin varieties (e.g., Beijing Mandarin) are expected to maintain the /n/ and /ŋ/ coda nasal contrast followed by the low vowel in Mandarin (Chen & Guion-Anderson, 2011; Wu et al., 2016).

LexTALE scores differed across the three groups (Chi-M = 55.81, Can-M = 72.83, Eng = 90.96), suggesting that Can-M had more advanced English lexical knowledge and higher L2 proficiency compared to Chi-M. The BLP quantified participants' language dominance. Questions included their self-reported language history, language usage, language proficiency, and language attitude towards Mandarin and English. The dominance score ranged from –218 to +218. The negative means of the BLP scores showed that both Chi-M and Can-M groups were Mandarin-dominant rather than English-dominant (Chi-M = –141.76, Can-M = –60.48), yet Chi-M had a greater degree of dominance in Mandarin than Can-M.

2.2. Materials

2.2.1. Speeded AX discrimination task (Experiment 1a)

All the auditory stimuli were recorded by an L1 female speaker of Canadian English. These audio stimuli and a table in the supplementary materials (Table S1) describing the basic acoustic dimensions of f0 and duration for the tokens are available in the OSF directory for this project.[3] With the small number of items used as stimuli, statistical comparison of these acoustic measures is not appropriate. To summarize qualitatively, however, the f0 of the items overlapped considerably. Vowel durations also overlapped, and the velar nasal tended to be longer in duration than the alveolar and bilabial nasals. Overall, none of these differences would conspire towards concerning acoustic confounds in the materials.

Nine different item pairs contained all the combinations of loVN sequences (i.e., [æm, æn, æŋ, ɑm, ɑn, ɑŋ]), which differed only in either the vowel (e.g., [æm] vs. [ɑm], [æn] vs. [ɑn]) or the nasal (e.g., [ɑn] vs. [ɑŋ], [æm] vs. [æn]). The stimuli used for the six same loVN item pairs were physically identical. In addition, the stimuli consisted of pairs of low vowels (e.g., [ɑ] vs. [æ]) and pairs that contained the mid front vowel [e] and one of the nasals (e.g., [em] vs. [en]) in separate blocks. We included these two types of pairs to examine how well low vowels and coda nasals were distinguished, independent from a loVN environment, where [e] was chosen as a non-low vowel. A full stimuli list is available in the supplementary materials (Table S6). The items in each different pair were presented in both orders, and each stimulus was presented three times. Each of the same pair stimuli was presented 3–9 times to match the total number of its "different" pair counterpart in each block. This led to 156 trials in total. The stimuli were presented in three blocks (i.e., vowel only, mid vowel+nasal, and low vowel + nasal) and were randomized within each block. The RMS intensity of all stimuli was normalized to 65 dB. Given the study was conducted online, listeners could adjust their audio to a comfortable listening level on their own devices.

---

[3] All supplementary materials, code and audio materials can be accessed on OSF: https://osf.io/nzsjy/.

2.2.2. Perceptual similarity rating task (Exp. 1b)

The same set of loVN sequences recorded for Experiment 1a was used in this task. As presented in **Table 2**, the stimuli contained nine critical pairs that differed only in either the nasal (e.g., [ɑn] vs. [ɑŋ]) or the vowel (e.g., [æm] vs. [ɑm]). To establish the full range of possible rating scales, we also included six non-critical pairs that differed in both the nasal and the vowel (e.g., [ɑn] vs. [æm]), and six physically identical pairs. Items in the critical and non-critical pairs were presented in both orders. The stimuli of all three pair types were presented three times, resulting in 108 trials in total. The order of all the stimuli was randomized, presented in one single block.

Type	**Condition**	**Item**
Critical	**Nasal**	e.g., [æm] – [æn], [ɑn] – [ɑŋ]
	Vowel	e.g., [æm] – [ɑm], [æn] – [ɑn]
Non-critical		e.g., [æm] – [ɑn]
Same		e.g. [æm] – [æm]

Table 2: Conditions and example items in the perceptual similarity rating task.

2.3. Procedure

The experiment was conducted online, implemented in jsPsych. Participants were asked to complete the experiment on their own laptop, with headphones, in one sitting. Once they provided their consent on the first page, participants moved to a headphone check which required accurate responses on five out of six trials (Woods et al., 2017). Participants who passed the headphone check then proceeded to the perception tasks. The order of the two tasks was counterbalanced across participants. In both perception tasks, participants heard two auditory items on each trial, and the tasks differed with respect to the durations of inter-stimulus interval (ISI) and what was asked of listeners.

In the AX discrimination task, the ISI between the two stimuli on each trial was 250 ms. Participants were asked to respond as quickly as possible whether the two items in a trial were the same or different. The response time window was limited to 3000 ms. To encourage quicker responses, a feedback screen with the overall accuracy rate and response time was given after each response. The relatively short ISI and speeded response time window had a low memory load and aimed to tap low-level phonetic processing (Werker & Logan, 1985; Johnson & Babel, 2010; Boomershine et al., 2008).

In the similarity rating task, participants heard two auditory items on each trial, with an ISI of 500 ms. Participants rated how similar the two items were on a scale of 1 to 5 (1 =

"very different," 5 = "very similar"). There was no limit on the response time window, and no feedback was given when the response was received. The relatively longer ISI was designed to tap a higher level of processing that accessed phonological representations (Werker & Logan, 1985; Johnson & Babel, 2010).

After participants completed the first task, they were asked to complete the LexTALE, which consisted of 60 testing trials in total. At the end of the study, participants in the L2 group filled out the language background questionnaire BLP (Birdsong et al., 2012) that quantified their language dominance, and the L1 group filled out a survey about their language background.

All the instructions presented to the L2 participant groups were written in English and simplified Chinese characters. The whole study lasted about 40 minutes.

2.4. Analysis and results

2.4.1. Speeded AX discrimination task (Exp. 1a)

2.4.1.1. Accuracy

Here, we report the results of the accuracy of the low vowel + nasal sequences, which is the focus of interest in this study. We include the results for low-vowel only and mid-vowel + nasal sequences in our supplementary materials (Table S2).[4] **Table 3** presents the mean discrimination accuracy calculated for low vowel + nasal sequences by listener group, including nasal-different pairs, vowel-different pairs, and the identical pairs.[5] As presented in **Table 3**, all language groups, on average, showed high accuracy in discriminating the loVN sequences. Nasal-different pairs had slightly lower average accuracy than vowel-different pairs.[6]

---

[4] In general, the results for these trials with low vowels only or mid-vowel + nasal sequences exhibit accurate responses. The vowel-only trials showed higher accuracy responses (> 94% for all groups) than the mid-vowel + nasal trials, where the nasal was different, which showed much lower accuracy and more varied accuracy across groups. All groups performed with lower accuracy in mid-vowel + nasal compared to low-vowel + nasal seqences.

[5] Accuracy was also separately calculated for each type of contrast, including pairs with one phonotactially licit sequence + one phonotactially illicit sequence for Mandarin listeners (e.g., [ɑŋ] – [ɑn]), and pairs with both illicit sequences (e.g., [æm] – [ɑm]), or both licit in the Same condition (i.e., [æn] – [æn]). All of the contrast types have very high accuracy (no less than 87%). This is reported in the supplementary materials (Table S3).

[6] We included the d' index of sensitivity in the supplementary materials (Table S4) on the OSF page as an additional possible way to represent perceptual sensitivity. The results on d' are in line with the reported mean accuracy below, such that all listener groups had very high d' on average (ENG = 3.64, Can-M = 3.59, Chi-M = 3.29).

	ENG	Can-M	Chi-M
Pairs differ by nasal coda	96.24% (0.19)	93.46% (0.24)	90.45% (0.29)
Pairs differ by vowel	96.97% (0.17)	99.14% (0.09)	97.22% (0.16)
Physically identical pairs	95.47% (0.21)	96.07% (0.19)	96.12% (0.19)

Table 3: By-group mean accuracy rate of accurately discriminating loVN (SD in parentheses), including three conditions (nasal-different pairs, vowel-different pairs, and physically identical pairs).

A Bayesian mixed-effects model was fitted in Stan using the brms package (Bürkner, 2017) in R (R Core Team, 2022) to analyze response accuracy on trials with low vowel + nasal sequences. The fixed effects included dummy-coded Condition (nasals different, vowels different, same, ref level: nasals different), Helmert-coded Group (ENG vs. mean of Can-M and Chi-M, Can-M vs. Chi-M), the Order of items in each pair (Order1 vs. Order2), and the interaction effect of Condition and Group. The outcome variable was dummy-coded Correct Responses (True = 1, False = 0). By-pair and by-subject random intercepts were included, as well as by-pair random slopes for Group and by-subject random slopes for Condition. A Bernoulli distribution was assumed for the correct responses. Weakly informative priors were used for all population-level parameters (centered at 0 with a 0.25 standard deviation). Hamiltonian Monte-Carlo sampling was used to draw samples from the posterior distribution (four chains, each with 2000 iterations and 1000 warm-up). There were no divergent transitions and the Rhat values were all < 1.01, suggesting well mixed chains. We report the mean of the posterior distribution, the 95% credible interval (CrI), and the probability of direction (PD). A 95% CrI that does not encompass 0 is considered strong evidence for a meaningful effect. The evidence for an effect is described as weak if CrI includes 0 but the PD that is greater than 95% with a particular direction (positive or negative) (Nicenboim & Vasishth, 2016).

As presented in **Table 4**, the model revealed strong evidence for a difference between items with different nasals and different vowels in their response accuracy on average across groups ($\beta = 1.05$, CrI = [0.02, 2.11], PD = 97.67%). There was no evidence for a difference between the English group and the mean of the two Mandarin groups ($\beta = 0.64$, CrI = [–0.21, 1.47], PD = 93.40%), and no evidence for a difference between the two Mandarin groups ($\beta = 0.65$, CrI = [–0.15, 1.44], PD = 94.33%) with respect to the differing nasals reference level. There was a meaningful interaction between the condition with differing vowel and the English versus Mandarin speaker groups ($\beta = –1.32$, CrI = [–2.44, –0.11], PD = 98.3%). This effect captures the empirical observation gleaned from **Table 3** that L1 Mandarin listeners were less accurate on the trials with different nasals than the trials with the different vowels. Though there are some small differences in accuracy, these results generally indicate that the discrimination accuracy was high for all listener groups. Nasal-different pairs had slightly lower discrimination accuracy than vowel-different pairs for the Mandarin listeners.

	β	**95% CrI**	**PD**
Intercept	3.76	[2.40, 5.09]	100%
ConditionSame	–0.04	[–1.08, 0.93]	52.85%
ConditionVowel	1.05	[0.02, 2.11]	97.67%
Group.Eng.vs.Man	0.64	[–0.21, 1.47]	93.40%
Group.Can-M.vs.Chi-M	0.65	[–0.15, 1.44]	94.33%
Order	–0.15	[–0.96, 0.68]	63.65%
ConditionSame:Group.Eng.vs.Man	–0.75	[–1.79, 0.35]	92.25%
ConditionVowel:Group.Eng.vs.Man	–1.32	[–2.44, –0.11]	98.30%
ConditionSame:Group.Can-M.vs.Chi-M	–0.65	[–1.67, 0.32]	90.38%
ConditionVowel:Group.Can-M.vs.Chi-M	0.74	[–0.53, 2.02]	86.95%

Table 4: Population-level or fixed-effect predictors for the Bayesian model for response accuracy in discriminating loVN in Experiment 1a. The β estimate, 95% Credible Interval (CrI), and Probability of Direction (PD) are reported.

2.4.1.2. Reaction times

Reaction times (RTs) were calculated from the onset of the second stimulus for each trial. Responses received before the onset of the second stimulus were removed from the data (1% of the responses). Reaction times to the different loVN pairs reflect how difficult it is to hear the difference between the two stimuli (Shepard et al., 1975). **Figure 1** shows RTs for correct responses to different pairs comparing nasal-different vs. vowel-different pairs across the three language groups. In these violin plots, the width of the data visualized for each group shows the density of the full data distribution, and the black dots show each group's median. Descriptive statistics reveal that for the nasal-different pairs, L1 English listeners have shorter RTs (mean = 1130.02 ms, SD = 402.97), compared to both of the L1 Mandarin groups, namely, Can-M (mean = 1344.25 ms, SD = 429.73) and Chi-M (mean = 1374.39, SD = 422.73). Similarly, L1 English listeners have shorter RTs for the vowel-different pairs (mean = 962.97, SD = 394.38), while L1 Mandarin listeners experience relatively longer RTs, namely, Can-M (mean = 1090.76 ms, SD = 428.69) and Chi-M (mean = 1132.07, SD = 480.86).

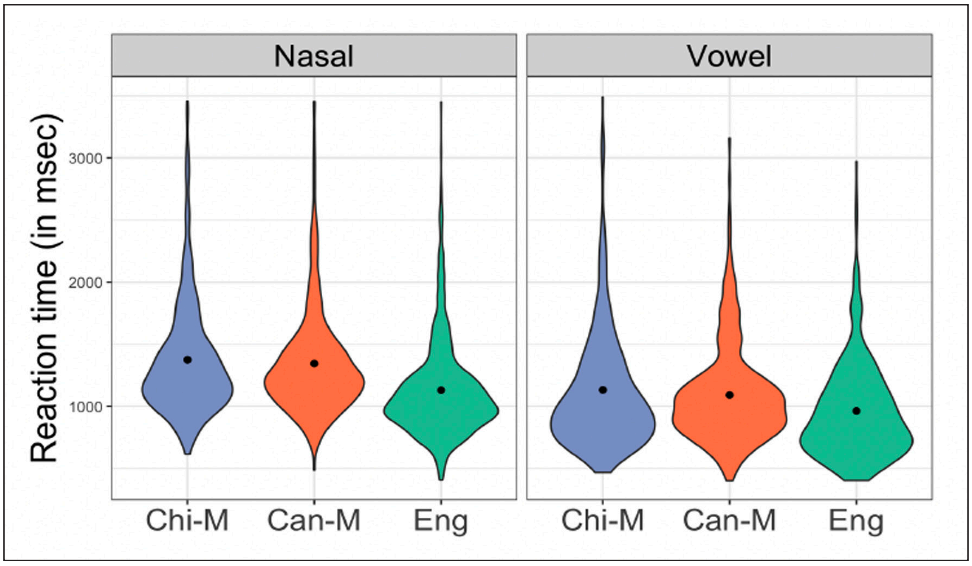

Figure 1: Reaction times to correct different pairs in AX task for different pair types (nasal vs. vowel) and listener groups.

	β	**95% CrI**	**PD**
Intercept	1294.77	[1190.21, 1399.22]	100%
Group.Eng.vs.Man	–231.97	[–357.22, –113.66]	100%
Group.Can-M.vs.Chi-M	–40.75	[–183.67, 113.10]	72.72%
ConditionVowel	–222.78	[–284.12, –162.01]	100%
Order	–8.09	[–62.99, 45.34]	62.08%
Group.Eng.vs.Man:ConditionVowel	78.92	[11.60, 143.33]	98.78%
Group.Can-M.vs.Chi-M:ConditionVowel	–14.08	[–97.04, 72.02]	63.62%

Table 5: Population-level or fixed-effect predictors for the Bayesian model for RTs in discriminating loVN in Experiment 1a. The β estimate, 95% Credible Interval (CrI), and Probability of Direction (PD) are reported.

We fit a Bayesian mixed-effects model on the trials with low vowel + nasal sequences, with weakly informative priors with normal distribution of (–500, 500) used for all parameters. Like the model for accuracy, fixed effects included dummy-coded Condition (nasals different, vowels different, ref level: nasals different), Helmert-coded Group, the interaction effect of Group and Condition, and Order. By-item and by-subject random intercepts were included, as well as by-item random slopes for Group and by-subject random slopes for Condition. There were no divergent transitions and the Rhat values were all < 1.01, suggesting well mixed chains.

The model results are summarized in **Table 5**. The model provided strong evidence that the RTs were longer for trials with different nasal pairs than vowel pairs ($\beta = -222.78$, CrI = [–284.12, –162.01], PD = 100%) for the English L1 reference level. We also observed strong evidence that the English group had shorter RTs than the mean of the two Mandarin groups in the nasal condition ($\beta = -231.97$, CrI = [–357.22, –113.66], PD = 100%), yet no evidence for a difference between the two Mandarin groups ($\beta = -40.75$, CrI = [–183.67, 113.10], PD = 72.72%). Nevertheless, we found strong evidence for an interaction effect between Group (ENG vs. mean of Chi-M and Can-M) and Condition ($\beta = 78.92$, CrI = [11.60, 143.33], PD = 98.78%), suggesting that the difference in RTs between English and the mean of the two Mandarin groups was greater for nasal pairs than vowel pairs (as seen in **Figure 1**). There was no evidence for an order effect ($\beta = -8.09$, CrI = [–62.99, 45.34], PD = 62.08%).

Together, these results show that all listeners accurately discriminated loVN sequences at a phonetic level. L2 listeners take longer to process the phonetic information in the signal, especially in trials that involve a contrast in the nasal coda, and these nasal trials also exhibit slightly lower (though still high) accuracy for L1 Mandarin listeners.

2.4.2. Perceptual similarity rating task (Experiment 1b)

We start by looking at the overall distribution of the similarity ratings across the three pair types. As expected, all listener groups perceived the same pairs as highly similar (mean = 4.77, SD = 0.72), whereas the non-critical pairs as relatively distinct (mean = 1.60, SD = 0.84), and the critical pairs in the middle of the two endpoint conditions (mean = 2.65, SD = 1.29). In the critical pairs, L1 English listeners (mean = 3.22, SD = 1.05) perceived nasal-different pairs as more similar compared to Can-Man (mean = 3.06, SD = 1.25) and Chi-Man (mean = 2.92, SD = 1.40). L1 English listeners (mean = 2.21, SD = 1.00) also perceived vowel-different pairs as more similar compared to Can-Man (mean = 1.64, SD = 0.97) and Chi-Man (mean = 1.55, SD = 0.88).

To further analyze the ratings of the critical pairs, we adopted a Bayesian mixed-effects ordinal regression (i.e., cumulative link) model. The model captures the cumulative likelihood of responses that fall in a certain range of ordered response categories. Compared to a linear regression model which assumes the dependent variables are continuous, the ordinal regression takes into account that the distributions across ordinal categories are uneven, and it can generate predictions that fall within the response scale (Bürkner & Vuorre, 2019; Kruschke & Liddell, 2018).

Fixed effects included dummy-coded Condition (nasal, vowel, ref level:nasal), Helmert-coded Group, the interaction effect of Group and Condition, and Order. The outcome variable was the rating score. By-item and by-subject random intercepts were included, as well as by-item random slopes for Group and by-subject random slopes for Condition. The link function parameter was set to "logit" to fit an ordered logistic regression model, and the threshold was "flexible" by default without adding any further constraints to the intercepts (Bürkner & Vuorre, 2019). We used the default uninformative priors for all parameters. There were no divergent transitions and the Rhat values were all < 1.01, suggesting well mixed chains.

We plotted the model-generated posterior distribution of average rating scores for nasal different pairs (**Figure 2a**) and vowel different pairs (**Figure 2b**). The density shows how likely data are observed along the rating score on the x-axis based on the model estimations. The model provided strong evidence that the listeners, on average, perceived the vowel pairs as less similar compared to the nasal pairs ($\beta = -2.97$, CrI = [–3.76, –2.11], PD = 100%), as presented in **Table 6**. However, we did not observe evidence of Group effects for nasal pair perception, as seen from **Figure 2a**. Specifically, there was no meaningful difference in the ratings between the English group and the mean of the two Mandarin groups ($\beta = 0.32$, CrI = [–0.51, 1.33], PD = 85.52%), nor between the two Mandarin groups ($\beta = 0.28$, CrI = [–0.50, 0.98], PD = 76.65%) in the nasal condition. For the interaction effects between Group and Condition, we found strong evidence that the differences in similarity ratings between the English group and the mean of the two Mandarin groups were greater for vowel pairs compared to nasal pairs ($\beta = 1.50$, CrI = [0.50, 2.54], PD = 99.83%). The interaction effect can be clearly seen when comparing **Figure 2a** and **Figure 2b**. There was no evidence that the differences in rating between Chi-M and Can-M vary for vowel pairs compared to nasal pairs ($\beta = -0.02$, CrI = [–1.17,1.18], PD = 51.85%).

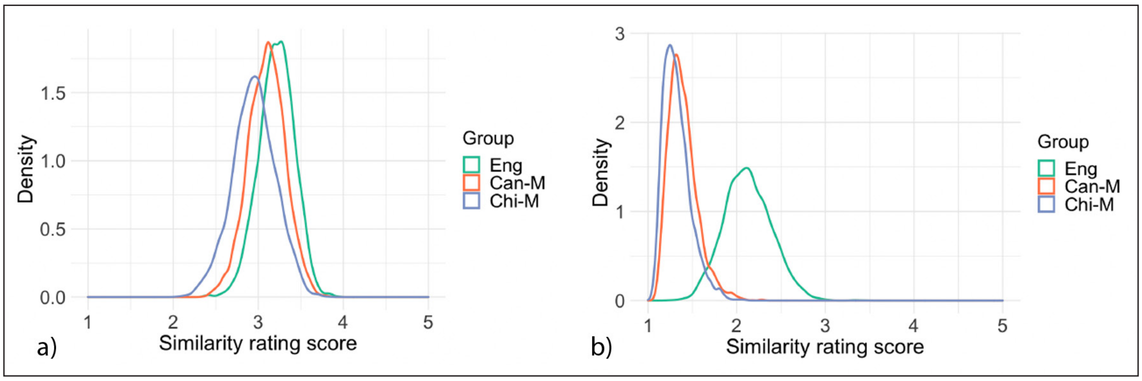

Figure 2: a) Average rating distributions for nasal-different pairs estimated by the model. **b)** Average rating distributions for vowel-different pairs estimated by the model.

	β	95% CrI	PD
Intercept[1]	–2.46	[–3.53, –1.34]	100%
Intercept[2]	–0.93	[–2.00, 0.17]	95.23%
Intercept[3]	0.41	[–0.66, 1.52]	77.12%
Intercept[4]	3.06	[1.98, 4.17]	100%
Group.Eng.vs.Man	0.32	[–0.51, 1.33]	85.52%
Group.Can-M.vs.Chi-M	0.28	[–0.50, 0.98]	76.65%
ConditionVowel	–2.97	[–3.76, –2.11]	100%
Order	0.02	[–0.64, 0.68]	51.80%
Group.Eng.vs.Man:ConditionVowel	1.50	[0.50, 2.54]	99.83%
Group.Can-M.vs.Chi-M:ConditionVowel	–0.02	[–1.17,1.18]	51.85%

Table 6: Population-level (or fixed-effect) predictors for the Bayesian model for the perceptual similarity ratings in Experiment 1b. The β estimate, 95% Credible Interval (CrI), and Probability of Direction (PD) are reported.

2.5. Interim discussion

The two perception studies examined how L1 Mandarin listeners processed L2 English loVN contrasts compared to L1 English listeners at different levels. The speeded AX discrimination task (Experiment 1a) revealed that both demographic groups of L1 Mandarin listeners demonstrated high accuracy at low-level processing, despite slightly lower accuracy and longer reaction times for nasal-different pairs compared to L1 English listeners.

The perceptual similarity rating task was designed with the goal of tapping phonological processing. According to Boomershine et al. (2008), Johnson & Babel (2010), and Huang & Johnson (2010), we would expect that L2 listeners perceive the English loVN contrasts with higher similarity compared to the L1 listeners, if one carries the assumption that their L1 knowledge affects their judgments of these sound sequences. However, we did not observe a meaningful difference in perceiving nasal-different non-native contrasts between the L1 and L2 listener groups; moreover, we see that the L2 listeners perceived the vowel-different English contrasts as more distinct compared to L1 listeners. Minimally, these results demonstrate that the allophony in Mandarin does not obliterate Mandarin-English bilinguals' ability to perceive these sound patterns as different.

The Mandarin-English bilinguals' ratings of the vowel pairs as *more* distinct than English listeners may appear counter-intuitive, considering the literature on perceptual similarity of phonemes versus allophones, but the current finding may be a novelty effect that enhances the acoustic distinction (Davis & Johnsrude, 2007). Such an interpretation would be like that proffered by Weber (2001) for Germans listening to Dutch. In a series of phoneme monitoring tasks, Weber observes that German listeners were faster at detecting an allophone in violation of a progressive fricative assimilation rule than an allophone in a phonotactically expected context. In our case, the presentation of a phonotactically illicit Mandarin sequence (e.g., *[ɑn] and *[æŋ]) prompts an auditory popout of the allophone, which highlights the acoustic distinction between the allophones, increasing Mandarin listeners' ratings of difference compared to those made by the English L1 listeners.

In Experiment 2, we designed an eye-tracking experiment testing how L1 Mandarin listeners process and encode an L2 English phonological contrast (i.e., loVN) at the lexical level. The eye-tracking technique is often known for its success in measuring unconscious, implicit responses, potentially avoiding the possibilities of triggering metalinguistic knowledge. We focus on word pairs with loVN that differ in the nasal, as they were harder to discriminate and take longer processing time compared to vowel-differed pairs as shown in Experiment 1a. The null effect in the difference of the two L2 proficiency groups suggests that the geographical categories (living in China vs. Canada) might not be the best predictor of L2 processing abilities, despite a difference in their LexTALE and BLP scores. More importantly, the categorical variable cannot fully reflect the individual variation in L2 experience or lexical capacities. Provided that vocabulary size can influence individual phonological development (Munson et al., 2011; Bundgaard-Nielsen et al., 2011, 2012), we included the LexTALE score (Lemhöfer & Broersma, 2012) as a predictor when analyzing L2 phono-lexical processing in Experiment 2 to better model the granularity in L2 phonological development.

3. Experiment 2

3.1. Hypotheses and predictions

Building upon the findings in Weber & Cutler (2004) and Darcy et al. (2013), this experiment examines how L1 Mandarin listeners process and encode L2 English phonological contrast (i.e., loVN) at the lexical level. The experiment uses eye-tracking with a Visual World Paradigm (VWP) design. Recordings of eye movements are used to examine lexical access in word processing tasks (Weber & Cutler, 2004; Desmeules-Trudel & Zamuner, 2023). In particular, fixation proportions to the target word and the phonological competitor in a VWP are closely mapped to their activation levels as spoken language unfolds over time (Magnuson, 2019). Compared to using pictures in a traditional visual world paradigm, the current study displays printed words as the visual stimuli, which avoids certain bias and disagreement towards the naming of the pictures, as well as allowing a wider range of word choice, including those that are not imageable (McQueen & Viebahn, 2007). More importantly, printed word displays have been shown to be more sensitive to phonological competition than using pictures (Huettig & McQueen, 2007; McQueen & Viebahn, 2007).

This experiment addresses the following research questions:

(1) Do L1 Mandarin listeners experience a greater competition effect compared to L1 English listeners? If so, at what stage in the online processing?
(2) Do L1 Mandarin listeners experience symmetric or asymmetric competition in processing English loVN sequences?
(3) Does a larger vocabulary predict less of a competition effect (i.e., more accurate phono-lexical processing)?

We hypothesize (H1) that L1 Mandarin listeners will show more lexical competition when processing English loVN sequences compared to L1 English listeners, akin to what was observed in Weber and Cutler (2004) and Cutler et al. (2006). We expect the competition effect is more likely to emerge at a later time window in online processing as opposed to an earlier one (Turner, 2022).

One limitation in the design of Weber and Cutler (2004) is that there was no baseline condition when comparing the degree of competition between the L1 and L2 listener groups. The fixation proportion to the target vs. competitor was compared in the two testing conditions without establishing a condition where both L1 and L2 groups were expected to share a similar level of competition. In the current study, we analyzed filler trials where the target and competitor words had phonological overlap in onsets that did not involve loVN sequences. This places us in a better position to conclude that the reported group difference was due to the difficulty of processing the particular illicit loVN contrast in question as opposed to a general deficit in L2 processing.

The current study also includes a *baseline* condition where both the target and the competitor words include a licit loVN to L1 Mandarin listeners ([æn] or [ɑŋ]). In comparison, in the two *critical* conditions, either the target or the competitor includes an illicit loVN or a licit loVN to L1 Mandarin listeners, following Weber and Cutler (2004). In this case, the baseline condition is expected to find the greatest competition effect for both language groups, as it contains a complete overlap of an identical loVN in the initial syllable. A larger group difference is expected in the critical conditions compared to the baseline condition.

The current study tests whether the competition effect as well as the encoding of phonological contrast is symmetric or asymmetric. If the phoneme is the mappable level of sound unit (H2a), then we would predict that L1 Mandarin listeners show a greater degree of a competition effect when the target contains an illicit loVN and the competitor contains a licit loVN, compared to the condition where the target contains a licit loVN and the competitor contains an illicit loVN. The

Mandarin vs. English listeners group difference is expected to be greater in the former condition (where the target is illicit and the competitor is licit) than the latter condition (where the target is licit and the competitor is illicit). Alternatively, if L1 allophones are mapped to phonemes in the L2 (H2b), then we would predict that L1 Mandarin listeners will not show asymmetric competition effects in the two critical conditions. There will be no meaningful language group differences between the two critical conditions.

Llompart (2021) finds that the phonolexical encoding accuracy of more advanced English learners (L1 German) is positively correlated with English vocabulary size, but that this is not the case for intermediate German L1 English learners. He posits that a certain threshold of lexical knowledge must be passed to improve the quality of L2 phonolexical encoding. Putting this into spoken word recognition, the lexical competition is resolved more quickly with greater language knowledge or ability (such as lexical development, reading abilities, and phonological processing skills, as discussed in Kutlu et al., 2024). We therefore predict that lexical competition will be resolved more quickly for individuals with larger estimated receptive vocabularies (H3).

3.2. Methodology

3.2.1. Participants

Twenty-four L1 Mandarin speakers who were students at the University of British Columbia (14 females, 10 males, aged 21–39 with a mean of 26) took part in the study. They were speakers of Northern varieties of Mandarin (e.g., Beijing Mandarin) and did not speak any other Chinese language. The L1 Mandarin speakers learned a North American variety of English as a second language, with the mean onset age of acquisition of 6.8 years old. All of them arrived in Canada after the age of 18 (age of arrival mean = 23.7, range = 18–37) and had lived in Canada for 2.5 years on average. The BLP (Birdsong et al., 2012) revealed that the mean score of language dominance was –94.77 (ranging from –162.77 to –49.13), suggesting that the participants were Mandarin-dominant bilingual speakers rather than English-dominant.

Twenty-three L1 North American English speakers who were students at the same Canadian university (13 females, 7 males, 3 non-binary, aged 18–32 with a mean of 22.3) took part in the study. Most of them also spoke languages other than English, such as French, Spanish, Hindi, Tagalog, German, etc. Crucially, they had no knowledge or proficiency in any Chinese languages, including Mandarin and Cantonese.

No participants reported any impairments of vision (other than corrected sight), hearing or speech that could possibly affect the results of the experiment. All participants were compensated with $15 CAD cash for their participation.

3.2.2. Materials

Thirty-six English words with the first syllable of the shape (C)loVN were chosen as target words, paired with another 36 words that differed in the nasal as their phonological competitors (except for the baseline condition). The target word and the competitor included either a licit loVN in Mandarin (e.g., [ɑŋ] *conquer*) or an illict loVN (e.g., [ɑn] *context*) in the first syllable. There were three levels of competition conditions regarding the type of loVN tested in the target and in the competitor, namely, (1) both the target and the competitor included a licit and identical loVN in Mandarin (i.e., ManMan, baseline); (2) the target included a Mandarin-illicit loVN and the competitor included a Mandarin-licit loVN (EngMan); (3) the target included a Mandarin-licit loVN and the competitor included an Mandarin-illicit loVN (ManEng).[7] The specific loVN sequences tested in the three conditions are listed in **Table 7**, as well as a few example item pairs. Each participant saw all three competition conditions. Each experimental trial appeared once in a given list.

Target-Competitor		**Competitor**	
		Licit in Mandarin	**Illicit in Mandarin**
Target	**Licit in Mandarin**	**ManMan**: /æn/-/æn/, /ɑŋ/-/ɑŋ/ (*brandy-branch*)	**ManEng**: /æn/-/æŋ/, /æn/-/æm/, /ɑŋ/-/ɑn/, /ɑŋ/-/ɑm/ (*sandal-sample*)
	Illicit in Mandarin	**EngMan**: /æŋ/-/æn/, /æm/-/æn/, /ɑn/-/ɑŋ/, /ɑm/-/ɑŋ/ (*context-conquer*)	----------

Table 7: The loVN sequences and example words tested in the three conditions. The first loVN or the item in each pair refers to the target and the second one refers to the competitor.

Among the 36 item pairs, participants were presented with 12 item pairs in each condition. Two lists were created for counterbalancing between the items used in the ManEng and EngMan conditions. In these conditions, participants saw each item pair in one of the two conditions. For instance, participants on List 1 were presented with *context* [kɑntɛkst] as the target word and *conquer* [kɑŋkɚ] as the competitor, and participants on List 2 were presented with *conquer* as the target word and *context* as the competitor. The full target-competitor item pairs used for the two lists are included in the supplementary materials (Table S8).

---

[7] Here we did not include a condition where both the target and competitor words involved Mandarin-illicit loVN. We are particularly interested in the mapping of L1 existing sound units in L2 processing, so the mapping between L1-illicit sequences is beyond the scope of this paper.

Two phonologically unrelated distractors were chosen for each target-competitor pair, and the four words were displayed together in each trial (Weber & Cutler, 2004). Sixty-four different words were used as distractors. The target, competitor and two distractors displayed in each trial were (nearly) matched in the number of graphemes and phonemes (McQueen & Viebahn, 2007). A small difference in the number of letters or phonemes (no more than two) was allowed, as the length of the four words could not be exactly matched in some trials within the selection criteria (Veivo et al., 2016). The frequencies of the targets, competitors and the distractors were controlled using the English Lexicon Project database (Balota et al., 2007), such that the frequency differences among the four words in all trials were not significant (all ps > 0.1). The four words in each set had no semantic relationships.

An additional 36 quadruplets were chosen for filler trials. In some filler trials, the four items had no overlap in the sound segments of the first syllable (e.g., *basket, perfume, helmet, couple*). Other trials included pairs of words that had some overlapping sounds in the first syllable (e.g., *lemon-letter*) and two phonologically unrelated words; either one of the words with overlapping initial sounds or one of the phonologically unrelated words was the target. The target item in the filler trials did not contain loVN in the first syllable. The four words in each filler trial were also (nearly) matched in the number of graphemes and phonemes.

All the target words were recorded in the carrier phrase "choose the word _" by an L1 female speaker of North American English. The sentences were read in a randomized order. One instance of the recorded carrier phrase was selected and used for all stimuli. A short prosodic pause of 400 ms was introduced between the carrier phrase and the onset of the target word using Praat (Boersma & Weenink, 2020).

3.2.3. Procedure

The laboratory session consisted of three parts: an eye-tracking experiment, an English lexical judgement task, and a language background questionnaire. The session was conducted at the Experimental Linguistics and Fieldwork Lab (ELF-lab) at the University of British Columbia. Participants were seated at a comfortable distance from a computer screen. At the beginning of the eye-tracking experiment, participants read written instructions in English, which explained the task. Participants' eye movements were monitored using an SR Research EyeLink Portable Duo (www.sr-research.com), with the sampling rate at 250 Hz. After the eye tracker was calibrated, they were presented with three practice trials, followed by 36 experimental trials and 36 filler trials (75 trials in total). Each participant was presented with experimental trials from one of the two lists. The experimental trials and the filler trials were presented in a pseudo-random order, such that each experiment trial was presented after a filler trial (Weber & Cutler, 2004).

In the beginning of each trial, participants saw a small cross in the center of the screen. A drift correction was performed when participants fixated on the cross. Once the system registered the look, the four printed words appeared on the screen. Using printed word displays instead of pictures in a Visual World Paradigm (McQueen & Viebahn, 2007; Veivo et al., 2016) provided a wider range of word choice in English that contained loVN, including items that were not picturable. Printed words have also been argued to be more sensitive to phonological competition than pictured items (Huettig & McQueen, 2007). The audio stimuli with the spoken instruction "choose the word" plus the target word was played 500 ms later after the visual display of the items, which gave participants a preview time to pre-scan the four words (Turner, 2022; c.f., McQueen & Viebahn, 2007). Participants pressed one of the four buttons on a button box to select the corresponding word that they heard. Once the press was received, the next trial was initiated. Participants' right eye was tracked throughout the experiment.

The printed words were presented in 24-point Times New Roman font, displayed at four fixed locations (see **Figure 3**). The horizontal distance between the center of the words on the left and right was 960 pixels (25.4 cm). The centers of the upper words were 540 pixels (14.3 cm) above those of the lower words. The positions of the four words in each trial were manually coded and thus counterbalanced across these four locations.

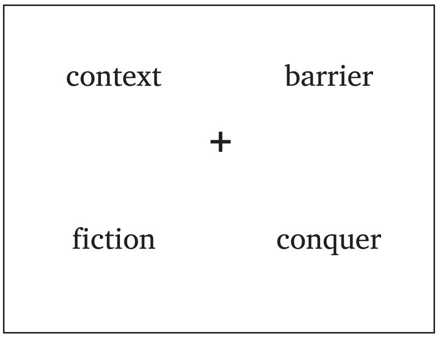

Figure 3: Example visual display presented to participants (target: context; competitor: conquer; distractors: barrier, fiction).

When the eye-tracking session was over, participants were asked to complete the LexTALE (Lemhöfer & Broersma, 2012) as an assessment of their lexical knowledge of English. On each trial, participants decided whether the sequence of letters they saw on the screen was an existing English word or a nonce word. There were 60 trials in total.

Finally, participants completed a questionnaire about their background. L1 Mandarin listeners filled out the BLP, whereas L1 English listeners completed a questionnaire about the proficiency and the history of the languages they spoke, as well as some demographic information. The whole session took about 40 minutes in total.

3.2.4. Statistical analysis

Fixation proportions to the targets, competitors, and the distractors were aggregated into 20 ms time bins using the R package *eyetrackingR* (Dink & Ferguson, 2015). Similar to the design in McQueen and Viebahn (2007), which also adopts items with varying length, a 400 × 300 pixels (10.58 cm × 7.94 cm) interest region centered on the middle of each word was predefined for all trials. A fixation was counted as occurring towards a given word if it fell within this region. All other fixations that fell outside of these regions, in this case, samples identified as blinks and saccades, were discarded. This constitutes 22.08% of the data in the time window of interest. The analysis of the time window of interest starts from 200 ms after the onset of the vowel in the target word stimuli and ends 1000 ms after the vowel onset. As it takes about 200 ms for a planned eye movement to launch (Fischer, 1992; Salverda et al., 2014), eye movements driven by the beginning of the acoustic signal should be reflected by fixations from 200 ms onwards.

To analyze and compare L1 vs. L2 processing in the time window of interest, a cluster-based permutation analysis (CPA) was conducted. CPA is a non-parametric test which reveals whether an effect is significant somewhere in a predefined interest period (Maris & Oostenveld, 2007; Ito & Knoeferle, 2023). It is well suited for analyzing the time-course eye-tracking data, given the dynamic nature of the data. CPA controls for autocorrelation between the data points in a time series and allows multiple comparisons across different time bins while maintaining statistical power (Ito & Knoeferle, 2023). However, CPA does not inform the specific onset or quantify the exact duration of an effect, as the *p*-value is calculated at cluster-level statistics rather than for a specific time bin; the CPA results simply reveal whether there is a significant difference *somewhere* within a temporal window (Maris & Oostenveld, 2007).

The CPA is based on a linear mixed effects regression model. An LME-based CPA model was fit to examine the effects of listener group (Mandarin L1 or English L1), competition condition (outlined in **Table 7**), and receptive vocabulary size for English (as measured by LexTALE). The categorical independent variables were dummy-coded Group (L1Man vs. L1Eng, ref level: L1Man) and Helmert-coded Condition (ManMan vs. mean of ManEng and EngMan, ManEng vs. EngMan). LexTALE scores were centered and standardized. By-item and by-subject random intercepts were included, as well as by item random slopes for Group and LexTALE, and by-subject random slopes for Condition.

The dependent variable of the model was the log-ratio of fixation proportions to the targets over the competitors using the formula *log ((fixation proportion on target + 0.5) / (fixation proportion on competitor + 0.5))*, to quantify a fixation bias toward the target vs. the competitor (Arai et al., 2007; a small amount of 0.5 was added to both the denominator and the numerator to avoid zero in the denominator). The higher the value of the log-ratio in the model results, the smaller the observed competition effect.

Figure 4 shows the empirical distribution of centered LexTALE scores by listener group. English listeners on average have more lexical knowledge of English than Mandarin listeners, despite some overlap in the middle attributed to high-proficiency L2 speakers. To address concerns about whether the group and LexTALE scores are correlated to a point of concern for model fitting, we conducted a collinearity test to check whether there is any correlation between the three independent variables (group, LexTALE score, and condition) in Model 1. We calculated the Variance Inflation Factors (VIFs) of the three independent variables using the package *car* in R. Since the maximum VIF is 1.802 (<10), the degree of collinearity is not problematic in this model (Sonderegger, 2023, p.122). See Table S5 in the supplementary materials for these results.

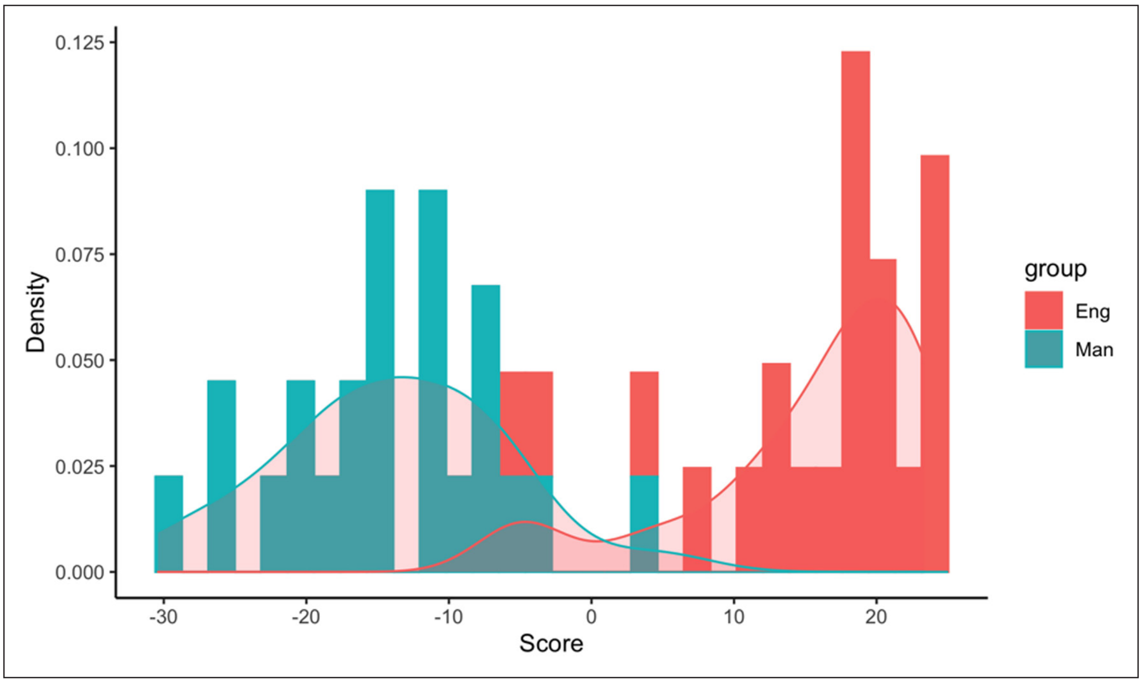

Figure 4: Distribution of LexTALE scores (mean centered) by L1-English and L1-Mandarin listener groups.

The *clusterperm.lmer* function in the *permutes* package (Voeten, 2022) was used to fit the model in R (R Core Team, 2021). Likelihood ratio tests were conducted at individual time bins against each predictor in the LME. Since the *clusterperm.lmer* function in the *permutes* package calculated likelihood ratio test statistics (*chi*-square) at individual time bins, model comparisons against each predictor (log-likelihood of the model with vs. without the predictor) were performed under the hood. Then permutation tests were performed at the cluster level to detect significant time windows based on the LRT statistics. A sum of individual test statistics (cluster mass statistic, CMS) was calculated in consecutive time bins. To generate a null hypothesis distribution for the clusters in the original data set for comparison, the label of the variable Condition was shuffled within subjects, and the label of Group (and LexTALE Score) was shuffled between subjects. The labels of Subject and Item remained unshuffled to preserve the structure of the data. The permutation was conducted 1000 times, and the *p*-value was obtained by comparing the cluster mass statistics of each cluster in the original data set with the null hypothesis distribution. Since the data is sampled into 20 ms time frames, clusters that span fewer than three frames (i.e., 60 ms) will not be considered (Hammerly et al., 2022; see discussions in Barr et al., 2014).

3.3. Results

The mean proportions of fixations to the four words for the 0–1000 ms time period starting from the vowel onset of the target word are plotted in **Figure 5** for the two groups of listeners and the three competition conditions.

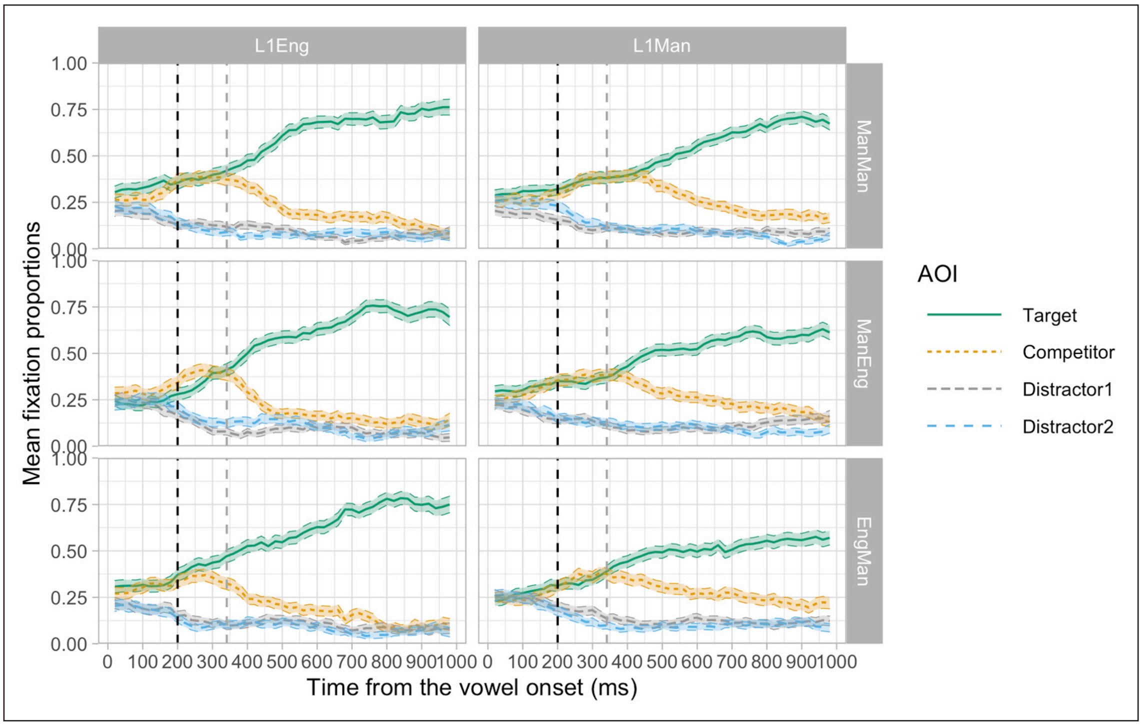

Figure 5: The mean proportions of fixations to the four words in 0–1000 ms starting from the vowel onset of the target word in the audio stimuli. The conditions are ManMan: baseline; ManEng: target = Mandarin-licit, competitor = Mandarin-illicit; EngMan: target = Mandarin-illicit, competitor = Mandarin-licit. Group is coded as L1Eng and L1Man. The dashed lines indicate 200 ms after the vowel onset (black) and 200 ms after the mean onset of the nasal (gray), where eye movement launches.

3.3.1. Model 1

The results of this model (see **Table 8**) reveal a significant main effect of group in a relatively late time window (positive cluster spanning from 680–980 ms; CMS = 600.32, p < 0.001). This indicates that L1 Mandarin listeners experienced a greater competition effect compared to L1 English listeners in the time span from 680 ms after the vowel onset until the end of the interest time period. Higher LexTALE scores predicted less competition effects (positive cluster spanning from 320 – 960 ms; CMS = 22.65, p < 0.001). This suggests that competition might be resolved more quickly when one has more robust phonological generalizations abstracted over the lexical knowledge learned. The interaction effect between group and LexTALE score was not significant throughout the 200–1000 ms interest period. This suggests that the effect of lexical knowledge does not differ between the two language groups.

Effect	**Cluster window (ms)**	**CMS**	**Sign**	***p*-value**
GroupL1Eng	680 – 980	600.32	1	* < 0.001
Lex_score	320 – 960	22.65	1	* < 0.001
Condition.Baseline.vs.Critical	420 – 980	29.21	–1	* < 0.001
Condition.ManEng vs.EngMan	460 – 680	11.43	1	* < 0.001
Lex_score:GroupL1Eng	—	—	—	—
GroupL1Eng: Condition.Baseline.vs.Critical	—	—	—	—
GroupL1Eng:Condition.ManEng vs.EngMan	200 – 660	26.48	1	* < 0.001
Lex_score: Condition.Baseline.vs.Critical	520 – 900	211.24	–1	* < 0.001
Lex_score:Condition.ManEng vs.EngMan	460 – 540	167.35	1	* < 0.001
GroupL1Eng:Lex_score: Condition.Baseline.vs.Critical	260 – 980	16.56	1	* < 0.001
GroupL1Eng:Lex_score:Condition.ManEng vs.EngMan	—	—	—	—

Table 8: Results of Model 1. All detected clusters are shown with the CMS and the significant p-value determined from the constructed null hypothesis distribution.

Regarding the main effects of condition, greater competition was observed in the baseline condition (ManMan) than the two critical conditions (negative cluster spanning from 420 – 980 ms; cluster mass = 29.21, p < 0.001). Mandarin listeners showed greater competition in the EngMan over the ManEng condition compared to English listeners, evidenced by both the significant main effect of Condition.ManEng vs.EngMan (positive, 460 – 680 ms, CMS = 11.43, p < 0.001) and interaction effect of group and Condition.ManEng vs.EngMan (negative, 200–660 ms, CMS = 26.48, $p < 0.001$). The amount of difference in the competition between the baseline and the critical condition did not differ by the two language groups (GroupL1Man:Condition. Baseline.vs.Critical was not significant). Nevertheless, LexTALE score significantly interacted with condition, namely, the effect of LexTALE score on the competition effect was greater in the two critical conditions compared to the baseline (negative, 520 – 900 ms, CMS = 211.04, $p < 0.001$). Moreover, the three-way interaction of group, LexTALE score and condition (Baseline. vs.Critical), was significant during 260 – 980 ms (positive; CMS = 16.56; $p < 0.001$), suggesting that the difference in the LexTALE score effect between baseline and the critical conditions was greater for Mandarin compared to English. The effect of the LexTALE score on the two critical conditions did not differ between the two language groups (GroupL1Eng:Lex_score:Condition. ManEng vs.EngMan was not significant). To better disentangle the three-way interaction, we ran additional models, as explained in Section 3.3.2.

But first, to eliminate a more general effect of the delayed L2 processing and more competition over L1 (not just due to the loVN sequences), we ran LME-based CPA models for two types of filler trials respectively. They were the filler trials that included phonologically overlapping onsets or rhymes in the targets and competitors (e.g., lemon vs. letter, which did not have loVN), as well as the filler trials that consisted of four phonologically unrelated words. The independent variable was dummy-coded Group (L1Man vs. L1Eng, ref level: L1Eng). By-item and by-subject random intercepts were included, as well as by-item random slopes for Group. The dependent variable was the log-ratio of fixation proportions to the targets over the competitors. Both models revealed that no significant cluster appeared throughout the time period of 200 ms to 1000 ms from the onset of the audio stimuli (CMSs < 160.61, *ps* > 0.05). This result further strengthens our interpretation that the L2 late competition effects only occurred in loVN processing, such that L2ers did not show more competition or slower processing in all other words compared to the L1 group.

3.3.2. Group-specific models

In the larger model (Model 1), we observed a three-way interaction effect between Group, LexTALE score, and the baseline versus critical conditions. This complex interaction occurred in a large time span from 260 ms to 980 ms. To better interpret and compare the effects of condition and lexical knowledge within each group, we fit separate models for each of the two groups. Each model included independent variables of centered and standardized LexTALE Score and Helmert-coded Condition (ManMan vs. mean of ManEng and EngMan, ManEng vs. EngMan), as well as the interaction effect between Condition and Group. By-item and by-subject random intercepts were included, as well as by-item random slopes for Score, and by-subject random slopes for Condition.

Table 9 and **Table 10** present the results of the two models for L1 English listeners and L1 Mandarin listeners, respectively. The effect of LexTALE score was significant for both L1 English (positive, 360 – 680 ms, CMS = 236.73, $p < 0.001$) and L1 Mandarin listeners (positive, 440

– 680 ms, CMS = 116.97, p < 0.001). This revealed that higher LexTALE scores predicted smaller competition effects for both listener groups, although this effect occurred slightly later for the L1 Mandarin group. Both listener groups experienced greater competition in the baseline condition compared to the critical conditions (L1Eng: negative, 200 – 980 ms, CMS = 12.03, p < 0.001; L1Man: negative, 380 – 680 ms, CMS = 20.59, p < 0.001). Among the critical conditions, L1 English listeners had greater competition for the ManEng condition compared to the EngMan condition very early on (as seen from Fig.6; negative, 220 – 340 ms, CMS = 59.59; p < 0.001), while L1 Mandarin listeners had greater competition in the EngMan condition compared to the ManEng condition (positive, 400 – 540 ms, CMS = 77.44, p < 0.001). For both groups, the condition effects occurred in a relatively short period of time windows (no more than 140 ms) and disappeared at a later period of processing. The two interaction effects between the LexTALE score and the two levels of conditions were significant for both groups (see **Tables 9** and **10** for details). Namely, the effect of LexTALE score on the competition effect was greater in the two critical conditions compared to the baseline. Moreover, the effect of LexTALE score was significantly different between the two critical conditions for both groups.

Effect	**Cluster window (ms)**	**CMS**	***p*-value**	**Sign**
Lex_score	360–680	236.73	* < 0.001	1
Condition.Baseline.vs.Critical	200–980	12.03	* < 0.001	–1
Condition.ManEng vs.EngMan	220–340	59.95	* < 0.001	–1
Lex_score: Condition.Baseline.vs.Critical	480–980	17.73	* < 0.001	–1
Lex_score: Condition.ManEng vs.EngMan	240–420	15.63	* < 0.001	1

Table 9: Results of the model for L1 English listeners. All detected clusters are presented with the CMS and the p-value.

Effect	**Cluster window (ms)**	**CMS**	***p*-value**	**Sign**
Lex_score	440–680	116.97	* < 0.001	1
Condition.Baseline.vs.Critical	380–680	20.59	* < 0.001	–1
Condition.ManEng vs.EngMan	400–540	77.44	* < 0.001	1
Lex_score: Condition.Baseline.vs.Critical	340–780	17.19	* < 0.001	–1
Lex_score:Condition.ManEng vs.EngMan	580–760	81.72	* < 0.001	1

Table 10: Results of the model for L1 Mandarin listeners. All detected clusters are presented with the CMS and the p-value.

3.4. Discussion

The eye-tracking experiment examined how L1 Mandarin listeners process and encode L2 English phonological contrasts (i.e., loVN) at the lexical level. The results of the larger model support Hypothesis 1, proposed in Section 3.1, that L1 Mandarin listeners experience a greater competition effect compared to L1 English listeners when processing English loVN sequences. We tested whether there was a symmetric or asymmetric competition effect in L2 processing in the encoding of the loVN contrast. Taking together the results of the cluster permutation models–those run on the whole data set and separated by language background group–we do not see strong evidence for the hypothesis of asymmetric competition (H2a), contra Weber and Cutler (2004) and Darcy et al. (2013) who have examined the mapping of a *novel* phoneme in the L2. When condition and receptive vocabulary size are taken into account to explain phono-lexical competition, a difference between the two critical conditions appeared in a relatively early (before 550 ms from the onset of the vowel) and short period of time window (no more than 140 ms) for both L1 and L2 groups, and disappeared at a later time period of processing. Although L2 listeners experienced greater competition in the EngMan condition than ManEng in the period of 400 ms to 540 ms, the competition was quickly resolved later where phonological processing is arguably involved to a larger degree (Turner, 2022). Overall, the results revealed that there is no distinction in the encoding of loVN patterns for Mandarin L1-L2 English learners, supporting hypothesis H2b that allophones are mapped from the L1 to the L2 (Flege & Bohn, 2021). L2 learners can reassign vowel allophones in the L1 to phonemes in the L2, in line with learners of a second dialect as previously observed in German et al. (2013).

Finally, the results support H3 such that a larger vocabulary size predicts more robust phono-lexical L2 processing. Akin to the previous studies on the relationship between phonological and lexical development in the L1 (Stoel-Gammon, 2011, and references therein), less competition was observed for listeners with larger estimated receptive vocabularies, as the phonological knowledge generalized across larger lexicons facilitates spoken word recognition for *both* L1 and L2 listeners. This is consistent with the findings in Llompart (2021) for advanced L2 learners that receptive vocabulary knowledge encourages phono-lexical encoding, leading to quicker competition resolution.

4. General Discussion

Experiment 1 establishes that Mandarin L1-English L2 listeners perceive the low vowel + nasal (loVN) sequences robustly, with minimal differences in comparison to English L1 listeners. When shifting listeners' task to the word-level in Experiment 2, however, we observe differences between Mandarin L1 and English L1 listener groups, in addition to effects of estimated vocabulary size across groups.

4.1. When does the L2 phono-lexical competition effect emerge in online processing?

Mandarin-English bilinguals show more competition with the loVN sequences than the English L1 listeners in Experiment 2, but this is not accompanied by a lack of perception skills. The results in Experiment 1 indicate that the Mandarin-English bilinguals were sensitive to the phonetic differences in the stimuli. At a basic level, these results demonstrate that the locus of L2 difficulties surfaces at the phono-lexical level, not the phonetic level (e.g., Darcy et al., 2013; Amengual, 2016). However, because we deploy a VWP, we are in a position where we can better identify the *when* question.

We observe a long-lasting effect (320–960 ms in the combined model, 350–680 ms in the L1 English model, and 440–680 ms in the L1 Mandarin model) of lexical knowledge, quantified with a test of receptive vocabulary size in Experiment 2, which suggests an influence of the lexicon across a time span that implicates both low-level phonetic processing and a higher-level phonological encoding. This role of the lexicon in the entire time course of spoken word recognition contrasts with the finer temporal locus of the listener group difference in the combined model. A difference in language background was found at a relatively later time window (680–980 ms). This suggests that L2 listeners differ from the L1 listeners in phonological encoding but less in terms of low-level phonetic sensitivity, which aligns with the findings of minimal group differences and high accuracy in the discrimination task in Experiment 1(a). Moreover, the null effect of the interaction between group and LexTALE suggests the influence of lexical knowledge on processing does not differ between L1 and L2 listeners. Namely, a larger receptive vocabulary size predicts more robust phono-lexical processing for both L1 and L2 listeners.

Previous studies on L2 online processing have suggested that the low-level processing of acoustic or phonetic information tends to occur in an earlier time window, whereas higher-level phonological encoding happens later (Turner, 2022; Escudero, 2009; McQueen, 2005). Turner (2022) reported that although L1 British English-L2 French listeners process French /y/ easily in the time window of 200–400 ms because of its acoustic similarity with British English /u/, they experience similar amounts of competition for both /y/ and /u/ after 600 ms, arguably when the phonological encoding has begun. The current findings align with Turner, suggesting that it is likely that L2 listeners in this study have little difficulty processing the low-level acoustic details in an early time period, yet more L2 difficulties lie in the phonological level of processing and encoding that arguably occur at a later time.

4.2. Symmetric competition and the remapping of allophones

Weber and Cutler (2004), Cutler et al. (2006) and Darcy et al. (2013) suggest that L2 learners tend to asymmetrically map the non-native contrast to a “dominant category,” which typically refers to a familiar L1 phonological category. Escudero and colleagues (2008), however, demonstrate that such asymmetries only occur with external evidence (in their case, orthographic information) that maps the phonological patterns; the learning of novel (non)words from an audio-only signal does not show preference for the dominant L1 category (Escudero et al., 2008). However, the results in the current study with real lexical items showed that the asymmetric encoding of English phonological contrast was quickly resolved by the Mandarin learners of English. Competition was symmetrical, especially in a later time window when the phonological knowledge is more involved in online processing. L2 listeners experience competition effects not only when the target consists of L1-ilicit loVN, but also the L1-licit ones, when the competitor contains illicit loVN. Since Escudero and colleagues were teaching Dutch learners of English nonwords and we use English words, differences in results across these experiments might reflect how listeners parse in the moment versus how they leverage their phonological knowledge in parsing known words.

Symmetry in Mandarin-English bilinguals processing of the loVN sequences is expected, if allophones are a unit that listeners can be deployed across languages, as is posited by the SLM-r (Flege & Bohn, 2021). The use of allophones in second language learning is also apparent in Llompart, Eger et al. (2021), who find that L2 learners of German process German words exhibiting free allophonic variation in ways that map the German allophones to their L1 phonemic categories. The reapplication of allophones from an L1 to an L2 unifies second language learning with what has been observed in experimental approaches to second dialect learning (German et al., 2013).

Although Darcy et al. (2013) reported asymmetric mapping in L2 lexical representations, no asymmetry was found for *advanced* L2 learners of German, in contrast with intermediate learners where the asymmetric mapping was observed. The absence of asymmetry in lexical processing suggests a learning stage in the process of establishing a "more efficient, target-like lexical access" and gradually encoding the phonological contrast in more detail (p. 42). In our current study, the L2 learners have been living in an English-speaking country for an average of 2.5 years and have been learning English for an average of 19.2 years since a mean age of 8.6. Similarly, the advanced learners in Darcy et al. are also immersed in a German-speaking environment for more than one year. Our participants' relatively higher level of proficiency in English and the substantial amount of English exposure may account for the absence of asymmetric competition. In comparison, the L2 learners in Weber and Cutler (2004) and Cutler et al. (2006) acquired English as a foreign language in a classroom setting for 7.8 years, beginning at a mean age of 11, and they lived in the Netherlands all their lives (p. 4). Such asymmetry in lexical encoding can be resolved with more L2 input as L2 proficiency improves.

An alternative interpretation of symmetric competition is to say that L2 learners have not encoded a phonemic contrast between the nasals at all, at least not in the loVN sequence.[8]

---

[8] We thank an anonymous reviewer for bringing up this alternative interpretation.

Escudero et al. (2008) interpreted a lack of asymmetry in their audio-only condition as evidence that their L1 Dutch-L2 English group had not established a lexical contrast between /ɛ/ and /æ/ (p. 358). Similar reasoning is found in Llompart and Reinisch (2017, 2020). In addition, the pattern observed in the current study also mirrors the results of Eger, Mitterer et al. (2019) (i.e., robust explicit discrimination, distinction not used in online processing), who examine L1-Italian-L2-German speakers' perception and production for German /h/ and /ʔ/ that do not have obvious counterparts in Italian, although those L2 learners had a much later onset of acquisition (at a mean age of 19.8; p. 5). For our sample of L1 Mandarin-English L2 listeners, we are hesitant to claim that they have not acquired a phonemic contrast in the English nasals. Given their level of English proficiency, we would fully anticipate this population to be able to meaningfully distinguish English minimal pairs like *ram, ran*, and *rang*, though we recognize we do not demonstrate this directly here. The empirical plot of eye-tracking data in **Figure 5** provides the time point where eye movement launches when driven by the acoustic cue of the nasal onset in the stimuli (gray dashed lines). L1 Mandarin listeners showed divergence between the target and the competitor at about the same time point (or slightly after) in ManEng and EngMan conditions, suggesting that they were able to rely on the nasal contrast to distinguish between words. However, L1 Mandarin listeners continued to experience a greater amount of competition at a later time point, perhaps due to L2 global fuzziness in phono-lexical encoding for low vowel + nasal sequences.

Nevertheless, it could be that the allophony leveraged from Mandarin to English is about the nasals and not the vowels. Although most of the literature on Mandarin phonology postulates that the nasal difference is phonemic and the vowels are allophonic, conditioned by the nasal (Duanmu, 2007), there is some evidence that the vowels tend to bear the contrast, as the nasal codas are prone to deletion in production (Zhang & Tessier, 2024). That the vowels may be phonemic in Mandarin and the nasals allophonic is, perhaps, also suggested in the results from Experiment 1. Mandarin listeners were more accurate and faster at responding to pairs that differed in vowels than nasals. Being better able to detect a difference in allophones than phonemes would be surprising, given the literature (e.g., Boomershine et al., 2008). Speculating that the vowels are phonemic and the nasals are allophonic in Mandarin is a rather bold claim, however, and the current study was not designed around adjudicating this difference. We note, however, that Nesbitt (2023) provides a theoretical space for gradual phonological emergence, which might prove useful to further inquiries into this line of research.

4.3. Global fuzziness in competition and the role of lexical development

We must also make sense of our findings that loVN sequences show more competition for Mandarin-English bilinguals than L1 English listeners. We interpret this observation of increased competition as evidence that non-native phonological contrasts are fuzzily encoded in the L2 lexical representation (Gor et al., 2021; Cook et al., 2016; Gor, 2018). Our results support the hypothesis of an overall imprecision in encoding L2 phonological contrasts at a global level, independent of local ambiguity in relation to the mapping to specific L1 phonemes (Gor et al., 2021; Cook et al., 2016; Gor, 2018; Darcy et al., 2024). Our results further reveal that a larger vocabulary size feeds more robust phonological generalizations, predicting less competition and less fuzziness in phono-lexical encoding. The results support the existing findings that lexical development facilitates more detailed phonological representations and more robust phono-lexical encoding of contrasts (Daidone & Darcy, 2021; Llompart, 2021). This complements recent work suggesting that changes in spoken word recognition are better understood as reflections of increased ability or knowledge of language (Kutlu et al., 2024).

The online use of abstract phonological knowledge should arise online from the relevant distributions based on their accumulated lexical experience (Cutler et al., 2010). The existing body of literature focuses on L2 fuzziness fed by mismatches between the L1 and L2 phonological systems (Gor et al., 2021), which is perhaps further fed by perceptual imprecision at the lexical level. The current findings suggest that the level of fuzziness in the phono-lexical representations decreases through word learning by simply increasing the size of the lexicon. Future research can examine other factors that contribute to L2 fuzziness, independent of the influence of the L1 inventory.

It is important to note that nasals in coda positions have relatively weak perception cues in general (Bond, 1981; Chen et al., 2012). Mandarin, in particular, has a relatively weak place of articulation cues for coda nasals, such that the coda closure is often attenuated or deleted, and the preceding vowel is nasalized and possibly lengthened in production (Chen, 2000; Fang 2004; Luo et al., 2020). In Mandarin, the place of the nasal can be cued from the backness of the low vowel, as only loVN with matched backness is permitted. If the listeners rely more on the vowel cue in predicting the nasal place in Mandarin, it may be the case that Mandarin listeners have a certain level of fuzziness in encoding the exact place of the nasal in loVN in their L1 as well. It is possible that the observed imprecise phonological encoding of loVN contrast exists in the L1 and is not just limited to L2 representation. These questions regarding the competition present in L1 vs. L2 representation warrant future study, to be considered alongside findings for asymmetry in L1 data as in Barrios and Hayes-Harb (2021).

4.4. Remaining questions

Orthography can play a role in L2 phono-lexical representation, as neatly illustrated by the contrasting results of Weber and Cutler (2004) and Escudero et al. (2008). The Ontogenesis Model proposed by Bordag et al. (2022) argues that there is a link established between the phonological and orthographic forms of L2 lexicon. Depending on the grapheme phoneme relations in the L1 and L2, listeners may benefit from the orthographic information in L2 phonological encoding when their L1 has congruent graphic-phonological correspondences (Hayes-Harb et al., 2010).

On the other hand, when a common grapheme is used in both the L1 and the L2, it might be associated with the same phonological category as in the L1, which may actually lead to the inaccurate mapping between the orthography and phonology in an L2 (Best & Tyler, 2007). In the current study, the lack of transparency in the spelling of the loVN in English, as well as a different grapheme-phoneme relation in Mandarin *pinyin* (the romanized writing system of Mandarin), can contribute to the L2 fuzziness in the encoding of the phonological form of loVN. For example, in English, nasals such as /ŋ/ are usually spelled as *ng* word finally, but, as *n*, sometimes word medially. In Mandarin *pinyin*, /ŋ/ is represented consistently as *ng* while /n/ as *n*, though not all Mandarin varieties produce the nasal codas with the same phone. Since the current eye-tracking study presented the orthographic form of the lexical items as the visual stimuli, the L2 mapping between orthographic and phonological forms can be activated, such that L2 learners may access a non-target-like phonological category because of the inconsistent grapheme-to-phoneme correspondence in English in comparison with Mandarin *pinyin* system. It is also worth noting that the current study focuses on the specific contrast of low vowel + nasal sequence in Mandarin and English, as the pattern raises interesting theoretical points to test that build upon the previous literature (e.g., Weber & Cutler, 2004; Cutler et al., 2006; Darcy et al., 2013). It is yet unknown how the results from the current study can be generalized to other patterns and languages.

The current study provides evidence for the abstract phonological knowledge stored in the lexicon; namely, the larger the lexicon, the better encoding and processing of L1 and L2 phono-lexical contrasts. However, it is very likely that listeners, especially L1 listeners, attend to the subphonemic cues in the auditory signals, such as the nasalized cues in the vowel, before the nasal information unfolds (e.g., Beddor, 2009). This seems to appear in the early-on asymmetric competition in the L1 English group (occurred in the 220–340 ms in **Table 9**), where the ManEng condition underwent more competition than EngMan condition (seen from **Figure 5**). L1 English listeners are sensitive to the coarticulatory vowel cues even before the start of the nasal in the stimuli. Given that vowels in North American English often have varying degrees of subphonemic cues to the upcoming consonant (Nesbitt, 2023), L1 English listeners may readily attend to the vowel differences that index the probability of a particular coda consonant.

5. Conclusions

In this study, sequential Mandarin-English bilinguals were shown to be similarly phonetically sensitive to low vowel nasal sequences that are phonemic in English but patterned as allophonic variants in Mandarin, as English L1 listeners in an AX discrimination task. The differences observed in discrimination were small and unrelated to the vowels. Moving up the ladder of abstraction, a visual world paradigm demonstrates that Mandarin-English bilinguals experience more competition for words containing the low vowel nasal sequence than English L1 listeners

in a spoken word recognition task. The timing of the competition for the Mandarin-English bilinguals suggests that the influence of Mandarin phonology affects phono-lexical encoding and not lower-level phonetic encoding. The symmetric L2 competition between the L1-licit and L1-illicit contexts suggests that L2 listeners repurpose L1 allophones to phonemes in the L2. Additionally, smaller receptive vocabulary size estimates predicted more competition for both Mandarin-English bilinguals and English L1 listeners. This suggests that the phonological knowledge generalized across larger lexicons facilitates spoken word recognition, resolving competition more quickly for first and second language learners.

Acknowledgements

We would like to thank the audience at the LabPhon 19 and INTERSPEECH 2023 for their invaluable feedback and questions about this research project. We appreciate feedback at multiple stages and help with piloting from the members of the Speech in Context Lab and Child Phonology Lab at the University of British Columbia, as well as all students enrolled in the LING 530K (psycholinguistics and eye-tracking) course in Winter Term 2 2023 taught by the third author. We are grateful to Roger Lo for their insightful comments and Márton Sóskuthy for the help with the statistical analysis. The study is funded by a SSHRC Insight Grant (435-2023-0169) to the last author, a UBC internal grant to the third author, and the UBC Arts Graduate Research Award to the first author.

Competing interests

The authors have no competing interests to declare.

Authors' contributions

SZ: Conceptualization, Methodology, Software, Investigation, Data curation, Statistical Analysis, Visualization, Writing – Original draft preparation, Writing – Reviewing and Editing, Writing – Revising.

MB: Conceptualization, Supervision, Writing – Original draft preparation, Writing – Reviewing and Editing, Writing – Revising.

CH: Methodology, Supervision, Writing – Reviewing and Editing.

A-MT: Supervision, Writing – Reviewing and Editing.

References

Amengual, M. (2016). The perception and production of language-specific mid-vowel contrasts: Shifting the focus to the bilingual individual in early language input conditions. *International Journal of Bilingualism*, *20*(2), 133–152. https://doi.org/10.1177/1367006914544988

Arai, M., Van Gompel, R. P. G., & Scheepers, C. (2007). Priming ditransitive structures in comprehension. *Cognitive Psychology, 54*(3), 218–250. https://doi.org/10.1016/j.cogpsych.2006.07.001

Babel, M., & Johnson, K. (2010). Accessing psycho-acoustic perception and language-specific perception with speech sounds. *Laboratory Phonology*, *1*(1), 179–205. https://doi.org/10.1515/labphon.2010.009

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English lexicon project. *Behavior Research Methods, 39*(3), 445–459. https://doi.org/10.3758/BF03193014

Barr, D. J., Jackson, L., & Phillips, I. (2014). Using a voice to put a name to a face: The psycholinguistics of proper name comprehension. *Journal of Experimental Psychology, 143*(1), 404–413. https://doi.org/10.1037/a0031813

Barrios, S., & Hayes-Harb, R. (2021). L2 processing of words containing English /æ/-/ɛ/ and /l/-/ɹ/ contrasts, and the uses and limits of the auditory lexical decision task for understanding the locus of difficulty. *Frontiers in Communication, 6*(144). https://doi.org/10.3389/fcomm.2021.689470

Beddor, P. S. (2009). A coarticulatory path to sound change. *Language*, *85*(4), 785–821. https://doi.org/10.1353/lan.0.0165

Berent, I., Steriade, D., Lennertz, T., & Vaknin, V. (2007). What we know about what we have never heard: Evidence from perceptual illusions. *Cognition*, *104*(3), 591–630. https://doi.org/10.1016/j.cognition.2006.05.015

Best, C. T., & Tyler, M. (2007). Nonnative and second-language speech perception. In O. Bohn and M. Munro (Eds.), *Language experience in second language speech learning* (pp. 13–34). John Benjamins. https://doi.org/10.1075/lllt.17.07bes

Birdsong, D., Gertken, L. M., & Amengual, M. (2012). *Bilingual language profile: An easy-to-use instrument to assess bilingualism*. COERLL, University of Texas at Austin. https://sites.la.utexas.edu/bilingual/

Boersma, P., & Weenink, D. (2020). *Praat: Doing phonetics by computer* (Version 6.1.28) [Computer Software]. https://www.praat.org

Bond, Z. (1981). Listening to elliptic speech: Pay attention to stressed vowels. *Journal of Phonetics, 9*(1), 89–96. https://doi.org/10.1016/S0095-4470(19)30929-5

Boomershine, A., Hall, K. C., Hume, E., & Johnson, K. (2008). The impact of allophony versus contrast on speech perception. In P. Avery, B. E. Dresher, & K. Rice (Eds.), *Contrast in Phonology: Theory, Perception, Acquisition* (pp. 143–172). De Gruyter Mouton. https://doi.org/10.1515/9783110208603.2.145

Bordag, D., Gor, K., & Opitz, A. (2022). Ontogenesis model of the L2 lexical representation. *Bilingualism: Language and Cognition*, *25*(2), 185–201. https://doi.org/10.1017/S1366728921000250

Broersma, M. (2012). Increased lexical activation and reduced competition in second-language listening. *Language and Cognitive Processes, 27*(7–8), 1205–1224. https://doi.org/10.1080/01690965.2012.660170

Bundgaard-Nielsen, R. L., Best, C. T., Kroos, C., & Tyler, M. D. (2012). Second language learners' vocabulary expansion is associated with improved second language vowel intelligibility. *Applied Psycholinguistics*, *33*(3), 643–664. https://doi.org/10.1017/S0142716411000518

Bundgaard-Nielsen, R. L., Best, C. T., & Tyler, M. D. (2011). Vocabulary size is associated with second-language vowel perception performance in adult learners. *Studies in Second Language Acquisition*, *33*(3), 433–461. https://doi.org/10.1017/S0272263111000040

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Bürkner, P.-C., & Vuorre, M. (2019). Ordinal regression models in psychology: A tutorial. *Advances in Methods and Practices in Psychological Science*, *2(*1), 77–101. https://doi.org/10.1177/2515245918823199

Byun, T. M. A., & Tessier, A. M. (2016). Motor influences on grammar in an emergentist model of phonology. *Language and Linguistics Compass*, *10*(9), 431–452. https://doi.org/10.1111/lnc3.12205

Chang, C. B. (2019). The phonetics of second language learning and bilingualism. In W. F. Katz & P. F. Assmann (Eds.), *The Routledge handbook of phonetics* (pp. 427–447). Routledge. https://doi.org/10.4324/9780429056253-16

Chen, M. Y. (2000). Acoustic analysis of simple vowels preceding a nasal in Standard Chinese. *Journal of Phonetics*, *28*(1), 43–67. https://doi.org/10.1006/jpho.2000.0106

Chen, Y., Kapatsinski, V., & Guion-Anderson, S. (2012). *Acoustic cues of vowel quality to coda nasal perception in Southern Min*. *Proceedings of Interspeech,* Portland, OR, United States*, 1412–1415*. https://doi.org/10.21437/Interspeech.2012-49

Cook, S. V., Pandža, N. B., Lancaster, A. K., & Gor, K. (2016). Fuzzy nonnative phonolexical representations lead to fuzzy form-to-meaning mappings. *Frontiers in Psychol*ogy, *7*. https://doi.org/10.3389/fpsyg.2016.01345

Curtin, S., Goad, H., & Pater, J. V. (1998). Phonological transfer and levels of representation: the perceptual acquisition of Thai voice and aspiration by English and French speakers. *Second Language Research*, *14*(4), 389–405. https://doi.org/10.1191/026765898674095369

Cutler, A., Eisner, F., McQueen, J. M., & Norris, D. (2010). How abstract phonemic categories are necessary for coping with speaker-related variation. In C. Fougeron, B. Kuhnert, M. d'Imperio, & N. Vallee (Eds.), *Laboratory Phonology*, *10*(91–111). De Gruyter Mouton. https://doi.org/10.1515/9783110224917.1.91

Cutler, A., Weber, A., & Otake, T. (2006). Asymmetric mapping from phonetic to lexical representations in second-language listening. *Journal of Phonetics*, *34*(2), 269–284. https://doi.org/10.1016/j.wocn.2005.06.002

Daidone, D., & Darcy, I. (2021). Vocabulary size is a key factor in predicting second language lexical encoding accuracy. *Frontiers in Psychology, 12*, 688356. https://doi.org/10.3389/fpsyg.2021.688356

Darcy, I., Daidone, D., & Kojima, C. (2013). Asymmetric lexical access and fuzzy lexical representations in second language learners. *The Mental Lexicon*, *8*(3), 372–420. https://doi.org/10.1075/ml.8.3.06dar

Darcy I., Dekydtspotter L., Sprouse, R. A., Glover, J., Kaden, C., McGuire, M., & Scott, J. H. (2012). Direct mapping of acoustics to phonology: On the lexical encoding of front rounded vowels in L1-English–L2-French acquisition. *Second Language Research, 28*(1)*,* 5–40. https://doi.org/10.1177/0267658311423455

Darcy, I., Llompart, M., Hayes-Harb, R., Mora, J. C., Adrian, M., Cook, S., & Ernestus, M. (2024). Phonological processing and the L2 mental lexicon: Looking back and moving forward. *Studies in Second Language Acquisition*, *47*(1), 361–387. https://doi.org/10.1017/S0272263124000482

Darcy, I., & Thomas, T. (2019). When blue is a disyllabic word: Perceptual epenthesis in the mental lexicon of second language learners. *Bilingualism: Language and Cognition*, *22*(5), 1141–1159. https://doi.org/10.1017/S1366728918001050

Davis, M. H., & Johnsrude, I. S. (2007). Hearing speech sounds: Top-down influences on the interface between audition and speech perception. *Hearing Research, 229*(1), 132–147. https://doi.org/10.1016/j.heares.2007.01.014

Desmeules-Trudel, F., & Zamuner, T. S. (2023). Spoken word recognition in a second language: The importance of phonetic details. *Second Language Research*, *39*(2), 333–362. https://doi.org/10.1177/02676583211030604

Dink, J. W., & Ferguson, B. (2015). *eyetrackingR* [Computer software]. An R Library for Eye-tracking Data Analysis. http://www.eyetrackingr.com

Duanmu, S. (2007). *The phonology of Standard Chinese*. Oxford University Press. https://doi.org/10.1093/oso/9780199215782.001.0001

Dupoux, E., Parlato, E., Frota, S., Hirose, Y., & Peperkamp, S. (2011). Where do illusory vowels come from? *Journal of Memory and Language*, *64*(3), 199–210. https://doi.org/10.1016/j.jml.2010.12.004

Eger, N. A., Mitterer, H., & Reinisch, E. (2019). Learning a new sound pair in a second language: Italian learners and German glottal consonants. *Journal of Phonetics*, *77*, 100917. https://doi.org/10.1016/j.wocn.2019.100917

Escudero, P. (2009). The linguistic perception of similar L2 sounds. In P. Boersma & S. Hamann (Eds.), *Phonology in Perception* (pp. 151–190). Mouton de Gruyter. https://doi.org/10.1515/9783110219234.151

Escudero, P., Hayes-Harb, R., & Mitterer, H. (2008). Novel second-language words and asymmetric lexical access. *Journal of Phonetics, 36*(2), 345–360. https://doi.org/10.1016/j.wocn.2007.11.002

Fang, Q. (2004). A Study on Nasal Coda Loss in Continuous Speech. *The 8th International Conference on Spoken Language Processing.* Jeju Island, Korea. https://doi.org/10.21437/Interspeech.2004-67

Fischer, B. (1992). Saccadic reaction time: Implications for reading, dyslexia and visual cognition. In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading* (pp. 31–45). Springer Verlag. https://doi.org/10.1007/978-1-4612-2852-3_3

Flege, J. E., & Bohn, O.-S. (2021). The revised speech learning model (slm-r). In R. Wayland (Ed.), *Second language speech learning: Theoretical and empirical progress* (pp. 3–83). Cambridge University Press. https://doi.org/10.1017/9781108886901.002

German, J. S., Carlson, K., & Pierrehumbert, J. B. (2013). Reassignment of consonant allophones in rapid dialect acquisition. *Journal of Phonetics, 41*(3–4), 228–248. https://doi.org/10.1016/j.wocn.2013.03.001

Gor, K. (2018). Phonological priming and the role of phonology in nonnative word recognition. *Bilingualism: Language and Cognition*, *21*(3), 437–442. https://doi.org/10.1017/S1366728918000056

Gor, K., Cook, S., Bordag, D., Chrabaszcz, A., & Opitz, A. (2021). Fuzzy lexical representations in adult second language speakers. *Frontiers in Psychology*, *12*, 732030. https://doi.org/10.3389/fpsyg.2021.732030

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds ‘l’ and ‘r’. *Neuropsychologia, 9*(3), 317–323. https://doi.org/10.1016/0028-3932(71)90027-3

Hammerly, C., Staub, A., & Dillon, B. (2022). Person-based prominence guides incremental interpretation: Evidence from obviation in Ojibwe. *Cognition*, *225*, 105122. https://doi.org/10.1016/j.cognition.2022.105122

Hansen, J. G. (2001). Linguistic constraints on the acquisition of English syllable codas by native speakers of Mandarin Chinese. *Applied Linguistics*, 22(3), 338–365. https://doi.org/10.1093/applin/22.3.338

Harnsberger, J. D. (2000). A cross-language study of the identification of non-native nasal consonants varying in place of articulation. *Journal of the Acoustical Society of America*, *108*(2), 764–783. https://doi.org/10.1121/1.429610

Harnsberger, J. D. (2001). On the relationship between identification and discrimination of non-native nasal consonants. *Journal of the Acoustical Society of America*, *110*(1), 489–503. https://doi.org/10.1121/1.1371758

Hayes-Harb, R. (2005). Optimal L2 speech perception: Native speakers of English and Japanese consonant length contrasts. *Journal of Language and Linguistics*, *4*(1), 1–29.

Hayes-Harb, R., & Masuda, K. (2008). Development of the ability to lexically encode novel second language phonemic contrasts. *Second Language Research*, *24*(1), 5–33. https://doi.org/10.1177/0267658307082980

Hayes-Harb, R., Nicol, J., & Barker, J. (2010). Learning the phonological forms of new words: Effects of orthographic and auditory input. *Language and Speech*, *53*(3), 367–381. https://doi.org/10.1177/0023830910371460

Huang, T., & Johnson, K. (2010). Language specificity in speech perception: Perception of Mandarin tones by native and nonnative listeners. *Phonetica, 67*(4), 243–267. https://doi.org/10.1159/000327392

Huettig, F., & McQueen, J. M. (2007). The tug of war between phonological, semantic and shape information in language-mediated visual search. *Journal of Memory and Language*, *57*(4), 460–482. https://doi.org/10.1016/j.jml.2007.02.001

Ito, A., & Knoeferle, P. (2023). Analysing data from the psycholinguistic visual-world paradigm: Comparison of different analysis methods. *Behavior Research Methods*, *55*(7), 3461–3493. https://doi.org/10.3758/s13428-022-01969-3

Johnson, K., & Babel, M. (2010). On the perceptual basis of distinctive features: Evidence from the perception of fricatives by Dutch and English speakers. *Journal of Phonetics*, *38*(1), 127–136. https://doi.org/10.1016/j.wocn.2009.11.001

Kruschke, J. K., & Liddell, T. M. (2018). Bayesian data analysis for newcomers. *Psychonomic Bulletin & Review*, *25*(1), 155–177. https://doi.org/10.3758/s13423-017-1272-1

Kutlu, E., Klein-Packard, J., Jeppsen, C., Tomblin, J. B., & McMurray, B. (2024). The development of real-time spoken and written word recognition derives from changes in ability, not maturation. *Cognition, 251*, 105899. https://doi.org/10.1016/j.cognition.2024.105899

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods*, *44*, 325–343. https://doi.org/10.3758/s13428-011-0146-0

Liu, Y. (2016). Oral gestural reduction of English nasal coda produced by Mandarin English-as-a foreign-language learners. *Journal of the Acoustical Society of America*, *140*(4), 3339–3339. https://doi.org/10.1121/1.4970656

Llompart, M. (2021). Phonetic categorization ability and vocabulary size contribute to the encoding of difficult second-language phonological contrasts into the lexicon. *Bilingualism: Language and Cognition*, *24*(3), 481–496. https://doi.org/10.1017/S1366728920000656

Llompart, M., Eger, N. A., & Reinisch, E. (2021). Free allophonic variation in native and second language spoken word recognition: The case of the German rhotic. *Frontiers in Psychology, 12*, 711230. https://doi.org/10.3389/fpsyg.2021.711230

Llompart, M., & Reinisch, E. (2017). Articulatory information helps encode lexical contrasts in a second language. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(5), 1040. https://doi.org/10.1037/xhp0000383

Llompart, M., & Reinisch, E. (2020). The phonological form of lexical items modulates the encoding of challenging second-language sound contrasts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(8), 1590. https://doi.org/10.1037/xlm0000832

Luo, M.,Van de Weijer, J., & Sloos, M. (2020). A perceptually-based OT approach to opacity in Mandarin Chinese nasal rhymes. In D. Chen & D. Bell (Eds.), *Explorations of Chinese Theoretical and Applied Linguistics* (pp. 8–43*)*. Cambridge Scholars Publishing.

Magnuson, J. S. (2019). Fixations in the visual world paradigm: Where, when, why? *Journal of Cultural Cognitive Science*, *3*(2), 113–139. https://doi.org/10.1007/s41809-019-00035-3

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. *Journal of Neuroscience Methods, 164*(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

McQueen, J. M. (2005). Speech perception. In K. Lamberts & R. Goldstone (Eds.), *The handbook of cognition* (pp. 255–275). Sage. https://doi.org/10.4135/9781848608177.n11

McQueen, J. M., & Viebahn, M. C. (2007). Tracking recognition of spoken words by tracking looks to printed words. *Quarterly Journal of Experimental Psychology*, *60*(5), 661–671. https://doi.org/10.1080/17470210601183890

Metsala, J. L., & Walley, A. C. (1998). Spoken vocabulary growth and the segmental restructuring of lexical representations: Precursors to phonemic awareness and early reading ability. In J. L. Metsala and L. C. Ehri (Eds.), *Word Recognition in Beginning Literacy* (pp. 89–120). Lawrence Erlbaum.

Munson, B., Edwards, J., & Beckman, M. E. (2011). Phonological representations in language acquisition: climbing the ladder of abstraction. In A. C. Cohn, C. Fougeron, & M. K. Huffman (Eds.), *The Oxford handbook of laboratory phonology* (pp. 288–309). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199575039.013.0012

Narayan, C. R. (2008). The acoustic–perceptual salience of nasal place Contrasts*. Journal of Phonetics*, *36*(1), 191–217. https://doi.org/10.1016/j.wocn.2007.10.001

Narayan, C. R., Werker, J. F., & Beddor, P. S. (2010). The interaction between acoustic salience and language experience in developmental speech perception: Evidence from nasal place discrimination. *Developmental Science*, *13*(3), 407–420. https://doi.org/10.1111/j.1467-7687.2009.00898.x

Nesbitt, M. (2023). Phonological emergence and social reorganization: Developing a nasal/æ/ system in Lansing, Michigan. *Language Variation and Change*, *35*(3), 273–297. https://doi.org/10.1017/S0954394523000182

Nicenboim, B., & Vasishth, S. (2016). Statistical methods for linguistic research: Foundational ideas—Part II. *Language and Linguistics Compass*, *10*(11), 591–613. https://doi.org/10.1111/lnc3.12207

R Core Team. (2022). *R: A Language and Environment for Statistical Computing.* [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Salverda, A. P., Kleinschmidt, D., & Tanenhaus, M. K. (2014). Immediate effects of anticipatory coarticulation in spoken-word recognition. *Journal of Memory and Language*, *71*(1), 145–163. https://doi.org/10.1016/j.jml.2013.11.002

Shepard, R. N., Kilpatric, D. W., & Cunningham, J. P. (1975). The internal representation of numbers. *Cognitive Psychology*, *7*(1), 82–138. https://doi.org/10.1016/0010-0285(75)90006-7

Sonderegger, M. (2023). *Regression Modeling for Linguistic Data*. MIT Press.

Stoel-Gammon, C. (2011). Relationships between lexical and phonological development in young children. *Journal of Child Language*, *38*(1), 1–34. https://doi.org/10.1017/S0305000910000425

Turner, J. (2022). Analysing the relationship between L2 production and different stages of L2 processing: Eye-tracking and acoustic evidence for a novel contrast. *Journal of Phonetics*, *91*, 101134. https://doi.org/10.1016/j.wocn.2022.101134

Veivo, O., Järvikivi, J., Porretta, V., & Hyönä, J. (2016). Orthographic activation in L2 spoken word recognition depends on proficiency: Evidence from eye-tracking. *Frontiers in Psychology*, *7*, 1120. https://doi.org/10.3389/fpsyg.2016.01120

Vihman, M. (2017). Learning words and learning sounds: Advances in language development. *British Journal of Psychology*, *108*(1), 1–27. https://doi.org/10.1111/bjop.12207

Vihman, M., & Croft, W. (2007). Phonological development: Toward a “radical” templatic phonology. *Linguistics, 45*(4), 683–725. https://doi.org/10.1515/ling

Voeten, C. (2022). *permutes: Permutation Tests for Time Series Data* (Version 2.8). [Computer software]. R Project.

Weber, A. (2001). Help or hindrance: How violation of different assimilation rules affects spoken-language processing. *Language & Speech, 44*(1), 95–118. https://doi.org/10.1177/00238309010440010401

Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language*, *50*(1), 1–25. https://doi.org/10.1016/S0749-596X(03)00105-0

Werker, J. F., & Curtin, S. (2005). PRIMIR: A developmental framework of infant speech processing. *Language Learning and Development, 1*(2), 197–234. https://doi.org/10.1080/15475441.2005.9684216

Werker, J. F., Gilbert, J. H., Humphrey, K., & Tees, R. C. (1981). Developmental aspects of cross-language speech perception. *Child Development, 52*(1), 349–355. https://doi.org/10.2307/1129249

Werker, J. F., & Logan, J. S. (1985). Cross-language evidence for three factors in speech perception. *Perception & Psychophysics*, *37*(1), 35–44. https://doi.org/10.3758/BF03207136

Woods, K. J., Siegel, M. H., Traer, J., & McDermott, J. H. (2017). Headphone screening to facilitate web-based auditory experiments. *Attention, Perception, & Psychophysics, 79*, 2064–2072. https://doi.org/10.3758/s13414-017-1361-2

Wu, M., Sloos, M., & Weijer, J. van de (2016). The perception of the English alveolar-velar nasal coda contrast by monolingual versus bilingual Chinese speakers. *The 10th International Symposium on Chinese Spoken Language Processing.* Tianjing, China. https://doi.org/10.1109/ISCSLP.2016.7918401

Zhang, S., & Tessier, A. M. (2024). Modeling the consequences of an L1 grammar for L2 production: Simulations, variation, and predictions. *Frontiers in Language Sciences*, *3*, 1327600. https://doi.org/10.3389/flang.2024.1327600