(Mis) perception of consonant clusters and short vowels in English as a foreign language

Abstract: Research addressing second language (L2) speech is expanding. Studies increasingly demonstrate that a learner's first language (L1) filters the L2 input, resulting in learners misperceiving what they have heard. This L1 filter can result in learners perceiving sounds not actually present in the input. We report on a study which explored English consonant clusters and short, unstressed vowel perception of 70 Arabic-, Mandarin-, Spanish-speaking foreign language learners and 19 native English speakers. These are the vowels which speakers from two of the L1s typically insert in their production of English to break up L1-disallowed consonant clusters and the schwa which is documented to cause both perception and production problems. Results show that participants misperceive stimuli containing consonant clusters and counterparts where clusters are broken up by epenthetic/prothetic elements. In line with Sakai, Mari & Colleen Moorman 2018. We call for the inclusion of such findings on perception in pedagogical advice on pronunciation.

Keywords: consonant clusters; illusory vowels and epenthesis; perceptual illusion; second language perception; short vowels and schwa

---

We thank our participants and teachers for taking part in the study. We thank the audience of our presentation, 'Play it by ear': A cross-linguistic study of perception of schwa and reduced epenthetic vowels in L2 English, at the British Association for Applied Linguistics Conference, Leeds 1st September 2017. We also express our sincere gratitude to the editor of IRAL for his support and the three anonymous reviewers for their insightful comments which helped enhance the clarity and depth of our paper. Thanks also goes to Antony Xinliang Jiang for discussing issues pertaining to Mandarin phonology with us.

---

*Corresponding author: Alex Ho-Cheong Leung, Department of Humanities, Northumbria University, Newcastle upon Tyne, UK, E-mail: alex.ho-cheong.leung@northumbria.ac.uk.
https://orcid.org/0000-0003-2175-9860
MarthaYoung-Scholten,ChloeNashandOliviaOuthwaite,SchoolofEnglishLiterature,Language and Linguistics, Newcastle University, Newcastle upon Tyne, Tyne and Wear, UK
Wael Almurashi and Saleh Ghadanfari, School of Education, Communication, and Language Sciences, Newcastle University, Newcastle upon Tyne, Tyne and Wear, UK

1. Introduction

Research into the acquisition of an L2 sound system has now caught up to work on other language domains and has become a lively field of inquiry whose evolution can be traced from Ioup and Weinberger's (1987) edited volume on interlanguage phonology through Hansen-Edwards and Zampini (2008), Gut (2009), Arabski and Wojtszek (2011), Gabryś-Barker (2011), Derwing and Munro (2015), Kennedy and Trofimovich (2017), Nyvad et al. (2019), Wayland (2021), Darcy (in prep.), and Shea (in prep.) which have introduced a range of issues, models, hypotheses and theories. The early lag in research relates to the observed inevitability of a foreign accent with increasing age and the assumption that its source was the L1 and purely articulatory. Since the 1980s, research has blossomed, examples of which also include the thriving decades-old triennial New Sounds conference series, emergence of new conference series (e.g. http://psllt.org/index.php/psllt/2018), and journals (e.g. The Journal of Second Language Pronunciation). Of course, non-native speakers can still be identified as such by their accents (Flege et al. 1995; Ioup 1984). Early L2 phonology research, largely driven by Lado's (1957) the Contrastive Analysis hypothesis, focused on aspects of the L1 differing from the L2. Initially concentrating on the production of segments and later moving on to perception, researchers introduced models seeking to account for where native language-based difficulties would and would not occur, e.g. the Speech Learning Model (Flege 1995) and Perceptual Assimilation Model (Best 1995). Under both, the learner misperceives the input due to L1-based miscategorization of L2 sounds which differ from the L1 but are similar and confused.[1]

Research now indicates that a foreign accent is neither based entirely on the learner's L1 nor is it simply articulatory (i.e., purely a production issue). Studies also indicate that phonological universals contribute to both younger and older learners' production (Eckman 1981; Li and Post 2014; Tarone 1979), a line of reasoning pursued in Major's Ontogeny Phylogeny Model (2008) and Optimality Theory (Hancin-Bhatt 2008). Mounting evidence indicates that universals also apply in the perception of target language input (e.g., Altenberg 2005; Broselow et al. 1998; Escudero and Boersma 2004).[2] Given the solid evidence demonstrating the inter-relatedness of

---

[1] L2 acquisition of phonology has long been observed to become more effortful upon closure of a purported critical period (Lenneberg 1967) though the age at which this period closes continues to be debated, e.g. at puberty (Oyama 1976), age 15 (Patkowski 1990) or age six (Long 1990); see Munoz and Singleton's (2011) overview. The idea of period of heightened resonance with the input continues to attract attention but because our study participants were exposed at similar ages, we do not address age.

[2] Studies also continue to confirm that older L2 learners tend not to acquire a native-like phonology (Abrahamsson and Hyltenstam, 2009) and research explores age alongside other internal and external influences on learners' interlanguage phonologies such as L1 use (Flege and MacKay,2004);

speech perception and production (e.g. Broselow and Kang 2013; Casserly and Pisoni 2010; Feng 2020; Huensch and Tremblay 2015; Shinohara and Iverson 2018),researchers have recently considered whether strategies such as epenthesis and deletion employed by learners to simplify L2 consonant clusters have a perceptual origin(Baker and Trofimovich 2006; see also Broselow and Kang 2013; Hansen-Edwards and Zampini 2008; Kissling 2015).

Considering a set of L1s different from previous studies (see below), and not only the vowels used in L1 epenthesis but also including schwa, our study explores 1) whether foreign language learners of English experience perceptual difficulties with consonant clusters, and 2) whether such difficulties (if present) are related to their L1 phonological system. The study investigated the perception of L2 English consonant clusters in conjunction with certain unstressed vowels that studies of production have long shown learners use to modify consonant clusters disallowed in their L1s, for example, as in (1a–c).[3] While learners also modify clusters through consonant deletion (common for Mandarin speakers), our focus is on vowel insertion, common in certain environments, in Spanish before a word-initial /s/ and Arabic, between consonants (in the example from Broselow, the Egyptian variety).

(1) a. Deletion: #CCV → CV, /dripi/ → [ripi]) (Lin 2003)

b. Prothesis: #CCV → VCCV, /speɪn/ → [espeɪn] (Carlisle 1998)

c. Epenthesis: #CCC → CVC, / flɔːr/ → [filɔːr] (Broselow 1983)

Participating in a perception experiment were 70 secondary school learners of English as a foreign language in China, Kuwait and Spain. The experiment probed whether they could detect differences between target clusters and clusters modified through the insertion of certain vowels. We wanted to see whether learners perceived the difference, for example, between 'Spain' and 'Espain', and 'floor' and 'filoor'.

In what follows, we consider how consonants align themselves in a given language before we discuss the relationship between production and perception, and then studies related to L2 learners' (mis)hearing what is not in the input, i.e., perceptual illusion. Next, we give the background of the three L1s in the study, outline our methodology and finally turn to data analysis and results. We conclude by laying out pedagogical implications of our study and studies similar to ours.

---

orthographic input (Bassettiet al.2015); social factors (Hansen2004); input (Piske and Young-Scholten 2009) and motivation and training (Bongaerts et al. 2000).

[3] Learners also substitute final consonants; for example, a final stop is devoiced and becomes less marked in this position, e.g. friend → frient (Simon 2010).

1.1 Consonant cluster and phonotactics

One source of difficulty with L2 syllables in languages such as English are L1s without consonant clusters or with different distribution of consonants in clusters. Earlier studies reveal that learners′ acquisition of a new syllable structure, regardlessofageofL2exposure,is also guided by universals which enable them to go beyond their L1 syllable structure (Broselow 1987; Eckman 1981; Tarone 1979).[4]In addition to use of strategies shown in (1), learners apply universal notions of maximization of onsets and markedness in favoring CV syllables in (1a) and (1c) and those ending with devoiced stops (endnote 3). The phonotactics of a given L1 also influence L2 acquisition in constraining what is allowed word-internally. For example, while neither Mandarin nor Japanese allows complex onsets, codas or final stops, only in Japanese are sequences of (identical and adjacent) stops (geminates) allowed word-internally as in multisyllabic words such as [kitta] ′cut; sliced′.[5]

1.2 Perception and production

The majority of L2 studies have aimed to explain patterns found in learners′ oral production. As argued above, we need not only probe how learners′ L1s affect their production but also bias their listening. Until recently, studies of speech production and speech perception had ″progressed more or less independently for more than 60 years″ (Casserly and Pisoni 2010: 629). Gut′s (2009) comprehensive review of studies of L2 speech published over 39 years in 16 international journals uncovers a mere 29 studies which ″describe the perceptual abilities of non-native speakers or focus on the relationship between perception and production″ (2009: 40). Emerging neurological evidence (e.g., Calabrese 2012) points to a close interaction between perception and production. Close interaction is supported by behavioral empirical research which, for example, demonstrates that adult L2 learners who struggle to produce certain sounds cannot perceive them either (Flege 1995; Strange 1995). A case in point is L1 Japanese

[4] We thank the reviewer for pointing out that motor theory (Liberman and Mattingly 1985) and/or simulation theory (Gambi and Pickering 2013) can inform this discussion. We have situated our discussion in universals based on our own theoretical background (see also Parker 2015).

[5] The study of loan word phonology provides another window on a language′s phonotactics. Where words resulting from borrowing conform to the phonological constraints of the L1, e.g. cup → cuppu in Japanese (Kubozono 1999), this is evidence that the L1 biases perception and in turn results in L1-based production (see Peperkamp and Dupoux 2003).

learners' struggles with L2 English /r/ and /l/; only one liquid exists in Japanese. However, the lateral is available phonetically through the existence of a flap [ɾ] allophone. Learners report difficulties hearing the difference between /r/ and /l/ in English, along with difficulty in producing them (Bradlow 2008). Sakai and Moorman's (2018) meta-analysis convincingly demonstrates that perception training can lead to improvement in production as well, once again suggesting the intertwined nature between the two domains (see the pedagogical implications section for elaboration). Other evidence supporting the inter-relatedness between speech perception and production includes the phonetic convergence observed among interlocutors through multi-turn exchanges (Casserly and Pisoni 2010) and speakers' fine-grained (often sub-conscious) sensory-motor adaptions of their own speech in response to external factors such as background noise (Houde and Jordan 2002). That said, one should be mindful that the relationship between the two domains in L2 speech learning is far from straightforward. In fact, "[t]he degree to which L2 perception and L2 production processes overlap, the nature of this overlap, and the manner in which the perception-production relationship differs between L1 and L2 learning remain some of the basic questions in research on L2 speech learning" (Chang 2019: 439).

1.3 L2 perception of consonant clusters

English words such as 'spell' are perceived as [espɛl] by Spanish speakers. This is a perceptual illusion, tantamount to the learner responding to a phantom vowel (e.g., Berent et al. 2007; Carlson et al. 2016; Davidson and Shaw 2012; Dehaene Lambertz et al. 2000; de Leeuw et al. 2019; Dupoux et al. 1999; Durvasula et al. 2018; Matthews and Brown 2004). Studies of adult L2 learners' perception of syllables employ discrimination tasks involving pairs of non-words with consonant clusters versus the same sequence into which vowels are inserted. Non-words are used to eliminate the need to control for participants' L2 mental lexicon and to avoid orthographic influence of aural-visual word familiarity. Results suggest misperception mirrors production.

Dupoux et al. (1999) is typical of studies in its use of an identification task where participants indicated whether they heard a vowel. 10 native Japanese and 10 French speakers listened to non-words containing a sequence of a vowel (V), first consonant (C1), the vowel u, a second, different consonant (C2) and vowel (V): VC1uC2V as in egudo, abuno, akumo. The u was successively shortened to millisecond durations of 72, 54, 36, 18 ms, and 0 (no vowel) to determine when participants would stop detecting a vowel. In Japanese but not in French, epenthesis is productive in the L1 and Japanese learners typically epenthesize when producing English clusters. The French participants stopped detecting shorter duration vowels but Japanese speakers continued to think they heard the vowel significantly more often than did the French speakers when the vowel they heard was increasingly shorter: 36, 18, and even 0 ms, when no vowel was present. The authors attributed this to the difference in phonological structures between the two languages: word-internal obstruent sequences such as /gd/,/dn/and/km/exist in French but not Japanese where there is always a vowel between them.[6]

Matthews and Brown′s (2004) study further explored Japanese learners′ perception of English clusters. The 12 speakers in the study perceived as a phantom vowel /ɯ/, productive in Japanese epenthesis. Stimuli included /kt/ word-medial clusters disallowed in Japanese but allowed in the L1 of the other group, 12 Thai speakers. Misperceptions were greatest for Japanese speakers when the interval between words was 1,500 ms but no such effect was observed for the Thais. Kabak and Idsardi (2007) is another relevant study, with Korean speakers. Like Japanese, epenthesis in Korean is productive and has also frequently been documented in L2 English speakers′ production. The 25 Korean speakers of English who listened to non-words responded similarly to those in the Japanese studies.

Confirming the importance of the position of epenthesis in the L1 is demonstrated in Gibson′s (2012) study of 50 Spanish learners of English aged 10 and 11 who counted syllables. A large proportion of participants counted an extra syllable in onset sequences of /s/ followed by a consonant. That is, they thought a vowel preceded the /s/ in the stimuli when it did not. A sub-sample of 30 of these learners took part in an oddity discrimination task and a forced-choice identification task to confirm that L1 position of the vowel is highly relevant in L2 perception.

In keeping with the discovery that interlanguage phonology is shaped by universals, researchers have found that listening is not just biased by the learner′s L1. For example, sonority plays a role. Cross-linguistically, the Sonority Sequencing Generalization with its sequence of stop-fricative-nasal-liquid-glideVOWEL-glide–liquid-nasal-fricative-stop relates to how consonants within clusters are distributed and Minimal Sonority Distance governs the similarity in sonority of adjacent consonants in a cluster (Selkirk 1982). Research shows that both notions of sonority account for variability found in learner data and go beyond models such as Flege′s SLM and Best′s PAM; see Broselow and Finer (1991), Hancin-Bhatt and Bhatt (1997), Broselow et al. (1998), and Hancin Bhatt (2000). Berent et al. (2007) argue that sonority also modulates how L1-disallowed clusters are perceived. They studied 16 Russian and 16 English participants′ perception of onsets in a syllable judgement task, and a further 30 Russian

[6] Dupoux et al. (2011) also consider whether this could be due to co-articulatory cues where the vowel affects the preceding and following consonants in the recording used for the 1999 experiment.

and 30 English speakers who took identity judgement and perceptual priming tasks. These two languages differ for clusters allowed in syllable-initial position with Russian allowing more complex clusters and phonotactic possibilities. Both groups were predicted to respond differently to clusters within which the sonority sequence varied. Results showed that even though English listeners were generally more susceptible to perceptual illusion across experimental conditions, Russians were affected by sonority but also prone to illusions in onsets not following the sonority generalization, with falling instead of rising sonority such as /lb/; their responses were slower for pairs such as lebif-lbif where the latter word violates the sonority sequence.

To further explore the role of the L1 in perception of L2 clusters, we carried out a study of adolescent learners of English from three native language backgrounds, two of which involve productive, language-internal epenthesis but in different environments and with different vowels, namely Kuwaiti Arabic and Spanish, and one – Mandarin – which disallows consonant clusters but because the language does not exhibit productive L1 prothesis or epenthesis, learners often delete consonants rather than insert vowels (Bayley 1996; Hansen-Edwards 2011), as in example (1a).

1.4 Syllable structures of Arabic, Spanish, and Mandarin Chinese

The three languages restrict consonants in onsets, codas and medial sequences in the ways shown in Table 1.

In Arabic, including in the Kuwaiti variety in our study, vowel insertion is highly productive; in Spanish, it is restricted to onsets beginning with /s/; it does not exist in Mandarin. Syllabification in Kuwaiti Arabic proceeds from the right edge of the word to the left, so disallowed clusters are broken up by the vowel [i]

Table 1: Arabic, Mandarin and Spanish syllable structure.

	Syllable complexity: onset	Syllable complexity: coda	Word-internal sequences	Vowel insertion
Arabic (Kuwaiti)	(C) C	CC	CC	Prothesis
Spanish	CC only if /p, t, k, b, d, g, f/ + liquid	Single C most common	If /s/+C, no epenthesis. Prothesis for /s/ becomes onset of word-initial next syllable s-clusters only	
	MacKenzie (2017)	CC if C[illegible] = /s/ MacKenzie (2017)		Cressey (1978)
Mandarin	Single C	(/ɻ/)*, /n/ and /ŋ/	none	none

*As part of the r-suffix (ɚ)/ r-coloring of the rhyme (see Duanmu 2007).

inserted to the left. For example, in a medial triconsonantal sequence (VCCCV), the first consonant is the coda of a syllable created with the preceding vowel and the last consonant is the onset of a syllable created with the following vowel. The middle (underlined) consonant is unattached to a syllable so [i] is inserted to its left: VC.iC.CV. Onset clusters exist, but the first consonant phonologically behaves as if it is not part of that word (i.e., extrametrical (Farwaneh 1995)). A syllable can therefore optionally be created to the left of an onset cluster: CC -> iCC through prothesis. This is productive in Kuwaiti Arabic, and unsurprisingly there is evidence in production for its transfer in L2 English (Broselow 1992: 25). The vowel schwa can also be an epenthetic vowel in this variety, particularly in the onset (Aquil 2011; Ayyad 2011; Watson 2002: 21).[7]

In Spanish, the addition of a vowel before a consonant occurs with /e/ with sC onsets. Note that this brings Spanish speakers' L1 onsets into conformity with the Sonority Sequencing Generalization which is violated if the segment following the /s/ is a stop such as /t/ or /p/, where sonority rises rather than falls. Prothesis in Spanish is a highly productive pattern of /e/sC rather than s/e/C (Eddington 2001).

Whether there are clusters in Mandarin is a matter of debate; in some accounts C + Glides are argued to exist, but in others they are considered a single consonant based on articulatory facts and phonological economy (see Duanmu 2007 for details about competing analyses).[8] For the purpose of our study, following Duanmu's account we assume C+Glide to be one consonant.[9]Studies of L2 English production by Mandarin speakers find that syllables are more likely to be modified, usually by deletion, when the L2 consonants are less similar to Mandarin (Best 1995; Brown 2000; Lan and Oh2012; Wang 1995). When vowel insertion is found, it appears to be constrained by word length and to be developmental (Eckman 1981; Major1987; Weinberger1994). In Wang's (1995) study of the production of codas by intermediate-level adults living in an English-speaking country, epenthesis was most frequently used with monosyllabic words to create the disyllabic syllables preferred in Mandarin (and less marked universally as well; Broselow et al. 1998). Speakers inserted vowels less often when di- and tri-syllabic words would have resulted in longer words.

---

[7] The occurrence of schwa as a phoneme in Kuwaiti Arabic and other Arabic varieties may be why some consider Arabic varieties to have a stress-timed rhythm (Aquil 2011; Demirezen 2010).

[8] Some accounts consider CG to share the same slot (e.g., Ao 1992; Cheung 1986), some consider that G shares the nuclear slot with V(Bao1990;Goh2000),while some argue that G has its own slot (see Duanmu 2007).

[9] "InChinese[sw]sounds like a single sound, in that the rounding of [w] starts at the same time as [s]. The simplest way to represent the difference, therefore, is to use [sw] (two sounds) for English and [s^w] (one sound) for Chinese (or C^G for Chinese CG in general) (Duanmu 2007: 80)." We should add that our testing instrument did not contain C + Glide clusters, hence, avoiding this debate and its associated complication. We intend to conduct a follow up study that targets this exact complication.

We included schwa as it is known to be difficult in production and in perception in terms of detection (e.g., Arboleda Guirao 2012; Demirezen 2010; Lacabex and Gallardo-del-Puerto 2020). It has, to date, rarely been tested in relation to perceptual illusions. Schwa does not exist in Spanish, it is a possible epenthetic vowel in Kuwaiti Arabic onsets and exists in Mandarin (Duanmu 2007),[10] but not as an epenthetic vowel. In none of these languages does schwa have the same status as in English, where it is a product of complex metrical phonology involving reduction of vowels in unstressed syllables. It is ubiquitous in spoken English and L2 learners with the amount of exposure as those in our sample might respond to not only to their L1 epenthetic vowels but also schwa in our test items, to which we now turn.

2. Methods and materials

2.1 Participants

We recruited participants who, as foreign language learners with negligible classroom and extra-classroom interaction with native speakers, would not have mastered English syllable structure yet. They would have had sufficient exposure to English to go beyond reliance on their L1. The sample was of a relatively less represented group in L2 speech research: secondary school students in their home countries. They had similar amounts of only classroom exposure, having typically started learning English in primary school with several weekly hours of instruction over 10–15 years. The sample comprised 23 Arabic speakers in Kuwait, 27 Spanish speakers in Spain and 20 Mandarin speakers in China, evenly split by gender. Participants were aged 15–19, in their final years of secondary school. 19 native, slightly older (18–24 years) English-speaking university students served as a control group.

The L2 participants completed a questionnaire on L1 background, age, gender and sources of exposure to and interaction with English and any other languages. 75 percent reported listening to music and watching television/films in English, but none had visited English-speaking countries as more than tourists, none had on-going contact with native English speakers or were proficient in other foreign languages and their teachers were rarely native English speakers. All reported normal hearing except for one Spanish participant; we did not exclude this person′s data as random intercepts for participants were included in the model to account for variability that might be due to different participants′ performances (i.e., idiosyncracy); see below.

The study design was ecologically valid by testing participants in a familiar situation, at their schools rather than in an offsite location such as a laboratory.

[10] See Duamu (2007) for details about the restricted distribution of schwa in Mandarin.

Embedding the tasks in regular classroom instruction, however, restricted the number of test items. Written consent was obtained from participants and their guardians for data collection, analysis and the sharing of anonymized results. Participation was entirely voluntary, unrelated to class marks. Participants were allowed to withdraw at any time; none did so.

2.2 Testing instrument

We adopted Matthews and Brown′s (2004) study design using an AX discrimination task, where the first (A) and second (X) member of a pair were the same or different. The stimuli were 63 non-words, following designs described above. The non-words represented English phonotactics (e.g. [bɹɪkbæt]), and the only sonority violation was when /s/ preceded a stop as it can in English (e.g. [speɪkɪ]). Participants listened to words with clusters in three positions: in syllable onset, word-medially and syllable coda. We selected three common consonant clusters/sequences for word-medial position and created non-words that otherwise began or ended with a single consonant or vowel. The clusters and sequences targeted vowel insertion sites in Arabic and in Spanish, using the vowels expected in that L1. All 70 participants heard examples with the L1 vowels where they would be inserted in Kuwaiti Arabic and Spanish and also heard the same examples with schwa.
Clusters were those disallowed in the three languages including /s/clusters specific to Spanish prothesis. Medial sequences involved a single syllable-final obstruent followed by one of the three clusters (e.g. [mæpθɹi:], [hɪkspæn]; see Appendix for full word list). In this design, participants′ inability to differentiate stimuli with an extra vowel from those without is taken as evidence of their perceptual difficulty with consonant clusters.

63 test pairs with and without inserted vowels were created. Each non-word generated seven test pairs: an identical match, two non-matches with insertion of schwa [ə], two non-matches with the insertion of the Spanish epenthetic vowel [e], and two non-matches with the insertion of Arabic epenthetic vowel [i]. Where a vowel was inserted, it preceded, followed, or broke up the consonant cluster or sequence depending on its position within the word. Table 2 shows how these were generated for each position; ′x′ represents the consonants in each cluster/sequence and inserted vowels are in IPA transcription. In this paper, we focus only on word initial and word-medial positions (42pairs) since we could not compare the Spanish participants′ results in word-final position given there is no environment for insertion triggered by /sC/.

To reduce the chance of participants guessing the purpose of the test, we created five non-words for 21 distractor items of a similar format. These included matched and non-matched pairs varying by more salient phoneme distinctions. Items were recorded by a native British English Received Pronunciation speaker (author 5), who co-developed the items, using a high-quality Tascam DR-05 device with a sampling rate of 44,100 Hz and 16 bit amplitude resolution in a soundproof room at a UK university. The tokens were randomized to create the final test order.

One consideration in studies of this nature is the amount of time between two items in a pair. There is experimental evidence which suggests that with age learners retain the auditory ability to discriminate sounds (Werker 1994 and more recent 'first exposure' studies; Gullberg et al. 2010). Relevant are studies that show

Table 2: Template for tokens generated in each word position.

Token	Inserted Vowel	Onset Cluster [xx]	Medial Cluster [x]	Coda cluster [xx]
?	Match	[xx] versus [xx]	[x] versus [x]	[xx] versus [xx]
?	Schwa [ə]	[xx] versus [əxx]	[x] versus [əx]	[xx] versus [xxə]
?		[xx] versus [xəx]	[x] versus [xə]	[xx] versus [xəx]
?	Spanish epenthetic vowel [e]	[xx] versus [exx]	[x] versus [ex]	[xx] versus [xxe]
?		[xx] versus [xex]	[x] versus [xə]	[xx] versus [xex]
?	Arabic epenthetic vowel [i]	[xx] versus [ixx]	[x] versus [ix]	[xx] versus [xxi]
?		[xx] versus [xix]	[x] versus [xi]	[xx] versus [xix]

participants can accurately perceive segments that do not exist in their L1when the task taps into raw auditory ability, when two different L2 sounds are heard close together, separated by a 250 ms interval. We sought to avoid this. Discrimination becomes more difficult when items are separated by longer intervals, e.g.,1500ms, because this taps into the high-order mental representations (phonology) that bias listening (Strange and Shafer2008). Some studies have manipulated the amount of time between stimuli,we reduced the chance of participants identifying a match or mismatch purely through acoustic memory by separating the two words in each pair with the carrier phrase 'I said', for example: 'Brickbat. I said, brickbat', with an average time of 2 s between the two words. This allowed time for participants to select their answers, but was short enough to recruit subconscious knowledge rather than metalinguistic knowledge gained in the classroom. Between each pair of words, there was then a 2 s pause.

To verify consistency and to identify human error in the recording, all tokens were judged auditorily by two native British English listeners with knowledge of phonetics of British English. In addition, all formant values were checked manually using Praat (Boersma and Weenink 1992–2019) to ensure the accuracy of the tokens. Some random tokens for the inserted vowels were also extracted from the recording, manually labelled, and analyzed using Praat, as shown in Figure 1. The

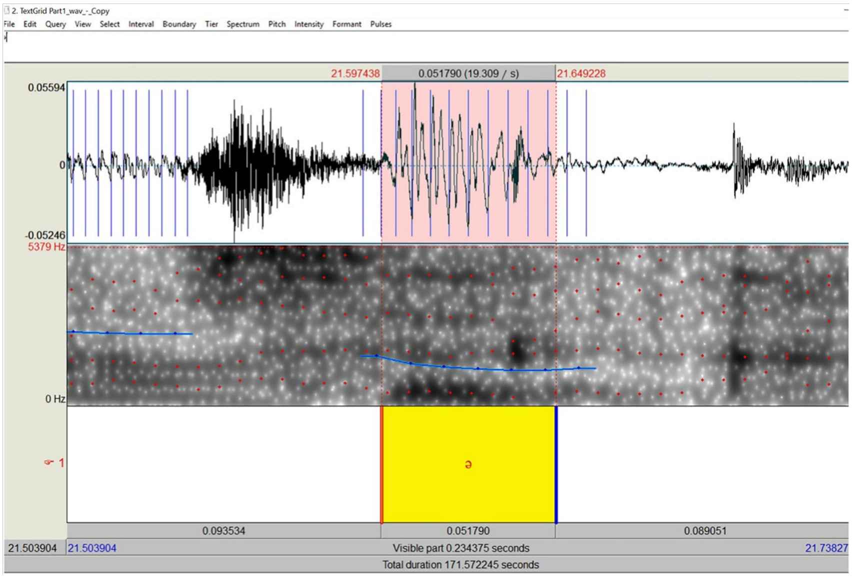

Figure 1: Praat analysis of word-medial schwa [ə] inserted post-consonantally in non-word 'mat<u>ə</u>kliff'. Average formant values in this case are F1 = 550 Hz and F2 = 1,785 Hz.

vowel formants were measured in the midpoint (the middle of the vowel) to eliminate the surrounding consonants' effects on the vowel. Formant values for schwa were also compared against previous averages established for this vowel in various positions by Flemming (2007). Our samples were consistent and corroborated these values, ensuring the validity of the testing instrument.

2.3 Procedure

Teachers known to the authors recruited participants in China, Kuwait and Spain. Teachers received identical instruction packs in English (and their native language if requested) with documents for conducting the experiment in a controlled manner on behalf of the researchers. This pack included the research objectives, access instructions to materials and audio files containing practice and test stimuli and consent forms. Teachers were instructed not to allow participants to confer with peers when taking the test en masse inside their classroom.[11] After testing was completed, teachers were

[11] One of the reviewers raised an interesting and important question about the environment in which the perception task took place, in particular the comparability between the native controls who completed the task individually and the target participants who took part en masse. We wished to emphasise

debriefed, information about the study, including the perceptual illusion effect, was shared with them and their students in writing as part of an awareness-raising exercise,[12]which was again an attempt to enhance the ecological validity of our study and to ensure that participants benefitted in some ways after having taken part.

This AX-discrimination task required participants to indicate whether items in a pair were ‘the same’ or ‘different’. To maintain participants’ concentration, the stimuli were split into three sections of 27–28 items. Between each section, there were questions from the background questionnaire in English. Before hearing the experimental tokens, teachers familiarized participants with the testing format by asking them to answer three practice questions. With the background questions, the entire test comprised 94 items and it took around20 minutes. The answer sheet was available online with an alternative user-friendly paper version if required.

We hypothesized the following:

1. Kuwaiti Arabic speakers will perform worse in prothesis than epenthesis with the vowels /i/ and /ə/.
2. Spanish speakers will perform worse in prothesis with /s/ clusters with the vowel /e/.
3. Mandarin speakers will show no differences in positions or across the three vowels.

3. Results

Scoring was based on response accuracy for correct answers of ‘same’ or ‘different’. Scores were then categorized into the relevant cluster groups. This generated binary data analyzed using Boolean statistics. The control group, who took the test online, were accurate 97% of the time in detecting the presence and absence of a vowel. In addition to the Praat measurements mentioned above, this confirmed the validity of the testing instrument.

To ensure that participants were paying attention, we also analyzed their responses to the distractor items, which included words with matched clusters and unmatched

---

ecological validity in this study by delivering tasks in the classroom, rather than in an environment unfamiliar to the participants (such as a speech lab). This inevitably has a trade-off. However, we would like to stress that the same set of stimuli was used with each class (.wav files recorded at 16 bit and 44,100 Hz sampling rate). A future study could be conducted whereby test conditions are controlled individually across all conditions.

[12] The information pack included the project description as well as questions drawing learners’ attention to the perceptual illusion phenomenon and form-focused listening exercises subsequent to testing. This is complemented by a discussion around accents.

clusters. Tables 3–6 show the proportion of correct and incorrect answers to the distractors in the three languages and for English. The fact that all groups were able to detect correct and incorrect answers from matched and unmatched clusters pairs demonstrates that participants paid attention while listening to the task items.

Table 3: Accuracy scores for distractor items by Kuwaiti participants.

Response	Matched	Unmatched
'Different'	13%	93%
'Same'	87%	7%

Table 4: Accuracy scores for distnctor items by Spanish participan s.

Response	Matched	Unmatched
'Different'	6%	88%
'Same'	94%	12%

Table5: Accuracy scores for distractors items by Mandarin Chinese participants.

Response	Matched	Unmatched
'Different'	10%	87%
'Same'	90%	13%

Table 4: Accuracy scores for distnctors items by English participants.

Response	Matched	Unmatched
'Different'	4%	97%
'Same'	96%	3%

Table 3 shows a high percentage of correct answers for the matched pairs (87% 'Same') and for the unmatched pairs (93% 'Different'). A Chi-square test revealed that the difference between responses to the matched and unmatched pairs was significant $X^2(1) = 279$, $p < 0.0001$

InTable4, Spanish participants' responses to both the matched and unmatched pairs were correct a high percentage of the time (94%'Same'); (88%'Different').The difference between the responses to the matched pairs and the unmatched pairs was significant $X^2(1) = 312$, $p < 0.0001$.

We can see from Table 5 that Mandarin Chinese participants had a noticeable proportion of correct responses to the matched pairs (90% ‘Same’) and to the unmatched pairs (87% ‘Different’). A Chi-square test showed that the difference was significant between responses to the matched and unmatched pairs, $X^2(1) = 308$, $p < 0.0001$

Table 6 shows similar results for the native speakers, and a Chi-square test revealed that the difference between responses to the matched and unmatched pairs was significant $X^2(1) = 328$, $p < 0.0001$.

Now we turn to participants′ responses to the experiment items, to unmatched pairs containing words with clusters and words with clusters broken up with either epenthesis or prothesis with different vowel types (/ə/, /i/ and/e/). We report mixed effects logistic regression analysis for Arabic, Spanish and Mandarin Chinese of their responses to the unmatched pairs with clusters and broken clusters. We used *glmer* function from the package lme4 (Bates et al. 2015) in the R software (R Core Team 2019; version 3.6.1) to conduct our analyses.

In the regression analysis, we looked at the effect of modification (epenthesis and prothesis), vowel type (/i/, /e/ and /ə/) and the interaction between modification and participants′ responses. First, we investigated whether the predictors (modification, vowel type and the interaction between the two) improved the models through likelihood ratio tests. This was achieved by model comparisons using the function "anova" in R. Specifically, to investigate whether modification improved the model's fit, we compared a model that contained modification and vowel type as the predictors to a model that contained only vowel type as the predictor. If the outcome of the ANOVA test was significant, dropping modification from the model would decrease the likelihood of the model, hence it was included in the regression analysis. The same procedure was applied to identify whether vowel type improved the model's fit. A model that included modification and vowel type as predictors was compared to a model that contained modification only as the predictor. As for the interaction term between modification and vowel type, a model that included modification, vowel type and an interaction term between modification and vowel type was compared to a model that contained modification and vowel type as the predictors only. The final model represents factors that improved the model's fit only. Speakers were included as random intercepts in the regression analysis.

3.1 Initial position

Table7 shows accuracy scores (proportion of correct answers and incorrect answers) for Arabic speakers, conditioned by modification (epenthesis between consonants, and prothesis preceding consonants) and vowel type. We were interested in when learners correctly detected stimuli pairs with and without vowels and what among the tested factors correlated with incorrect responses.

Table7: Kuwaitis'correctandincorrectanswersbymodificationandvoweltypeininitialposition.

	Epenthesis CvC	Prothesis vCC	i	e	ə
'Different' Correct answers	76%	53%	66%	71.4%	52.8%
'Same' Incorrect answers	24%	46%	34%	28.5%	47.2%

Likelihood ratio tests showed that modification had a significant effect on the model's fit, i.e., significant effect on participants' responses, $X^2(1) = 27.7$, $p < 0.0001$. Vowel type also had a significant effect on participants' responses $X^2(1) = 20.5$, $p < 0.0001$. The interaction term was not significant $X^2(2) = 0.9$, $p = 0.6$. Thus, our final model includes modification and vowel type as the predictors with speakers as random intercepts. Table

8 summarises the model's output. Estimates with a negative sign indicate that a certain factor was more likely to cause correct (Different) answers, while estimates with a positive sign indicate that a certain factor was more likely to cause incorrect (Same) answers. Note that p-values in the model's output are computed by the Wald test. The model shows that/e/compared to /ə/, which was at the reference level, was more likely to cause correct answers, with an estimate of −0.8294, and the difference was significant $p < 0.003$. /i/ compared to /ə/ was more likely to cause correct answers, with an estimate of −0.5945, and the effect was significant $p < 0.03$. We also conducted a pairwise t-test to compare all vowels. The test revealed that the difference between /e/ and /ə/ was significant $p < 0.001$, the difference between /i/ and /ə/ was significant $p < 0.001$ and the difference between /e/ and /i/ was also significant $p < 0.001$.

For modification, the model shows that prothesis compared to epenthesis, which was at the reference level, was more likely to lead to incorrect answers, with an estimate of 1.1967, and the effect was significant $p < 0.0007$.

Figure 2 shows the predicted probabilities of incorrect answers of vowel type and modification based on the model's outcome. We used package effects (Fox et al. 2020) to obtain probabilities from the model's estimates (which represented log-odds) and to plot the probabilities (Figure 3).

The probability of incorrect answers was higher for /ə/ than /e/ and /i/, and the probability of incorrect answers for prothesis was higher than in epenthesis. These results partially confirm hypothesis 1 in that Kuwaiti participants performed worse with pairs modified by prothesis. However, though these participants struggled with /i/ as predicted, it was /ə/ which was significantly more likely to be associated with incorrect answers.

Table 9 illustrates accuracy scores for Spanish participants (proportion of correct answers and incorrect answers), again conditioned by modification position and vowel type.

Table 8: Summary of regression results for Kuwaiti participants in initial position.

Fixed effects						Random effects
	Estimate	Standard error	95% Confidence intervals	z valu e	P-value	Standard deviation
Intercept	-0.872	0.3308	{-1.50, -0.22}	-2.63	< 0.008	1.75
Vowel /e/	-0.8294	0.2802	{-1.37, -0.280}	-2.96	< 0.003	
Vowel /i/	-0.5945	0.2831	{-1.14, -0.03}	-2.10	< 0.03	
Modification(Prothesis)	1.1967	0.2336	{0.73, 1.65}	5.12	< 0.0007	

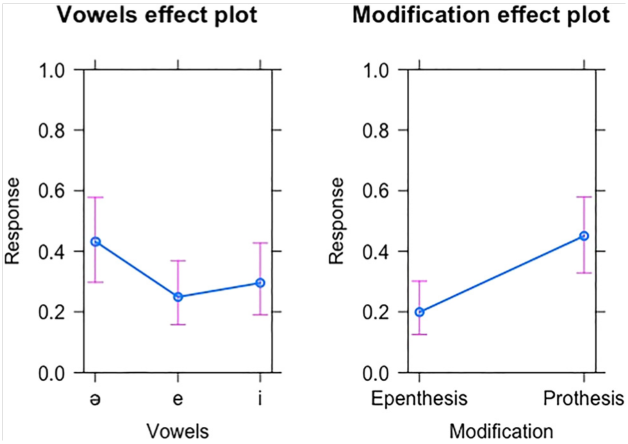

Figure2: Probabilities of incorrect answers (y-axes) for vowel types and modifications in Kuwaiti data. Pink bars represent confidence intervals.

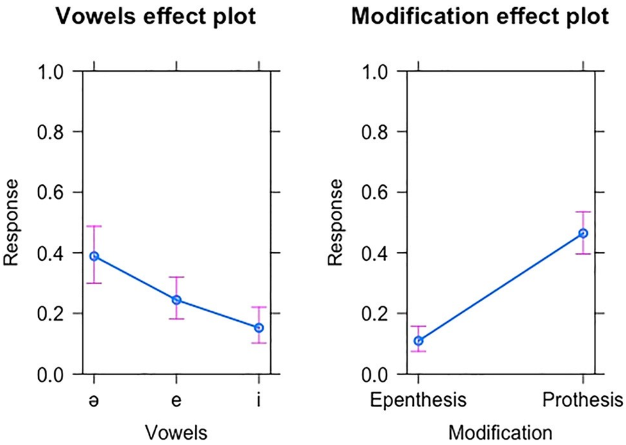

Figure 3: Probabilities (y axis) of incorrect responses in the Spanish data for vowel type and modification effects. Pink bars represent confidence intervals.

Table9: Spanish speakers' correct and incorrect answers by modification and vowel type in initial position.

	Epenthesis CvC	Prothesis vCC	i	e	ə
'Different'	88.4%	52.3%	80.5 %	73.7 %	54.5 %

Correct answers					
'Same' Incorrect answers	11.5%	47.6%	19.4 %	26.2 %	45.4 %

The likelihood ratio test shows that vowel type had a significant effect on the participants′ responses $X^2(4) = 25.03$, $p < 0.0001$. Modification also had a significant effect $X^2(3) = 80.01$, $p < 0.0001$. The interaction between modification and vowel type was not significant, $X^2(2) = 1.2$, $p = 0.08$. The final model′s output is summarizedinTable10.Themodelshowsthat/e/comparedto/ə/wasmorelikely to result in correct answers, with an estimate of −0.6788, and the effect was significant $p < 0.01$. /i/ compared to /ə/ was also more likely to result in correct answers, with an estimate of −1.2673, and the effect was significant $p < 0.00001$. A pairwise t-test analysis revealed that the difference between /e/ and /ə/ was significant, $p < 0.01$, the difference between /i/ and /ə/ was significant, $p < 0.001$, and the difference between /e/ and /i/ was significant $p < 0.0005$ too. As for modification, the model shows that prothesis compared to epenthesis was more likely to result in incorrect answers, with an estimate of 1.9623, and the effect was significant $p < 0.0001$. Figure 3 provides probabilities of incorrect answers in Spanish data based on the model′s outcome.

/ə/ was associated with more incorrect answers than the other vowels and prothesis was associated with more incorrect answers than epenthesis. These results provide partial support for hypothesis 2 in that Spanish participants performed worse with prothesis pairs, but once again /ə/ seemed to be causing more problems than other vowels investigated.

Turning now to the Mandarin data, Table 11 shows the proportion of correct and incorrect answers by modification position and vowel type,

The likelihood ratio test demonstrates that there was a significant effect for vowel type on participants′ responses, $X^2(2) = 28.1$, $p < 0.03$. There was no significant effect for modification, $X^2(1) = 0.28$, $p = 0.5$, nor for an interaction, $X^2(1) = 1.4$, $p = 0.09$. Table 12 summarizes the final model′s output (Figure 4). /e/ compared to /ə/ was more likely to result in correct answers, with an estimate of −0.5463, and the effect was significant $p < 0.03$. /i/ compared to /ə/ was more likely to cause correct answers, with an estimate of −0.6662, and the effect was significant

Table 10: Results of regression analysis for Spanish participants in initial position.

Fixed effects						Ran dom effects
	Esti mate	Stan dard error	95 % Confid	z value	P - value	Stan dard

			ence intervals			deviation
Intercept	-1.4161	0.2615	{-1.92, -0.90}	-5.14	< 0.0001	0.223
Vowel /e/	-0.6788	0.2671	{-1.20, -0.155}	-2.54	< 0.01	
Vowel /i/	-1.2673	0.2911	{-1.83, -0.69}	-4.35	< 0.00001	
Modification(Prothesis)	1.9623	0.2488	{1.47, 2.44}	7.88	< 0.001	

Table 11: Mandarin speakers' correct and incorrect answers by modification and vowel type in initial position.

	Epenthesis CvC	Prothesis vCC	i	e	ə
'Different' Correct answers	51.1%	52.4%	57.5%	55%	41%
'Same' Incorrect answers	48.8%	47.5%	42.5%	44%	59%

$p < 0.01$. A pairwise t-test revealed that the difference between /e/ and /ə/ was significant, $p < 0.001$, the difference between /i/ and /ə/ was significant, $p < 0.001$ but the difference between /i/ and /e/ was not significant, $p = 0.7$.

The probability of /ə/ causing incorrect answers was higher than the other vowels. Even though we did not make specific directional predictions regarding Mandarin participants and that the proportion of accuracy scores seem to point to a rather equivocal performance overall, the regression analysis suggests that /ə/was more likely to be significantly associated with incorrect answers compared to /i/ and /e/.

3.2 Medial position

We followed the same analytical procedure for medial positions as we did for initial positions. Table 13 illustrates accuracy scores for Kuwaiti participants.

The likelihood ratio test demonstrates that there was no significant effect for vowel type, $X^2(1) = 8.1$, $p = 0.4$, nor for modification, $X^2(1) = 4.1$, $p = 0.6$. There was no significant interaction between the two factors, $X^2(2) = 5.2$, $p = 0.2$. Given that no factor had a significant effect on participants' responses, the output of the regression analysis is not provided.

Table 12: Results of regression analysis for Mandarin Chinese participants in initial position.

Fixed effects						Random effects
	Estimate	Standard error	95% Confidence intervals	z value	P-value	Standard deviation
Intercept	0.3640	0.2033	{-1.34, -0.27}	1.79	= 0.07	0.002
Vowel /e/	-0.5463	0.2637	{-3.46, -0.56}	-2.07	< 0.0001	
Vowel /i/	-0.6662	0.2747	{-2.18, -0.44}	0.015	< 0.0001	

Table 13: Kuwaitis' correct and incorrect answers by modification and vowel type in medial position.

	Epenthesis CCvC	Prothesis CvCC	i	E	ə
'Different' Correct answers	70.3%	68%	65%	76.4%	69.2%
'Same' Incorrect answers	29.6%	32%	35%	23.5%	30.7%

As for the Spanish data, Table 14 shows accuracy scores for correct and incorrect answers.

Likelihood ratio tests show that there was no significant effect for vowel type, $X^2(1) = 5.1$, $p = 0.4$. The effect of modification was significant, $X^2(2) = 16.5$, $p < 0.0001$.

There was no significant interaction between the vowel type and modification, $X^2(1) = 2.1$, $p = 0.8$. Table 15 summarizes the final model output.

The model shows that prothesis compared to epenthesis was more likely to result incorrect answers, with an estimate of −1.5924, and the effect was significant $p < 0.002$. Figure 5 shows predicted probability of incorrect answers for modification.

Table14: Spanish speakers' correct and incorrect answers modification and vowel type in medial position.

	Epenthesis CCvC	Prothesis CvCC	i	e	ə
'Different' Correct answers	79.7%	95%	83.2%	81.4%	82.4%
'Same' Incorrect answers	20.2%	5%	16.7%	18.5%	17.5%

Table 15. *Regression results of the Spanish data in medial position*

Fixed effects						Random effects
	Estimate	Standard error	95 % Confidence intervals	z value	P-value	Standard deviation
Intercept	-1.3995	0.1505	{-2.37, -0.47}	-9.29	< 0.0001	0.318
Modification(Prothesis)	-1.5924	0.5299	{0.22, 4.32}	-3.00	< 0.002	

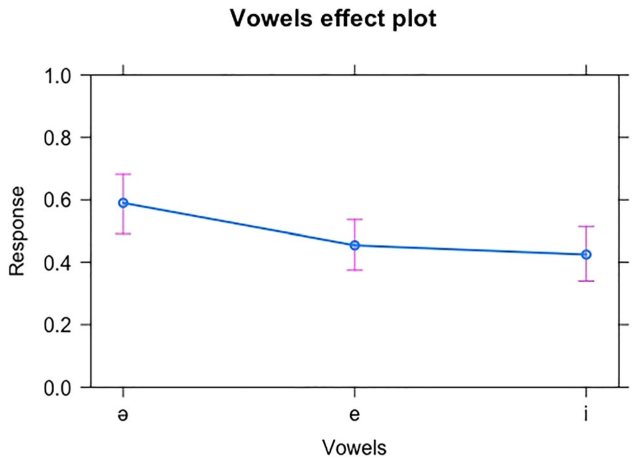

Figure4: Probabilities of incorrect answers due to vowel types in Mandarin Chinese data, which turned to be significant. Pink bars indicate confidence intervals.

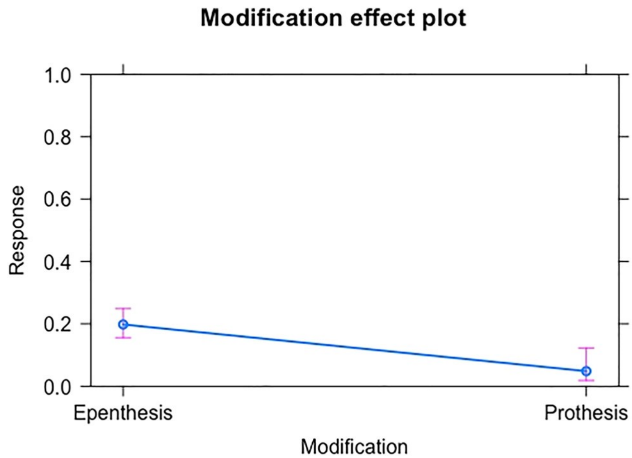

Figure 5: Probabilities (y-axis) of incorrect answers for Spanish data in medial position due to modification, which was significant. Pink bars indicate confidence intervals.

As can be seen, the probability for incorrect answers was higher for epenthesis than prothesis.

And finally, Table 16 shows accuracy scores in proportion for correct and incorrect answers for the Mandarin speakers.

Likelihood ratio tests show that there was no significant effect for vowel type, $X^2(1)= 3.2$, $p = 0.7$,norformodification, $X^2(1)= 4.1$, $p =0.5$.Therewasnosignificant

Table 16: Mandarin speakers' correct and incorrect answers modification and vowel type in medial position.

	Epenthesis CCvC	Prothesis CvCC	i	e	ə
'Different' Correct answers	42.1%	56.6%	49.1%	47%	38.3%
'Same' Incorrect answers	57.8%	43.3%	50.8%	53%	61.6%

two-way interaction between vowel type and modification, $X^2(1) = 3.4$, $p = 0.3$. Again, since no factor had a significant effect, the output of the regression analysis is not provided.

Taken together with the descriptive statistics provided for each group, the above regression models, on the whole, suggest that participants had various difficulties perceiving the pairs. Different targeted vowels posed slightly different challenges for the groups, with schwa standing out as the one that was significantly associated with more incorrect answers in initial position. The confusion also depended partly on where the extra vowel was introduced (i.e., modification type: epenthesis vs. prothesis). We offer further interpretations in the discussion below.

4. Discussion

By juxtaposing results from speakers of three different native languages, we are able to show that a speaker's bringing L2 consonant clusters into conformity with their L1 syllable structure indeed applies to perception and this is shown when we tap their higher-order phonological representations. Participants struggle to distinguish stimuli with an extra vowel which breaks up consonant clusters from stimuli without a prothetic/epenthetic vowel. This is not random; cluster simplification at the perceptual level relates to how the L1 operates in terms of position of prothetic/epenthetic vowel and the specific vowel involved. These findings corroborate studies reviewed above (e.g., Berent et al. 2007; Dupoux et al. 1999; Gibson 2012; Kabak and Idsardi 2007; Matthews and Brown 2004).

For the Kuwaiti Arabic speakers, their responses to the pairs of onset clusters versus onset clusters preceded or broken up by an L1 vowel confirm our hypothesis that perception is constrained by L1 syllable structure in terms of where prothesis takes place; a vowel is added before the cluster rather than breaking it up. Kuwaitis have a marked range of error when perceiving pairs like ʻklobcatʼ versus ʻiklobcatʼ as well as pairs involving schwa. Following Matthews and Brown (2004), when L2 learners misperceive such pairs, it is likely that they think they hear a vowel that brings the cluster into conformity with their L1. According to the site of modification of onset clusters, the noticeable proportion of errors in perceiving them is due to an illusory vowel to the left of the word-initial consonant. Perceiving the cluster in this way is the repair mechanism employed by speakers that brings the onset cluster into a well-formed prosodic shape in their interlanguage English. Speakers showed difficulty in correctly hearing pairs with schwa in initial position given its L1 status in that position but were better at doing so for the same clusters in medial position.

Spanish speakers were unable to perceive the difference between onset clusters that started with /s/ from those which started with /e/, the vowel involved in their L1 prothesis. Moreover, schwa also caused significant confusion. This could be because speakers also treat sC as a sonority violation. In keeping with this universal principle, they are perhaps at the developmental stage outlined in Majorʼs Ontogeny Phylogeny Model (Major 2008) where L1 transfer effects fade as universals emerge. This could also be an effect of increasing exposure to schwa in the input. Since Spanish speakers do not insert vowels in other positions in their production, evidence that they had no other problems perceiving different pairs was expected. What remains to be investigated is how the complexity of clusters interacts with perceptual illusion, for example, whether non-L1 clusters violating the sonority generalization cause more confusion than those that do not (as in Berent et al. 2007).

For the Mandarin speakers, who could not distinguish between pairs in either onsetormedialposition,wespeculatethatresponsesreflectnothearingavowelin the first place rather than detecting a phantom or illusory vowel.[13] As Mandarin speakers prefer in their interlanguage English the disyllabic words their native language prefers – and because such words are also universally less marked (Broselow 2009) – in the trisyllabic test words created by vowel insertion such as ʻhekispanʼ versus disyllabic ʻhekspanʼ, speakers subconsciously omitted, did not hear, the vowel. Further studies are needed to verify the psychological reality of these two L1-based responses.

Thediscoverythatthevowelschwahascreatedproblemsforallourparticipants from all three L1s is also noteworthy, confirming what researchers have found regarding the

[13] The reader needs to be aware that vowels amid clusters are the result of epenthesis, which renders such vowels as epenthetic vowels. However, we are addressing such vowels as ʻnonepentheticʼ only because epenthesis is not available in the phonology of Mandarin speakers.

difficulty learners face. For example, Lacabex and Gallardo-del-Puerto (2020) note the challenge schwa poses to learners, especially in pedagogical contexts where there is no relevant explicit instruction. The results we report above show that instances of schwa in perceptual misperception in comparison to L1-epenthesis vowels deserve more attention.

4.1 Pedagogical implications

Our study's purpose was to show how the perception of typical foreign language learners of English – with years of classroom exposure but little/no naturalistic exposure to native speakers –is still biased by their L1 and by universals when they deal with English syllables. Recall that the control group showed a 97% success rate on the test while none of the L2 learners approached this rate. This demonstrates that the processing of acoustic signals in part depends on speakers' phonological system (cf. Iverson et al. 2003). For the learners in our study, this system is clearly not the same as the native speakers'. This leads to acknowledging that acquiring an L2 phonology must involve developing target-like perception. The teacher might begin his/her thinking about perception and how it might influence production by surmizing that learners are grappling with a more complex syllable structure than in their L1 not simply with short and unstressed vowels. We have shown that there are differences in the perception of /i/, /e/ and /ə/ and that this is based on the learner's L1 phonology. Thus, while learners have problems, this is not simply a question of failing to hear a given minimal unit of sound but – depending on their L1 in various ways – also of 'hearing' them when they are not present in the input.

Learners' misperceptions as reported above could indeed partially explain the production difficulties they encounter with consonant clusters. What learners produce is, at least in part, a result of what they perceive, and this is influenced by what is and is not possible in their L1s for syllable onsets and medial sequences (all three L1s) and allowable number of syllables in a word (only Mandarin) as well as by what is less and more marked universally. The speakers in our study processed the L2 words in terms of their L1 phonology and in terms of an interlanguage English phonology favoring unmarked forms such as disyllabic words. Given that the interplay between L1 influence and universals changes as learners progress (Major 2008), a longitudinal study of learners' developing perception could test predictions and its results would be of value to teachers to raise their awareness of what to expect over time from their learners.

With respect to teaching pronunciation, in addition to numerous books providing guidance, there has been a proliferation of accent reduction tips on the internet since the early 90s.[14] A decade ago, the empirical basis of such resources and tips was questioned given that "empirical, classroom-based research is nearly nonexistent in relation to pronunciation pedagogy" (Baker and Murphy 2011: 37; see also Kennedy and Trofimovich 2017). Where research has demonstrated the benefits of pronunciation teaching on learners' development of segmental and suprasegmental features (see Lee et al. 2015; Thomson and Derwing 2015), our study contributes to research which points to benefits of perceptual training for learners' production (e.g., Lee et al. 2020). However, as Murphy and Baker note, it is possible that "some of the more interesting resources were not all that widely read, assimilated, and applied by classroom teachers. As in many fields, it takes time for specialists' contributions to influence wider audiences" (2015: 36). The potential lack of uptake and integration of the robust academic research findings by materials in public circulation is a rather unfortunate state of affairs.

Pronunciation teaching has lagged and continues to lag behind (Derwing and Munro 2015), and popular advice tends not to have moved beyond Lado's Contrastive Analysis hypothesis that L1–L2 differences (including no clusters vs. clusters) lead to a foreign accent (Swan 2001). The best advice from Lado was to review the L1 facts as currently understood by researchers. Though often neglected in materials written for popular consumption, this should include perception (for exceptions see Field 2003; Wilson 2003) and should go beyond production improvement via minimal pair exercises and mechanical drilling,[15] in the foreign language classroom (see Bailey 2020). There is a need to go beyond explicit instruction (Kennedy and Trofimovich 2017).

Viewed in conjunction with existing literature, our findings suggest that some of the well- documented production difficulties with consonant clusters can indeed be traced back to confusion and illusion in the perceptual domain (see also Best and Tyler 2007). Some have called for an initial focus on listening and perception (Baese-Berk

[14] Not surprisingly, the advice assumes learners aspire to a native accent and ignores the idea that the target should be a lingua franca variety and the aim should be fluency and intelligibility (Jenkins 2012; Kenworthy 1987; see also Levis 2018). This issue is beyond the scope of this paper, but readers interested can refer to Gao and Weinberger (2018) which investigated how segmental/ syllable errors contribute to accentedness. See also Magen (1998).

[15] Drills are associated with earlier, behaviourist's beliefs about language learning.

and Samuel 2016; Postovsky 1974). We concur with recent strictly controlled and classroom-based research that advocates a more integrated approach based on the idea that perception and production are mutually reinforcing (e.g., Kissling 2018; Sakai and Moorman 2018; see also Inceoglu 2016; Hardison 2021 for multi-modal training). In an integrated approach, the teaching of pronunciation could be embedded in skills development sessions (Bailey 2020; Kennedy and Trofimovich 2017). Perception training, including that of consonant clusters, can be among the array of tools, e.g., communicative-oriented practice (Mora and Levkina 2017), awareness raising exercise, form-focused instructions (Saito and Lyster 2012) that teachers employ.

APPENDIX: Word list

[mæpθɹi:] versus [mæpθɹi:]	[bɪkæsp] versus [bɪkæsep]
[klɒbkæt] versus [iklɒbkæt]	[wɜɹdɛlm] versus [wɜɹdɛlim]
[hɪkspæn] versus [hɪksipæn]	[bɹɪkbæt] versus [ibɹɪkbæt]
[mætklɪf] versus [mætəklɪf]	[θɹufɛl] versus [θɹufɛn]
[gɛɹækst] versus [gɛɹæpst]	[bɹɪkbæt] versus [biɹɪkbæt]
[ɹʌʃbɹɪn] versus [ɹʌʃəbɹɪn]	[hɪkspæn] versus [hɪkispæn]
[bɪkæsp] versus [bɪkæsp]	[bɪkæsp] versus [bɪkæspe]
[kæɹənt] versus [kæɹəntə]	[gɛɹækst] versus [fɛɹækst]
[gɛɹækst] versus [gɛɹæpst]	[kæɹənt] versus [kæɹənet]
[mætklɪf] versus [mætkilɪf]	[ɹʌʃbɹɪn] versus [ɹʌʃbiɹɪn]
[ɹʌʃbɹɪn] versus [ɹʌʃibɹɪn]	[hɪkspæn] versus [hɪkespæn]
[wɜɹdɛlm] versus [wɜɹdɛlm]	[bæfslə] versus [kæfslə]
[hɪkspæn] versus [hɪkspæn]	[wɜɹdɛlm] versus [wɜɹdɛlme]
[bɹɪkbæt] versus [əbɹɪkbæt]	[kænbɪft] versus [kænbɪft]
[kænbɪft] versus [kænbɪfs]	[speɪkɪ] versus [espeɪkɪ]
[speɪkɪ] versus [sipeɪkɪ]	[bæfslə] versus [bæfslən]
[kæɹənt] versus [kæɹənt]	[bɹɪkbæt] versus [beɹɪkbæt]
[bɪkæsp] versus [bɪkæspi]	[hɪkspæn] versus [hɪksəpæn]
[kænbcft] versus [tænbɪft]	[klɒbkæt] versus [əklɒbkæt]
[θɹufɛl] versus [kɹufɛl]	[gɛɹækst] versus [gɛɹækst]
[hɪkspæn] versus [hɪksepæn]	[mætklɪf] versus [mæteklɪf]
[kæɹənt] versus [kæɹənte]	[bɪkæsp] versus [bɪkæspə]
[kænbɪfs] versus [tænbɪft]	[mækθɹi:] versus [bæpθɹi:]
[speɪkɪ] versus [ispeɪkɪ]	[ɹʌʃbɹɪn] versus [ɹʌʃbɹɪn]
[kæɹənt] versus [kæɹənit]	[speɪkɪ] versus [səpeɪkɪ]

[mæpθɹi:] versus [bæpθɹi:]	[hɪkspæn] versus [hɪkəspæn]
[bɪkæsp] versus [bɪkæsip]	[mætklɪf] versus [mætiklɪf]
[bɹɪkbæt] versus [ebɹɪkbæt]	[bæfslən] versus [kæfslə]
[mætklɪf] versus [mætkəlɪf]	[wɜɹdɛlm] versus [wɜɹdɛlmə]
[θɹufɛl] versus [θɹufɛl]	[bɹɪkbæt] versus [bəɹɪkbæt]
[wɜɹdɛlm] versus [wɜɹdɛlem]	[bæfslə] versus [bæfslə]
[ɹʌʃbɹɪn] versus [ɹʌʃebɹɪn]	[mætklɪf] versus [mætklɪf]
[kæɹənt] versus [kæɹənət]	[klɒbkæt] versus [eklɒbkæt]
[ɹʌʃbɹɪn] versus [ɹʌʃbeɹɪn]	[mætklɪf] versus [mætkelɪf]

(continued)

[wɜɹdɛlm] versus [wɜɹdɛlmi]	[θɹufɛn] versus [kɹufɛl]
[gɛɹæpst] versus [fɛɹækst]	[klɒbkæt] versus [kəlɒbkæt]
[mæpθɹi:] versus [mækθɹi:]	[speɪkɪ] versus [əspeɪkɪ]
[ɹʌʃbɹɪn] versus [ɹʌʃbəɹɪn]	[bɹɪkbæt] versus [bɹɪkbæt]
[bɪkæsp] versus [bɪkæsəp]	[klɒbkæt] versus [kelɒbkæt]
[speɪkɪ] versus [sepeɪkɪ]	[wɜɹdɛlm] versus [wɜɹdɛləm]
[klɒbkæt] versus [klɒbkæt]	[speɪkɪ] versus [speɪkɪ]
[kæɹənt] versus [kæɹənti]	[klɒbkæt] versus [kilɒbkæt]

References

Abrahamsson, Niclas & Kenneth Hyltenstam. 2009. Age of onset and nativelikeness in a second language: listener perception versus linguistic scrutiny. Language Learning 59. 249–306.

Altenberg, Evelyn. 2005. The judgment, perception, and production of consonant clusters in a second language. International Review of Applied Linguistics in Language Teaching 43. 53–80.

Ao, Benjamin X. P. 1992. The non-uniqueness condition and the segmentation of the Chinese syllable. Working Papers in Linguistics (Ohio State University) 42. 1–25.

Aquil, Rajaa. 2011. Empirical evidence: Stress as a perceptual unit in Cairene spoken Arabic. In Ellen Broselow & Ouali Hamid(eds.),Perspective on Arabic linguistics XXII–XXIII: Papers from the Annual Symposia on Arabic linguistics, 1–20. Amsterdam: John Benjamins.

Arabski, Janusz & Adam Wojtaszek. 2011. The acquisition of L2 phonology. Bristol: Multilingual Matters.

Arboleda Guirao, Inmaculada de Jesús. 2012. Do people selectively interpret their schwa versus syllabic consonant perception on the basis of their background, experience and expectations? Revista de Lingüística y Lenguas Aplicadas 7. 23–35.

Ayyad, Hadeel. 2011. Phonological development of typically developing Kuwaiti Arabic-speaking preschoolers. CA: University of British Columbia PhD thesis.

Baese-Berk, Melissa M. & Arthur G. Samuel. 2016. Listeners beware: Speech production may be bad for learning speech sounds. Journal of Memory and Language 89. 23–36.

Bailey, Kathleen.2020. Teaching listening and speaking in second and foreign language contexts. London: Bloomsbury.

Baker, Amanda & John Murphy. 2011. Knowledge base of pronunciation teaching: Staking out the territory. TESL Canada Journal 28. 29–50.

Baker, Wendy & Pavel Trofimovich. 2006. Perceptual paths to accurate production of L2 vowerls: The role of individual differences. International Review of Applied Linguistics in Language Teaching 44. 231–250.

Bao, Zhiming. 1990. Fanqie languages and reduplication. Linguistic Inquiry 21(3). 317–350.

Bassetti, Bene, Paola Escudero & Rachel Hayes-Harb. 2015. Second language phonology at the interface between acoustic and orthographic input. Applied Psycholinguistics 36(1). 1–6.

Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. Journal of Statistical Software 671. 1–48.

Bayley, Robert. 1996. Competing constraints on variation in the speech of adult Chinese learners of English. In Robert Bayley & Dennis Preston (eds.), Second language acquisition and linguistic variation, 97–120. Amsterdam: Benjamins.

Berent, Iris, Donca Steriade, Tracy Lennertz & Vaknin Vered. 2007. What we know about what we have never heard: Evidence from perceptual illusions. Cognition 104(3). 591–630.

Best, Catherine. 1995. A direct realist perspective on cross-language speech perception. In Winifred Strange (ed.), Speech perception and linguistic experience: Issues in crosslanguage research, 171–204. Timonium: York Press.

Best, Catherine & Michael Tyler. 2007. Nonnative and second language speech perception: Commonalities and complementarities. In Bohn Ocke-Schwen & Murray Munro (eds.), Language experience in second language speech learning, 13–34. Amsterdam: John Benjamins.

Boersma, Paul & David Weenink. 1992–2019. Praat: Doing phonetics by computer.

Bongaerts, Theo, Susan Mennen & Frans van der Slik. 2000. Authenticity of pronunciation in naturalistic second language acquisition: The case of very advanced late learners of Dutch as a second language. Studia Linguistica 54. 298–308.

Bradlow, Ann R. 2008. Training non-native language sound patterns: Lessons from training Japanese adults on the English/ɹ/-/l/ contrast. In Jette G. Hansen-Edwards & Mary L. Zampini (eds.), Phonology and second language acquisition, 287–308. Amsterdam: John Benjamins.

Broselow, Ellen. 1983. Non-obvious transfer: On predicting epenthesis errors. In Susan Gass & Larry Selinker (eds.), Language transfer in language learning, 269–280. Rowley: Newbury House.

Broselow, Ellen. 1987. An investigation of transfer in second language phonology. In Georgette Ioup & Steven Weinberger (eds.), Interlanguage phonology: The acquisition of second language sound system, 261–278. Rowley: Newbury House.

Broselow, Ellen. 1992. Parametric variation in Arabic dialect phonology. In Ellen Broselow, Mushira Eid & John McCarthy (eds.), Perspectives on Arabic linguistics IV: Papers from the fourth annual symposium on Arabic linguistics, 7–45. Amsterdam: John Benjamins.

Broselow, Ellen. 2009. Stress adaptation in loanword phonology: Perception and learnability. In Boersma Paul & Silke Hamann (eds.), *Phonology in perception*, 191–234. Berlin: Mouton De Gruyter.

Broselow, Ellen & Daniel Finer.1991. Parameter setting in second language phonology and syntax. *Second Language Research* 7. 35–59.

Broselow, Ellen, Su-I Chen & Chilin Wang. 1998. The emergence of the unmarked in second language phonology. *Studies in Second Language Acquisition* 20. 261–280.

Broselow, Ellen & Yoonjung Kang. 2013. Phonology and speech. In Julia Herschensohn & Martha Young-Scholten (eds.), *The Cambridge handbook of second language acquisition*, 529–553. Cambridge: Cambridge University Press.

Brown,Cynthia.2000. The interrelation between speech perception and phonological acquisition from infant to adult. In John Archibald (ed.), *Second language acquisition and linguistic theory*, 4–63. London: Blackwell Publishers Inc.

Calabrese, Andrea. 2012. Auditory representations and phonological illusions: A linguist's perspective on the neuropsychological bases of speech perception. *Journal of Neurolinguistics* 25. 355–381.

Carlisle, Robert S. 1998. The acquisition of onsets in a markedness relationship: A longitudinal study. *Studies in Second Language Acquisition* 20. 245–260.

Carlson, Matthew, Matthew Goldrick, Michael Blasingame & Angela Fink. 2016. Navigating conflicting phonotactic constraints in bilingual speech perception. *Bilingualism: Language and Cognition* 19. 939–954.

Casserly, Elizabeth & David B. Pisoni. 2010. Speech perception and production. *WIREs Cognitive Science* 1. 629–647.

Chang, Charles. 2019. The phonetics of second language learning and bilingualism. In William F. Katz & Peter F. Assmann (eds.), *The Routledge handbook of phonetics*, 427–447. London: Routledge.

Cheung, Kwan-Hin. 1986. *The phonology of present-day Cantonese*. London: University of London Doctoral Dissertation.

Cressey, William. 1978. *Spanish phonology and morphology: A generative view*. Washington: Georgetown University Press.

Darcy, Isabelle. In prep. *Second language laboratory phonology: An experimental and methodological guide*. New York: Routledge.

Davidson, Lisa & Jason A. Shaw. 2012. Sources of illusion in consonant cluster perception. *Journal of Phonetics* 40. 232–248.

Dehaene-Lambertz, Ghislaine, Emmanuel Dupoux & A. Gout. 2000. Electrophysiological correlates of phonological processing: A cross-linguistic study. *Journal of Cognitive Neuroscience* 12. 635–647.

de Leeuw, Esther, Linnaea Stockall, Dimitra Lazaridou-Chatzigoga & Celia Gorba Masip. 2019. Illusory vowels in Spanish–English sequential bilinguals: Evidence that accurate L2 perception is neither necessary nor sufficient for accurate L2 production. *Second Language Research*. https://doi.org/10.1177/0267658319886623.

Demirezen, Mehmet. 2010. The specification of the difficulties in the perception and articulation of the English schwa phoneme by Turkish English teachers and students. Procedia– Social and Behavioral Sciences 2. 1572–1576.

Derwing, Tracey M. & Murray J. Munro. 2015. Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research. Amsterdam: John Benjamins.

Duanmu, San. 2007. The phonology of standard Chinese, 2nd edn. Oxford: OUP.

Dupoux, Emmanuel, Erika Parlato-Oliveira, Sonia Frota, Yuki Hirose & Sharon Peperkamp. 2011. Where do illusory vowels come from? Journal of Memory and Language 64. 199–210.

Dupoux, Emmanuel, Kazuhiko Kakehi, Yuki Hirose, Christophe Pallier & Jacques Mehler. 1999. Epenthetic vowels in Japanese: A perceptual illusion? Journal of Experimental Psychology: Human Perception and Performance 25. 1568–1578.

Durvasula, Karthik, Ho-Hsin Huang, Sayako Uehara, Qian Luo & Yen-Hwei Lin. 2018. Phonology modulates the illusory vowels in perceptual illusions: Evidence from Mandarin and English. Laboratory Phonology: Journal of the Association for Laboratory Phonology 9. 1–27.

Eckman, Fred. 1981. On the naturalness of interlanguage phonological rules. Language Learning 31. 195–216.

Eddington, David. 2001. Spanish epenthesis: Formal and performance perspectives. Studies in the Linguistic Sciences 31. 33–53.

Escudero, Paola & Paul Boersma. 2004. Bridging the gap between L2 speech perception research and phonological theory. Studies in Second Language Acquisition 26. 551–585.

Farwaneh, Samira. 1995. Directionality effects in Arabic dialect syllable structure. Utah: University of Utah.

Feng, Zeyu. 2020. Effects of identification and pronunciation training methods on L2 speech perception and production: Training adult Japanese speakers to perceive and produce English/r/-/l/. Studies in Applied Linguistics and TESOL 20(2). 57–83.

Field, John. 2003. Promoting perception: Lexical segmentation in L2 listening. ELT Journal 57(4). 325–333.

Flege, James. 1995. Second language speech learning: Theory, findings, and problems. In Winifred Strange (ed.), Speech perception and linguistic experience: Issues in crosslanguage research, 233–277. Baltimore: York Press.

Flege, James & Ian MacKay. 2004. Perceiving vowels in a second language. Studies in Second Language Acquisition 26. 1–34.

Flemming, Edward. 2007. The phonetics of Schwa vowels. http://web.mit.edu/flemming/www/paper/schwaphonetics (accessed 30 May 2017).

Fox, John, Sanford Weisberg, Michael Friendly, Jangman Hong, Robert Anderson, David Firth & Steve Taylor & the R Core Team.2020 .Effect displays for linear, generalized linear, and other models. Available at: https://cran.r-project.org/web/packages/effects/effects.pdf.

Gabryś-Barker, Danuta. 2011. Phonological issues in second language acquisition studies: Focus areas and implications for EL instructional practices. In Janusz Arabski & Wojtaszek Adam (eds.), The acquisition of L2 phonology, 123–137. Bristol: Multilingual Matters.

Gambi, Chiara & Martin Pickering. 2013. Prediction and imitation in speech. Frontiers in Psychology 4. 340.

Gao, Zhiyan & Steven Weinberger. 2018. Which phonetic features should pronunciation instructions focus on? An evaluation on the accentedness of segmental/syllable errors in L2 speech. Research in Language 16. 135–154.

Gibson, Mark. 2012. Perception-based vowel insertion by native Spanish-speaking learners of English. TIPA. Travaux Interdisciplinaires sur la Parole et le Langage 28. 1–22.

Goh, Yeng-Seng. 2000. Beijinghua shi danyinjie yuyan de zhiyi [Is Beijing Mandarin a monosyllabic language?]. Contemporary Linguistics 2(4). 231–247. [吳英成, 2000, 北京話是單音節語言的置疑. 當代語言學 2.4. 231–47].

Gullberg, Marianne, Leah Roberts, Christine Dimroth, Veroude Kim & Indefrey Peter. 2010. Adult language learning after minimal exposure to an unknown natural language. Language Learning 60. 5–24.

Gut, Ulrike. 2009. Non-native speech: A corpus-based analysis of phonological and phonetic properties of L2 English and German. Berlin: Peter Lang.

Hancin-Bhatt, Barbara. 2000. Optimality in second language phonology: Codas in Thai ESL. Second Language Research 16. 201–232.

Hancin-Bhatt, Barbara. 2008. Second language phonology in optimality theory. In Jette G. Hansen-Edwards & Mary L. Zampini (eds.), Phonology and second language acquisition, 117–146. Amsterdam: John Benjamins.

Hancin-Bhatt, Barbara & Rakesh Bhatt. 1997. Optimal L2 syllables: Interaction of transfer and developmental effects. Studies in Second Language Acquisition 19. 331–378.

Hansen, Jette G. 2004. Developmental sequences in the acquisition of L2 syllable codas: A preliminary study. Studies in Second Language Acquisition 26. 85–124.

Hansen-Edwards, Jette.2011. Deletion of /t,d/ and the acquisition of linguistic variation by second language learners of English. Language Learning 61. 1256–1301.

Hansen-Edwards, Jette G. & Mary L. Zampini. 2008. Phonology and second language acquisition. Amsterdam: John Benjamins.

Hardison, Debra. 2021. Multimodal input in second-language speech processing. Language Teaching 54. 206–220.

Houde, John & Michael Jordan. 2002. Sensorimotor adaptation of speech I: Compensation and adaptation. Journal of Speech, Language and Hearing Research 45. 295–310.

Huensch, Amanda & Annie Tremblay. 2015. Effects of perceptual phonetic training on the perception and production of second language syllable structure. Journal of Phonetics 52. 105–120.

Inceoglu, Solène. 2016. Effects of perceptual training on second language vowel perception and production. Applied Psycholinguistics 37. 1175–1199.

Ioup, Georgette. 1984. Is there a structural foreign accent? A comparison of syntactic and phonological errors in second language acquisition. Language Learning 34. 1–15.

Ioup, Georgette & Steven Weinberger. 1987. Interlanguage phonology. Rowley: Newbury House Publishers.

Iverson, Paul, Patricia K. Kuhl, Reiko Akahane-Yamada, Eugen Diesch, Yoh'ich Tohkura, Andreas Kettermann & Claudia Siebert. 2003. A perceptual interference account of acquisition difficulties for non-native phonemes. Cognition 87. 47–57.

Jenkins, Jennifer. 2012. Current perspectives on teaching world Englishes and English as a lingua franca. TESOL Quarterly 40. 157–181.

Kabak, Bariş & William Idsardi. 2007. Perceptual distortions in the adaptation of English consonant clusters: Syllable structure or consonantal contact constraints? Language and Speech 50. 23–52.

Kennedy, Sara & Pavel Trofimovich. 2017. Pronunciation acquisition. In Shawn Loewen & Masatoshi Sato (eds.), Routledge handbook of instructed second language acquisition, 260–279. London: Routledge.

Kenworthy, Joanne. 1987. Teaching English pronunciation. London: Longman.

Kissling, Elizabeth. 2015. Phonetics instruction improves learners' perception of L2 sounds. Language Teaching Research 19. 254–275.

Kissling, Elizabeth. 2018. Pronunciation instruction can improve L2 learners' bottom-up processing for listening. The Modern Language Journal 102. 653–675.

Kubozono, Haruo. 1999. Nihongo-no onsei (Japanese phonetics). Japan: Iwanami Shoten.

Lacabex, Gomez Esther & Francisco Gallardo-del-Puerto. 2020. Explicit phonetic instruction versus implicit attention to native exposure: Phonological awareness of English schwa in CLIL. International Review of Applied Linguistics in Language Teaching 58. 419–442.

Lado, Robert. 1957. Linguistics across cultures. Michigan: University of Michigan Press.

Lan,Yizhou&SunyoungOh.2012.Comparative L2 English [r] cluster production by Cantonese and Mandarin speakers. In Paper presented at International conference on bilingualism and comparative linguistics. Hong Kong: Chinese University of Hong Kong.

Lee, Bradford, Luke Plonsky & Kazuya Saito. 2020. The effects of perception- versus production based pronunciation instruction. System 88. 1–13.

Lee, Junkyu, Juhyun Jang & Luke Plonsky. 2015. The effectiveness of second language pronunciation instruction: A meta-analysis. Applied Linguistics 36. 345–366.

Lenneberg, Eric. 1967. Biological foundation of language. London: Wiley.

Levis, John. 2018. Intelligibility, oral communication and the teaching of pronunciation. Cambridge: CUP.

Li, Aike & Brechtje Post. 2014. L2 acquisition of prosodic properties of speech rhythm: Evidence from L1 Mandarin and German learners of English. Studies in Second Language Acquisition 36. 223–255.

Liberman, Alvin & Ignatius Mattingly. 1985. The motor theory of speech perception revised. Cognition 21. 1–36.

Lin, Yuh-Huey.2003. Interphonology variability: Sociolinguistic factors affectingL2 simplification strategies. Applied Linguistics 24(4). 439–464.

Long, Michael. 1990. Maturational constraints on language development. Studies in Second Language Acquisition 12. 251–285.

MacKenzie, Ian. 2017. The linguistics of Spanish. http://www.staff.ncl.ac.uk/i.e.mackenzie/ index.html (accessed September 2017).

Magen, Harriet. 1998. The perception of foreign-accented speech. Journal of Phonetics 26. 381–400.

Major, Roy C. 1987. A model for interlanguage phonology. In Georgette Ioup & Steven Weinberger (eds.), Interlanguage phonology, 101–124. Rowley: Newbury House Publishers.

Major, Roy C. 2008. Transfer in second language phonology: A review. In Jette Hansen & Mary Zampini (eds.), Phonology and second language acquisition, 63–94. John Benjamins: Amsterdam.

Matthews, John & Cynthia Brown. 2004. When intake exceeds input: Language specific perceptual illusions induced byL1 prosodic constraints. International Journal of Bilingualism 8(1).5–27.

Mora, Joan & Mayya Levkina. 2017. Task-based pronunciation teaching and research: Key issues and debates. Studies in Second Language Acquisition 39. 381–399.

Munoz, Carmen & David Singleton. 2011. A critical review of age-related research on L2 ultimate attainment. Language Teaching 44. 1–35.

Murphy, John & Amanda Baker. 2015. History of ESL pronunciation teaching. In Marnie Reed & John Levis (eds.), The handbook of English pronunciation, 36–65. London: Wiley-Blackwell.

Nyvad, AnneMette, Michaela Hejná, Anders Højen, Anna Bothe Jespersen & Mette Hjortshøj Sørensen (eds.). 2019. A sound approach to language matters – In honor of Ocke-Schwen Bohn. Aarhus: Dept. of English, School of Communication & Culture, Aarhus University.

Oyama, Susan. 1976. A sensitive period for the acquisition of a non-native phonological system. Journal of Psycholinguistic Research 5. 261–285.

Parker, Steve. 2015. Sounding out sonority. Language and Linguistics Compass 11. https://doi. org/10.1111/lnc3.12248.

Patkowski,Mark.1990. Age and accent in a second language: A reply to James Emil Flege. Applied Linguistics 11. 73–89.

Peperkamp, Sharon & Emmanuel Dupoux.2003. Reinterpreting loanword adaptations: The role of perception. Proceedings of the 15th International Congress of Phonetic Sciences, 367–370. Barcelona: Causal Productions.

Piske, Thorsten & Martha Young-Scholten. 2009. Input matters in SLA. Bristol: Multilingual Matters.

Postovsky, Valerian A. 1974. Effects of delay in oral practice at the beginning of second language learning. Modern Language Journal 58. 229–239.

R Core Team. 2019. R: A language and environment for statistical computing (version 3.6.1). Vienna, Austria: R Foundation for Statistical Computing.

Saito, Kazuya & Roy Lyster. 2012. Effects of form-focused instruction and corrective feedback on L2 pronunciation development of /ɹ/ by Japanese learners of English. Language Learning 62. 595–633.

Sakai, Mari & Colleen Moorman. 2018. Can perception training improve the production of second language phonemes? A meta-analytic review of 25 years of perception training research. Applied Psycholinguistics 39. 187–224.

Selkirk, Elisabeth. 1982. The syllable. In Harry vander Hulst & Norval Smith (eds.), The structure of phonological representations: Part 2, 337–384. Dordrecht: Foris.

Shea, Christine. Experimental Research in Multilingual Speech Perception. New York: Routledge, in prep.

Shinohara, Yasuaki & Paul Iverson. 2018. High variability identification and discrimination training for Japanese speakers learning English/r/–/l/. Journal of Phonetics 66. 242–251.

Simon, Ellen. 2010. Phonological transfer of voicing and devoicing rules: Evidence from L1 Dutch and L2 English conversational speech. Language Sciences 32. 63–86.

Strange, Winifred. 1995. Speech perception and linguistic experience: Issues in cross-language speech research. Baltimore: York Press.

Strange, Winifred & Valerie L. Shafer. 2008. Speech perception in second language learners: The re-education of selective perception. In Jette Hansen-Edwards & Mary Zampini (eds.), 153–192.

Swan, Michael. 2001. Learner English: A teacher's guide to interference and other problems. Cambridge: Cambridge University Press.

Tarone, Elaine. 1979. Interlanguage as chameleon. Language Learning 29(2). 181–191.

Thomson, Ron & Tracey Derwing. 2015. The effectiveness of L2 pronunciation instruction: A narrative review. Applied Linguistics 36. 326–344.

Wang, Chilin. 1995. The acquisition of English word-final obstruents by Chinese speakers. New York: University of New York.

Watson, Janet. 2002. The phonology and morphology of Arabic. Oxford: Oxford University Press.

Wayland, Ratree. 2021. Second language speech learning: Theoretical and empirical progress. Cambridge: CUP.

Weinberger, Steven H. 1994. Functional and phonetic constraints on second language phonology. In Mehmet Yavas (ed.), First and second language phonology, 283–302. San Diego: Singular Publishing Group.

Werker, Janet F. 1994. Cross-language speech perception: Development change does not involve loss. In Judith C. Goodman & Howard C. Nusbaum (eds.), The development of speech perception: The transition from speech sounds to spoken words, 93–120. Cambridge, Massachusetts: MIT Press.

Wilson, Magnus. 2003. Discovery listening: Improving perceptual processing. ELT Journal 57(4). 335–343.