

Research Article

Joan C. Mora* and Isabelle Darcy

Individual differences in attention control and the processing of phonological contrasts in a second language

<https://doi.org/10.1515/phon-2022-0020>

Received July 13, 2022; accepted May 20, 2023; published online June 21, 2023

Abstract: This study investigated attention control in L2 phonological processing from a cognitive individual differences perspective, to determine its role in predicting phonological acquisition in adult L2 learning. Participants were 21 L1-Spanish learners of English, and 19 L1-English learners of Spanish. Attention control was measured through a novel speech-based attention-switching task. Phonological processing was assessed through a speeded ABX categorization task (perception) and a delayed sentence repetition task (production). Correlational analyses indicated that learners with more efficient attention switching skill and faster speed in correctly identifying the target phonetic features in the speech dimension under focus could perceptually discriminate L2 vowels at higher processing speed, but not at higher accuracy rates. Thus, attentional flexibility provided a processing advantage for difficult L2 contrasts but did not predict the extent to which precise representations for the target L2 vowels had been established. However, attention control was related to L2 learners' ability to distinguish the contrasting L2 vowels in production. In addition, L2 learners' accuracy in perceptually distinguishing between two contrasting vowels was significantly related to how much of a quality distinction between them they could make in production.

Keywords: attention control; English; L2 consonant perception and production; L2 vowel perception and production; Spanish

***Corresponding author: Joan C. Mora**, Department of Modern Languages and Literatures and English Studies, Faculty of Philology and Communication, Universitat de Barcelona, Gran Via de les Corts Catalanes 585, 08007 Barcelona, Spain, E-mail: mora@ub.edu. <https://orcid.org/0000-0001-6071-2008>
Isabelle Darcy, Department of Second Language Studies, Indiana University, Bloomington, IN, USA, E-mail: idarcy@indiana.edu. <https://orcid.org/0000-0003-3889-2414>

1 Introduction

Speaking and understanding a second language (L2) is a complex, cognitively demanding task. L2 users at all levels of competence normally need to put more effort when using their L2 than when using the language they grew up speaking at home (first language or L1) in some or all of the linguistic domains (morphology, syntax, vocabulary, pronunciation) needed to function fluently in everyday communication. This is because in the L1, many of the processes involved in using language (e.g. grammatical and phonological encoding and decoding; lexical activation, selection and retrieval; articulation) are characterized by automaticity and processing efficiency and occur fluently and effortlessly. By contrast, in the L2 such processes are less automatic, require effortful processing and usually result in dysfluent language use (Segalowitz 2010). This is particularly notorious in the case of instructed adult L2 learners in classroom environments with very limited exposure to authentic L2 input and few opportunities for meaningful language use beyond a few hours of instruction per week (Muñoz 2014). The limitations of this kind of language learning experiences are especially striking in the domain of L2 phonology for most learners, because sustained L2 input is necessary for learners to improve L2 phonological processing as well as to establish precise phonetic representations for L2 sounds, which would allow them to acquire the segmental contrasts of the L2 and develop an L2 phonological system (Tyler 2019).

Although exceptional outcomes in phonology have been reported for some learners, arguably due to a combination of learning styles and cognitive (aptitude and talent), psychological (motivation and strong sense of L2 self) and experiential factors (e.g., length of residence) (Moyer 1999, 2014), most adult L2 learners struggle with L2 pronunciation and, without pronunciation instruction, tend to see only modest improvements in comprehensibility or accentedness over time. In addition, previous studies assessing L2 phonological acquisition in naturalistic, classroom and lab training contexts have shown large inter-learner variation in performance (Bradlow et al. 1999; Derwing and Munro 2013; Golestani and Zatorre 2009; MacKay et al. 2001). The sources of this variability have been widely investigated and attributed to a myriad of linguistic, contextual and learner variables: the extent to which the L1 and the L2 differ phonetically, age-related factors, amount and quality of L2 input, frequency and amount of L1 and L2 use, motivation, or language learning aptitude (see Munro and Bohn 2007; Piske et al. 2001 for a review). Among these factors, age of onset of L2 learning, input quality and quantity and amount of L2 use have been shown to explain a substantial amount of variance in L2 phonological acquisition in immersion settings (Flege 2008), whereas aptitude-related factors remain under-researched in both naturalistic and classroom learning contexts.

Executive control functions (e.g., memory, attention, inhibition) constitute one source of aptitude-related inter-learner variation in L2 phonological acquisition (e.g., Darcy et al. 2016; Ghaffarvand Mokari and Werner 2019), as they underlie the effective functioning of speech processing mechanisms both in L1 and L2. Recent research (Saito et al. 2019, 2020, 2021) also suggests that general auditory processing skills may explain L2 speech development. In the present study we investigate the cognitive mechanism of attention control, and more specifically attentional flexibility (i.e., attentional switching skill) as a source of individual differences in L2 speech perception and production in instructed L2 learners.

1.1 Attention control and language processing

The executive network of the human brain, also known as *executive control* or *executive function*, is responsible for a set of cognitive control mechanisms (*executive functions*) that allow an individual to function efficiently in terms of self-control, problem solving, task shifting, action planning and goal implementation (Petersen and Posner 2012). Such mechanisms include the updating and mental manipulation of information in working memory (*updating*), selectively attending to information under focus while inhibiting irrelevant information (*inhibiting*), and efficiently shifting attention between tasks or representations (*shifting*) (Miyake and Friedman 2012), all of which are implicated in speech processing and language comprehension and production, and consequently in second language acquisition (SLA). Phonological short-term memory, the subcomponent of working memory responsible for temporarily holding auditory verbal information in working memory (*updating*), is implicated in L2 vocabulary acquisition (Speciale et al. 2004), L2 grammar learning (French and O'Brien 2008; Kormos and Sáfár 2008) and L2 speech perception (Darcy et al. 2015; MacKay et al. 2001). Inhibitory control is responsible for a bilinguals' control of language interference by inhibiting the language not in use (Green 1998). It is thus related to L2 phonological development, leading to lower levels of L2 phonological influence on the L1 in long-term immersion (Lev-Ari and Peperkamp 2013, 2014) as well as less interference from the L1 in L2 phonological processing in instructed SLA (Darcy et al. 2016). However, much less research has investigated attention control as a source of inter-learner variability in L2 phonological acquisition despite its potentially central role (Ellis 2006; Robinson 1995). Since attention control determines the extent to which linguistic form can be attended to while meaning is being processed (Van Patten 2004) it could play a mediating role between input and acquisition.

Attention control (*shifting*) has been shown to explain a significant amount of variance in L2 learners' proficiency (Segalowitz and Frenkiel-Fishmann 2005) and

speaking fluency (Taube-Schiff and Segalowitz 2005). Attention also relates to general processing mechanisms involved in the perception and production of speech. For example, it guides auditory processes during speech perception by focussing processing resources on the relevant information, and by allowing listeners to select the acoustic information that is critical for appropriately interpreting auditory events during oral communication (Baese-Berk et al. 2015; Mattys and Wiget 2011). Attention shifting has also been shown to facilitate perceptual learning, predicting listeners' skill in understanding an unfamiliar accent (Janse and Adank 2012) and seems connected to processing speed in native speakers' tonal discrimination (Ou et al. 2015; Ou and Law 2017). Taken together, attention shifting appears to contribute to experience-related quality differences in the nature of phonological representations in long-term memory (Heald and Nusbaum 2014). In speech production, attention skills are important in word planning processes (Sikora et al. 2016) and in resolving the selection of cross-linguistically co-activated linguistic representations in bilinguals (Kroll et al. 2008).

Given that individuals vary in attentional capacity (Petersen and Posner 2012) and in the use they make of their attentional resources (Wager et al. 2006), inter-learner differences in phonological attainment may be partly due to individual differences in attention control. Learners must be able to shift their attentional focus flexibly between various phonetic cues and phonological dimensions (e.g., segmental duration and quality, distributional constraints, pitch changes) during phonological processing as spoken messages unfold in time in a way that is specific of the L2 and may differ from the L1. For example, in English segmental duration needs to be attended to as a primary cue to the voicing of word-final devoiced obstruents (longer /eɪ/ in *plays* than *place*) but is normally negligible as a cue in the identification of vowels (longer /i:/ in *beat* than *bit*), where vowel quality is the primary cue. Thus, learners of English need to develop attentional flexibility in the L2-specific use of segmental duration as a phonological cue in the phonology of English.

Although indirect evidence of the implication of attention in L2 speech learning may be found in the effectiveness of directing learners' attention towards specific phonetic dimensions during phonetic training (Guion and Pederson 2007) or acoustic cue manipulations (Iverson et al. 2005), evidence of a direct relationship between attention control skills and L2 speech learning is still inconclusive. For example, some phonetic training studies have found an association between auditory selective attention and accuracy gains in perceiving target phonological contrasts (Mora and Mora-Plaza 2019; Oliveira 2020), but others have not (Ghaffarvand Mokari and Werner 2019). One study examined whether differences in attention control predicted accuracy of L2 phono-lexical encoding but did not find a relationship (Daidone and Darcy 2021). The present study extends this line of research by focussing on the relationship between attention switching skill (measured through a novel speech-

based attention switching task) and L2 phonological processing in the perception and production of L2 sound contrasts.

2 The present study

The goal of the present study is to explore the relationship between attention control and L2 phonological processing from an individual differences perspective. We hypothesized that a more efficient attention control may enhance the processing of acoustic-phonetic information in the input by bringing relevant (L2-specific) acoustic information to the foreground during speech processing while keeping irrelevant information in the background, which would lead to more accurate processing of L2 phonological categories in perception and production. We examine the relationship between performance on a domain-specific (speech) measure of attention control and measures of L2 phonological perception and production for two groups of late-onset L2 learners: a group of L1-English learners of Spanish and a group of L1-Spanish learners of English. Following previous research in the domain of grammar (Segalowitz and Frenkiel-Fishmann 2005), we chose to test attention switching skill through a domain-specific task (speech) that aimed at capturing L2 learners' individual differences in attention control during the processing of two types of phonetic information required in L2 speech learning: a specific language-independent segmental aspect of speech sounds (nasal vs. non-nasal) and a set of language-specific phonetic differences characterising a sound sequence (Spanish-like phonetics vs. English-like phonetics). As the phonetic and phonological dimensions that need to be attended to for successful phonological development are complex and mostly language-specific, we think of efficient attention control as a built-in *cue enhancement device* by means of which the appropriate relevant phonetic cues are brought to the perceptual foreground in L2 speech processing. We assume that learners do not only need to learn to attend to the relevant phonetic cues or dimensions when processing L2 speech sound contrasts (i.e. they need to learn the specific phonetic cue-weighting of linguistically relevant phonetic dimensions such as voicing, duration, or spectral information in the L2), they also need to learn to bring a specific phonetic dimension to the attentional foreground in one context and to the attentional background in another. L2 learners' ability to efficiently switch their attention between a specific language-independent segmental aspect of speech sounds (e.g. presence vs. absence of nasal resonance for nasal consonants) and a set of language-specific phonetic differences characterising a sound sequence (differential segmental phonetic properties of Spanish-like vs. English-like speech, e.g. VOT in oral stops or unstressed vowel reduction) is a way to obtain a measure of attention switching skill in a speech processing context that would closely resemble L2

learners' use of their attentional skills in processing L2 phonetic information. None of the currently available methods of assessing attention control have exclusively targeted phonological dimensions (but see Darcy et al. 2015; Safronova 2016), which we deemed crucial in establishing a link between attention control and phonological acquisition during L2 phonological processing. In this study, we have developed a novel, fully phonologically oriented version of an attention control task with the aim of observing L2 learners' efficiency in the use of their attentional flexibility resources during the phonological processing of L2 auditory stimuli.

3 Methods

We obtained measures of attention control with our attention shifting task and examined how these related to measures of learners' L2 speech perception (ABX categorical discrimination task) and production (delayed sentence repetition). We also obtained demographic background information from learners and, as vocabulary size may partly determine L2 phonological competence (Bundgaard-Nielsen et al. 2011), we also estimated their receptive vocabulary size as a phonologically-related measure of overall proficiency (Uchihara and Clenton 2020) that we could control for. Participants included in the study had passed a pure-tone audiometry test (Reilly et al. 2007). A retrieval-induced inhibition task and a working memory (serial non-word recognition) task were also administered but are not reported here.

We tested L2 speech perception bi-directionally (Darcy et al. 2016), i.e., both learner groups were tested on English and Spanish stimuli, so that the English and Spanish stimuli served as control stimuli for the L1-English and L1-Spanish learners, respectively, and L1-English learners served as controls for the Spanish-learners' performance on the English stimuli, and vice versa. This design enhances generalizability because of the language-independent nature of any potential effects and correlations. For L2 speech production we used baseline measures from two groups of L1-English ($n = 7$) and L1-Spanish ($n = 6$) speakers recruited in the US and Spain, respectively, and who grew up speaking their L1 (English or Spanish) at home from birth. These speakers reported having undergone a primarily monolingual language learning experience. They had studied foreign languages at school but reported only using their L1 (English or Spanish) on a daily basis and stated they could not speak languages other than their L1 (English or Spanish) in a fluent manner. A limitation of this approach is that learners' performance was being compared to L1-English or L1-Spanish baselines that do not necessarily correspond to the input L2 learners are exposed to when acquiring the L2 through formal instruction or exposure to media.

The testing procedures were similar for both learner groups. They did the production task first, followed by the attention control task, the perception task, the

vocabulary size test and the demographic and language background questionnaires. The testing session lasted 90 min approximately, including breaks between tasks. L2-Spanish learners were tested in a psycholinguistics laboratory at Indiana University in Bloomington (USA), whereas L2-English learners were tested (2–4) in the phonetics laboratory at the University of Seville (Spain). Participants were compensated for their participation either through a small payment or a USB-memory drive.

3.1 Participants

Participants were 21 Spanish learners of English (L2-English) and 19 English learners of Spanish (L2-Spanish) (see Table 1 for demographics). The learners’ language background was determined in the call for participation stating that we were looking for native speakers of Spanish (in Spain) and of English (in the US). In addition, in the language background questionnaire participants were asked to fill in, several questions were included to determine whether they had been raised in either Spanish- or English-speaking homes in a Spanish- or English-speaking environments, with Spanish and English, respectively, as their only language of exposure. Current L2-use was estimated asking participants to choose from 5 L2-use intensity levels (0 = 0 %, 1 = 1–25 %, 2 = 26–50 %, 3 = 51–75 %, and 4 = 76–100 %) in 9 L2-use situations (e.g., conversations with friends, in internet chats, while shopping). that would produce an L2-use maximum score of 36, corresponding to an estimated L2 use of 76–100 %. Participants also self-evaluated their L2 proficiency on a 5-point scale (1 = very

Table 1: Means (standard deviations in parentheses) for participants’ demographic variables.

Variable	L2 English (n = 21)		L2 Spanish (n = 19)	
	M	SD	M	SD
Age at testing	23.3	4.8	20.3	1.0
Motivation (1–9)	6.0	0.6	5.6	0.4
Current L2 use (max. 36)	16.4	5.4	9.4	6.9
Self-rating (1–5)	4.0	0.5	4.0	0.6
Residence abroad (weeks)	4.7	9.2	11.4	28.1
Years of study	11.7	2.6	8.7	2.8
Age of first L2 exposure	8.2	3.1	8.4	4.3
Age of first L2 use	13.1	4.8	10.5	4.0
Gender (% female)	66.6		84.2	
X_Lex	4,257	485	3,715	602

poor, 5 = very well) in speaking, understanding, reading and writing the L2, and we used the 4 ratings to compute an average score by participant. Motivation to learn or use the L2 was assessed through 9 statements (e.g., “I enjoy learning new words and new ways of saying things in English/Spanish”) participants reacted to by selecting a level on a Likert-type agreement scale (1 = strongly agree, 9 = strongly disagree). A mean motivation score was obtained by averaging the scale score chosen for each statement. Vocabulary size was estimated through the Spanish and the English versions of a 120-item yes/no vocabulary size test (X_Lex, Meara and Milton 2003), which yields an estimate of receptive vocabulary size up to 5,000 words.

Overall, compared to the L2-English learners, L2-Spanish learners were younger, spoke the L2 less, had studied their L2 for less time and were slightly less motivated to learn their L2. However, both groups were comparable in terms of the age of onset of L2 learning, the age at which they started to use their L2, how long they had resided abroad and their self-reported level of L2 knowledge. However, L2-English learners had a significantly larger L2 vocabulary size than the L2-Spanish learners ($t(38) = -3.143$, $p = 0.003$), which might reflect a group difference in overall proficiency.

3.2 Attention control task

We developed a novel *speeded set-switching* task to measure attention control. In this task test trials were comprised of 10 nasal-initial nonwords, and 10 non-nasal-initial nonwords. All nonwords were disyllabic with a 'CVCV structure. We chose to mainly use shared phonological categories for English and Spanish so as not to disadvantage one group over another by having to process too many unfamiliar sounds. Therefore, we created nonwords such as “saso”, which could be pronounced distinctly in Spanish ([ˈsaso]) and English ([ˈsæsəʊ]). These 20 nonwords were recorded with English and Spanish phonetics by two female balanced early bilinguals who spoke Mexican Spanish and American English, so that voice identity could not be used to determine the stimulus language. These speakers reported having been raised speaking both Spanish and English in Spanish-English families, they had lived in both Spanish- and English-speaking environments (Mexico, Spain and the US) for extended periods of time, and they did not have a perceptually detectable Spanish accent when speaking English or an English accent when speaking Spanish.

The two phonological dimensions used, *nasality* and *L1 phonetics* can be considered comparable in difficulty across both participant L1s. *Nasality*, probes whether a stimulus initial sound is a nasal sound (/n/ or /m/) or not, whereas *L1 phonetics* probes whether a stimulus is produced with L1 or L2 phonetics. We chose these two dimensions because they trigger a fast, automatic decision (phoneme or

Table 2: Nonword stimuli sample.

	Spanish	English		Spanish	English
<i>noma</i>	[ˈnoma]	[ˈnəʊmə]	<i>pigo</i>	[ˈpiɣo]	[ˈpʰɪgəʊ]
<i>nole</i>	[ˈnole]	[ˈnəʊleɪ]	<i>dofe</i>	[ˈdofe]	[ˈdəʊfeɪ]
<i>niso</i>	[ˈniso]	[ˈnɪsəʊ]	<i>saso</i>	[ˈsaso]	[ˈsæəsəʊ]

accent detection, respectively) which can only be based on the phonetic properties of the stimuli, since they were all phonotactically legal non-words in the participants’ L1 and L2. For example, participants had to decide on the presence of a nasal resonance at the beginning of the word, such as [ˈnole] as opposed to [ˈsaso], or on the English-like diphthongal realization of a vowel, such as [ˈdoufeɪ] as opposed to Spanish-like monophthongal [ˈdofe] (Table 2).

Participants were asked to answer one of two possible questions: “Nasal?” versus “English?” (or “Nasal?” vs. “Spanish?” for L1-Spanish speakers) with respect to an auditory stimulus by pressing one of two assigned computer keys (*yes* or *no*). An experimental trial consisted of a fixation cross displayed for 500 ms, followed by the question (e.g., “Nasal?”) displayed for 500 ms, followed by an auditory stimulus (e.g., [ˈnofe], spoken with Spanish phonetics). The task was administered with the software *DMDX* (Forster and Forster 2003). After a warm-up phase of 16 trials, and 8 practice trials on which feedback for accuracy (correct! or wrong) and speed (e.g., too slow! or 1800 ms) was provided, participants completed 82 trials.

Switch (S) trials, those showing a different question from the previous trial, alternated predictably with repeat (R) trials, those showing the same question as the previous trial, in SRSR sequences (Monsell 2003). Switch trials required participants to refocus their attention onto a different dimension and were expected to induce a switching cost, whereas repeat trials provided a baseline reaction time. The audio files were randomly ordered to match a SRSR sequence, resulting in two lists, one for each L1 with the only restriction that two “similar” tokens (e.g., /dofe/ spoken in Spanish and English) could not follow each other. Tokens from either voice were randomly assigned to a roughly equal number of items in each list.

3.3 L2 perception: speeded ABX discrimination task

In the speeded ABX categorization task (e.g., Gottfried 1984). Participants heard a sequence of three stimuli and had to identify the last stimulus (X) as either the same as A or B. The stimuli consisted of trisyllabic non-words in both Spanish and English

Table 3: Phonetic realization of sample stimuli.

Stimulus type	Language	Contrast type	Contrast	Stimulus A	Stimulus B
Test	Spanish	Vowel	/e/-/ei/	[fa'neða]	[fa'neiða]
		Consonant	/d/-/r/	[sa'ðeβo]	[sa'reβo]
	English	Vowel	/i/-/ɪ/	[fə'ni:dɪ]	[fə'nɪdɪ]
		Consonant	/ʃ/-/tʃ/	[sə'ʃi:dən]	[sə'tʃi:dən]
Control	Spanish	Vowel	/a/-/i/	[lu'pito]	[lu'pato]
		Consonant	/t/-/d/	[ga'taso]	[ga'ðaso]
	English	Vowel	/a/-/i/	[lə'pʰi:dɪk]	[lə'pʰædɪk]
		Consonant	/t/-/d/	[gə'tʰæfɪn]	[gə'dæfɪn]

with the structure CV.'CV.CV(C) (e.g., [fa'neða]). Physically different tokens produced by the two female early balanced bilinguals (Mexican Spanish and American English) were used in each trial: one voice for stimuli A and B and the other for X. Thus, learners had to correctly identify whether X contained the same vowel as item A or item B by comparing realizations of the same nonword produced by two different voices. If learners are able to correctly identify the target contrasting vowels or consonants as being the same in two items spoken by two different voices, we interpret this to indicate that learners have developed distinct phonetic category representations for the target L2 vowels or consonants at a pre-lexical phonological level. Therefore, a significant correlation between attention control and L2 learners' performance in the ABX task would indicate that learners with stronger attention control skills are more likely to have developed distinct phonetic categories for the difficult L2 sounds we targeted. All participants heard all Spanish and English stimuli, in two separate blocks. The L2 contrasts for L1-Spanish learners were L1 contrasts for the L1-English learners, and vice versa (Table 3). All of the L2 contrasts were deemed to pose learning difficulties for L2-Spanish (see Díaz and Simonet 2015, for /e/-/ei/; Rose 2010, for /d/-/r/) and L2-English learners (see Morrison 2009, for /i/-/ɪ/; Anrrich 2007, for /ʃ/-/tʃ/).

In total, four nonword pairs per condition were tested; each pair was repeated in four combinations (ABA, ABB, BAA, and BAB), yielding a total of 128 trials, 64 for each stimulus language. Trials were assigned to two blocks according to stimulus language (English-Spanish or vice-versa), and block order was counterbalanced across participants. Within each block, trials were randomized. If a participant made no response within 2500 ms, the next trial was initiated. The task was administered on a PC through headphones using the presentation software *DMDX* (Forster and Forster 2003), and took about 15 min to complete. We computed % correct accuracy scores to gauge L2 learners' ability to qualitatively distinguish the L2 sound contrasts in

perception and response time (RT) scores to gauge L2 learners' efficiency at perceptually processing the L2 sound contrasts, which reflected the robustness of their phonologically encoding. We expected group performance to be more accurate on L1 than L2 test contrasts and to be at ceiling for control contrasts, which would allow us to attribute differences in performance on the test condition to the L1 or L2 of the contrasts (rather than the stimulus language).

3.4 L2 production: delayed sentence-repetition task

In order to assess participants' production of the L2 contrasts tested perceptually in the ABX task we administered a delayed sentence repetition task (Trofimovich and Baker 2006) (see Appendix A). Participants performed 16 trials in the L2 in a recording booth. A trial consisted of the auditory and orthographic presentation of a question (prompt) and a following answer (response) 250 ms later, after which the prompt was presented again auditorily only (with a 500 ms delay) for the participant to repeat the previously heard response. The aim of this production task was to elicit the production of the same target sounds we tested in perception in a way that sound productions would reflect the nature of the phonetic categories learners have developed. We assumed that eliciting the target sounds in L2 words embedded in meaningful sentences presented in mini-dialogue format (prompt 1 → response 1; prompt 2 → response 2) where prompt 1 and response 1 are produced by different voices and prompt 2 intervenes between response 1 and the participants' response 2 (repeating response 1) would avoid participants focussing on the target word, which would facilitate eliciting the target vowels and consonants in a context that would closely resemble an L2 communicative context. We deemed this elicitation procedure would enhance the production of the target sounds in such a way that they would reflect the stage of development of the learners' L2. Although we cannot completely discard the possibility that participants would mimic the segmental content of the utterance, both the delay between the response and its repetition, and the intervening prompt between the response to be repeated and its repetition from memory, would minimize the possibility of direct mimicry as well as enhance attention to meaning rather than to segmental form (Trofimovich and Baker 2006). The stimuli in both languages were recorded by the two female balanced early bilinguals of Mexican Spanish and American English and were normalized for amplitude. In half of the prompt-response sets, one voice was used for the prompt token, and the other was used for the response tokens, and the reverse was done for the remaining sets. We elicited four pairs of words for each of the two contrasts embedded in 16 response sentences in L2-Spanish (/e/-/ej/: *maceta-aceite*, *pena-peina*, *reno-reino*, *vente-veinte*;

and /d/-/r/: *cada-cara, moda-moras, oda-oras, todos-toros*) and the same in L2-English (/i/-/ɪ/: *cheap-chips, feet-fit, seat-sit, sheep-ship*; and /ʃ/-/tʃ/: *shake-cheque, sheep-cheap, shows-chose, shops-chops*). The task took 5–7 min to complete.

Vowel production accuracy measures were based on the size of the Euclidean distance between the contrastive vowels and were contrast-specific. A larger Euclidean distance between two contrastive vowels represented a larger qualitative distinction between them in production, which was interpreted as an indication of higher production accuracy in contrastiveness (Melnik-Leroy et al. 2022). For the L2-Spanish monophthong-diphthong contrast /e/-/ei/, three measurement points (MP) were placed 20 %, 50 % and 80 % into the vowels, and the mean values for F1, F2, and f_0 were extracted from a 10 ms window centred at the three MPs. These frequency measures were first converted to Bark (B), and then a Bark-distance metric was computed by subtracting B0 from B1 (B1–B0) for tongue height and B1 from B2 (B2–B1) for degree of tongue fronting (Bohn and Flege 1990). We measured the amount of formant movement in the vowel by computing the Euclidean distance between the 20 % and the 50 % MPs and between the 50 % and the 80 % MPs. Then we added up the two Euclidean distances and used this spectral distance score as a measure of formant movement, as represented on the Bark-normalized vowel space. Higher formant movement indicates a diphthongized vowel (English-like), lower movement corresponds to a more Spanish-like monophthong. We also assessed whether the duration of the monophthong /e/ and the diphthong /ei/ were comparable across speaker groups by computing a duration difference score (in ms) and a duration difference ratio (e.g., /ei/ was 1.4 times longer than /e/) between e/ and /ei/ that would index how well learners could distinguish the monophthong from the diphthong in production.

For the L2-English /i:/ versus /ɪ/ contrast, F1, F2 and f_0 were extracted from a 15 ms window centred at the midpoint of the steady-state portion of the second formant of the vowel. The Euclidean distance between the contrasting vowels on a Bark-normalized vowel space was used as a measure of accuracy in qualitatively differentiating the two vowels, so that a larger distance was interpreted as a more English-like distinction between the vowels. Because Spanish learners of English have been shown to also rely on duration cues in distinguishing this tense-lax vowel contrast, unlike L1 English speakers who rely primarily on spectral cues (Escudero and Boersma 2004), a duration difference score (in ms) and a duration difference ratio (e.g., /i:/ was 1.1 times longer than /ɪ/) between /i:/ and /ɪ/ were computed as a measure of accuracy in quantitatively differentiating the two vowels.

For all consonant contrasts production accuracy was measured categorically (score 0–8) by visually and auditorily inspecting the spectrograms. For the L2-Spanish /d/-/r/ contrast an accurate realization of Spanish intervocalic /d/ was identified as a spirantized [ð], whereas accurate realizations of intervocalic /r/ had to

consist of a single-closure tap with very short constriction duration. For the L2-English /f/-/tf/ contrast realizations had to be palato-alveolar and show presence (/tf/) or absence (/f/) of a closure phase in the spectrogram.

4 Results

4.1 Attention control

As the descriptives in Table 4 below show, the attention switching task worked as expected in that both groups responded faster to repeat than to switch trials, regardless of question type. This indicates that switching dimensions (L1 or nasality) had a response time cost in milliseconds based on which a switching cost score (the difference between switch and repeat RTs) can be obtained as a measure of attention control.

A mixed-effects model was fitted to the response speed data for correct responses (in SPSS 25) with the factors *L2-group* (Spanish, English), *trial type* (switch, repeat), *dimension* (L1, nasal) and *stimulus language* (L1, L2) and their interactions as fixed effects. The random effects structure that did not lead to a convergence error

Table 4: Mean RT and accuracy (proportion correct) by trial type, dimension and L2 learner group.

	Group	Trial type	Dimension	Mean	SE	95 % CI	
						Lower	Upper
RT (ms)	L2 Spanish (<i>n</i> = 19)	Switch (S)	L1	1,147	46.5	1,056	1,238
			Nasal	1,104	46.2	1,013	1,194
		Repeat (R)	L1	1,061	46.3	971	1,152
			Nasal	983	46.0	893	1,073
	L2 English (<i>n</i> = 21)	Switch (S)	L1	1,212	44.5	1,125	1,299
			Nasal	1,158	44.3	1,071	1,245
		Repeat (R)	L1	1,161	44.3	1,074	1,248
			Nasal	1,107	43.9	1,021	1,193
Accuracy (proportion correct)	L2 Spanish (<i>n</i> = 19)	Switch (S)	L1	0.832	0.030	0.764	0.883
			Nasal	0.945	0.014	0.910	0.967
		Repeat (R)	L1	0.904	0.021	0.856	0.938
			Nasal	0.952	0.013	0.920	0.972
	L2 English (<i>n</i> = 21)	Switch (S)	L1	0.780	0.034	0.706	0.839
			Nasal	0.873	0.024	0.818	0.913
		Repeat (R)	L1	0.858	0.026	0.800	0.901
			Nasal	0.941	0.014	0.907	0.963

and provided a better fit of the data to the mixed-effects model (i.e. the lowest Akaike's information criterion AIC) included random intercepts for *subject* and *item*, and a random slope for *dimension* by *subject*. The significance threshold was set at $p = 0.05$ and in pairwise contrasts adjusted via sequential Bonferroni for all analyses. The visual analysis of residuals confirmed that the model was a satisfactory fit for the data structure. The parameter estimates are presented in Appendix B-1.

These analyses revealed significant main effects of *trial type* ($F(1, 2816) = 27.4$, $p < 0.001$) and *dimension* ($F(1, 2816) = 8.42$, $p = 0.004$). The main effect of *L2-group* did not reach significance ($F(1, 2816) = 2.16$, $p = 0.141$), suggesting that both groups did not differ significantly from one another in overall response speed. The *L2-group* \times *trial type* interaction reached significance ($F(1, 2816) = 9.07$, $p = 0.003$) because L2-Spanish learners were overall faster than L2-English learners on both repeat and switch trials (a difference that according to pairwise contrasts approached significance for repeat trials, but not for switch trials: $t(2816) = -1.93$, $p = 0.054$ and $t(2816) = -0.976$, $p = 0.329$, respectively). Crucially, both groups were significantly slower on switch than repeat trials (L2-Spanish: $t(2816) = 6.00$, $p < 0.001$; L2-English: $t(2816) = 2.84$, $p = 0.005$) and no other interaction involving *trial type* reached significance, suggesting that participants were slower on switch than repeat trials irrespective of *dimension* and *stimulus language*. The *L2-group* \times *stimulus language* interaction reached significance because whereas L2-English learners responded slightly more slowly to L2 stimuli than L1 stimuli ($t(2816) = -4.38$, $p < 0.001$), L2-Spanish learners did not ($t(2816) = 0.974$, $p = 0.330$). All other interactions turned out to be non-significant (all $F_s < 1.3$, all $p_s > 0.24$), suggesting that for both groups of learners responded faster on repeat than on switch trials for both dimensions and responded faster to the nasality dimension than to the L1 dimension regardless of trial type (see Table 4).

The dimensions *L1* and *Nasal* are therefore not equivalent in terms of RT. While this could be due to the position of the elements in the stimuli permitting the decision (initial for Nasal but anywhere for L1, yielding a faster RT in the first case), it could also be due to the dimensions differing in the complexity of the acoustic cues listeners had to process. Nasality cues are clear and well-defined, whereas cues indicating L1 versus L2 are multiple and therefore likely more complex to process. An analysis of response accuracy allows teasing apart this question. If the RT difference is due to a difference in cue complexity, accuracy scores might also be lower for the L1 dimension (more complex) than for the nasality dimension (less complex), indicating that both dimensions would not be fully comparable in terms of processing difficulty and cognitive complexity.

A mixed-effects model (binary logistic regression with a binomial distribution) was fit to the response accuracy data with the factors *L2-group* (Spanish, English), *trial type* (switch, repeat) and *dimension* (L1, nasal) and *stimulus language* (L1, L2) and their double interactions as fixed effects with random intercepts for *subject* and *item* (see

parameter estimates in Appendix B-2). The outcome of these analyses revealed a significant main effect of *L2 group* ($F(1, 3228) = 4.43, p = 0.036$) because overall L2-Spanish learners were significantly more accurate than L2-English learners were (0.92 vs. 0.87), in addition to being faster (see above). The main effects of *trial type* ($F(1, 3228) = 14.4, p < 0.001$) and *dimension* ($F(1, 3228) = 36.8, p < 0.001$) were significant, while neither the effect of *stimulus language* ($F(1, 3228) = 1.56, p = 0.211$) nor any of the interactions reached significance (all $F < 2.0$, all $p > 0.16$). Thus, the accuracy data follow the pattern of results for response speed; participants were significantly more accurate on repeat than switch trials for both questions (L1 and Nasality) and were significantly more accurate when responding to the nasality dimension than to the L1 dimension on both switch and repeat trial types (see Table 4).

These results suggest that switch costs were of a smaller magnitude in the nasal dimension than in the L1 dimension because nasality was an easier to process acoustic cue than L1/L2 phonetics. Therefore, the significant RT difference between the two dimensions was not only due to the position of the cues, but to a difference in complexity between dimensions. Consequently, because averaging RTs from the two dimensions might hide potential effects, we opted for computing two separate shift cost scores for each participant, one for each dimension, defined as the RT difference between switch and repeat trials, separately for the nasality and L1 conditions. We also used averaged RTs on switch trials only (separately for the nasality and L2 dimensions) as a general individual differences measure of how quickly L2 learners could re-focus their attention on a different dimension.

4.2 L2 perception: ABX discrimination

The results of the discrimination task (as shown in Figure 1) indicate that, as expected, participants performed at ceiling on L1 and L2 control contrasts and on L1 test contrasts, whereas they showed more difficulty in correctly discriminating L2 test contrasts (/i:/-/ɪ/ and /ʃ/-/tʃ/ for L2-English learners, /e/-/ej/ and /d/-/r/ for L2-Spanish learners; see Figure 1 and Table 6). An exception to this is the high accuracy rate L2-English learners obtained for /ʃ/-/tʃ/ ($M = 92, CI = 88-95$). We therefore decided to use the L2 vowel test scores as an index of individual performance in discrimination.

Mixed-effects models were fitted to the accuracy (binary logistic regression with a binomial distribution) and RT data with *L2-group* (Spanish, English), *stimulus language* (English, Spanish), and *condition* (control vs. test) and their interactions as fixed effects, with random intercepts for *subject* and *item*. The parameter estimates for each model (accuracy and RT) are presented in Appendices B3 and B4.

For accuracy, tests of fixed effects showed significant main effects of *L2-Group* ($F(1, 5112) = 7.92, p = 0.005$), *stimulus language* ($F(1, 5112) = 5.99, p = 0.014$) and *condition*

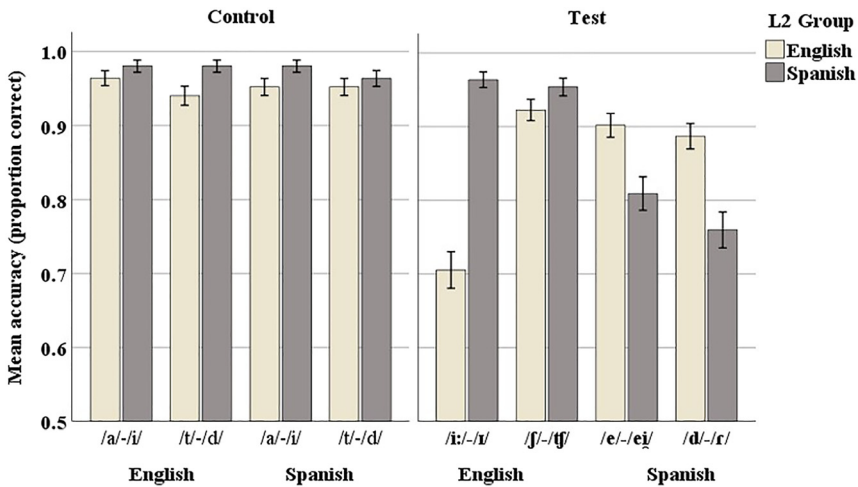


Figure 1: Mean accuracy rate (%) for control (left panel) and test (right panel) contrasts in both languages, indicated on the x-axis, as a function of learners' L2. Light bars represent L2 English learners, dark grey bars represent L2 Spanish learners (error bars = 95 % CI).

($F(1, 5112) = 62.5, p < 0.001$), and a significant $L2\text{-group} \times \text{stimulus language} \times \text{condition}$ triple interaction ($F(1, 5112) = 17.81, p < 0.001$). We analysed this interaction by splitting the data set by condition (i.e. test vs. control items) and fitting mixed-effects models with $L2\text{-group}$, stimulus language and their interaction as fixed effects, with random intercepts for *subject* and *item*. For test items, the main effects of $L2\text{-group}$ ($F(1, 2556) = 5.14, p = 0.023$) and stimulus language ($F(1, 2556) = 9.25, p = 0.002$) as well as their interaction ($F(1, 2556) = 86.6, p < 0.001$) reached significance. As expected, this interaction arose because both L2-Spanish ($t(2556) = 5.76, p < 0.001$) and L2-English ($t(2556) = -2.83, p = 0.003$) learners were significantly more accurate (proportion correct scores) on L1 than L2 test stimuli (Spanish or English). For control items, as expected, the main effects of $L2\text{-group}$ ($F(1, 2556) = 3.47, p = 0.063$) and stimulus language ($F(1, 2556) = 1.19, p = 0.730$) and their interaction ($F(1, 2556) = 1.19, p = 0.730$) did not reach significance, as all accuracy rates were at ceiling ($M > 0.94$).

$L2\text{-Group}$ effects, as predicted, emerged for the test condition because both groups, overall, were less accurate when listening to L2 stimuli than L1 stimuli. L2-Spanish learners were more accurate with English test contrasts ($M = 0.967$), and less so with Spanish test contrasts ($M = 0.803$), a significant ($t(5112) = 5.79, p < 0.001$) mean difference of 0.164 ($CI = 0.108\text{--}0.220$). Conversely, L2-English learners were more accurate when listening to Spanish test stimuli ($M = 0.906$) compared to English test stimuli ($M = 0.830$), a significant mean difference of 0.076 ($CI = 0.027\text{--}0.125$; $t(5112) = 3.02, p = 0.003$). For the control condition, both L2 groups performed equally

Table 5: Mean accuracy (proportion correct) by learner group, stimulus language and contrast.

L2 group	Stimulus type	Stimulus language	Contrast	M	SE	95 % CI	
						Lower	Upper
L2-Spanish	Control	English	/a/-/i/	0.982	0.008	0.956	0.993
			/t/-/d/	0.983	0.008	0.957	0.993
		Spanish	/a/-/i/	0.983	0.008	0.958	0.993
			/t/-/d/	0.969	0.012	0.936	0.985
	Test	English	/i/-/ɪ/	0.972	0.011	0.941	0.987
			/ʃ/-/tʃ/	0.959	0.014	0.921	0.979
		Spanish	/e/-/ej/	0.833	0.038	0.746	0.894
			/d/-/r/	0.781	0.045	0.681	0.856
		English	/a/-/i/	0.969	0.011	0.937	0.985
			/t/-/d/	0.950	0.015	0.909	0.973
L2-English	Control	Spanish	/a/-/i/	0.959	0.013	0.923	0.979
			/t/-/d/	0.959	0.013	0.923	0.979
	Test	English	/i/-/ɪ/	0.716	0.050	0.610	0.803
			/ʃ/-/tʃ/	0.933	0.019	0.886	0.962
		Spanish	/e/-/ej/	0.918	0.022	0.864	0.952
			/d/-/r/	0.903	0.025	0.843	0.942

well on Spanish stimuli (Mean difference: 0.019; $t(5112) = 1.69$, $p = 0.092$), but on the English stimuli the L2-English learners were slightly less accurate (0.956) than the L2-Spanish learners (0.983; $t(5112) = 2.44$, $p = 0.015$). To sum up, none of the control contrasts can be said to have posed perceptual difficulties to learners, all accuracy rates being $M > 0.956$, whereas when performing on L2 test contrasts participants did show perception difficulties (see Table 5).

For RTs, tests of fixed effects revealed a significant main effect of *stimulus language* ($F(1, 4667) = 14.8$, $p < 0.001$), and significant *L2-group* \times *stimulus language* ($F(1, 4667) = 152.4$, $p < 0.001$) and *L2-Group* \times *condition* \times *stimulus language* ($F(1, 4667) = 18.2$, $p < 0.001$) interactions, but neither the overall effect of *L2-group* ($F(1, 4667) = 2.16$, $p = 0.142$) nor that of *condition* ($F(1, 4667) = 0.05$, $p = 0.822$) reached significance. These effects overall partly parallel the accuracy data. When splitting the data set by *condition* we found that although the main effects of *L2-group* ($F(1, 2204) = 2.00$, $p = 0.157$) and *stimulus language* ($F(1, 2204) = 0.129$, $p = 0.719$) did not reach significance in the test condition, their interaction did ($F(1, 2556) = 118.5$, $p < 0.001$). As expected this interaction arose because both the L2-Spanish learners ($t(2204) = -118.4$, $p < 0.001$) and the L2-English learners ($t(2204) = 130.1$, $p < 0.001$) were more efficient (i.e., they obtained faster RTs) when processing L1 than L2 stimuli on the test condition. However, whereas in the control condition the main effect of

L2-group did not reach significance ($F(1, 2463) = 2.29, p = 0.130$), both the main effect of *stimulus language* ($F(1, 2463) = 28.1, p < 0.001$) and the *L2-group* \times *stimulus language* interaction did ($F(1, 2463) = 38.2, p < 0.001$). This is because whereas L2-Spanish learners were equally efficient on English and Spanish control stimuli ($t(2463) = 1.04, p = 0.299$), L2-English learners' RTs were slower on English than on Spanish stimuli ($t(2463) = 7.91, p < 0.001$), in accordance with their slightly lower accuracy on this condition. This might be attributed to the large RT variability of the L2-English group on the control condition.

We next examined accuracy rates for the 8 phonological contrasts separately. A mixed-effects model (binary logistic regression with a binomial distribution) was fitted to the accuracy data (Table 5) with the factors *L2-group* (Spanish, English) *contrast* and their interaction as fixed effects, random intercepts for *subject* and *item*, and a random slope for contrast by subject (see Appendix B-5 for parameter estimates).

Tests of fixed effects revealed a main effect of *L2-group* ($F(1, 5104) = 5.23, p = 0.022$), a main effect of *contrast* ($F(7, 5104) = 10.7, p < 0.001$) and a significant *L2-Group* \times *contrast* interaction ($F(7, 5104) = 9.74, p < 0.001$). Accuracy rates were significantly lower on test than on control contrasts for both groups, mainly due to the four L2 test contrasts in each group, for which accuracy rates were lower ($M = 81.2\%$) than for the control contrasts ($M = 93.5\%$). Bonferroni-adjusted pairwise contrasts revealed that performance of the two learner groups differed significantly for the test contrasts /i/-/ɪ/ ($t(5104) = 5.17, p < 0.001$), /e/-/eɪ/ ($t(5104) = -2.14, p = 0.032$) and /d/-/r/ ($t(5104) = -2.61, p < 0.009$), but not /ʃ/-/tʃ/ where L2-English learners ($M = 92.8$) did not differ from L1-English learners ($M = 95.9$) significantly ($t(5104) = 1.18, p = 0.236$), suggesting that this particular contrast posed no perceptual difficulty for the L2-English learners. The L2-Spanish and L2-English learner groups did not differ significantly from one another on any of the control contrasts (English /a/-/i/: $t(5104) = 1.03, p = 0.304$; Spanish /a/-/i/: $t(5104) = 1.61, p = 0.107$; English /t/-/d/: $t(5104) = 1.96, p = 0.050$; Spanish /t/-/d/: $t(5104) = 0.602, p = 0.547$), although L2-English learners found the English /t/-/d/ contrast substantially harder to discriminate than L1-English learners did ($M = 94.5$ and $M = 98.3$, respectively).

Except for the /ʃ/-/tʃ/ contrast, the results of the ABX task indicate specific perception difficulties with the L2 test contrasts for both learner groups. The amount of variability in L2 learners' ability to discriminate L2 contrasts indicated by the 95 % CIs in Table 5 (L2-Spanish: /e/-/eɪ/ = 0.746–0.894, /d/-/r/ = 0.681–0.856; L2-English: /i/-/ɪ/ = 0.610–0.803) suggests that their performance on the L2 vowel contrasts (/e/-/eɪ/ for L2-Spanish learners and /i/-/ɪ/ for L2-English learners) can be used as a valid index of individual differences in L2 perception. We consequently opted for using the vowel accuracy scores as a measure of performance accuracy in L2 speech perception for the individual differences analyses.

Table 6: Mean accuracy scores by learner group and contrast (standard deviations in parentheses).

Group	Contrasts and measures									
	Vowels								Consonants	
		Duration difference (ms)		Duration ratio (ms)		Euclidean distance (Bark)			Accuracy score (0–8)	
		M	SD	M	SD	M	SD		M	SD
L2-Spanish (n = 19)	/ej/-/e/	15	13	1.17	0.14	1.18	0.55	/d/-/t/	4.08	2.38
L1-Spanish (n = 6)	/ej/-/e/	29	12	1.36	0.16	3.19	0.70	/d/-/t/	7.83	0.41
L2-English (n = 21)	/i:/-/ɪ/	6.5	14	1.09	0.17	0.61	0.36	/tʃ/-/ʃ/	7.12	1.18
L1-English (n = 7)	/i:/-/ɪ/	16	11	1.20	0.15	3.71	0.75	/tʃ/-/ʃ/	8.00	0.00

4.3 L2 production: delayed sentence repetition

In general, the production data show the expected pattern of results, with L2 learners obtaining lower accuracy and greater variability in scores than the L1 speaker controls did (Table 6).

Because we computed individual vowel duration and vowel quality scores per speaker based on four productions of each of the contrasting vowels per language (L1 and L2), and because the vowel contrasts were quantitatively and qualitatively different for L2-English (/i:/-/ɪ/) and L2-Spanish learners (/e/-/ej/), they were not directly comparable. Therefore, we assessed vowel production accuracy separately for each of the two L2 learner groups.

In a first set of analyses, we examined the extent to which the vowel quality and duration differences between the vowels in each contrast in L2 learners’ productions were comparable to those of L1-speaker controls. For L2-Spanish learners, the formant movement (added Euclidean distances between measurement points) and duration measurements of /e/ and /ej/ were entered as dependent measures in an ANOVA with *Vowel* (/e/ vs. /ej/) and *Group* (L2-Spanish, L1-Spanish) as independent variables. For the formant movement measure, these analyses revealed significant main effects of *Vowel* ($F(1,23) = 66.65, p < 0.001, \eta^2 = 0.74$) and *Group* ($F(1,23) = 38.78, p < 0.001, \eta^2 = 0.63$), and a significant *Vowel* \times *Group* interaction ($F(1,23) = 41.50, p < 0.001, \eta^2 = 0.64$). This interaction arose because (according to Bonferroni-adjusted pairwise comparisons) L1-Spanish speakers produced significantly more formant movement than the L2-Spanish

learners did in the diphthong /ej/ ($M_{\text{diff}} = 2.01$ Bark, $SE = 0.274$, $p < 0.001$) but not in the monophthong /e/ ($M_{\text{diff}} = 0.66$ Bark, $SE = 0.161$, $p = 0.686$). For duration, the main effect of *Group* did not reach significance ($F(1,23) = 1.36$, $p = 0.256$, $\eta^2 = 0.06$), but the main effect of *Vowel* did ($F(1,23) = 52.32$, $p < 0.001$, $\eta^2 = 0.67$), as did the *Vowel* \times *Group* interaction ($F(1,23) = 5.27$, $p = 0.031$, $\eta^2 = 0.19$). The interaction arose because although for both L2-Spanish learners and L1-Spanish speakers diphthongs were significantly longer than monophthongs ($p < 0.001$), the difference was much larger for L1-Spanish speakers ($M_{\text{diff}} = 29.82$ ms, $SE = 5.46$) than it was for L2-Spanish learners ($M_{\text{diff}} = 15.45$ ms, $SE = 3.07$), who also produced /e/ significantly shorter than native speakers did ($M_{\text{diff}} = 13.76$ ms, $SE = 5.94$, $p = 0.030$).

For L2-English learners, the Bark-converted frequencies (B1–B0 for height and B2–B1 for fronting distance metrics) and duration of /i:/ and /ɪ/ were entered as dependent measures in an ANOVA with *Vowel* (/i:/ vs. /ɪ/) and *Group* (L2-English, L1-English) as independent variables. For both tongue height (B1–B0) and fronting (B2–B1) these analyses revealed main effects of *Vowel* (B1–B0: $F(1,26) = 147.2$, $p < 0.001$, $\eta^2 = 0.85$; B2–B1: $F(1,26) = 301.8$, $p < 0.001$, $\eta^2 = 0.92$) and *Group* (B1–B0: $F(1,26) = 11.21$, $p = 0.002$, $\eta^2 = 0.30$; B2–B1: $F(1,26) = 26.10$, $p < 0.001$, $\eta^2 = 0.50$), and a significant *Vowel* \times *Group* interaction (B1–B0: $F(1,26) = 67.74$, $p < 0.001$, $\eta^2 = 0.72$; B2–B1: $F(1,26) = 190.4$, $p < 0.001$, $\eta^2 = 0.88$). The interaction arose because – although both L2-English learners and L1-English speakers produced /i:/ with significantly higher tongue position than /ɪ/ (B1–B0: 1.57 vs. 1.86 and 1.63 vs. 3.13, respectively), and both groups produced /i:/ similarly (B1–B0: $M_{\text{diff}} = 0.06$, $SE = 0.237$, $p = 0.802$; B2–B1: $M_{\text{diff}} = 0.33$, $SE = 0.286$, $p = 0.262$) – L2-English learners produced /ɪ/ with a significantly higher (B1–B0: $M_{\text{diff}} = -1.27$, $SE = 0.184$, $p < 0.001$) and more advanced (B2–B1: $M_{\text{diff}} = 2.67$, $SE = 0.217$, $p < 0.001$) tongue position than L1-English speakers did. In addition, L1-English speakers made a much larger qualitative distinction between /i:/ and /ɪ/ (B1–B0: $M_{\text{diff}} = 29.82$, B2–B1: $M_{\text{diff}} = -1.50$) than L2-English learners did (B1–B0: $M_{\text{diff}} = 15.45$, B2–B1: $M_{\text{diff}} = -0.28$). In a nutshell, L2-English learners' production of /ɪ/ was qualitatively not very different from /i:/. In terms of duration, the main effect of *Vowel* reached significance ($F(1,26) = 15.16$, $p = 0.001$, $\eta^2 = 0.37$) because both L2-English learners and L1-English speakers produced /i:/ with longer duration than /ɪ/ (88.4 ms vs. 81.8 ms and 101.6 ms vs. 85.0 ms, respectively), but neither the main effect of *Group* ($F(1,26) = 1.42$, $p = 0.244$, $\eta^2 = 0.05$) nor the *Vowel* \times *Group* interaction ($F(1,26) = 2.85$, $p = 0.104$, $\eta^2 = 0.10$) was significant.

In a second set of analyses, we assessed whether L2 learners could qualitatively and quantitatively distinguish the vowels in the target vowel contrasts (/e/-/ej/ for L2-Spanish learners; /i:/-/ɪ/ for L2-English learners) to the extent that L1 speakers did. For L2-Spanish learners, we computed the difference in amount of tongue movement between /ej/ and /e/ (Euclidean distances in Bark; see Table 6) and submitted it to an independent samples *t*-test, which showed that L1-Spanish speakers produced a significantly larger difference in formant movement between /e/ and /ej/ than L2-Spanish learners did ($t(23) = -6.44$,

$p < 0.001$). We did the same for the duration ratio measure, and found L1-Spanish speakers to produce a significantly larger duration ratio for /e/-ej/ than L2-Spanish learners did ($t(23) = -2.71, p < 0.012$). For L2-English learners, we computed the Euclidean distance between /i:/ and /ɪ/ in Bark (see Table 6). An independent samples *t*-test showed that L1-English speakers produced a significantly larger distinction in quality (i.e. a larger Euclidean distance) between /i:/ and /ɪ/ than L2-English learners did ($t(26) = -14.6, p < 0.001$), but L2 learners did not differ from L1 speakers on the duration ratio measure for /i:/-ɪ/ ($t(26) = -1.54, p = 0.135$).

Finally, as regards the consonant contrasts, L2 Spanish learners obtained an average score of 4.08 ($SD = 2.38$) representing 51 % of target-like productions for the Spanish /d/-t/ contrast, whereas L1 Spanish speakers performed at ceiling (7.83, $SD = 0.4$). For the English /tʃ/-ʃ/ contrast, L2 English learners obtained an average score of 7.12 ($SD = 1.32$) representing an 89 % of target-like productions, whereas L1 English speakers' productions were 100 % accurate, as expected.

4.4 Relationship between attention and L2 perception and production

The main goal of the current study is to explore the relationship between individual differences in attention control and L2 phonological processing for the perception (categorical discrimination) and production (delayed sentence repetition) of difficult L2 phonological contrasts in two groups of L2 learners: Spanish learners of English (/i:/-ɪ/, /tʃ/-ʃ/) and English learners of Spanish (/e/-ej/, /d/-t/). However, Spanish and English learners' performance on the consonant contrasts was not comparable because L2-English learners had no difficulty with the English /tʃ/-ʃ/ contrast in either perception or production. This is likely due to the presence of a [ʃ] variant of the Spanish phoneme /tʃ/ in Andalusian Spanish (Regan 2020) coexisting with standard [tʃ] in the location where data was collected (Seville). Therefore, we gauged L2 phonological processing in perception and production through the vowel data only (/i:/-ɪ/ for L2-English learners and /e/-ej/ for L2-Spanish learners). In perception and in production, these two contrasts revealed substantial performance variation across learners.

In perception we used the ABX discrimination accuracy scores as a measure of L2 learners' ability to perceptually distinguish between /i:/ and /ɪ/ (L2-English learners) or between /e/ and /ej/ (L2-Spanish learners) and the ABX discrimination RT scores as a measure of processing speed of the quality difference between the target vowels. Faster RTs were deemed to reflect a more robust encoding of the target phonological contrast. In production, we used a unified measure of Bark-normalized spectral distances between L2 vowels to estimate the accuracy of the vowel quality

contrast, and we used duration ratios between L2 vowels to estimate the accuracy of the vowel quantity contrast. That is, we assess L2 English learners' ability to qualitatively and quantitatively distinguish /i:/ from /ɪ/ and L2 Spanish learners' ability to qualitatively and quantitatively distinguish /e/ from /ej/ (see Table 7). However, spectral distances and duration ratios between L2-English /i:/ and /ɪ/ and L2-Spanish /e/ and /ej/ may not be directly comparable. In fact, the magnitude of the spectral distance for the L2-English monophthongal vowel contrast (/i:/-ɪ/, $M = 1.18$ Bark) was much larger than that of the L2-Spanish monophthong-diphthong contrast (/e/-ej/, $M = 0.61$ Bark). Therefore, to make production accuracy measures comparable for all L2 learners, we computed individual z-scores of vowel production accuracy (spectral distances and duration ratios) based on the L1 speakers' means and standard deviations of the learners' corresponding L2. The relationship between attention control and phonological processing measures was explored for all L2 learners by using their attention control switch trial and switch cost RT scores (separately by the nasality and L1 phonetics dimensions) and the phonological processing scores of ABX accuracy and speed for perception and normalized spectral distance and duration ratio z-scores for production (see Table 8).

As the receptive vocabulary size of the L2-English learners was significantly larger than that of the L2-Spanish learners and this might be indicative of a between-groups difference in L2 proficiency (Uchihara and Clenton 2020), which might have affected the L2 phonological processing measures, we first examined whether vocabulary size was associated to any of the attention and L2 phonological processing measures. Shapiro-Wilk tests of normality (and visual inspection of histograms and Q-Q plots) indicated that proficiency, attention and L2 phonological processing scores were normally distributed (all $p > 0.05$, $W_s > 0.95$), except for the spectral distance score ($W(40) = 0.943$, $p = 0.045$). We therefore used *Spearman's-rho* for correlational analyses involving this variable, and *Pearson's-r* correlation coefficients for all other variables. Vocabulary size was not associated to either L2 vowel perception (ABX), production (spectral distances), or our attention measures, so we did not include it as a co-variate in the correlations.

Interestingly, ABX accuracy was significantly related to spectral distances ($r_s = 0.421$, $p = 0.007$), indicating an association between L2 learners' ability to distinguish between the contrasting vowels perceptually and their ability to produce a quality distinction between them in production. Although our perception task (categorical ABX discrimination of nonword items) taps into a pre-lexical phonological level of processing and our production task taps into a lexical semantic level of processing (elicitation of L2 words embedded in meaningful sentences), and are therefore not equivalent, we interpret this association to suggest that L2 learners who had developed more robust phonetic representations for the contrasting L2 sounds could also make a larger quality distinction between them in production.

Table 7: Mean scores for L2 learners' proficiency, attention, and phonological processing measures.

Measures	L2-English (n = 21)			L2-Spanish (n = 19)			L2 learners (n = 40)		
	M	SD	95 % CI lower-upper	M	SD	95 % CI lower-upper	M	SD	95 % CI lower-upper
X_Lex	4,257	486	4,036 4,478	3,716	602	342 4,006	4,000	603	3,807 4,193
Switch (L1) (ms)	1,213	212	1,116 1,309	1,141	163	1,063 1,220	1,179	191	1,118 1,240
Switch (N) (ms)	1,161	261	1,043 1,280	1,106	154	1,032 1,180	1,135	216	1,066 1,204
Switch cost (L1)	-56	59	-83 -29	-83	70	-116 -49	-69	65	-89 -48
Switch cost (N)	-55	86	-94 -16	-123	89	-166 -80	-87	93	-117 -58
ABX accuracy (%)	80.4	6	77 83	88.7	10	84 94	84.3	9	81 87
ABX speed (ms)	1,240	231	1,135 1,345	1,097	159	1,020 1,173	1,172	210	1,105 1,239
Spectral distance	-4.1	0.5	-4.3 -3.9	-2.8	0.8	-3.2 2.5	-3.5	0.9	-3.8 -3.2
Duration ratio	-0.7	1.1	-1.2 -0.2	-1.2	0.9	-1.6 -0.7	-1.0	1.0	-1.3 -0.6

Table 8: Correlation coefficients between proficiency, attention and phonological processing scores.

Correlation coefficients (<i>n</i> = 40)	L2 phonological processing							
	Perception				Production			
	ABX accuracy		ABX speed		Spectral distance (z-score)		Duration ratio (z-score)	
	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>	<i>r_s</i>	<i>p</i>	<i>r</i>	<i>p</i>
Proficiency								
X_Lex	−0.208	0.197	0.195	0.227	−0.209	0.195	0.051	0.755
Attention								
Switch (L1)	−0.280	0.080	0.475*	<0.001	−0.053	0.747	−0.069	0.670
Switch (nasality)	−0.135	0.407	0.532*	<0.001	−0.020	0.901	−0.082	0.615
Switch cost (L1)	−0.183	0.258	0.111	0.496	−0.218	0.177	−0.083	0.609
Switch cost (nasality)	−0.026	0.873	−0.033	0.840	−0.334 ⁺	0.035	0.261	0.103

Note: Asterisks (*) indicate p-values that remain significant after adjustment for multiple comparisons using Benjamini & Hochberg’s False Discovery Rate procedure, whereas (†) indicates a p-value that becomes non-significant after correction.

Previous research has found perception to be more closely related to production *within* rather than *across* pre-lexical and lexical processing levels (Melnik-Leroy et al. 2022), but even within a pre-lexical processing level employing equivalent tasks in perception and production, a relationship between the two is not always attested (e.g., Kartushina et al. 2022; see Kartushina et al. 2022; Kato and Baese-Berk 2020; Melnik-Leroy et al. 2022; Nagle and Baese-Berk 2022, for discussion on the relationship between perception and production modalities).

Significant medium-strength correlations were found between L2 learners’ speed in adjusting to a new dimension in the attention switching task and in ABX discrimination speed, suggesting that individual differences in speed of processing underlie performance in both tasks, that is, deciding on the presence of a nasal resonance or L1 phonetics in a context of switching dimensions may require the same underlying processing skills as deciding on the identity of the vowel in an ABX trial where the target vowel could randomly appear in position A or B of the triad. This was corroborated by the fact that RTs on the repeat trials in the attention switching task also correlated significantly with ABX speed (L1: $r = 0.547$, $p < 0.001$; *nasality*: $r = 0.554$, $p < 0.001$). However, none of the attention measures were significantly associated with ABX accuracy, suggesting that attention control did not explain variance in how accurately L2 learners perceived the target vowel contrasts. A significant, though weak correlation, emerged between the switch cost measure (in the nasality condition), a measure of attentional flexibility, and the spectral distance score, indicating a weak tendency for L2 learners with stronger attention control to

be better able to qualitatively distinguish the target L2 vowels in production. Significance tests for each measure were adjusted for multiple comparisons using Benjamini and Hochberg's (1995) False Discovery Rate procedure, at the 0.05 level for 5 simultaneous comparisons (p -values). For ABX speed, both significant correlations remain so after correction (the new significance threshold being 0.01 after FDR correction). For spectral distance, the correction places the p -value of the correlation between switch cost(nasality) and z -score above the significance threshold of 0.01.

5 Discussion and conclusions

We set out to explore the connection between individual differences in attention control (attention switching skill) and L2 phonological processing in perception and production. We conceptualized attention switching in terms of a cognitive skill functioning as a built-in “cue enhancement device” during L2 phonological processing that allows learners to efficiently extract the relevant language-specific segmental phonetic features of L2 sounds while bringing others to the perceptual background. We assessed individual differences in attention control through a novel speech-based task (adapted from the task switching paradigm) that required learners to switch their focus of attention between segmental speech dimensions: nasality (nasal vs. non-nasal) and language-specific phonetics (L1 vs. L2); we also measured learners' phonological processing in perception and production. We expected L2 learners with stronger attention switching skill to have developed more accurate L2 phonological representations during L2 learning (irrespective of learning history or target L2) based on their enhanced ability to attend to and extract the relevant segmental phonetic properties of the L2 sound system. Increased accuracy of L2 sound representation (L2 vowels) was expected to result in increased perceptual discrimination ability for difficult L2 sound contrasts (higher ABX discrimination scores) and increased ability to qualitatively distinguish contrasting sounds in production (larger spectral distance scores between L2 vowels).

Attention switching scores were related to processing speed in the perceptual discrimination task, and to the spectral distance between vowels in production. This suggests attention switching skill plays a role in L2 phonological processing, despite being modest in size. Our findings clearly indicate that those learners who were more efficient (i.e., faster) at focussing their attention on a given speech dimension in the attention task were also faster at discriminating the target vowel contrasts. Contrary to our expectations however, we did not find an association between this attention switching measure and discrimination *accuracy*. The association between response speed in the attention and discrimination tasks might simply be due to the potential consistency of individual differences in working memory or phonological

memory capacity (which we did not test in the current study) across the two tasks, or to the similar speech processing requirements of both tasks (deciding on the quality of a phonetic feature in a context of switching speech dimensions is similar to deciding on the identity of a vowel with respect to contrasting vowels that randomly switch positions in a triad), or both. The lack of relationship between attention switching skill and discrimination accuracy suggests that individual differences in attention switching, even when assessed through a speech-based task, do not predict L2 sound discrimination skills at intermediate proficiency levels. Further research needs to confirm whether this is indeed the case. For example, one study using speech-based attention control tasks found auditory selective attention (but not attention switching skill or auditory inhibition) to be related to L2 learners' *gains* in ABX discrimination accuracy after high-variability phonetic training (Mora and Mora-Plaza 2019). These studies are not directly comparable to ours though, since our task doesn't measure learning gains, but rather the state of phonological knowledge *after* it has been learnt. It is also plausible that the implication of attention skills in L2 phonological development, together with other sources of individual differences in L2 phonological processing, such as inhibition (Darcy et al. 2016) or general auditory processing skills (Saito et al. 2019, 2020, 2021) may contribute to a different extent at different stages of acquisition, playing a larger role at initial stages (or during initial learning) than at the intermediate proficiency level we targeted in the current study.

Interestingly, discrimination accuracy (which was unrelated to attention switching) was significantly related to spectral distances between the same contrasting vowel in production. These spectral distance scores were in turn related to attentional flexibility (albeit weakly), suggesting that attention switching skill may be more directly implicated in speech production than in speech perception. One way to explain these seemingly diverging findings is to consider the nature of the tasks we used to measure L2 perception and production. The perception task used nonwords, which likely enhanced a phonetic processing mode and made the processing of acoustic differences between the contrasting A and B items in the ABX trials easier than if the target vowels had been embedded in confusable lexical minimal-pair words (Ortega et al. 2021; Thomson and Derwing 2016). Because the task doesn't involve meaning and is lower in cognitive complexity, it may not allow differences in attention skill to influence performance because it reflects underlying phonological knowledge which has already been established in the past. While differences in attention may impact the way by which phonological representations are established at the time of learning (e.g. speed of learning or initial precision), they do not interact with the outcome during testing, which measures underlying phonological knowledge (suggesting that tasks such as the ABX are indeed good tasks to measure underlying phonological knowledge independently of attentional abilities). By comparison, the production task made use of L2 lexical words embedded in meaningful sentences that

had to be repeated after a delay and intervening speech material (a prompt) that forced learners to repeat them from memory after having processed their meaning. The meaning-focused nature and the higher cognitive complexity of this task might have allowed individual differences in attention control to play a role, due to this task's increased attentional demands compared to ABX discrimination. As a result, L2 learners with efficient attention skills might have been better able to pay attention to the quality of the target sounds in the words that made up the sentences during the production task. Given the nature of this task and its primary focus on meaning during repetition from memory, it is uncertain whether learners could in fact focus on the phonetic features of the target difficult L2 sounds.

Finally, it is important to acknowledge that our vowel production accuracy measures are mainly based on a measure of distinctiveness, rather than one of “nativelikeness” or how close their vowel production was to that of L1 speakers. Thus, although our L2 learners' ability to produce a larger acoustic distance between two contrastive vowels was interpreted an indication of a larger qualitative distinction between them in production (as in recent research, e.g. Melnik-Leroy et al. 2022) and, by hypothesis, an indication of higher production accuracy, a more “accurate” production does not necessarily imply a more target-like quality in production. The potential difference between a measure of contrastiveness and one based on distance from L1 speakers' productions in indexing L2 learners' development of the target L2 vowel representations was partly overcome by using L1 speakers' means and standard deviations of the learners' corresponding L2 to compute the z-scores of vowel production accuracy. However, it is uncertain to what extent a measure of accuracy computed as the Euclidean distance between L2 learners' productions and those of L1 speakers might have resulted in comparable results as regards the role of attention control in predicting L2 vowel production accuracy.

In this study, we examined the extent to which attention switching skill is associated with L2 phonological processing in perception and production. Future research investigating the role of attention control in L2 phonological acquisition would benefit from exploring additional attentional skills such as auditory selective attention and from doing so with learners at different proficiency levels and within longitudinal research designs, both through lab- and classroom-based studies.

Acknowledgments: We would like to thank Paola Rodrigues, Tanya Flores, Diana Arroyo, Ana Fernandez, Maggie Peters, and Fiona Pannett for help in stimuli recording and Danielle Daidone, Amanda Rabideau, Elena Safronova, Eva Cerviño-Povedano, Marina Barrio Parra and M. Heliadora Cuenca Villarín for help with testing and data processing. We thank two anonymous *Phonetica* reviewers for their very insightful comments and suggestions. Any remaining errors are our own.

Research funding: This study was supported by Grant-in-Aid (Indiana University) and grants FFI2013-47616-P (Spanish Ministry of Economics and Competitiveness), PID2019-107814GB-I00 (MCIN/AEI/10.13039/501100011033) (Spanish Ministry of Science, Innovation and Universities), and grants 2014SGR1089 and 2017SGR560 (Generalitat de Catalunya).

Author contributions statements: The corresponding author Joan C. Mora and co-author Isabelle Darcy are responsible for the conceptualization and design of the study and carried out the participant recruitment, data collection and analysis and writing up of the manuscript.

Statement of ethics: The participants in the study gave their written informed consent to participate in the study. The study protocol adhered to the good practices of data collection, anonymization, processing, and storage and was approved by the Institutional Review Boards of the University of Barcelona (IRB0003099) and Indiana University (IRB00000222; study 1210009699).

Conflict of interest statement: The authors have no conflicts of interest to declare.

References

- Anrrich, Graciela M. 2007. *Substitutions for English consonants by adult speakers of Cuban Spanish. (Doctoral dissertation, Georgetown University)*. Georgetown University.
- Baese-Berk, Melissa Michaud, Tessa Bent, Stephanie Borrie & Megan McKee. 2015. Individual differences in perception of unfamiliar speech. In The Scottish Consortium for ICPHS 2015 (ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*, 1–4. Glasgow, UK: the University of Glasgow. <http://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0460.pdf>.
- Benjamini, Yuav & Yosef Hochberg. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B (Methodological)* 57(1). 289–300.
- Bohn, Ocke-Schwen & James Emil Flege. 1990. Interlingual identification and the role of foreign language experience in L2 vowel perception. *Applied Psycholinguistics* 11. 303–328.
- Bradlow, Ann R., Reiko Akahane-Yamada, David B. Pisoni & Yoh'ichi Tohkura. 1999. Training Japanese listeners to identify English/r/and/l: Long-term retention of learning in perception and production. *Perception & Psychophysics* 61(5). 977–985.
- Bundgaard-Nielsen, Rikke L., Catherine T. Best & Michael D. Tyler. 2011. Vocabulary size matters: The assimilation of second-language Australian English vowels to first-language Japanese vowel categories. *Applied Psycholinguistics* 32(1). 51–67.
- Daidone, Danielle & Isabelle Darcy. 2021. Vocabulary size is a key factor in predicting second language lexical encoding accuracy. *Frontiers in Psychology* 12(2769). 1–16.
- Darcy, Isabelle, Joan C. Mora & Danielle Daidone. 2016. The role of inhibitory control in second language phonological processing. *Language Learning* 66(4). 741–773.
- Darcy, Isabelle, Hanyong Park & Chung-Lin Yang. 2015. Individual differences in L2 acquisition of English phonology: The relation between cognitive abilities and phonological processing. *Learning and Individual Differences* 40. 63–72.

- Derwing, Tracey M. & Murray J. Munro. 2013. The development of L2 oral language skills in two L1 groups: A 7-year study. *Language Learning* 63(2). 163–185.
- Díaz, Miriam & Miquel Simonet. 2015. Second language acquisition of Spanish/e/and/ei/by native English speakers. *Hispania* 98(4). 750–761.
- Ellis, Nick C. 2006. Selective attention and transfer phenomena in L2 acquisition: Contingency, cue competition, salience, interference, overshadowing, blocking, and perceptual learning. *Applied Linguistics* 27(2). 164–194.
- Escudero, Paola & Paul Boersma. 2004. Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition* 26(4). 551–585.
- Flège, James E. 2008. Give input a chance. In Thorsten Piske & Martha Young-Scholten (eds.), *Input matters in SLA*, 175–190. Bristol: Multilingual Matters.
- Forster, Kenneth I. & Jonathan C. Forster. 2003. DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods* 35(1). 116–124.
- French, Leif M. & Irena O'Brien. 2008. Phonological memory and children's second language grammar learning. *Applied Psycholinguistics* 29(3). 463–487.
- Ghaffarvand Mokari, Payam & Stefan Werner. 2019. On the role of cognitive abilities in second language vowel learning. *Language and Speech* 62(2). 260–280.
- Golestani, Narly & Robert J. Zatorre. 2009. Individual differences in the acquisition of second language phonology. *Brain and Language* 109(2–3). 55–67.
- Gottfried, Terry L. 1984. Effects of consonant context on the perception of French vowels. *Journal of Phonetics* 12(2). 91–114.
- Green, David W. 1998. Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition* 1(2). 67–81.
- Guion, Susan G. & Eric Pederson. 2007. Investigating the role of attention in phonetic learning. In Ocke-Schwen Bohn & Murray J. Munro (eds.), *Language experience in second language speech learning: In honor of James Emil Flège*, 57–77. Amsterdam: John Benjamins.
- Heald, Shannon L. M. & Howard C. Nusbaum. 2014. Speech perception as an active cognitive process. *Frontiers in Systems Neuroscience* 8. 1–15.
- Iverson, Paul, Valerie Hazan & Kerry Bannister. 2005. Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r/-/l/ to Japanese adults. *The Journal of the Acoustical Society of America* 118(5). 3267–3278.
- Janse, Esther & Patti Adank. 2012. Predicting foreign-accent adaptation in older adults. *The Quarterly Journal of Experimental Psychology* 65(8). 1563–1585.
- Kartushina, Natalia, David Soto & Clara Martin. 2023. Metacognition in second language speech perception and production. *Language Learning* 73(2). 508–542.
- Kato, Misaki & Melissa Michaud Baese-Berk. 2020. The effect of input prompts on the relationship between perception and production of non-native sounds. *Journal of Phonetics* 79. 100964.
- Kormos, Judit & Anna Sáfár. 2008. Phonological short-term memory, working memory and foreign language performance in intensive language learning. *Bilingualism* 11(2). 261–271.
- Kroll, Judith F., Susan C. Bobb, Maya Misra & Taomei Guo. 2008. Language selection in bilingual speech: Evidence for inhibitory processes. *Acta Psychologica* 128(3). 416–430.
- Lev-Ari, Shiri & Sharon Peperkamp. 2013. Low inhibitory skill leads to non-native perception and production in bilinguals' native language. *Journal of Phonetics* 41(5). 320–331.
- Lev-Ari, Shiri & Sharon Peperkamp. 2014. The influence of inhibitory skill on phonological representations in production and perception. *Journal of Phonetics* 47. 36–46.

- MacKay, Ian R. A., Diane Meador & James Emil Flege. 2001. The identification of English consonants by native speakers of Italian. *Phonetica* 58(1–2). 103–125.
- Mattys, Sven L. & Lukas Wiget. 2011. Effects of cognitive load on speech recognition. *Journal of Memory and Language* 65(2). 145–160.
- Meara, Paul & James Milton. 2003. *X Lex, The Swansea Levels Test*. Newbury, UK: Express Publishing.
- Melnik-Leroy, Gerda Ana, Rory Turnbull & Sharon Peperkamp. 2022. On the relationship between perception and production of L2 sounds: Evidence from Anglophones' processing of the French/*u*–*ʏ*/contrast. *Second Language Research* 38(3). 581–605.
- Miyake, Akira & Naomi P. Friedman. 2012. The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science* 21(1). 8–14.
- Monsell, Stephen. 2003. Task switching. *Trends in Cognitive Sciences* 7(3). 134–140.
- Mora, Joan C. & Ingrid Mora-Plaza. 2019. Contributions of cognitive attention control to L2 speech learning. In Anne Mette Nyvad, Michaela Hejná, Anders Højen, Anna Bothe Jespersen & Mette Hjortshøj Sørensen (eds.), *A sound approach to language matters – In honor of Ocke-Schwen Bohn*, 477–499. Aarhus, Denmark: Aarhus University.
- Morrison, Geoffrey Stewart. 2009. L1-Spanish speakers' acquisition of the English /i/-/ɪ/ contrast II: Perception of vowel inherent spectral change. *Language and Speech* 52(4). 437.
- Moyer, Alene. 1999. Ultimate attainment in L2 phonology. *Studies in Second Language Acquisition* 21(1). 81–108.
- Moyer, Alene. 2014. Exceptional outcomes in L2 phonology: The critical factors of learner engagement and self-regulation. *Applied Linguistics* 35(4). 418–440.
- Muñoz, Carmen. 2014. Contrasting effects of starting age and input on the oral performance of foreign language learners. *Applied Linguistics* 35(4). 463–482.
- Munro, Murray J. & Ocke-Schwen Bohn. 2007. The study of second language speech: A brief overview. In Ocke-Schwen Bohn & Murray J. Munro (eds.), *Language Experience in second language speech learning: In honor of James Emil Flege*, 3–11. Amsterdam: John Benjamins.
- Nagle, Charles L. & Melissa M. Baese-Berk. 2022. Advancing the state of the art in L2 speech perception-production research: Revisiting theoretical assumptions and methodological practices. *Studies in Second Language Acquisition* 44(2). 580–605.
- Oliveira, Diana. 2020. *Auditory selective attention and performance in high variability phonetic training: The perception of Portuguese stops by Chinese L2 learners*. Minho, Portugal: Universidade do Minho Unpublished doctoral dissertation.
- Ortega, Mireia, Ingrid Mora-Plaza & Joan C. Mora. 2021. Differential effects of lexical and non-lexical high-variability phonetic training on the production of L2 vowels. In Anastasia Kirkova-Naskova, Alice Henderson & Jonás Fouz-González (eds.), *English pronunciation instruction: Research-based insights*, 327–355. Amsterdam: John Benjamins.
- Ou, Jinghua & Sam-Po Law. 2017. Cognitive basis of individual differences in speech perception, production and representations: The role of domain general attentional switching. *Attention, Perception, & Psychophysics* 79(3). 945–963.
- Ou, Jinghua, Sam-Po Law & Roxana Fung. 2015. Relationship between individual differences in speech processing and cognitive functions. *Psychonomic Bulletin & Review* 22(6). 1725–1732.
- Petersen, Steven E. & Michael I. Posner. 2012. The attention system of the human brain: 20 years after. *Annual Review of Neuroscience* 35. 73–89.
- Piske, Thorsten, Ian R. A. MacKay & James E. Flege. 2001. Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics* 29(2). 191–215.

- Regan, Brendan. 2020. Intra-regional differences in the social perception of allophonic variation: The evaluation of [tʃ] and [ʃ] in Huelva and Lepe (Western Andalucía). *Journal of Linguistic Geography* 8(2). 82–101.
- Reilly, Jamie, Vanessa Troiani, Murray Grossman & Rthur Wingfield. 2007. An introduction to hearing loss and screening procedures for behavioral research. *Behavior Research Methods* 39(3). 667–672.
- Robinson, Peter. 1995. Attention, memory, and the “noticing” hypothesis. *Language Learning* 45(2). 283–331.
- Rose, Marda. 2010. Differences in discriminating L2 consonants: A comparison of Spanish taps and trills. In Matthew T. Prior, Yukiko Watanabe & Sang-Ki Lee (eds.), *Selected proceedings of the 2008 Second Language Research Forum: Exploring SLA perspectives, positions, and practices*, 181–196. Somerville, MA: Cascadilla Proceedings Project.
- Safronova, Elena. 2016. *The role of cognitive ability in the acquisition of second language perceptual competence*. Barcelona: Universitat de Barcelona Unpublished doctoral dissertation.
- Saito, Kazuya, Hui Sun & Adam Tierney. 2019. Explicit and implicit aptitude effects on second language speech learning: Scrutinizing segmental, prosodic and temporal sensitivity and performance via behavioral and neurophysiological measures. *Bilingualism: Language and Cognition* 22(5). 1123–1140.
- Saito, Kazuya, Hui Sun & Adam Tierney. 2020. Domain-general auditory processing as a perceptual-cognitive anchor of L2 pronunciation learning in adulthood: A longitudinal study. *Applied Psycholinguistics* 41. 1083–1123.
- Saito, Kazuya, Yui Suzukida, Mai Tran & Adam Tierney. 2021. Domain-general auditory processing partially explains L2 speech learning in classroom settings: A review and generalization study. *Language Learning* 71(3). 669–715.
- Segalowitz, Norman. 2010. *Cognitive bases of second language fluency*. New York: Routledge.
- Segalowitz, Norman & Sarah Frenkiel-Fishman. 2005. Attention control and ability level in a complex cognitive skill: Attention-shifting and second language proficiency. *Memory and Cognition* 33(4). 644–653.
- Sikora, Katarzyna, Ardi Roelofs, Daan Hermans & Harry Knoors. 2016. Executive control in spoken noun-phrase production: Contributions of updating, inhibiting, and shifting. *The Quarterly Journal of Experimental Psychology* 69(9). 1719–1740.
- Speciale, Giovanna, Nick C. Ellis & Tracey Bywater. 2004. Phonological sequence learning and short-term store capacity determine second language vocabulary acquisition. *Applied Psycholinguistics* 25(2). 293–321.
- Taube-Schiff, Marlene & Norman Segalowitz. 2005. Linguistic attention control: Attention shifting governed by grammaticized elements of language. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 31(3). 508–519.
- Thomson, Ron I. & Tracey M. Derwing. 2016. Is phonemic training using nonsense or real words more effective? In John Levis, Huong Le, Ivana Lucic, Evan Simpson & Sonca Vo (eds.), *Proceedings of the 7th pronunciation in second language learning and teaching conference*, 88–97. Ames, IA: Iowa State University.
- Trofimovich, Pavel & Wendy Baker. 2006. Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition* 28(1). 1–30.
- Tyler, Michael D. 2019. PAM-L2 and phonological category assimilation in the foreign language classroom. In Anne Mette Nyvad, Michaela Hejrná, Anders Højen, Anna Bothe Jespersen & Mette Hjortshøj Sørensen (eds.), *A sound approach to language matters – In honor of Ocke-Schwen Bohn*, 607–630. Aarhus, Denmark: Aarhus University.

- Uchihara, Takumi & Jon Clenton. 2020. Investigating the role of vocabulary size in second language speaking ability. *Language Teaching Research* 24(4). 540–556.
- VanPatten, Bill (ed.). 2004. *Processing instruction: Theory, research, and commentary*. Mahwah, NJ: Erlbaum.
- Wager, Tor D., John Jonides & Edward E. Smith. 2006. Individual differences in multiple types of shifting attention. *Memory & Cognition* 34(8). 1730–1743.

Supplementary Material: This article contains supplementary material (<https://doi.org/10.1515/phon-2022-0020>).