Contributions of Cognitive Attention Control to L2 Speech Learning

Joan C. Mora and Ingrid Mora-Plaza
Universitat de Barcelona

Abstract
This study examined the effect of cognitive attention control on L2 phonological development from an individual differences perspective. L1-Catalan/Spanish learners of L2-English were trained on the perception and production of English /æ/-/ʌ/ and /iː/-/ɪ/ through AX discrimination, identification immediate repetition tasks. Learners' gains in L2 phonological development were assessed through L2 perception (ABX discrimination and lexical decision) tests. Additionally, we obtained individual measures of auditory selective attention, auditory attention switching and auditory inhibition and a measure of overall L2 proficiency. Results revealed robust gains in L2 perception for both target contrasts. Auditory selective attention scores were significantly related to learners' gains in /æ/-/ʌ/ perception, and attention switching skills to performance in the ABX discrimination tests. Overall the results highlight the role of cognitive attention control in L2 speech learning.

1. Introduction

Speaking in a second language (L2) requires listeners and speakers to efficiently switch their focus of attention between competing linguistic cues as required by the context of communicative interaction. Exercising successful control of attention in L2 use is therefore essential for communication, but the human attentional system is of limited capacity (Petersen & Posner, 2012) and individuals vary in how efficiently they can shift, focus and maintain their attention when using a L2 (Wager, Jonides, &

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*** (pp. 477-499). Dept. of English, School of Communication & Culture, Aarhus University.

Smith, 2006). In addition, whereas the use of attention control in language processing is highly automatized in one's first language, it appears to be relatively effortful and inefficient in a L2 (Segalowitz, 2010), especially at lower levels of proficiency. This suggests that individual differences in cognitive attention control, as well as in other cognitive components of executive functioning (e.g. inhibitory control and working memory), may have a substantial impact on second language acquisition and may therefore constitute relevant sources of variability that can help explain the large inter-learner variability commonly associated with L2 phonological development.

Previous research has shown that inter-learner differences in attentional capacity may have an overall impact on L2 learning either enhancing or impairing lexical, grammatical or phonological development (Segalowitz & Frenkiel-Fishman, 2005). However, cognitive attention control may not impact all linguistic domains to the same extent. For example, it may impact efficiency of L2 processing by regulating the shift of focus between form and meaning when processing utterances, or by selectively attending to various linguistic dimensions (phonology, morphology, grammar or semantics), or by focusing attention on a specific relevant feature within a linguistic dimension (e.g. duration differences in phonological encoding). In addition, languages also differ in what linguistic features speakers' attentional resources need to be allocated to and in how attention is variably allocated to these linguistic features. In L2 phonological processing, inefficient use of this attentional skill may cause perceptual difficulties for adult L2 learners who may fail to apply some L2-specific cue-weighting when phonologically encoding L2 sounds (Bohn, 1995; Flege, 1995).

Attention control has also been shown to be related to general mechanisms involved in the perception and production of speech by guiding auditory processes during speech perception. It allows listeners to focus their processing resources on the relevant acoustic information and to select the acoustic information that is critical for appropriately interpreting the auditory input during oral communication (Akeroyd, 2008; Astheimer, Berkes, & Bialystok, 2016; Baese-Berk et al., 2015). The use of such attention control mechanisms, which requires both attention-switching skill and the ability to selectively attend to a single dimension or feature during speech processing (Astheimer & Sanders, 2009; Bialystok, Craik, & Luk, 2012), facilitates perceptual learning and L2 learners' skill in processing L2 phonological contrasts (Ou, Law & Fung, 2015).

In the domain of L2 phonological acquisition, better attention switching skills have been associated to enhanced performance in L2 phonological processing tasks (Darcy, Mora & Daidone, 2014; Mora & Darcy, 2016). Learners with better attention control may thus be better able to make use of the phonological features embedded in L2 speech input to guide their perceptual learning process. Few studies to date have examined the relationship between cognitive attention control and L2 perception gains obtained through phonetic training, with apparently mixed findings. For example, Ghaffarvand Mokari and Werner (2018) recently examined the role of attention control (measured through the Stroop task) in vowel learning after two weeks of perceptual training on English vowels through discrimination and identification tasks and found no significant associations between attention scores and perceptual learning. However, Hazan and Kim (2010), investigated predictors of phonetic training benefits in the context of phonetic training and found that attention switching, as measured through one of the components of the Test of Everyday Attention (TEA) (Robertson, Ward, Ridgeway, & Nimmo-Smith, 1994), correlated significantly with gains in word identification. Altogether these findings suggest that the impact of cognitive attention control in L2 speech learning is still not well understood. In addition, the role of the various components of attention control (attention switching, selective attention and inhibition) in L2 phonological acquisition are largely under-researched, especially as measured in the auditory domain. In the present study we measure L2 learners' efficiency in the use of their attentional resources through attention control tasks that require participants to recruit their attentional resources in the processing of L1 and L2 speech.

The goal of the present study is to explore the relationship between cognitive attention control and L2 phonological development. We examined individual differences in three subcomponents of cognitive attention control in the auditory domain (selective attention, attention switching and inhibition) and related these scores to L1-Spanish learners' performance and gains in perceptual sensitivity to two difficult L2-English vowel contrasts (/æ/-/ʌ/ and /iː/-/ɪ/) on which they had been trained through discrimination and identification tasks.

2. Methods

L1-Spanish English learners were tested on their ability to accurately perceive and produce two difficult L2 vowel contrasts (/æ/-/ʌ/ and / iː/-/ɪ/) before and after four 45-minute phonetic training sessions. We assessed L2 perception through ABX discrimination and lexical decision tasks. L2 production was assessed through delayed repetition tasks. The L2 phonetic training consisted of AX discrimination and identification tasks (perception) and two immediate repetition tasks (production). All the tasks were administered in *DmDx* (Forster & Forster, 2003) on laptop computers using noise-cancelling headphones. We used three auditory attention control tasks involving the learners' L1 and L2 to assess individual differences in their attentional skills: an auditory selective attention task, an auditory attention switching task, and an auditory inhibition task. In addition, we obtained a measure of overall L2 proficiency through an elicited imitation task. In the present chapter we report on the learners' perceptual performance only.

2.1 Participants

The participants in the study (*N*=17, 14 female) were Catalan-Spanish bilingual undergraduate learners of English who participated in this research for course credit. They had learnt English mainly through formal instruction at school and had limited weekly exposure to English (Table 1). They could all speak Catalan and Spanish but varied in degree of dominance (6 Catalan-dominant, 4 Spanish-dominant,7 balanced), which was not expected to affect their perception and production of the target vowel contrasts, as both /æ/-/ʌ/ and /iː/-/ɪ/ are mapped onto the same L1 Spanish and Catalan vowels (/a/ and /i/, respectively).They reported having no speech or hearing pathologies.

Measure	*M*	*SD*
Age at testing (years)	22.06	9.33
Age of onset of L2 learning (years)	7.35	5.02
L2 instruction (years)	14.53	2.66
Spoken L2 input / output (hours per week)[1]	22.61 / 11.14	11.29 / 6.87
Self-estimated proficiency (1=very poor-9=native-like)[2]	6.32	1.11

[1] L2 use with native and non-native speakers in hours per week.

[2] Averaged self-estimated ability to speak spontaneously, understand, read, write and pronounce English.

Table 1. Participants' demographics

2.2 Materials

The materials used in the training and testing consisted of 32 nonwords and 16 words for each one of the 4 vowels in the target vowel contrasts (/æ/, /ʌ/, /iː/, /ɪ/). They were elicited in carrier phrases (*I say X, I say X again*) read by 3 female (F1, F2, F3) and 3 male (M1, M2, M3) native speakers of Southern British English. Carrier phrases were digitally recorded in a soundproof booth and the best of the two target items in each carrier phrase was excised and normalized for amplitude. Half of the nonwords (16) were 3-syllable nonwords containing the target vowels in stressed position (*fadattick* /fə'dætɪk/, *faduttick* /fə'dʌtɪk/, *fadeetick* /fə'diːtɪk/, *fadittick* /fə'dɪtɪk/) and half of them were 1-syllable nonwords consisting of the same stressed syllable (*datt* /dæt/, *dutt* /dʌt/, *deet* /diːt/, *ditt* /dɪt/). The consonants preceding and following the stressed vowel in the target CVC syllables varied in place of articulation and voicing, and so did the consonants in the initial and final unstressed syllables, which always had a weak unstressed vowel (either /ə/ or /ɪ/; e.g. C/ə/-CVC-/ɪ/C). Half of the words were 1-syllable common English words (*cap, cup, feet, fit*) and half were common 2-syllable words (*ankle, uncle, feeling, filling*). Stimuli from 4 of the 6 speakers (F1, F2, M1, M2) were used in the training, whereas the remaining two voices (F3, M3) were used in the testing only. During the training participants were exposed to 1- and 3-syllable nonwords only, whereas the testing included both nonwords and words. This was done to test whether the phonetic training based on nonwords (and therefore void of lexical meaning) was effective in improving the perception of the target vowel contrasts in known lexical items.

2.3 Phonetic training

The phonetic training sessions consisted of 4 45-minute training sessions, 2 sessions per week with a day in between. The order of the training tasks was consistent across all 4 sessions: AX discrimination, identification, and immediate repetition. Participants were trained on the production and perception of the two target vowel contrasts separately in two blocks within every session. The blocks were counterbalanced across participants. The stimuli were distributed across the training sessions in order of increasing linguistic complexity, so that participants were exposed to 1-syllable nonwords only in sessions 1 and 2 and to 3-syllable nonwords only in sessions 3 and 4. However, participants were exposed to all 4 speakers (F1, F2, M1, M2) within a single session in every training task to ensure exposure to speaker variability. The order in which participants were trained on the two target contrasts (/æ/-/ʌ/ and /iː/-/ɪ/) was counterbalanced within participants across sessions. The same nonwords by the same speakers were used for perception and production training.

2.3.1 AX Discrimination

In the AX discrimination tasks participants were exposed to a total of 1152 minimal pair trials, 576 trials per vowel contrast, that is, 144 trials per contrast in each session. All AX trials consisted of two nonwords produced by two different voices presented with a 500ms inter-stimulus interval (ISI). Participants were exposed to the same number of *same* (AA, BB) and *different* (AB, BA) trials, where the 4 speakers' voices appeared equally frequently in all positions producing the two members of each vowel contrast in all possible orders. Out of the 144 trials per contrast in one session (6 minimal pair nonword pairs x 4 trial orders =24 trials), half of the nonword pairs, 72 trials (6 voice combinations x 3 minimal pair nonword pairs x 4 trial orders) corresponded to trials with an initial female voice combination (F1-F2, F1-M1, F1-M2, F2-F1, F2-M1, F2-M2), whereas 72 trials corresponded to initial male voice combinations (M1-M2, M1-F1, M1-F2, M2-M1, M2-F1, M2-F2). The nonword pairs participants were exposed to were different in each training session so that they were exposed to a total of 12 different 1-syllable nonword pairs in sessions 1 and 2 and a total of 12 different 3-syllable nonword pairs in sessions 3 and 4.

During training participants were instructed to decide, as fast and as accurately as they could, whether the two (non)words they heard were the *same* (i.e. contained the same stressed English vowel) or *different* by pressing a designated labelled key on the computer keyboard. These instructions were provided in English orally by the researchers and in written form on the computer screen. Participants performed 6 practice trials to ensure they understood the task, after which they could ask questions if they had doubts. The 144 trials were presented in fully randomized order. After 72 trials participants could take a break if they wished. Participants received visual feedback for error ("**Correct!**" or "**Wrong!**") and response latency (the RT in milliseconds: "**1056**").

2.3.2 Identification

The identification tasks were performed immediately after the discrimination tasks in each one of the training sessions. They were constructed to provide identification training on the same nonwords participants had previously been trained on through discrimination. A total of 192 identification training trials were included in the training, 48 per session (6 nonword pairs, i.e. 12 trials x 4 voices = 48 trials). Participants performed 4 practice trials, after which they were asked to identify 48 nonwords presented randomly. Each nonword was presented auditorily only as two pictures containing labels for each one of the target vowels in the contrast appeared simultaneously on the left and right side of the screen. The labels included orthographic, phonetic and visual semantic representations (standardized line drawings) of the words *cap, cup, feet, fit*. Participants selected the label corresponding to their selected response option by pressing a designated labelled key on the computer keyboard and received the same type of feedback they had received during the discrimination training.

2.4 Pre- and post-tests

2.4.1 ABX Discrimination

In order to assess learners' L2 perception, a speeded categorical ABX discrimination test was administered. Trials were created by combining (non)words into ABX triads with a 500ms ISI (e.g. A=*fadattick*-B=*faduttick*-X=*fadattick*). Participants were instructed to decide, as accurately and as fast as they could, whether the last (non)word in the triad contained the same stressed vowel as the first (A) or the second word (B) by selecting a key labelled as A or B on the computer keyboard. A, B and X were always produced by a different speaker. A and B were always produced by the speakers who had provided the stimuli for the training (F1, F2, M1, M2). These speakers' voices appeared the same number of times in all A and B positions. The last word in the trial was always produced by two speakers participants had not been exposed to during the training (F3 or M3). Different voices were used within each trial to ensure that participants made a decision based on the phonological categorization of the stimuli while disregarding indexical phonetic variability between nonwords coming from the speakers' voices.

For each contrast participants were presented with 64 experimental trials testing the target contrasts (/æ/-/ʌ/ and /iː/-/ɪ/) and 16 control trials testing vowel contrasts that were not expected to pose perceptual difficulty to learners (/æ/-/iː/ and /uː/-/ɒ/). The 64 experimental trials per contrast consisted of 8 3-syllable nonword pairs and 8 real word pairs (4 1-syllable and 4 2-syllable words) presented in all 4 possible orders (ABB, ABA, BAA, BAB). Half of the nonwords were trained in the discrimination and identification tasks and half were untrained. None of the real English words had previously appeared in the training. Untrained nonwords were included to test for generalization to new nonword items. Untrained real words were included to test whether phonetic training based on non-lexical items (nonwords) was effective in modifying sensitivity to the same target vowel contrasts in a lexical context and consequently had the power of modifying already existing phono-lexical representations where the target vowel contrasts might have been previously misrepresented or not properly encoded phonologically.

Before doing the test participants performed 8 practice trials during which they received visual feedback for error and response latency, as explained above. The 80 test trials were presented in fully randomized order. If a participant made no response within 2500 milliseconds, the next trial was initiated. The response latencies in milliseconds measured from the onset of the third nonword in the triad were used as a measure of speed. Both the accuracy and speed measures were meant to reflect perceptual sensitivity to the contrasts being tested.

2.4.2 Lexical decision

A lexical decision test (Darcy, 2018) was used to obtain a measure of L1-Catalan learners' perceptual sensitivity to the L2-English contrasts /æ/-/ʌ/ and /iː/-/ɪ/ in a lexical context, reflecting the extent to which L2 learners had accurately encoded these phonological contrasts lexically. Participants were asked to decide, as accurately and as fast as possible, whether a sequence of sounds presented auditorily (as spoken by a male native speaker of English) constituted an English word or not. Control trials were distractor sound sequences consisting of 34 monosyllabic and disyllabic English words (*cake, jumping*) and 34 English nonwords (*peef, sagreem*). Test trials consisted of 28 English words containing the target test vowels (*map, sun, clean, gift*) and 28 English nonwords created by substituting the target test vowels by their contrasting counterparts (*mup, san, clin, geeft*). Native-like sensitivity to the /æ/-/ʌ/ and /iː/-/ɪ/ contrasts would therefore be reflected in correctly identifying both test words and nonwords. We calculated, for every participant and testing time (pre-test and post-test), average accuracy rate and RT scores per contrast (including both words and nonwords) separately for test and control trials, as well as two individual measures of perceptual sensitivity to the contrast based on accuracy rates for words and nonwords. The first one is a *d-prime* (*d'*) score, $d'=(z(H))-z(FA)$, where H (hit rate) is the proportion of test words correctly identified as words and FA (false alarm rate) is the proportion of test nonwords incorrectly identified as words. The second is an adjusted accuracy measure called *delta* (δ), which we computed as the average difference between performance on control trials and test trials (test-control accuracy rates).

2.5 Cognitive attention control tasks

2.5.1 Auditory selective attention

The English learners performed auditory selective attention tasks in their L1 (Catalan) and in their L2 (English) based on single-talker competition (Humes, Lee, & Coughlin, 2006). Each task consisted of 64 trials of pairs of sentences (target vs. competitor). The two sentences in a pair were always different, one spoken by a male voice and one by a female voice, and presented simultaneously (e.g. male: *Ready CHARLIE go to BLUE SIX now*; female: *Ready TIGER go to RED EIGHT now*). All of the Catalan and English sentences were normalized for duration to 1700ms. In every trial, a call signal (e.g. TIGER) appearing on the screen previous to the auditory presentation of the sentence, cued the voice participants had to attend to for correctly identifying 1 of 4 colours and 1 of 8 digits visually presented on the screen. Individual ASA scores were computed by adding up all correctly identified colours and digits in each one of the two tasks up to a maximum score of 128.

2.5.2 Auditory attention switching

A measure of L1 attention switching skill (RT and accuracy switching costs) was obtained through a task that required participants to attend to either the duration (quantity) or the voice (quality) of L1 (Catalan) vowels presented in isolation (Safronova, 2016; Safronova & Mora, 2013). This task was designed as an auditory version of Segalowitz & Frenkiel-Fishman's (2005) linguistic version of the task-switching paradigm (Monsell, 2003) and aimed at providing a measure of attentional flexibility for speech dimensions. Participants were required to shift focus of attention from segmental duration (long vs. short) to voice quality (female vs. male) in the perception of vowel sounds. Several tokens of the Catalan vowels /i e ɛ a ɔ o u/ produced by a male and a female speaker on a falling pitch were manipulated using the PSOLA algorithm in Praat (Boersma & Weenink, 2009) to create long (500ms) and short (200ms) versions of the 7 vowels (7 x 2 x 2 = 28 stimuli). Eight identical copies of each stimulus (28 x 8 = 224 trials) were randomly presented to participants over headphones for categorization (*long, short, female, male*) after three separate practice blocks (long vs. short duration; female vs. male voice; duration + voice in alternating runs). Participants used designated labelled keyboard keys to categorize a vowel sound as *long* or *short* when a speaker icon appeared in any of the two top boxes of a framework of 4 square boxes and to label a vowel sound as *female* or *male* when a speaker icon appeared in any of the two bottom boxes of the framework. Speaker icons appeared predictably in clockwise fashion around the framework at the onset of the auditory stimuli. Trials alternated predictably between duration (D) and voice quality dimensions (V) creating a sequence of *repeat* (same dimension as preceding stimulus) and *switch* (different dimension from preceding trial) trials. Participants were expected to obtain lower accuracy and speed scores on switch than on repeat trials due to the cost associated with having to refocus attention on a different acoustic dimension. The switching cost (the difference between switch and repeat RTs) was used as a measure of attention control, so that the smaller the switching costs, the stronger the attention control.

2.5.3 Auditory inhibition

An auditory inhibition task based on Filippi, Leech, Thomas, Green, & Dick (2012) and Filippi, Karaminis, & Thomas (2014) was used to obtain a measure of auditory inhibition. Participants were presented with 72 pairs of sentences binaurally over headphones, one was always produced by a male voice (e.g. *the dog is chasing the cat*) and the other one by a female voice (e.g. *the dog is chased by the cat*). The sentences, which could be in English or in Catalan, were produced by 4 speakers, 1 male and 1 female native speaker of each language. They were recorded in a sound-proof booth and normalized for amplitude and duration (2000ms). The 72 trials were presented in two 36-trial blocks. In block 1, participants were instructed to attend to the female voice only, and in block 2 to attend to the male voice. Blocks 1 and 2 were counterbalanced across participants. The participants' task was to decide which of two animals in the sentence that was being attended to (*bird, bull, cat, cow, dog, frog, goat, horse, parrot, seal, snake, wolf*) did the action (*bite, chase, eat, grab, scare, scratch*) by selecting one of two response keys corresponding to one of the two animal pictures appearing on the screen. Twenty-four trials consisted of pairs of L1-L1 (12) or L2-L2 (12) sentences whereas 48 trials consisted of L1-L2 sentence pairs. In the L1-L2 trials the voice that had to be attended to (*target*) was always a voice speaking in English, so that participants were forced to inhibit the sentence in their L1 (competitor) in order to correctly identify the animal doing the action (e.g. target: *the dog is chased by the cat* vs. competitor: *el gos persegueix el gat*). Half of the target English sentences in the L1-L2 trials were produced by a male voice and half by a female voice, half were in the active and half in the passive, and half of the correct responses corresponded to animals appearing on the right side of the screen and half on the left. Trials where the L1 had to be attended to and the L2 inhibited were not included in order to keep the task short. We obtained two measures of auditory inhibition accuracy and RT from this task, overall *general* measures based on all 72 trials in the test, and measures based on L1-L2 trials only (those where the L1 had to be inhibited) to measure L1 inhibition.

2.6 L2 proficiency

Overall L2 proficiency was assessed through an elicited imitation task and a receptive vocabulary size test. The elicited imitation task was originally designed by Ortega, Iwashita, Rabie and Norris (2002) for a cross linguistic study on syntactic complexity measures. It includes 30 test sentences ranging from 7-17 syllables constructed to include high frequency vocabulary items, a range of syntactic complexity, and typical grammatical features known to challenge instructed learners. The sentences were produced by a female native speaker of English and were presented auditorily only over headphones for delayed repetition. Participants were instructed to repeat each sentence as accurately as they could (and as much of the sentence as they could) after a 250ms *beep* signal, which occurred 2000ms after the sentence end. Participants had 6.8 seconds to repeat the sentence after the *beep*. The learners' productions were recorded onto a digital recorder and assessed for accuracy following Ortega *et al*'s (2002) rubric, where each sentence received a score from 0 to 4 as a function of how much of it was repeated and the type of inaccuracies and missing unrepeated material. Individual scores could therefore range 0-120 points.

3. Results

We first present the results of the pre- and post-tests for perception (ABX discrimination and Lexical Decision) and then those of the cognitive attention control tasks. When response latencies (RTs) are reported, these correspond to RTs screened for accuracy (only including correct responses) and extreme values (2.5 standard deviations below or above each subject's mean).

3.1 Perceptual learning

The results of the ABX discrimination tests showed robust improvement from pre-test to post-test for the two test vowel contrasts, both in response accuracy and speed (see Table 2 for overall results and Table 3 for results by word type). A series of ANOVAs with *Trial Type* (Test, Control) and *Testing Time* (T1=pre-test, T2=post-test) as within-subjects factors revealed, for accuracy, significant main effects of *Trial Type* (/æ/-/ʌ/: $F(1, 16)=298.14$, $p<.001$, $\eta^2=.949$; /iː/-/ɪ/: $F(1, 16)=90.93$, $p<.001$, $\eta^2=.850$) and *Testing Time* (/æ/-/ʌ/: $F(1, 16)=4.56$, $p=.048$, $\eta^2=.222$; /iː/-/ɪ/: $F(1, 16)=11.89$, $p=.003$, $\eta^2=.426$) for both vowel contrasts, suggesting that control contrasts, as expected, were significantly easier to discriminate than test contrasts, and that correct discrimination rates improved from pre- to post-test. The *Trial Type* x *Testing Time* interaction, however, was significant (/æ/-/ʌ/: $F(1, 16)=5.79$, $p=.028$, $\eta^2=.266$; /iː/-/ɪ/: $F(1, 16)=9.46$, $p=.007$, $\eta^2=.372$), as gains from pre- to post-test did not reach significance for control trials (/æ/-/ʌ/: $t(16)=-2.10$, $p=.837$; /iː/-/ɪ/: $t(16)=-2.10$, $p=.837$). A similar pattern of results was obtained for response speed, with significant main effects of *Trial Type* (/æ/-/ʌ/: $F(1, 16)=298.14$, $p<.001$, $\eta^2=.949$; /iː/-/ɪ/: $F(1, 16)=90.93$, $p<.001$, $\eta^2=.850$) and *Testing Time* (/æ/-/ʌ/: $F(1, 16)=4.56$, $p=.048$, $\eta^2=.222$; /iː/-/ɪ/: $F(1, 16)=11.89$, $p=.003$, $\eta^2=.426$) for both vowel contrasts, suggesting that control contrasts, as expected, could be discriminated faster than test contrasts, and participants were significantly faster at doing so at post-test than at pre-test. Again, a significant *Trial Type* x *Testing Time* interaction arose, as gains in speed were much smaller for control than for test items.

			Accuracy				*RT*			
Trial Type	*Contrast*	Test	*M*	*SD*	*Min.*	*Max.*	*M*	*SD*	*Min.*	*Max.*
Test	/æ/-/ʌ/	T1	.644	.066	.53	.75	966	153	657	1160
		T2	.730	.077	.56	.91	799	172	597	1284
	/iː/-/ɪ/	T1	.605	.112	.42	.81	1002	159	724	1252
		T2	.732	.114	.55	.95	832	173	635	1278
Control	/æ/-/ʌ/	T1	.915	.112	.56	1.00	899	137	650	1114
		T2	.922	.084	.69	1.00	780	172	529	1262
	/iː/-/ɪ/	T1	.911	.143	.44	1.00	885	162	589	1109
		T2	.963	.058	.81	1.00	729	179	540	1228

Table 2. Mean accuracy (proportion of correct responses) and response latencies (RT) in the ABX discrimination test at pre-test (T1) and post-test (T2) by trial type and vowel contrast.

In order to assess whether these general learning outcomes were generalizable to untrained test nonwords and words, we examined trainees' performance at pre-test and post-test for untrained test nonwords and words. As shown in Table 3 below, gains in accuracy and speed were consistent across all item types. We submitted the accuracy and RT scores to a series of ANOVAs with *Testing Time* (T1=pre-test, T2=post-test) and *Word Type* (nonword, word) as within-subjects factors. These analyses revealed significant main effects of *Testing Time* (/æ/-/ʌ/: $F(1, 16)=43.72$, $p<.001$, $\eta^2=.732$; /iː/-/ɪ/: $F(1, 16)=29.31$, $p<.001$, $\eta^2=.647$) and *Word Type* (/æ/-/ʌ/: $F(1, 16)=15.21$, $p=.001$, $\eta^2=.487$; /iː/-/ɪ/: $F(1, 16)=11.89$, $p=.003$, $\eta^2=.426$) on accuracy, and significant main effects of *Testing Time* (/æ/-/ʌ/: $F(1, 16)=175.52$, $p<.001$, $\eta^2=.911$; /iː/-/ɪ/: $F(1, 16)=15.73$, $p=.001$, $\eta^2=.496$) and *Word Type* (/æ/-/ʌ/: $F(1, 16)=15.73$, $p=.001$, $\eta^2=.496$; /iː/-/ɪ/: $F(1, 16)=23.57$, $p<.001$, $\eta^2=.596$) on speed. None of the interactions reached significance. This showed that participants were more accurate and faster at discriminating the target vowel contrasts in untrained real English words than in untrained nonwords and that they improved significantly from pre-test to post-test both in discrimination accuracy and speed, confirming the effectiveness of the treatment.

			Accuracy				*RT*			
Word Type	*Contrast*	*Test*	*M*	*SD*	*Min.*	*Max.*	*M*	*SD*	*Min.*	*Max.*
Trained nonwords	/æ/-/ʌ/	T1	.591	.130	.25	.81	1094	189	753	1363
		T2	.720	.103	.50	.88	861	154	669	1294
	/iː/-/ɪ/	T1	.536	.166	.19	.81	1120	185	798	1431
		T2	.702	.176	.38	.94	922	174	707	1287
Untrained nonwords	/æ/-/ʌ/	T1	.518	.078	.38	.69	1060	176	657	1338
		T2	.577	.140	.25	.81	897	210	663	1414
	/iː/-/ɪ/	T1	.562	.134	.25	.81	1079	185	775	1370
		T2	.683	.153	.44	1.00	922	206	689	1459
Untrained words	/æ/-/ʌ/	T1	.733	.107	.53	.88	884	140	605	1071
		T2	.812	.115	.63	.97	736	175	528	1211
	/iː/-/ɪ/	T1	.661	.113	.47	.88	930	146	665	1119
		T2	.772	.096	.63	.94	753	169	557	1214

Table 3. Mean accuracy (proportion of correct responses) and response latencies (RT) in the ABX discrimination test at pre-test (T1) and post-test (T2) by word type and vowel contrast.

Given the consistency of the overall improvement in discrimination accuracy and speed of the target contrasts for both words and nonwords we computed accuracy and speed gain scores based on all test trials in the ABX discrimination task (Table 4) to be able to relate individual differences in attention control to individual gains in discrimination accuracy and speed.

	Accuracy				*RT*			
Contrast	*M*	*SD*	*Min.*	*Max.*	*M*	*SD*	*Min.*	*Max.*
/æ/-/ʌ/	.086	.067	-.05	.17	-167	143	-369	170
/iː/-/ɪ/	.126	.103	-.08	.36	-170	116	-391	25

Table 4. Mean accuracy (proportion of correct responses) and response latency (RT) gains in the ABX discrimination test by vowel contrast.

L2 learners' performance on the lexical decision task showed that, as expected, test words (*map*) were identified correctly at much higher accuracy rates (79-85%) than test nonwords (*mup*; 39-50%), whereas control words (86%) and nonwords (76%) where identified at similar accuracy rates. Similarly, test words were identified faster (1260-1304ms) than test nonwords (1387-1453 ms). Large differences between control and test items were obtained for test nonwords (76% vs. 39-50%, respectively), whereas for words differences between control and test items were very small (86% vs. 79-85%). Improvement in accuracy and speed between pre-test and post-test, however, was relatively small and only observable for test nonwords (5% for /æ/-/ʌ/ and 4.2% for /iː/-/ɪ/). The measures of perceptual sensitivity to the contrasts obtained through this task (*d'* and δ) showed a similar pattern of results (Table 5).

Contrast	*Measure*	*d-prime* (*d'*)				*delta* (δ)			
/æ/-/ʌ/	T1	.97	.81	-.57	2.91	.16	.08	.00	.33
	T2	.93	.47	.18	2.03	.18	.09	-.01	.33
	Gain	.016	.08	-.14	.14	-.007	.09	-.17	.17
/iː/-/ɪ/	T1	.91	.84	-.40	3.27	.16	.07	.02	.30
	T2	1.10	.91	.00	3.27	.18	.08	.00	.28
	Gain	.010	.10	-.25	.21	-.005	.10	-.12	.29

Table 5. Mean *d-prime* (*d'*) and *delta* (δ) pre-test and post-test scores and gains by vowel contrast.

3.2 Cognitive attention control

Participants obtained slightly higher accuracy scores in the Catalan version of the auditory selective attention task (*AudSelAtt*) than they did in the English version (Table 6). This difference did not reach significance ($t(16)$= .968, p=.348), but both scores were only moderately correlated (r=.442, p=.075), suggesting that individual differences in auditory selective attention were not consistent across the two tasks within participants.

In the auditory attention switching task (*AudAttSw*), as expected, participants were less accurate and slower at identifying the duration (*long* or *short*) and voice quality (*male* or *female*) in the vowels on switch trials (86%, 865ms) than on repeat trials (90%, 726ms). The overall error rate was low (10-13%), suggesting that these perceptual dimensions posed no difficulty to listeners (Table 6). Because RTs were measured from stimulus onset, participants took longer to respond to a duration trials than to voice trials, as whereas voice quality could be immediately identified from the beginning of the stimulus, the decision on duration required participants to wait for the duration of a short vowel (200ms). Consequently, we used an adjusted RT measure obtained by subtracting 200ms from the original RTs. We submitted the accuracy and adjusted RT scores to a series of ANOVAs with *Dimension* (duration, voice) and *Trial Type* (switch, repeat) as within-subjects factors. These analyses yielded a significant main effect of *Trial Type* $F(1, 16)=12.31$, $p=.003$, $\eta^2=.435$) and a non-significant main effect of *Dimension* ($F(1, 16)=.317$, $p=.581$, $\eta^2=.019$) on accuracy, suggesting that participants were equally accurate on both dimensions but made significantly more errors on switch than on repeat trials. For response speed (RTs), the ANOVA revealed significant main effects of both *Dimension* ($F(1, 16)=20.72$, $p<.001$, $\eta^2=.564$) and *Trial Type* ($F(1, 16)=45.65$, $p<.001$, $\eta^2=.741$), because participants were slower at deciding on the duration of a vowel than on whether it was produced by a male or a female speaker. None of the interactions reached significance. We used the switch cost measure as an index of attention switching skill.

In the auditory inhibition task (*AudInh*), the results showed that target sentences were processed slightly less accurately in L1-L1 sentence pairs (69-77%) than in L2-L2 (78-86%) or L2-L1 sentence pairs (80-89%), especially when the voice of the competing sentence was male. In L2-L2 and L2-L1 sentence pairs accuracy was lower when the voice of the competing sentence was female (78-80%) than when it was a male voice (86-89%), indicating that, when the L2 is attended to, a female voice is harder to inhibit than a male voice (irrespective of whether the female voice is speaking in the participants' L1 or L2). We submitted the aggregated scores for accuracy (proportion of correct responses) and RTs to a series of ANOVAs with *Language* (L2-L1, L2-L2, L1-L1) and *Target Voice* (Male, Female) as within-subject factors. The results of these analyses showed, for accuracy, a significant main effect of *Language* ($F(2, 15)=6.29$, $p=.010$, $\eta^2=.456$), a non-significant main effect of *Target Voice* ($F(1, 16)=0.38$, $p=.546$, $\eta^2=.023$) and a significant *Language* x *Target Voice* interaction ($F(2, 15)=6.29$, $p=.044$, $\eta^2=.340$). The interaction arose because whereas for L2-L1 sentence-pair trials with competitor sentences spoken by a female speaker obtained lower acccuracy rates than those spoken by a male speaker ($t(16)=-2.56$, $p=.021$), such a difference did not reach significance for L2-L2 ($t(16)=-1.07$, $p=.299$) and L1-L1 sentence pairs ($t(16)=1.03$, $p=.316$). Also, whereas *Language* had a significant main effect on response accuracy when attending to a female voice ($F(2, 15)=8.92$, $p=.003$, $\eta^2=.543$), this effect did not reach significance when attending to a male voice ($F(2, 15)=.440$, $p=.652$, $\eta^2=.055$). For response latencies, however, the ANOVA yielded significant main effects of *Language* ($F(2, 15)=18.24$, $p=.001$, $\eta^2=.709$) and *Target Voice* ($F(1, 16)=7.36$, $p=.015$, $\eta^2=.315$), and a non-significant *Language* x *Target Voice* interaction ($F(2, 15)=1.41$, $p=.274$, $\eta^2=.159$). We calculated *general* accuracy and RT inhibition scores across all language combinations and voices and a more specific score based on L2 learners' performance on L2-L1 trials only (Table 6), that is, trials where the L2 had to be attended to and the L1 had to be inhibited (*L1 inhibition*).

Task	*Conditions*	*M*	*SD*	*Min.*	*Max.*
AudSelAtt	Catalan (L1)	101.18	10.90	86	118
	English (L2)	98.47	10.932	79	114
AudAttSw	Switch	865	224	485	1338
	Repeat	726	221	445	1356
	Switch Cost	139	85	-17	283
AudInh	General (accuracy)	.803	.100	.54	.95
	L1 inhibition (accuracy)	.850	.102	.54	.94
	General (RT)	2392	386	1786	2989
	L1 inhibition (RT)	2338	314	1829	2861

Table 6. Mean scores in the attention control tasks: *AudSelAtt* (accuracy score 0-128), *AudAttSw* (adjusted RT in milliseconds) and *AudInh* (proportion of correct responses and RT in milliseconds)

3.3 Relationship between cognitive attention control and perceptual learning

Perception scores at pre-test, as expected, were related to the overall proficiency measure. L2 learners with higher scores in the elicited imitation task were better able to discriminate the target vowels /æ/-/ʌ/ ($r=.434$, $p=.082$) and /iː/-/ɪ/ ($r=.590$, $p=.013$) and also showed higher sensitivity to these contrasts (*d'* scores) in the lexical decision task (/æ/-/ʌ/: $r=.597$, $p=.011$; /iː/-/ɪ/: $r=.551$, $p=.022$), but proficiency was unrelated to gain scores.

Before assessing the contribution of cognitive attention control skills to L2 speech learning we explored the relationship between the various attention control measures (*AusSelAtt, AudAttSw, AudInh*). These analyses revealed an association between learners' auditory selective attention and auditory inhibition skills (Table 7), suggesting that the stronger their ability to focus their attention on a target voice in the presence of a competing voice in the *AudSelAtt* task, the better they could inhibit a competing voice in their first and second language in the *AudInh* task (Table 7). Thus, both these tasks appear to require participants to resort to the same underlying attentional resources. Interestingly, learners' switching costs in the *AudAttSw* task were strongly related to their ability to inhibit a voice in the L1 while attending to a voice speaking in the L2 (L1 inhibition), suggesting that learners with better auditory attentional flexibility (i.e. attention switching skills) were better able to inhibit their L1 when attending to the L2.

		AudAttSw		*AudInh*							
		RT		Accuracy				RT			
		Switch Cost		General		L1 Inhibition		General		L1 Inhibition	
		r=	*p=*	*r=*	*p=*	*r=*	*p=*	*r=*	*p=*	*r=*	*p=*
AudSelAtt	Catalan	-.396	.115	.578	.015	.649	.005	-.159	.543	.069	.794
	English	-.263	.308	.670	.003	.475	.054	-.576	.015	-.520	.033
AudAttSw	Switch Cost			-.410	.102	-.627	.007	.311	.224	.252	.328
	Switch							.807	<.001	.842	<.001
	Repeat							.698	.002	.757	<.001

Table 7. Pearson-r correlation coefficients between the attention control measures (shaded cells indicate significance).

Both ABX discrimination accuracy and RT gains of the two target contrasts were correlated with one another (*r*=.549, *p*=.022 and *r*=.723, *p*=.001, respectively), as they were in the lexical decision task (*r*=.541, *p*=.025), indicating that individual gain sizes were of similar magnitude for the /æ/-/ʌ/ and /iː/-/ɪ/ contrasts. We next assessed the relationship between the attention control measures and L2 learners' perception scores. We ran these analyses both for T1 perception scores and T1-T2 gains. For ABX discrimination accuracy, the results revealed significant moderately strong correlations between learners' perception gains and auditory selective attention, reaching significance for the /æ/-/ʌ/ contrast (*r*=.522, *p*=.031) and approaching significance for the /iː/-/ɪ/ contrast (*r*=.441, *p*=.076). This suggests that auditory selective attention predicts a considerable amount of variance (about 27%) in how much learners could benefit from the training. No significant associations were found between ABX discrimination gains and attention switching (*AudAttSw*) or inhibition (*AudInh*) scores. However, pre-test ABX discrimination accuracy scores were significantly related to the auditory attention switching measure (*r*=-.510, *p*=.037) and the L1 inhibition measure (*r*=.520, *p*=.032) for the /æ/-/ʌ/ contrast, suggesting that at pre-test both attention switching skill and L1 inhibition skills predicted a significant amount of variance (about 25%) in the learners' ability to discriminate the /æ/-/ʌ/ contrast. In addition, the RT L1 inhibition measure was strongly related to pre-test RT scores for both the /æ/-/ʌ/ (*r*=.619, *p*=.008) and /iː/-/ɪ/ (*r*=.701, *p*=.002) contrasts, explaining more than 40% of the variance in the discrimination response speed at pre-test. Finally, the relationship between the attention switching cost measure and the *d'* gain scores (gains in perceptual sensitivity) for the /æ/-/ʌ/ contrast approached significance (*r*=-.476, *p*=.063), and reached significance in the case of the adjusted δ accuracy gain measure in the lexical decision task (*r*=-.594, *p*=.015), suggesting that attention switching skill may be implicated in effecting changes in the lexical encoding of phonological contrasts. It should be noted, however, that perception gains between pre-test and post-test measured through the lexical decision task did not reach significance.

4. Discussion and conclusion

The main aim of the present study was to explore the contribution of cognitive attention control to L2 phonological development. We tested a group of L1-Catalan learners of L2 English on their attention control skills and trained them on the perception and production of two difficult L2 vowel contrasts (/æ/-/ʌ/ and /iː/-/ɪ/) through minimal-pair nonwords. We then assessed their gains and related them to the attention control measures.

The results revealed robust improvement from pre-test to post-test for both contrasts in response accuracy and speed, as well as consistent generalization effects to untrained nonwords and words. A major finding regarding phonetic training gains is that training based exclusively on minimal-pair nonwords, and therefore void of lexical content, led to improvement in the perception of minimal-pair words participants had not been trained on, suggesting that improvement in perceptual sensitivity to phonetic contrasts at the phonetic perceptual level may effect changes and lead to improvement in corresponding phono-lexical representations exploiting the same contrasts. However, the lexical decision task, which provided a measure of sensitivity to the /æ/-/ʌ/ and /iː/-/ɪ/ contrasts encoded lexically, only revealed little (and non-significant) improvement in sensitivity to the contrasts in nonwords. These apparently contradictory findings may result from the nature of the lexical decision task and the stage of development of the learners' L2 phonology. In a lexical decision task improvement in performance is based on the participants' ability to identify nonwords based on the phonological distinction between two members of a contrast in a lexical context, a task that required our learners to have accurately encoded the /æ/-/ʌ/ and /iː/-/ɪ/ contrasts lexically in their phonologies. Further research is needed to explore the efficiency of phonetic training in developing or changing phono-lexical representations. In particular, it would be interesting to carry out a follow-up training study based on words (rather than nonwords) and assess generalization to new lexical items through a lexical decision task.

The relationship between the cognitive attention control measures revealed an association between learners' performance on the auditory selective attention and the auditory inhibition tasks. Although the former did not require test-takers to inhibit a language through attention to voice, both tasks were based on voice competition and apparently required the recruitment of similar attentional resources. Similarly, participants' attention switching skills were related to their ability to inhibit a voice in the L1 when attending to L2 speech, suggesting that attention switching is implicated in L2 speech processing.

As regards the relationship between the cognitive attention control and the L2 vowel perception measures, a moderately strong correlation between L2 gains in the perception of the /æ/-/ʌ/ contrast and auditory selective attention suggests that learners' ability to focus their attention to specific speech dimensions is related to L2 phonological acquisition, confirming previous findings (Darcy et al., 2014; Safronova, 2016). Stronger associations between attention control and gains in L2 perception could have surfaced for tendencies identified in the current study with a slightly larger sample size.

The present study has comtributed to research on individual differences in L2 speech learning suggesting that cognitive attention control plays an important role in L2 speech learning. The fact that attention control explains a substantial amount of variance in L2 vowel perception has important implications for L2 pronunciation instruction beyond phonetic training. In particular, cognitive attention control is likely to play an important role in the context of communicative language teaching where recent research (Gurzynski-Weiss, Long, & Solon, 2017) has shown that meaning-oriented tasks with a focus on phonetic form making L2 pronunciation essential for task resolution is effective in developing L2 speech perception and production.

Acknowledgments

We would like to thank Isabelle Darcy for kindly sharing the lexical decision task with us, Natalia Wisniewska for helping recruit participants and collect data, and Zhifei Zhang for help in data collection. This study was supported by grant 2017SGR560 from the Catalan government.

References

Akeroyd, M. A. (2008). Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults. *International Journal of Audiology, 47*, S53-S71.

Astheimer, L. B., & Sanders, L. D. (2009). Listeners modulate temporally selective attention during natural speech processing. *Biological Psychology, 80*(1), 23-34.

Astheimer, L. B., Berkes, M., & Bialystok, E. (2016). Differential allocation of attention during speech perception in monolingual and bilingual listeners. *Language, Cognition and Neuroscience, 31*(2), 196-205.

Baese-Berk, M., Bent, T., Borrie, S., & McKee, M. (2015). Individual Differences in Perception of Unfamiliar Speech. In *Proceedings of the 18th International Congress of the Phonetic Sciences*.

Bialystok, E., Craik, F. I., & Luk, G. (2012). Bilingualism: consequences for mind and brain. *Trends in Cognitive Sciences, 16*(4), 240-250.

Boersma, P., & Weenink, D. (2018). Praat: doing phonetics by computer [Computer program]. Version 6.0.40, retrieved from http://www.praat.org/

Bohn, O.-S. (1995). Cross-language perception in adults: First language transfer doesn't tell it all. In W. Strange (Ed.), Speech perception and linguistic experience: Issues in cross-language research (pp. 379-410). Timonium, MD: York Press.

Darcy, I. (2018). Executive functions and L2 phonological development: a study in multiplicity. Paper presented at the *SimPhon.Net Workshop*, 4-6 June 2018, Stuttgart, Germany.

Darcy, I., Mora, J. C., & Daidone, D. (2014). Attention control and inhibition influence phonological development in a second language. In *Proceedings of the International Symposium on the Acquisition of Second Language Speech*. Concordia Working Papers in Applied Linguistics, 5, 115-129.

Filippi, R., Karaminis, T., & Thomas, M.S. (2014). Language switching in bilingual production: Empirical data and computational modelling. *Bilingualism: Language and Cognition, 17*(2), 294-315.

Filippi, R., Leech, R., Thomas, M. S., Green, D. W., & Dick, F. (2012). A bilingual advantage in controlling language interference during sentence comprehension. *Bilingualism: Language and Cognition, 15*(4), 858-872.

Flege, J. E. (1995). Second-language Speech Learning: Theory, Findings, and Problems. In W. Strange (Ed.), Speech perception and linguistic experience: Issues in cross-language research (pp. 229-273). Timonium, MD: York Press.

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, 35*(1), 116-124.

Gurzynski-Weiss, L., Long A. Y., & Solon M. (2017). TBLT and L2 Pronunciation: Do the benefits of Tasks Extend beyond Grammar and Lexis? *Studies in Second Language Acquisition, 39*, 213-224.

Hazan, V., & Kim, Y. H. (2010). Can we predict who will benefit from computer-based phonetic training? In *Online Proceedings of the INTERSPEECH 2010 Satellite Workshop on Second Language Studies: Acquisition, Learning, Education and Technology (L2WS 2010)*, Tokyo, Japan.

Humes, L. E., Lee, J. H., & Coughlin, M.P. (2006). Auditory measures of selective and divided attention in young and older adults using single-talker competition. *The Journal of the Acoustical Society of America, 120*(5), 2926-2937.

Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences, 7*(3), 134-140.

Mora, J. C. & Darcy, I. (2016). The relationship between cognitive control and pronunciation in a second language. In T. Isaacs & P. Trofimovich (Ed.), Second Language Pronunciation Assessment: Interdisciplinary Perspectives (pp. 95-120). Bristol, UK: Multilingual Matters.

Ortega, L., Iwashita, N., Norris, J. M., & Rabie, S. (2002). An investigation of elicited imitation tasks in crosslinguistic SLA research. In *Second Language Research Forum*, Toronto.

Ou, J., Law, S. P., & Fung, R. (2015). Relationship between individual differences in speech processing and cognitive functions. *Psychonomic Bulletin & Review, 22*(6), 1725-1732.

Petersen, S. E., & Posner, M. I. (2012). The attention system of the human brain: 20 years after. *Annual Review of Neuroscience, 35*, 73-89.

Robertson, I. H., Ward, T., Ridgeway, V., & Nimmo-Smith, I. (1994). The test of everyday attention (TEA). *Bury St Edmunds: Thames Valley Test Company.*

Safronova, E. & Mora, J. C. (2013). Attention control in L2 phonological acquisition. In A. Llanes Baró, L. Astrid Ciro, L. Gallego Balsà and R. M. Mateus Serra (Ed.), Applied Linguistics in the Age of Globalization (pp. 384-390). Lleida: Edicions de la Universitat de Lleida.

Safronova, E. (2016) *The Role of Cognitive Ability in the Acquisition of Second Language Perceptual Competence*. Unpublished PhD Thesis. Universitat de Barcelona.

Segalowitz, N. (2010). *Cognitive Bases of Second Language Fluency*. London, UK: Routledge.

Segalowitz, N., & Frenkiel-Fishman, S. (2005) Attention control and ability level in a complex cognitive skill: attention-shifting and second language proficiency. *Memory and Cognition, 33*, 644-653.

Wager, T. D., Jonides, J., & Smith, E. E. (2006). Individual differences in multiple types of shifting attention. *Memory & Cognition, 34*(8), 1730-1743.