

High-Variability Phonetic Training Under Different Conditions: Individual Differences in Auditory Attention Control



Ingrid Mora-Plaza, Mireia Ortega, and Joan C. Mora

Abstract Cognitive attention control guides auditory processes during speech processing but its contribution to L2 speech learning remains under-researched. This study examined the interaction between individual differences in auditory selective attention (ASA) and attention switching (ASW), and the effectiveness of high-variability phonetic training (HVPT) administered under different stimuli and presentation conditions to improve L2 learners' sensitivity to an L2 vowel contrast and its lexical encoding. Catalan-Spanish bilingual adult learners of English ($N = 102$) were randomly assigned to eight HVPT groups and trained in four 35-min sessions on the perception and production of English /æ/ and /ʌ/ through identification, discrimination, and immediate repetition tasks. Learners' gains were assessed through ABX discrimination and delayed word repetition tasks. Lexical encoding was tested through lexical decision and delayed sentence repetition tasks. We measured ASA through a single-talker competition paradigm and ASW through a novel speech-based version of the alternating-runs task-switching paradigm. Results showed that ASA was often related to pre-test (T1) and post-test (T2) perception accuracy but unrelated to either production accuracy or T1-T2 perceptual and production gains. However, ASW was related to /æ/ and /ʌ/ perception and production gains, but this varied as a function of stimuli type and presentation condition.

Keywords Attention control · HVPT · L2 vowel perception · L2 vowel production · Lexical encoding

I. Mora-Plaza (✉) · M. Ortega · J. C. Mora
University of Barcelona, Barcelona, Spain
e-mail: imoraplaza@ub.edu

M. Ortega
e-mail: m.ortega@ub.edu

J. C. Mora
e-mail: mora@ub.edu

1 Introduction

Although cases of exceptional L2 phonological acquisition have been attested in the Second Language Acquisition (SLA) literature (Moyer, 2014), most L2 learners struggle with L2 pronunciation, especially in instructed foreign language learning contexts where opportunities for L2 exposure and use are generally scarce. Experience-related factors that have been shown to explain inter-learner variation in L2 pronunciation learning in immersion settings, such as amount of L1 and L2 use, age of onset of L2 learning, length of residence, L2 input quantity and quality, among others (Flege, 2009; Munro & Bohn, 2007), have been shown to play a modest role in instructed SLA (Cebrian, 2006). However, both in immersion and instructed foreign language settings, individual differences in L2 phonological attainment cannot fully be accounted for by experience-related factors alone. Socio-psychological factors such as motivation, anxiety, or willingness to communicate (Kormos, 2017) as well as cognitive and aptitude-related factors such as auditory processing (Saito et al., 2020), working memory, inhibition, and attention (Darcy et al., 2014; Ghaffarvand Mokari & Werner, 2019; Lev-Ari & Peperkamp, 2014) also play a role.

Given the myriad of factors affecting L2 phonological acquisition over time and their interaction with L2 learners' individual differences, identifying, isolating, and quantifying the independent contribution of specific cognitive variables (e.g., attention control) to L2 speech learning becomes a challenging research objective. Two features make laboratory-based phonetic training an optimal testing ground: (a) variability in the extent to which learners benefit from it, and (b) full control over the type and amount of input learners receive (Golestani & Zatorre, 2009). Under such conditions, gains in perception and production can be directly related to independent measures of cognitive control.

The current study sets out to explore the role of cognitive attention control in L2 speech learning by examining the interaction between individual differences in auditory selective attention (ASA) and auditory attention switching (ASW) skills, and the effectiveness of high-variability phonetic training (HVPT) administered under different stimuli and presentation conditions. We focused on L1-Spanish/Catalan advanced learners' perception and production of English /æ/-/ʌ/ and its lexical encoding.

2 Literature Review

2.1 Phonetic Training

Most previous phonetic training research has used either perception (Bradlow, 2008) or production training methods (Kartushina et al., 2015). In perception, identification training has generally been found to lead to larger gains than discrimination training (Carlet & Cebrian, 2019), but few studies have combined discrimination

and identification training (Shinohara & Iverson, 2018) or perception and production training tasks in a HVPT paradigm (Wong, 2013). Additionally, phonetically-oriented training with nonwords has been shown to lead to larger gains than training with words because non-lexical materials allow learners to focus on the phonetic properties of the training stimuli while avoiding interference from lexically misrepresented phonetic forms (Ortega et al., 2021; Thomson & Derwing, 2016). Auditory attention control skills may potentially have a differential impact on training gains under phonetically- and lexically-oriented conditions. For example, as hypothesized in the current study, ASA may play a fundamental role in phonetically-oriented training, allowing learners to more easily extract the relevant phonetic properties that distinguish the target vowels /æ/ and /ʌ/. On the other hand, ASW, which involves inhibiting phonetic dimensions not under focus, may be more relevant in lexically-oriented training, where learners are trained on phono-lexical forms that are not likely to match their own representations.

Some training conditions have been shown to lead to greater gains. For instance, the presence of noise during training has been proved to have the effect of degrading the intelligibility of the speech signal (Mattys et al., 2012), but at the same time, it may help learners focus their attention on the more robust phonetic properties distinguishing the target contrast (Cooke & García-Lecumberri, 2018), and in production training it may lead to hyper-articulated speech (Hazan & Baker, 2011), which may enhance learners' ability to distinguish the target vowels in production. Audiovisual phonetic training has been shown to be superior to auditory-only training in training L2 sound contrasts (Hazan et al., 2005), and visual feedback has proved particularly effective in training the production of L2 vowels (Kartushina et al., 2015).

2.2 *Attention Control in L2 Speech Learning*

Attention control is implicated in speech processing and language comprehension and production (Miyake & Friedman, 2012) and in second language acquisition (Segalowitz & Frenkiel-Fishman, 2005). Both ASW and ASA skills allow listeners to selectively attend to specific acoustic dimensions during speech processing and to focus their processing resources on the auditory information that is relevant for language decoding processes to work efficiently (Astheimer et al., 2016). ASA skills, additionally, allow listeners to selectively attend to a single acoustic dimension or feature during speech processing, thus facilitating perceptual learning and the processing of L2 phonological contrasts (Ou et al., 2015). Phonetic training is effective in training learners to attend to speech dimensions and L2-specific acoustic cues not attended to in their native language (Iverson et al., 2005), suggesting that attention control skills may be an important source of individual differences in L2 phonetic training.

Research on the role of attention control in L2 phonetic training is scarce and has produced mixed results. For example, Kim and Hazan (2010) found ASW skills to be related to training gains in naïve L1-English speakers trained to perceive a

novel Korean stop voicing contrast. Mora and Mora-Plaza (2019) trained L1-Spanish learners in the perception and production of two L2-English vowel contrasts (/æ/-/ʌ/ and /i:/-/ɪ/). They found ASA to explain gains in the perception of one contrast (/æ/-/ʌ/), but not the other (/i:/-/ɪ/) and ASW was related to accuracy of performance in perceptual discrimination tasks, but unrelated to perception training gains. In the same line, Ghaffarvand Mokari and Werner (2019) found attention control to be unrelated to training gains for L1-Azerbaijani learners of English.

3 The Study

The main aim of this study is to examine the extent to which individual differences in auditory attention control can explain inter-learner variability in training gains for a challenging L2 vowel contrast. We chose the /æ/-/ʌ/ contrast because it is a difficult L2 contrast for L1-Spanish and L1 Spanish-Catalan bilingual learners of English alike (Rallo-Fabra & Romero, 2012), as both English vowels are perceptually mapped onto a single L1 low central vowel category /a/ in Spanish and Catalan, although /æ/ is a slightly better perceptual match for Spanish and Catalan /a/ than English /ʌ/ (Cebrian, 2019; Cebrian et al., 2011). To maximize potential training gains, we used a comprehensive HVPT paradigm that included two perception and one production task in every training session (see Sect. 4.3.1). Finally, to investigate potential interactions of cognitive attention control with training conditions requiring differential use of attentional resources, we trained learners with nonwords or with words. We also trained them with or without noise, and with or without visual monitoring. Based on Cooke and García-Lecumberri (2018), we expected learners with stronger auditory attention control skills to be better able to focus attention on the target vowels during stimuli repetition in the presence of masking noise. Additionally, we assessed the potential benefits of visual monitoring (watching one's own lips) during production training (with and without noise). Based on Hardison (2018), strong auditory attention control should allow learners to benefit from visual cues enhanced through the presence of masking noise.

The following research questions guided our investigation:

1. Does HVPT improve the perception and production of /æ/ and /ʌ/?
2. Does HVPT improve the lexical encoding of the /æ/-/ʌ/ contrast?
3. Do individual differences in auditory attention control explain variance in training gains?
4. To what extent does auditory attention control interact with training conditions to explain training gains?

4 Methods

4.1 Participants

The participants were 116 Spanish-Catalan bilingual undergraduate learners of English (see Table 1 for demographics) randomly assigned to one of eight different experimental training groups ($N = 102$) or to an untrained control group ($N = 14$; Table 2). One-way ANOVAs with *Training Group* as the independent variable confirmed that the experimental groups were comparable in L2 proficiency, $F(7,93) = 0.688$, $p = 0.681$, and L2 vocabulary size, $F(7,88) = 0.436$, $p = 0.877$. All participants reported having no speech or hearing pathologies.

4.2 Materials

The testing and training word and nonword stimuli contained the target vowels /æ/ and /ʌ/ as produced by six southern British English speakers (3 females, 3 males). They were elicited in carrier phrases (*I say X, I say X again*), recorded in a soundproof booth, excised, and normalized for amplitude in Praat (Boersma & Weenink, 2020). Four voices were used in the training and two of them (1 female, 1 male) were used for the testing stimuli only. Training stimuli were high-variability monosyllabic CVC nonword (8) and word (8) minimal pairs with the target vowels in eight different phonetic environments (e.g., *chang* /tʃæŋ/, *chung* /tʃʌŋ/, *mad* /mæd/, *mud* /mʌd/). Testing stimuli consisted of 12 monosyllabic CVC nonword minimal pairs (6 trained, 6 untrained) and 18 monosyllabic CVC word minimal pairs (6 trained, 12 untrained), plus 16 words which were presented in isolation and in the context of a sentence.

4.3 Procedure

Participants completed a language background questionnaire, and then they were trained individually in four 35-min sessions in a quiet lab, twice per week for two consecutive weeks (see training tasks in Sect. 4.3.1) and pre-and post-tested immediately before and after the training (see testing tasks in Sect. 4.3.2). Participants' cognitive attention control was measured in Session 2 (see cognitive control attention tasks in Sect. 4.3.3). Finally, participants' L2 proficiency was assessed in Session 3 via an elicited imitation (EI) test (Ortega et al., 2002) consisting of 30 sentences varying in length (7–19 syllables) and grammatical complexity. Participants had to repeat the sentences from memory after a 2000 ms delay. They also completed a yes/no vocabulary knowledge test (X/Y Lex; Meara & Miralpeix, 2006) that provided a measure of receptive vocabulary size (0–10,000 words). Figure 1 displays the distribution of training and testing tasks, and the attention control and L2 proficiency tasks.

Table 1 Participants' demographics

Measure	G1		G2		G3		G4		G5	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Age at testing (years)	24.7	9.8	23.9	10.1	21.2	2.3	21.7	4.6	22.2	7.4
L2 learning onset age (years)	7.2	2.9	7.3	4.1	6.1	1.9	5.2	2.2	6.1	2
Spoken L2 input	18.1	11	13.8	7.5	10.9	3.6	11.6	5.9	8.1	4.9
Spoken L2 output	7.1	5.3	5.7	3.9	5.9	3.3	5.9	5.6	2.9	2.6
L2 use ^a	9.8	3.8	14.2	6.5	12	7.4	12.2	6.9	13.9	9.7
Vocab. size (0–10,000 words) ^b	6431	1243	6491	1442	6215	1124	6579	1200	6277	1310
L2 proficiency (0–120 points) ^c	93.3	11.6	89.3	21.4	94.9	12.1	97.1	10.9	93.6	18.9
Self-estimated proficiency (1 = very poor—9 = native-like) ^d	6.7	0.8	6	1.5	7.2	1.0	7	0.7	6.8	1.2
Measure	G6		G7		G8		Control			
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
Age at testing (years)	20.4	1.1	22.6	10.2	22.3	7.4	26.7	7.1		
L2 learning onset age (years)	4.5	2.6	5.6	1.8	5.8	1.9	7.7	6.5		
Spoken L2 input	16.3	9	11	5.9	15.6	8.8	12.6	9.7		
Spoken L2 output	5.8	3.7	4.9	2.6	7.6	4.7	4.9	4.3		
L2 use ^a	13	4	17.1	10.9	10.5	4.4	12.5	6.2		
Vocab. size (0–10,000 words) ^b	5891	1310	6408	1196	5923	1141	–	–		
L2 proficiency (0–120 points) ^c	94.5	15.9	100.2	15.6	89.6	9.1	–	–		
Self-estimated proficiency (1 = very poor—9 = native-like) ^d	6.3	2.3	6.9	1.4	6.4	1	6.4	2		

^aL2 use with native and non-native speakers in hours per week

^bObtained through the X/Y Lex test (Meara & Miralpeix, 2006)

^cObtained through the Elicited Imitation task (Ortega et al., 2002)

^dAveraged self-estimated ability to speak spontaneously, understand, read, write and pronounce English

Table 2 Participant groups and training conditions

Group	N	Production training conditions		
		Stimuli Type	Monitoring	Listening
G1	13	Nonwords (NW)	Visual (V)	Noise
G2	11			Silence
G3	13		Auditory-only (A)	Noise
G4	14			Silence
G5	13	Words (W)	Visual	Noise
G6	12			Silence
G7	14		Auditory-only	Noise
G8	12			Silence

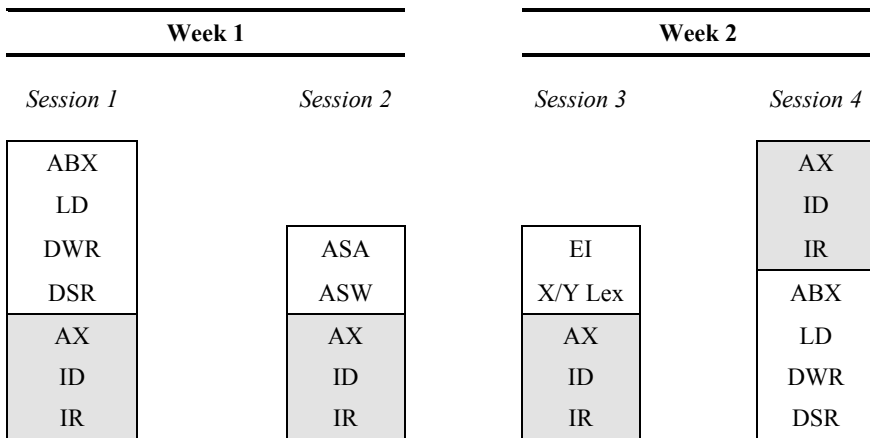


Fig. 1 Distribution of testing and training tasks (shading identifies training tasks)

4.3.1 Phonetic Training

The eight training groups differed in the type of stimuli they were trained on (nonwords or words) and the conditions in which they were administered during production training (with or without noise and/or visual monitoring) (Table 2).

In each of the four training sessions learners were trained perceptually through AX discrimination and identification tasks, and productively through an immediate repetition task (in this order, see Fig. 1).

- **AX Discrimination (AX):** Participants heard two stimuli (ISI = 500 ms) and decided (as fast and accurately as they could) whether the second vowel in the stimuli (X) contained the same English vowel as the first (same) or not (different). Participants responded to four practice trials and 96 test trials in every session (96 × 4 = 384 trials) to which they received feedback on accuracy and response

latency in milliseconds. The task contained the same number of same (AA, BB) and different trials (AB, BA), and combined a female and a male voice within trials. This perception task was included as a complement to identification training (Shinohara & Iverson, 2018) to increase learners' sensitivity to the primary acoustic cues qualitatively distinguishing /æ/ from /ʌ/ (1st and 2nd formant frequencies) and to improve their pre-categorical processing.

- **Identification (ID):** Participants heard one stimulus and identified (as fast and accurately as they could) whether it contained the vowel in the word *cap* or in the word *cup* by pressing a designated key on the keyboard matching the corresponding word, which appeared (together with its phonetic transcription and a picture representing it) on the bottom left or right side of the screen. Participants responded to four practice trials and 32 test trials in every session ($32 \times 4 = 128$ trials) and received feedback as in the AX task. This perception task was intended to improve category representations for /æ/ and /ʌ/ and their categorical processing in order to enhance generalization across contexts and talkers (Sadakata & McQueen, 2013).
- **Immediate Repetition (IR):** Participants heard the same stimuli as those in the ID task and were asked to repeat them twice as accurately as they could focusing on the vowel sound. They heard one stimulus, had 2000 ms to repeat it, then they heard it again, and had 2000 ms more to repeat it again. This procedure allowed learners to monitor their own productions. Participants responded to four practice trials and 32 test trials in every session ($32 \times 4 = 128$ trials). The training conditions for this task varied depending on the experimental group (Table 2) in terms of stimuli type (nonwords vs. words) and presentation condition (with or without noise and visual monitoring). This production task was included to allow participants to implement articulatory changes in the production of the contrast as they learned to perceptually discern /æ/ from /ʌ/. In this task, masking noise was included to enhance the production of clear speech in the auditory-only condition and to enhance attention to articulatory visual cues in the visual monitoring condition.

4.3.2 Testing

Vowel perception and production was pre- and post-tested through an ABX discrimination task and a delayed word repetition (DWR) task, respectively. The lexical encoding of the target vowel contrast was pre- and post-tested in perception and production through a Lexical Decision (LD) task and a delayed sentence repetition task (DSR), respectively (see Fig. 1).

- **ABX Discrimination (ABX):** Participants heard three stimuli in a row (ISI = 500 ms) and decided within 2500 ms (as fast and accurately as they could) whether the third one (X) contained the same vowel as the first (A) or the second (B) stimulus. Participants responded to a total of 136 trials: 30 test trials in four orders (ABA, ABB, BAB, BAA) = 120; and 8 control trials (/æ/-i:/, /ʌ/-i:/).
- **Delayed Word Repetition (DWR):** Participants repeated the words and nonwords they heard after a tone signal presented 1500 ms after stimulus onset. This delayed presentation procedure avoided repetition from sensory memory and ensured the

elicited stimuli reflected participants' vowel representations. To test for generalization effects, the testing stimuli contained trained and untrained words and nonwords in two different untrained voices (1 female, 1 male).

- **Lexical Decision (LD):** Participants heard the stimuli in a novel female speaker's voice and decided whether they were real or fake English words. Out of the 56 trials in the test, half were fillers (e.g., *lake*), and the other half were 14 word (e.g., *map*, *sun*) and 14 nonword (e.g., *mup*, *san*) test trials with an equal number of /æ/ and /ʌ/ items (half words and half nonwords). We used the proportion of correctly identified nonwords (e.g., *mup* or *san*) as a measure of perceptual sensitivity to the target contrast in a lexical context.
- **Delayed Sentence Repetition (DSR):** Participants silently read a sentence appearing on the screen (e.g., *He looked at the map to find his way*) targeting an /æ/ or /ʌ/ word (e.g., *map*), then they heard the sentence without reading it, and then waited 1500 ms for a tone signal to repeat it from memory. Sixteen sentences in untrained voices (1 female, 1 male) were repeated twice. Vowels elicited this way were deemed to reflect their corresponding category representations as encoded in the learners' mental lexicon.

4.3.3 Cognitive Attention Control

In Session 2, participants carried out two cognitive attention control tasks (see Fig. 1).

- **Auditory Selective Attention (ASA)** (Humes et al., 2006): This task consisted of 64 trials of pairs of English sentences (target vs. competitor). The two sentences in a pair were always different, one spoken by a female voice and the other by a male voice and were presented simultaneously through both ears. In every trial, a word signal (e.g., CHARLIE) appeared on the screen cueing the voice participants had to pay attention to in the sentences they would hear simultaneously (e.g., "*Ready Charlie go to blue six now*" + "*Ready Tiger go to red four now*"). Participants identified 1 of 4 colours and 1 of 8 digits visually presented on the screen (e.g., *blue* and *six* for the word signal CHARLIE). In this way, one of the voices and spoken sentences had to be attended to in order to correctly identify the colour and digit while the other was inhibited. Scores could range 0–128, one point for correctly identified colour and digit.
- **Auditory Attention Switching (ASW):** This task required participants to attend to either the duration (quantity) or the voice (quality) of L1 Catalan vowels (Safronova & Mora, 2013). Tokens of seven isolated Catalan vowels /i e ε a ɔ o u/ produced by a male and a female speaker were manipulated in Praat (Boersma & Weenink, 2020) to create short (200 ms) and long (500 ms) versions of the seven vowels. Eight identical copies of each stimulus ($28 \times 8 = 224$ trials) were randomly presented to participants over headphones for categorization as either long/short or male/female. The location of a speaker icon appearing predictably in clockwise fashion together with each auditory stimulus in one of four boxes cued the dimension to be attended to: long/short when appearing in one of the two top boxes, male/female when appearing in one of two bottom boxes.

Within-dimension (repeat trials) response times (RTs) were expected to be shorter than across-dimension (switch trials) RTs. A shorter switch-cost RT score (switch RT minus repeat RT) reflected stronger ASW skills.

The perception and production tasks and the ASW test were administered in *DmDx* (Forster & Forster, 2003), the ASA test in *Inquisit* (Draine, 1999). Participants' productions were recorded at a sampling frequency of 44.1 kHz on Marantz PMD-661 digital recorders with an external Shure SM58 voice microphone.

4.4 Data Analysis

For the ABX and LD tasks, we obtained accuracy and RT scores. RT scores included correct responses only and were screened to exclude RTs 2.5 *SDs* below or above each subject's mean. For the DWR and DSR tasks, we computed vowel production accuracy scores as the spectral distance between participants' vowel production and the average of the same vowels in the same items as produced by the six native speakers whose voices were used in the testing. Vowel frequency measures (f_0 , F1, F2) were extracted in Praat (Boersma & Weenink, 2020) from a 10-ms window centred at the midpoint of the steady-state portion of the target vowels. Extreme values above or below 3 *SDs* from each participant's mean were replaced with the mean value for that vowel in the same testing time. To minimize age, gender, and vocal tract size differences, frequency values in Hertz (Hz) were converted to Bark (B), and then a Bark-distance normalization procedure was used to provide speaker-independent estimates of vowel quality. The difference in Bark between F1 and f_0 (B1-B0) estimated vowel height, whereas the difference between F2 and F1 (B2-B1) estimated vowel frontness (Bohn & Flege, 1990).

Scores from all tasks were fitted to Generalized Linear Mixed Models (GLMMs) in SPSS 25, with *Testing Time* (T1, T2), *Group* (G1-G9), and *Vowel* ($/\text{æ}/$, $/\Delta/$) as fixed effects, and *Subject* and *Item* as random factors. To assess the relationship between attention control and training gains, we aggregated the scores by subject and ran *Pearson-r* correlations.

5 Results

First, we present the results by group in terms of the effects of training on participants' sensitivity to the contrast (ABX and DWR) and its lexical encoding (LD and DSR). Second, we report the results on the relationship between cognitive attention control (ASA and ASW) and perception and production training gains and performance.

5.1 Training Effects on /æ/ and /ʌ/ Perception and Production

In general, vowel perception and production accuracy (ABX and DWR) improved for all groups (Table 3), and the lexical encoding (LD and DSR) of the contrast did, too, but to a lesser extent, except for the control group (G9), who did not show improvement in any testing task.

For ABX accuracy, the GLMM revealed a significant main effect of *Testing Time*, $F(1,28524) = 203.352, p < 0.001$, and *Vowel*, $F(1,28524) = 254.430, p < 0.001$, and a significant *Group* \times *Testing Time* \times *Vowel* interaction, $F(8,28524) = 2.787, p = 0.004$. This interaction arose because only G3 (NW + A + noise), G4 (NW + A + silence), G6 (W + V + silence), and G7 (W + A + noise) significantly improved on both vowels (see Tables 2 and 3). No other main effects or interactions reached significance.

Table 3 Descriptive statistics for ABX (proportion of correct responses), LD (proportion of correctly identified nonwords), DWR and DSR (spectral distances in Bark between learners' and native speakers' productions), by vowel and group. Shading indicates improvement ($M = mean, SD = standard deviation$)

	ABX								DWR							
	/æ/				/ʌ/				/æ/				/ʌ/			
	T1		T2		T1		T2		T1		T2		T1		T2	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
G1	.78	.41	.83	.37	.70	.45	.74	.44	1.99	1.56	1.77	1.38	2.16	1.52	1.96	1.42
G2	.80	.40	.83	.37	.70	.46	.78	.41	1.38	0.89	1.31	0.95	1.48	0.96	1.52	0.98
G3	.71	.45	.81	.39	.68	.46	.73	.44	1.58	1.30	1.33	0.97	1.80	1.30	1.80	1.36
G4	.75	.43	.83	.37	.65	.47	.75	.43	1.60	1.38	1.32	1.04	1.92	1.46	1.66	1.28
G5	.77	.42	.86	.34	.71	.45	.75	.44	1.47	1.31	1.58	1.37	1.72	1.50	1.70	1.40
G6	.78	.41	.85	.35	.70	.46	.81	.39	1.39	1.12	1.18	0.84	1.49	1.06	1.34	0.89
G7	.77	.42	.81	.38	.67	.47	.79	.41	1.35	1.16	1.30	1.10	1.48	1.13	1.46	1.13
G8	.80	.40	.83	.37	.67	.47	.78	.41	1.28	1.02	1.30	0.96	1.49	1.00	1.44	1.02
G9	.77	.42	.77	.37	.64	.46	.65	.45	1.24	0.94	1.38	1.17	1.45	1.04	1.57	1.13
	LD								DSR							
	/æ/				/ʌ/				/æ/				/ʌ/			
	T1		T2		T1		T2		T1		T2		T1		T2	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
G1	.55	.50	.62	.48	.68	.47	.68	.47	2.12	1.46	1.97	1.35	2.56	1.84	2.25	1.62
G2	.60	.49	.62	.48	.61	.49	.64	.48	1.58	1.25	1.57	0.96	1.67	1.28	1.69	1.32
G3	.54	.50	.57	.49	.62	.49	.66	.47	1.50	1.01	1.91	1.44	2.19	1.67	2.28	1.69
G4	.57	.50	.52	.50	.64	.48	.64	.48	2.28	2.02	1.37	1.10	2.63	1.84	1.65	1.02
G5	.55	.50	.63	.48	.65	.48	.66	.48	1.51	1.22	1.57	1.41	1.55	1.04	1.62	1.46
G6	.62	.49	.68	.46	.63	.48	.63	.48	1.44	1.17	1.53	1.12	1.70	1.31	1.60	1.12
G7	.58	.49	.63	.48	.71	.46	.70	.46	1.71	1.45	1.47	1.02	2.28	1.71	1.89	1.49
G8	.54	.50	.60	.49	.68	.47	.64	.48	1.22	0.83	1.44	1.03	1.56	0.97	1.74	1.41
G9	.52	.50	.52	.49	.71	.45	.71	.46	1.60	1.13	1.99	1.43	1.53	1.31	1.95	1.56

For the DWR spectral distance scores, the GLMM revealed a significant main effect of *Testing Time*, $F(1,18050) = 23.480$, $p < 0.001$, and *Vowel*, $F(1,18050) = 11.358$, $p = 0.001$, and a significant *Testing Time* \times *Group* interaction, $F(8,18050) = 7.996$, $p < 0.001$, and *Group* \times *Vowel* interactions, $F(8,18050) = 3.018$, $p = 0.002$. Bonferroni-adjusted pairwise comparisons indicated that the *Testing Time* \times *Group* interaction arose because three of the four groups trained with nonword stimuli (G1, G3 and G4) and only one of the four trained with word stimuli (G6, W + V + silence) produced both target vowels more accurately than the other groups.

For LD accuracy, the GLMM revealed a significant main effect of *Testing Time*, $F(1,6376) = 4.645$, $p = 0.031$, and a significant *Group* \times *Vowel* interaction, $F(8,6376) = 2.652$, $p = 0.007$. None of the other fixed factors or interactions reached significance.

For the DSR spectral distance scores, no significant main effects were found, but the *Testing Time* \times *Group*, $F(8,3708) = 10.488$, $p < 0.001$, and *Group* \times *Vowel* interactions, $F(8,3708) = 3.956$, $p < 0.001$, turned out to be significant. Bonferroni-adjusted pairwise comparisons showed that only group G4 (NW + A + silence) produced the /æ/ significantly more accurately at post-test, as it was also the case in the DWR task.

Overall, the results show that the HVPT improved learners' discriminability of the L2 vowel contrast (ABX and DWR tasks), but little improvement was obtained in the lexical encoding of the contrast (DSR and LD tasks). Production gains were very modest, but groups trained with nonwords (G1, G2, G3, G4) gained significantly more than groups trained with words (G5, G6, G7, G8).

5.2 Attention Control and L2 Training Gains

Participants obtained a mean score of 94.60 ($SD = 16.14$, $Range = 52-125$) in the ASA task. In the ASW task, as expected, participants were significantly less accurate, $t(26206) = -7.326$, $p < 0.001$, and slower, $t(22771) = 30.759$, $p < 0.001$, on switch trials (Acc: $M = 0.88$, $SD = 0.326$; RT: $M = 976.44$ ms, $SD = 350.09$) than on repeat trials (Acc: $M = 0.91$, $SD = 0.290$; RT: $M = 840.53$ ms, $SD = 316.42$). Their attention switch-cost score ($M = 139.36$, $SD = 90.95$) was used in the correlation analyses.

Overall, correlational analyses failed to reveal an association between learners' gains in L2 vowel perception and production and the attention control measures, suggesting that gain sizes were unrelated to individual differences in attention control. Only a weak correlation, $r = 0.279$, $p = 0.004$, arose between ASA and DWR gains. Correlational analyses conducted separately by group yielded a similar picture. ASA was unrelated to any of the gain measures in all training groups. Nevertheless, ASW scores were strongly associated with some of the gain measures for some of the groups (Table 4).

Table 4 Pearson-*r* correlation coefficients between ASW and L2 perception and production gains (shaded cells indicate significance)

	/æ/						/ʌ/					
	ABX		LD		DWR		ABX		LD		DWR	
	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>
G1	-.552	.050	-.237	.435	-.440	.133	-.390	.188	-.577	.059	-.159	.603
G2	-.096	.780	-.188	.580	-.418	.201	-.836	<.001	-.345	.298	-.200	.555
G3	.091	.768	-.248	.413	.088	.775	-.242	.426	.280	.354	-.374	.209
G4	.152	.604	.127	.665	-.420	.135	.016	.957	-.582	.029	-.577	.039
G5	.331	.270	-.279	.355	-.071	.817	-.423	.150	-.027	.929	.104	.734
G6	-.713	.009	-.683	.014	.224	.484	-.014	.996	-.049	.880	.252	.430
G7	.004	.988	.244	.053	-.538	.047	.346	.247	-.103	.725	-.033	.915
G8	-.099	.760	.193	.548	.266	.404	-.028	.931	.002	.988	-.119	.713

- ASW explained gain differences in the production of /æ/ in the DSR task ($p < 0.001$) for G2 (NW + V + silence).
- ASW was significantly correlated with gains in perceptual discrimination (ABX) ($p = 0.009$) and lexical encoding (LD) ($p = 0.014$) of /æ/ for G6 (W + V + silence).
- ASW explained a 55% of variance in the lexical encoding measure (LD) of /ʌ/ for G3 (NW + A + noise) and a 29% of variance in the production of words containing /æ/ for G7 (W + A + noise).
- Learners with stronger ASW skills in G4 (NW + A + silence) produced the L2 vowel /ʌ/ in the DWR and DSR significantly more accurately than those with poorer attention control (moderately strong correlations).

In sum, attention control (ASA and ASW) was not strongly related to gains in L2 vowel sensitivity and lexical encoding, but it helped in the conditions that required higher attentional demands (G2, G3, G6, G7).

Since as a whole attention control appeared to be unrelated to training gains, we explored whether it was related to individual differences in performance in the perception and production tasks at both testing times. Here we found that ASA was significantly related to ABX accuracy at T1 (/æ/: $r = 0.533$, $p < 0.001$; /ʌ/: $r = 0.508$, $p < 0.001$) and at T2 (/æ/: $r = 0.464$, $p < 0.001$; /ʌ/: $r = 0.473$, $p < 0.001$), explaining 21–28% of variance in participants' sensitivity to the target contrast, whereas ASW was only weakly related to ABX accuracy at T1 (/ʌ/: $r = -0.226$, $p = 0.022$). No significant associations were found between ASA or ASW and LD, DWR or DSR scores at T1 or T2. Therefore, ASA correlates strongly with ABX discrimination, which requires learners to perceptually discern between competing L2 vowel qualities by selecting one stimulus over another within every trial.

6 Discussion

Overall, HVPT was effective at improving trainees' discrimination of /æ/-/ʌ/ in perception and production (RQ1). Phonetically-oriented training through nonwords (unbiased by learners' lexical representations) led to larger gains in production than training through words, supporting previous findings (Ortega et al., 2021; Thomson & Derwing, 2016). However, trainees did not improve the lexical encoding of the contrast (RQ2). Longer HVPT combined with extended meaningful use of the L2 exploiting the target contrast in communicative tasks may be necessary for advanced learners to modify the lexical encoding of a phonological contrast.

Concerning the relationship between auditory attention control and L2 perception and production gains (RQ3), neither ASA nor ASW explained individual differences in training gains. In fact, we expected attention control to explain little variance in gains for groups that had obtained relatively small gains. Only ASW scores were found to be related to gains in L2 vowel learning, and only for some of the groups (G2, G3, G4, G6 and G7). It seems that learners' ability to switch between vowel quality

and quantity explained learning gains especially for those who had been trained on either visual or background noise conditions. However, contrary to our expectations, ASW skills were unrelated to gains when learners were trained under the most demanding condition (visual monitoring + noise). Further research is needed to investigate this lack of relationship.

Concerning RQ4, ASA correlated strongly with learners' T1 and T2 scores in the ABX task, indicating that ASA enhanced learners' ability to discern between the target vowels, supporting previous findings (Mora & Mora-Plaza, 2019). However, neither ASA nor ASW were found to consistently interact with the training conditions in explaining gains, possibly due to training gains being relatively small within groups and testing not including any of the conditions implemented in the training. These findings suggest that further research should examine the role of attention control in learners' performance within training sessions from an individual differences perspective. Attention control may be more directly implicated in learners' actual training performance in perceptual discrimination and identification, as well as in the production tasks, during which the noise and visual monitoring conditions were present.

7 Pedagogical Implications

7.1 Implications for Phonetic Training

The present study demonstrates that HVPT helps learners better categorize vowels produced by different L2 speakers, and improves their L2 phonetic skills by helping them place the indexical information in the input (speakers' voice quality) in the perceptual background, thus enhancing the development of L2 phonetic categories during perceptual learning (Best, 2011). Moreover, HVPT may help learners develop pronunciation learning strategies in identifying new words from new speakers that can be transferred to production, thus contributing effectively to L2 pronunciation learning.

Pronunciation practice outside the laboratory could be provided through computer-assisted pronunciation training applications. These applications are designed to draw learners' attention to sounds and minimize attention to meaning, are interactive and entertaining, and involve immediate corrective feedback. For example, the *English Accent Coach* (Thomson, 2018), which was designed using a principled, research-based approach, showed to effectively improve pronunciation (Thomson, 2011). This website may improve speech comprehensibility and intelligibility without production practice. It also allows endless research possibilities as teachers and researchers could collaborate remotely, monitoring the effect of perceptual training and its impact on pronunciation.

7.2 *Implications for Pronunciation Teaching*

Cognitive attention control is likely to play an important role in the context of communicative language teaching. Meaning-oriented tasks where attention is directed to phonetic form have been shown to be effective in developing L2 speech perception and production skills (Gurzynski-Weiss et al., 2017).

Given that attention to phonetic features is necessary for pronunciation learning, teachers should ensure that students have as much exposure as possible to L2 speech that preserves phonological contrasts between L2 phonemes. One way of achieving this is to first provide explicit pronunciation practice through the use of nonwords (Mora & Levkina, 2017) and then progressively incorporate communicative tasks that require learners to use contrasting L2 sounds in real words (Tyler, 2019). Teachers could gradually change their focus-on-form tasks to real-world task-based pronunciation teaching tasks. This may be possible through the use of map tasks using words (Solon et al., 2017) or realistic problem-solving tasks that make the target phonological features essential for task completion and orient learners' attention to L2 phonological elements through the manipulation of task features (i.e., \pm task complexity) (Mora-Plaza et al., 2018).

8 Conclusion

The present study has contributed to research on individual differences in L2 speech learning by exploring the role of auditory attention control in the phonetic training of L2 vowels. Based on prior research, it was hypothesized that training learners to exploit their attentional resources in phonetic form-focused pronunciation tasks to learn to perceive L2 phonological contrasts may prove a successful strategy to improve L2 pronunciation. Our study shows that Catalan-Spanish bilingual adult learners of English improved their ability to discriminate /æ/-/ʌ/ in perception and production tasks after receiving phonetic training, and that their production gains were larger when the training was through nonwords rather than through words. Yet, their lexical encoding of the contrast did not improve, and neither ASA nor ASW explained individual differences in training gains. Longer phonetic training with communicative tasks that draw attention to form may be necessary for advanced learners to modify the lexical encoding of a phonological contrast. For example, pair work involving minimal-pair based spot-the-difference tasks performed in noise might provide effective classroom training in auditory attentional skills that learners may find useful for L2 implicit perceptual learning through exposure to L2 oral input. Further research should empirically test the pedagogical value of manipulating auditory attentional demands to promote L2 pronunciation learning.

The present study is subject to several limitations. Sample sizes were small (11–14 per group). The visual monitoring and noise training conditions were implemented

during production training only; they should have also been included during perception training. Finally, we tested production without visual monitoring or masking noise irrespective of training condition. In addition, as many of the target sources of individual differences are likely to be related to one another (e.g., auditory processing skills are likely to be related to cognitive attention control), it would be convenient to include as many potentially related variables in a single study as possible. This would allow researchers to statistically assess the joint and unique contribution of predictor variables while controlling for the confounding effects of mediating ones. Finally, further research is needed to investigate the role of attention control within each training session to observe whether attention plays a role during training.

References

- 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. <https://doi.org/10.1080/23273798.2015.1083114>
- Best, C. T. (2011). Devil or angel in the details? Perceiving phonetic variation as information about phonological structure. In J. Romero & M. Riera (Eds.), *The phonetics-phonology interface: Representations and methodologies* (pp. 3–32). John Benjamins. <https://dx.doi.org/10.1075/cilt.335.01bes>
- Boersma, P., & Weenink, D. (2020). *Praat: Doing phonetics by computer* (Version 6.1.09) [Computer software].
- Bohn, O. S., & Flege, J. E. (1990). Interlingual identification and the role of foreign language experience in L2 vowel perception. *Applied Psycholinguistics*, 11(3), 303–328. <https://doi.org/10.1017/S0142716400008912>
- Bradlow, A. R. (2008). Training non-native language sound patterns: Lessons from training Japanese adults on the English /r/-/l/ contrast. In J. G. Hansen Edwards & M. L. Zampini (Eds.), *Phonology and second language acquisition* (pp. 287–308). John Benjamins. <https://dx.doi.org/10.1075/sibil.36.14bra>
- Carlet, A., & Cebrian, J. (2019). Assessing the effect of perceptual training on L2 vowel identification, generalization and long-term effects. In A. M. Nyvad, M. Hejná, A. Højen, A. B. Jespersen, & M. H. Sørensen (Eds.), *A sound approach to language matters: In honor of Ocke-Schwen Bohn* (pp. 91–119). Aarhus University. <https://dx.doi.org/10.7146/aul.322.218>
- Cebrian, J. (2006). Experience and the use of duration in the categorization of L2 vowels. *Journal of Phonetics*, 34(3), 372–387. <https://doi.org/10.1016/j.wocn.2005.08.003>
- Cebrian, J. (2019). Perceptual assimilation of British English vowels to Spanish monophthongs and diphthongs. *Journal of the Acoustical Society of America*, 145(1), EL52–EL58. <https://dx.doi.org/10.1121/1.5087645>
- Cebrian, J., Mora, J. C., & Aliaga-Garcia, C. (2011). Assessing crosslinguistic similarity by means of rated discrimination and perceptual assimilation tasks. In M. Wrembel, M. Kul, & K. Dziubalska-Kołaczyk (Eds.), *Achievements and perspectives in SLA of speech, New Sounds 2010* (Vol. 1, pp. 41–52). Peter Lang. ISBN: 9783631607220
- Cooke, M., & Garcia-Lecumberri, M. (2018). Effects of exposure to noise during perceptual training of non-native language sounds. *The Journal of the Acoustical Society of America*, 143(5), 2602–2610. <https://doi.org/10.1121/1.5035080>
- Darcy, I., Mora, J. C., & Daidone, D. (2014). Attention control and inhibition influence phonological development in a second language. *Proceedings of the 7th International Symposium on the Acquisition of Second Language Speech, New Sounds 2013: Concordia Working Papers in Applied Linguistics*, 5, 115–129. <https://hdl.handle.net/2022/22863>

- Draine, S. (1999). *Inquisit* (Version 5.0.14.0) [Computer software]. Millisecond Software. <https://www.millisecond.com/>
- Flege, J. E. (2009). Give input a chance! In T. Piske & M. Young-Scholten (Eds.), *Input matters in SLA* (pp. 175–190). Multilingual Matters.
- Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods*, 35(1), 116–124. <https://doi.org/10.3758/BF03195503>
- Ghaffarvand Mokari, P., & Werner, S. (2019). On the role of cognitive abilities in second language vowel learning. *Language and Speech*, 62(2), 260–280. <https://doi.org/10.1177/0023830918764517>
- Golestani, N., & Zatorre, R. J. (2009). Individual differences in the acquisition of a second language phonology. *Brain and Language*, 109(2–3), 55–67. <https://doi.org/10.1016/j.bandl.2008.01.005>
- Grzybowski-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(2), 213–224. <https://doi.org/10.1017/S0272263117000080>
- Hardison, D. M. (2018). Effects of contextual and visual cues on spoken language processing: Enhancing L2 perceptual salience through focused training. In S. M. Gass, P. Spinner, & J. Behney (Eds.), *Salience in second language acquisition* (pp. 201–220). Routledge.
- Hazan, V., & Baker, R. (2011). Acoustic-phonetic characteristics of speech produced with communicative intent to counter adverse listening conditions. *The Journal of the Acoustical Society of America*, 130(4), 2139–2152. <https://doi.org/10.1121/1.3623753>
- Hazan, V., Sennema, A., Iba, M., & Faulkner, A. (2005). Effect of audiovisual perceptual training on the perception and production of consonants by Japanese learners of English. *Speech Communication*, 47(3), 360–378. <https://doi.org/10.1016/j.specom.2005.04.007>
- 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. <https://doi.org/10.1121/1.2354070>
- Iverson, P., Hazan, V., & Bannister, K. (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. <https://doi.org/10.1121/1.2062307>
- Kartushina, N., Hervais-Adelman, A., Frauenfelder, U. H., & Golestani, N. (2015). The effect of phonetic production training with visual feedback on the perception and production of foreign speech sounds. *The Journal of the Acoustical Society of America*, 138(2), 817–832. <https://doi.org/10.1121/1.4926561>
- Kim, Y. H., & Hazan, V. (2010). Individual variability in the perceptual learning of L2 speech sounds and its cognitive correlates. In K. Dziubalska-Kołaczyk, M. Wrembel, & M. Kul (Eds.), *Proceedings of the 6th International Symposium on the Acquisition of Second Language Speech, New Sounds 2010* (pp. 251–256). Poznań, Poland. ISBN: 978-83-928167-9-9
- Kormos, J. (2017). The effects of specific learning difficulties on processes of multilingual language development. *Annual Review of Applied Linguistics*, 37, 30–44. <https://doi.org/10.1017/S026719051700006X>
- Lev-Ari, S., & Peperkamp, S. (2014). The influence of inhibitory skill on phonological representations in production and perception. *Journal of Phonetics*, 47, 36–46. <https://doi.org/10.1016/j.wocn.2014.09.001>
- Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes*, 27(7–8), 953–978. <https://doi.org/10.1080/01690965.2012.705006>
- Meara, P. M., & Miralpeix, I. (2006). *Y_Lex: The Swansea advanced vocabulary levels test* (Version 2.05) [Computer software]. Lognostics. <https://www.lognostics.co.uk/tools/>
- Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science*, 21(1), 8–14. <https://doi.org/10.1177/0963721411429458>

- Mora, J. C., & Levkina, M. (2017). Task-based pronunciation teaching and research: Key issues and future directions. *Studies in Second Language Acquisition*, 39(2), 381–399. <https://doi.org/10.1017/S0272263117000183>
- Mora, J. C., & Mora-Plaza, I. (2019). Contributions of cognitive attention control to L2 speech learning. In A. M. Nyvad, M. Hejná, A. Højen, A. B. Jespersen, & M. H. Sørensen (Eds.), *A sound approach to language matters: In honor of Ocke-Schwen Bohn* (pp. 477–499). Aarhus University. <https://dx.doi.org/10.7146/aul.322.218>
- Mora-Plaza, I., Mora, J. C., & Gilabert, R. (2018). Learning L2 pronunciation through communicative tasks. In J. Levis (Ed.), *Proceedings of the 9th Pronunciation in Second Language Learning and Teaching Conference*, ISSN 2380-9566, University of Utah, September, 2017 (pp. 174–184). Ames, IA: Iowa State University.
- Moyer, A. (2014). What's age got to do with it? Accounting for individual factors in second language accent. *Studies in Second Language Learning and Teaching*, 4(3), 443–464. <https://dx.doi.org/10.14746/ssllt.2014.4.3.4>
- Munro, M. J., & Bohn, O.-S. (2007). The study of second language speech: A brief overview. In O.-S. Bohn & M. J. Munro (Eds.), *Language experience, second language learning: In honor of James Emil Flege* (pp. 3–11). John Benjamins. <https://dx.doi.org/10.1075/llt.17.06mun>
- Ortega, L., Iwashita, N., Norris, J. M., & Rabie, S. (2002, October 3–6). *An investigation of elicited imitation tasks in crosslinguistic SLA research* [Conference presentation]. Second Language Research Forum, Toronto, Canada.
- Ortega, M., Mora-Plaza, I., & Mora, J. C. (2021). Differential effects of lexical and non-lexical high-variability phonetic training on the production of L2 vowels. In A. Kirkova-Naskova, A. Henderson, & J. Fouz-González (Eds.), *English pronunciation instruction: Research-based insights* (pp. 328–355). John Benjamins. <https://dx.doi.org/10.1075/aaals.19.14ort>
- 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. <https://dx.doi.org/10.3758/s13423-015-0839-y>
- Rallo-Fabra, L., & Romero, J. (2012). Native Catalan learners' perception and production of English vowels. *Journal of Phonetics*, 40(3), 491–508. <https://doi.org/10.1016/j.wocn.2012.01.001>
- Sadakata, M., & McQueen, J. M. (2013). High stimulus variability in nonnative speech learning supports formation of abstract categories: Evidence from Japanese geminates. *The Journal of the Acoustical Society of America*, 134(2), 1324–1335. <https://doi.org/10.1121/1.4812767>
- Safronova, E., & Mora, J. C. (2013). Attention control in L2 phonological acquisition. In A. Llanes Baró, L. Astrid Ciro, L. Gallego Balsà, & R. M. Mateus Serra (Eds.), *Applied linguistics in the age of globalization* (pp. 384–390). Edicions de la Universitat de Lleida.
- Saito, K., Kachlicka, M., Sun, H., & Tierney, A. (2020). Domain-general auditory processing as an anchor of post-pubertal L2 pronunciation learning: Behavioural and neurophysiological investigations of perceptual acuity, age, experience, development, and attainment. *Journal of Memory and Language*, 115, 104168. <https://doi.org/10.1016/j.jml.2020.104168>
- Segalowitz, N., & Frenkiel-Fishman, S. (2005). Attention control and ability level in a complex cognitive skill: Attention shifting and second-language proficiency. *Memory & Cognition*, 33(4), 644–653. <https://doi.org/10.3758/BF03195331>
- Shinohara, Y., & Iverson, P. (2018). High variability identification and discrimination training for Japanese speakers learning English/r/–/l/. *Journal of Phonetics*, 66, 242–251. <https://doi.org/10.1016/j.wocn.2017.11.002>
- Solon, M., Long, A. Y., & Gurzynski-Weiss, L. (2017). Task complexity, language-related episodes, and production of L2 Spanish vowels. *Studies in Second Language Acquisition*, 39(2), 347–380. <https://doi.org/10.1017/S0272263116000425>
- Thomson, R. I. (2011). Computer assisted pronunciation Training: Targeting second language vowels: Perception improves pronunciation. *CALICO Journal*, 28(3), 744–765. <https://dx.doi.org/10.11139/cj.28.3.744-765>
- Thomson, R. I. (2018). *English Accent Coach* (Version 2.3) [Computer software]. <https://www.englishaccentcoach.com/>

- Thomson, R. I., & Derwing, T. M. (2016). Is phonemic training using nonsense or real words more effective? In J. Levis, H. Le., I. Lucic, E. Simpson, & S. Vo (Eds.), *Proceedings of the 7th Pronunciation in Second Language Learning and Teaching Conference*, ISSN 2380-9566, Dallas, TX, October 2015 (pp. 88–97). Ames, IA: Iowa State University.
- Tyler, M. D. (2019). PAM-L2 and phonological category assimilation in the foreign language classroom. In A. M. Nyvad, M. Hejná, A. Højen, A. B. Jespersen, & M. H. Sørensen (Eds.), *A Sound approach to language matters: In honor of Ocke-Schwen Bohn* (pp. 607–630). Aarhus University. <https://dx.doi.org/10.7146/aul.322.218>
- Wong, J. W. S. (2013). The effects of perceptual and/or productive training on the perception and production of English vowels /ɪ/ and /i:/ by Cantonese ESL learners. In F. Bimbot, C. Cerisara, C. Fougeron, G. Gravier, L. Lamel, P. Pellegrino, & P. Perrier (Eds.), *Proceedings of the 14th Annual Conference of the International Speech Communication Association, Interspeech 2013* (pp. 2113–2117). ISCA.

Ingrid Mora-Plaza is Ph.D. candidate and Lecturer in Phonetics and Phonology at the University of Barcelona, Spain. She is member of the Grup de Recerca en Adquisició de Llengües (GRAL) and L2 Speech Research groups and investigates the effectiveness of L2 phonetic training and inter-individual differences in proficiency, working memory, and auditory attention control. She is also interested in applying task-based principles to the teaching of L2 pronunciation.

Mireia Ortega is Lecturer in the Department of Modern Languages and Literatures at the University of Barcelona, Spain, where she obtained a Ph.D. in English Applied Linguistics. Her research interests include L2 pronunciation and teaching, phonetic training methods, crosslinguistic influence, and learners' individual differences. She is member of the Grup de Recerca en Adquisició de Llengües (GRAL) and L2 Speech Research groups.

Joan C. Mora is Associate Professor in the Department of Modern Languages and Literatures and English Studies at the University of Barcelona, Spain. His research has examined the acquisition of L2 phonology and the role of contextual and individual factors in the development of L2 speech and oral fluency.