The relationship between perception and production of Brazilian Portuguese vowels in European Spanish monolinguals

Jaydene Elvin[1, 2], Daniel Williams[1, 2, 3] and Paola Escudero[1]

[1] The MARCS Institute, Western Sydney University, Australia
[2] ARC Centre of Excellence for the Dynamics of Language, Canberra, ACT, Australia
[3] Linguistics Department, Area of Excellence – Cognitive Sciences, University of Potsdam, Germany
j.elvin@westernsydney.edu.au; daniel.williams@uni-potsdam.de; paola.escudero@westernsydney.edu.au

ABSTRACT: This study investigates the relationship between non-native perception and production of Brazilian Portuguese (BP) vowels by eight European Spanish monolinguals. Participants' accuracy on a non-native discrimination task was used to predict performance in non-native production. We also investigated the acoustic similarity between participants' non-native and native vowel productions. The findings indicate a perception-production link: non-native vowels which were perceptually difficult to discriminate were also less stable and more varied when produced. These findings support the claim of the Second Language Linguistic Perception model (Escudero, 2005; van Leussen & Escudero, 2015) that learners initially perceive and produce the sounds of a second language according to the acoustic properties of the sounds produced in their native language.

Keywords: non-native perception; non-native production; interrelation.

RESUMEN: *Relación entre la percepción y la producción de las vocales del portugués brasileño por parte de monolingües de español europeo*. – Este estudio examina la relación entre la percepción y la producción de las vocales del portugués brasileño por ocho hablantes monolingües de español europeo. Los resultados de la estrategia de discriminación de las vocales por parte de los sujetos del estudio se utilizaron para predecir los de la estrategia de producción por parte de hablantes no nativos de portugués. También investigamos las similitudes acústicas entre la producción de las vocales de la lengua propia de los participantes (el español) y las del portugués. Los resultados indican que existe una relación entre la percepción y la producción de las vocales; en concreto, las vocales que presentaron dificultades de percepción se produjeron de manera menos estable y con más variabilidad en la segunda lengua (el portugués, en este caso). Los resultados se corresponden con el modelo propuesto por la teoría de la Second Language Linguistic Perception (Escudero, 2005; van Leussen & Escudero, 2015) que sostiene que, en los estados iniciales de la adquisición, un individuo percibe y produce los sonidos del segundo idioma con las propiedades acústicas de los sonidos del idioma propio o primera lengua.

Palabras clave: percepción de vocales no nativas; producción de vocales no nativas; interrelación entre ambas.

1. INTRODUCTION

One of the ultimate goals for many second language (L2) learners is to be able to perceive the L2 in a native-like manner and to produce speech without a discernible foreign accent. However, adult learners rarely achieve this goal (Rallo Fabra & Romero, 2012), as there are a number of factors that contribute to the degree of foreign-accented speech. The age at which one begins to learn the L2 is a commonly investigated factor that has been shown to contribute to the degree of foreign-accented speech. In fact, it is said that a "critical period" exists for one to acquire the sounds of a new language and when this period has passed, the ability to attain native-like pronunciation is lost (Scovel, 1969, 2000; Elliot, 1995; Flege, 1995).

For example, Scovel (1969) provides a review of studies that suggest the possible inability to produce native-like speech after puberty could be related to the onset

of cerebral dominance. Flege (1995) commented on the critical period hypothesis and further suggested that in order for language acquisition to be effective, it must occur before the establishment of the hemispheric specialisation of language functions. This claim is supported by Elliot (1995) who found evidence that listeners' pronunciation accuracy was related to right hemispheric specialisation. That is, these authors found evidence that right-specialised individuals appeared to have better pronunciation, but this finding seems to relate specifically to tasks involving spontaneously produced speech and not necessarily for tasks that required repetition or the reading of words in isolation. Elliot (1995) proposes that different types of hemispheric specialisation may relate to pronunciation accuracy in different types of pronunciation tasks.

Thus, the critical period is generally thought to last through childhood, with the cut-off being puberty and, once this period has passed, language development by adults will be much slower. This claim is supported by many studies that show that learners who acquire their L2 later in life seem to have much stronger foreign accents than L2 speakers who have learned their L2 in childhood (e.g., Flege, Munro, & Mackay, 1995; Flege, 1991, 1995; Johnson & Newport, 1991; Piske, MacKay, & Flege, 2001). However, the "critical period" hypothesis has received some counterevidence, as some studies have shown that some L2 learners are able to speak the L2 without any detectable accent (e.g., Bongaerts, Van Summeren, Planken, & Schils, 1997; Moyer, 1999).

Language experience is also thought to influence foreign-accented speech and is often reflected in the individual's length of residence in an L2-speaking region. It is thought that if the learner has had more experience with the target language or has spent more time in the L2 environment, their pronunciation of the L2 is likely to be more native-like. Indeed, some studies have shown an association between amount of language experience and degree of foreign accent (e.g., Baker & Trofimovich, 2006; Bohn & Flege, 1990; Flege, 1995).

However, in empirical terms, results are not always consistent or straightforward. For example, Baker & Trofimovich (2006) found that the relationship between perception and production differed depending on the participant's length of residence in the L2 environment. In particular, the authors suggest that perception and production may be aligned at the initial and advanced stages of learning, but in the intermediate stages there is a misalignment between the two abilities. They further suggest that this misalignment could be related to the variations in the learners' amount and type of L2 experience.

Their findings also supported the aforementioned claim that the degree of foreign accent depends on age of acquisition because the authors found that learners who were exposed to the L2 in childhood were the only learners to perceive and produce the L2 sounds with native-like accuracy. However, as with the research on age of acquisition, not all studies have found the same association between language experience and the degree of foreign-accented speech (e.g., Flege, 1988; Moyer, 1999).

Piske et al. (2001) suggest that these discrepancies could be related to the fact that length of residence only provides a rough index of overall L2 experience and a more longitudinal design may be necessary.

Another contributing factor to the degree of foreign-accented speech is language use. That is, learners who speak their L1 frequently are likely to have a much stronger foreign accent than those who use it infrequently. For example, Flege, Frieda and Nozawa (1997) found that Italian speakers who continued to frequently use their L1 spoke English with a stronger foreign accent than those who rarely spoke their L1. But as with findings relating to language experience and age of acquisition, not all have found a similar effect of language use, as some studies (e.g., Elliott, 1995) found little to no effect of language use.

Furthermore, factors such as motivation (e.g., professional or social desire to produce the L2 correctly) and language learning aptitude are also thought to contribute to the degree of foreign-accented speech. For example, Smit (2002) found that EFL students' pronunciation achievements were positively influenced by motivational factors and Elliot (1995) also found that an individual's attitude or concern toward their own pronunciation abilities also influenced their accuracy in pronunciation. However, in both studies the authors acknowledge that other factors also contributed to the learners' L2 pronunciation abilities. This supports the claim by Piske et al. (2001) that, although these studies have shown some effect of motivation, they have not been able to provide evidence that these factors automatically lead to accent-free L2 speech.

With respect to language learning aptitude, Piske et al. (2001) also state that an aptitude for language learning as a result of musical ability or the ability to mimic has not yet been identified as a significant and independent predictor of the degree of foreign-accented speech. Piske et al. (2001) also indicate the need for future studies that investigate whether language learning aptitude is something that one is born with or develops as a result of other factors which have not yet been identified.

Most important to the present study is the fact that the strength or degree of a foreign accent is indeed influenced by the learner's L1 and, in particular, the learner's ability to perceive sounds in the L2. Models of speech perception, such as the Speech Learning Model (SLM, Flege, 1995), the Perceptual Assimilation Model (PAM, Best, 1994, 1995) and its extension to PAM-L2 (Best & Tyler, 2007), and the Second Language Linguistic Perception model (L2LP, Escudero, 2005; van Leussen & Escudero, 2015) share the common assumption that listeners filter and categorise the sounds of the L2 according to the existing categories in their own native language. As a result, pronunciation problems in the L2 are thought to be a result of an individual's difficulty to distinguish between L2 sound contrasts as influenced by the L1.

Indeed, the aforementioned models of speech perception each account for the link between perception and production. For example, the SLM was developed to ac-

count for the limitations of a learner's ability to perceive and produce native-like sounds due to experience and age related limitations. According to the SLM, a learner's ability to produce native-like sounds is largely dependent on how these sounds are perceived in relation to their L1. Although there is no explicit account for the link between perception and production in PAM and its extension to PAM-L2, the model posits that learners are able to detect articulatory information in the speech they perceive and therefore it assumes that a common articulatory metric is shared between perception and production.

The L2LP model proposes a direct link between perception and production as it states that at the initial state of learning, an individual's perception of L2 sounds should closely match the acoustic properties of the sounds as they are produced in the learner's native language. Thus, if a learner perceives L2 sounds as instances of their own L1 sounds, they should also produce those L2 sounds using acoustic properties similar to their L1 sounds.

There are a number of studies that support these theoretical claims and have identified a link between perception and production (e.g., Flege, Bohn, & Jang, 1997; Levy & Law, 2010; Morrison, 2003, 2006; Rallo Fabra & Romero, 2012; Rauber, Escudero, Bion, & Baptista, 2005). In particular, these studies suggest that a learner's perception of the L2 influences their production of the L2 (Levy & Law, 2010; Llisterri, 1995; Morrison, 2003). However, the relation between the two abilities is rather complex and still a matter of debate. This seems due to the fact that there is not a well-established link in terms of empirical evidence (Levy & Law, 2010). In fact, some studies have found an opposite pattern where production precedes perception. An example of this is documented in Sheldon & Strange (1982), where they found that some Japanese learners of English were more accurate in producing the English /r/-/l/ contrast even though they had lower accuracy scores when perceiving the contrast.

These inconclusive results may be related to the problematic nature of the methodology used to investigate the interrelation between the two abilities, as pointed out by Levy & Law (2010). These authors posit that part of the difficulty in measuring the relationship between perception and production is the fact that the task demands are different in perception and production studies and there are different techniques employed to assess these abilities. The authors further posit that in order to reliably assess the relationship between these abilities, analyses should be based on individual performance using the same participants in both tasks and ensuring that the methodology in each task is comparable and controlled.

The present study aims to assess the perception-production relationship in the acquisition of Brazilian Portuguese (BP) vowels by European Spanish (ES) individuals at the initial stage of L2 learning. To control for the aforementioned methodological concerns in the investigation of the interrelation between the two abilities, we used the same stimuli for both tasks and the perception and production data were collected from the same participants. We further collected the participants' own native production data to investigate the acoustic similarity between their own native vowel categories and their non-native production of BP.

The L2LP theoretical framework is applicable to the present study as the framework was developed to directly account for the perception-production link at the onset or initial stage of learning. That is, according to the model's *optimal perception hypothesis,* all native listeners are equipped with a perception grammar, that is, a system that allows them to map acoustic information in speech onto phonological representations (Colantoni, Steele, & Escudero, 2015; Escudero & Boersma, 2004). The L2LP model further proposes that L2 learners have separate perception grammars for their L1 and L2 and that at the initial stages of learning, the L2 perception grammar will be a copy of their L1 perception grammar. As learning takes place, learners will update and reorganise their L2 perception grammar to more closely match that of the target language.

According to the L2LP model, the link between perception and production is one where learners should initially perceive and produce the sounds of the new language in the same way that they would perceive and produce sounds in their own native language. Importantly, perception must be in place before L2 production can develop because learners who initially fail to detect a L2 contrast in speech perception will also fail to reliably produce acoustically distinct sounds for that L2 contrast in speech production. Also, while the initial ability to detect a L2 contrast indicates a contrast will be produced in the L2, it does not guarantee the contrast will be produced in a native-like manner because learners' perception grammars may not yet be geared towards the specific L2 categories. Thus, learners will only start to accurately produce a L2 contrast once their L2 perception grammars have been updated to account for the contrast as it is actually produced in the L2.

Following the L2LP model, predictions for non-native or initial L2 perception and production can be made through a comprehensive investigation of the acoustic similarity between native vowel categories and those in the specific target language variety. If our results confirm these predictions, we would expect that the ES participants will have BP vowel productions that are acoustically more similar to the properties of the closest vowel(s) in Spanish than to the target vowels produced by native BP speakers. Additionally, we would expect that participants should be able to produce two separate vowels for BP vowel contrasts that are perceptually easy to discriminate, but not for perceptually difficult BP contrasts.

2. METHOD

2.1. Participants

The present study reports a subset of 8 Spanish functional monolinguals (4 male) selected from a larger group reported in Elvin (2016). Participants were all born and

raised in Madrid, Spain, and aged between 18 and 30. The participants reported little to intermediate knowledge of English, but did not use it in their daily lives. They had little to no knowledge of any other foreign language, including the target language, namely Portuguese. They were recruited from universities around the Universidad Nacional de Educación a Distancia (UNED) and received €30 for their time. All participants provided informed consent in accordance with the ethical protocols in place at the UNED.

2.2. Stimuli and procedure

At the beginning of each session, participants completed a native production task in order to assess the acoustic similarity between their own native vowel production and their non-native vowel production. In this task, participants read pseudo-words in the /fVfo/ context containing one of the five Spanish vowels, namely, /i e a o u/. Each vowel was repeated 10 times per vowel for a total of 50 tokens. Participants then completed the non-native perception and production tasks which were counterbalanced.

The auditory stimuli for the non-native discrimination and repetition tasks consisted of naturally produced BP pseudo-words in the /fVfe/ context and selected from the Escudero, Boersma, Rauber, & Bion (2009) corpus. These BP pseudo-words were produced in isolation by five male and five female monolingual Brazilian Portuguese speakers from São Paolo. The vowel in the first syllable was always stressed and corresponded to one of the seven oral BP vowels, namely /i e ɛ a ɔ o u/.

The non-native perception task reported in Elvin (2016) consisted of an auditory two-alternate forced choice task in the XAB format run on a laptop using the E-Prime 2.0 program. In this task, participants listened to the three words using headphones and were required to make a decision as to whether the first word they heard sounded more like the second or the third. That is, three stimulus items were presented per trial. The second (A) and third (B) items were always from different BP vowel categories and the first item (X) was the target item for which a matching decision was required. On each trial, X was always one of the 70 target BP words and the A and B stimuli were always the 7th male and 7th female speaker from the Escudero et al. (2009) corpus to avoid any confusion of overlapping target and response categories. Furthermore, the order of the A and B responses was counterbalanced (namely, XAB and XBA). Each trial consisted of one of the six BP contrasts, namely /a/-/ɔ/, /a/-/ɛ/, /i/-/e/, /o/-/u/, /e/-/ɛ/ and /o/-/ɔ/, with a total of 40 trials per contrast.

To elicit non-native vowel production data, a non-native repetition task was administered to the participants. Although reading tasks are common in L2 production studies (e.g., Flege, 1987; Flege, Bohn et al., 1997; Flege, Mackay, & Meador, 1999; Morrison, 2003, 2006), the learners in these studies all had some level of experience with the L2. We instead used the non-native repetition

task as it was the most appropriate task for monolingual speakers, who unlike the aforementioned studies, had no experience with BP. In this task, participants were instructed to immediately repeat the word that they had heard into a headset microphone. There was a total of 70 /fVfe/ target words (7 vowels x 10 speakers), as well as three additional nonsense words by each speaker (/pipe/, /kuke/ and /sase/), included as filler items. Thus, in the non-native production task we had a total of 100 BP word tokens (70 target and 30 fillers).

2.3. Data analysis

We first segmented the native and non-native vowel tokens in the target words using WebMaus (Kisler, Schiel, & Sloetjes, 2012). This is an online tool used to automatically segment and label speech sounds. To ensure the accuracy of the automatically generated start and end boundaries, they were all manually checked and adjusted. Formant measurements for each vowel token were extracted at three time points (25 %, 50 % and 75 %) following the optimal ceiling method reported in Escudero et al. (2009) to ensure comparability across both the target and native language. In the optimal ceiling method, for every vowel, per speaker, the "optimal ceiling" is chosen as the one that yields the least amount of variation for both the first (F1) and second (F2) formant values within the set number of annotated tokens for the vowel. Formant ceilings ranged between 4500 Hz and 6500 Hz for females, and between 4000 Hz and 6000 Hz for males.
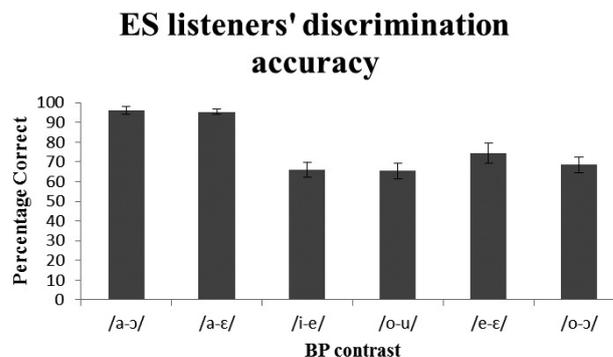
3. RESULTS

3.1. Non-native discrimination

Figure 1 shows the averaged accuracy scores across the six BP contrasts for the subset of eight ES participants previously reported in Elvin (2016).

The results indicate that on average listeners performed close to ceiling on the BP contrasts /a/-/ɔ/ and /a/-/ɛ/. However, their discrimination accuracy for the re-

Figure 1: Discrimination accuracy across the six BP contrasts.

ES listeners' discrimination accuracy

maining four contrasts was much lower. In particular, they had overall lower accuracy for BP /i/-/e/ and /o/-/u/. If there is a link between perception and production, where perception precedes production, we would predict that participants should be able to produce two separate vowel categories for the contrasts with higher discrimination accuracy. However, for the BP contrasts with overall lower accuracy, we would expect participants to produce each vowel in these categories with similar acoustic properties or, in other words, as one vowel category.

3.2. Non-native production

As previously mentioned, the L2LP model suggests that at the initial state of learning, non-native production should be acoustically more similar to the learner's production of their own native vowels than the target vowel categories. Figures 2 and 3 show the mean F1 and F2 values of the ES male and females' non-native production of the seven BP vowels, together with the mean F1 and F2 values of their own native vowel productions and the target BP vowels.

3.2.1. Acoustic similarity between non-native and native vowel production

Visual inspection of both vowel plots indicates that on average, the ES participants produced non-native vowel categories that fall between their own L1 and the target L2 vowel categories. Interestingly, there are some cases (e.g., the ES females' production of BP /a/, /ɛ/ and /ɔ/) where the averaged F1 and F2 values were similar to that of the averaged target BP values. Importantly, many of the averaged non-native male and female BP vowel pro-

Figure 2: The average F1 and F2 values for the ES males own native vowels (black) and their non-native production of the BP vowels (grey), as well as the target BP vowels reported in Escudero et al. (2009) (black, with circles).

Figure 3: The average F1 and F2 values for the ES females own native vowels (black) and their non-native production of the BP vowels (grey), as well as the target BP vowels reported in Escudero et al. (2009) (black, with circles).
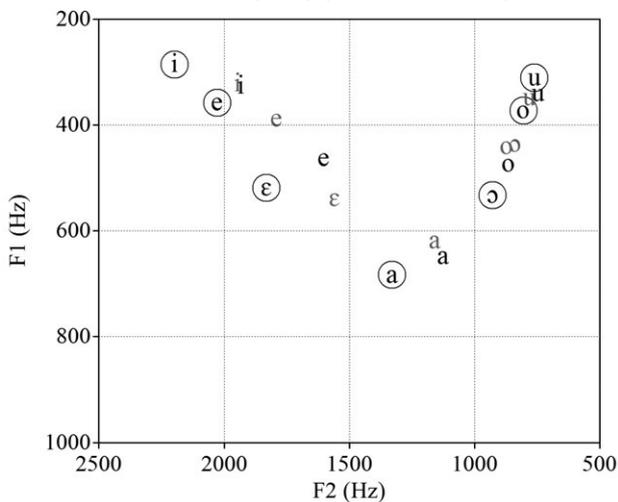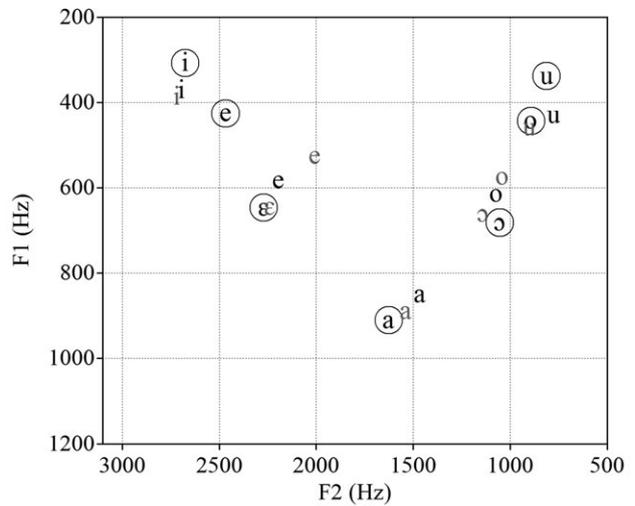

ductions appear to be acoustically closer to their own native ES vowels than the target BP vowels. This finding is indeed in line with the L2LP model's claim that learners will initially perceive and produce the L2 in the same way that they perceive and produce vowels in their own native language.

To confirm our visual inspection of the location of the vowels (target, native and non-native) in the F1 and F2 vowel space, we calculated the average Euclidean Distances (ED) between vowels (in Hz) as a quantitative measure of cross-linguistic acoustic similarity where smaller values indicate greater degree of similarity. Table 1 shows the average Euclidean Distance between the participants' non-native production of the seven target BP vowels (males and females separated) and the first and second acoustically closest native ES vowel categories. The table also shows the Euclidean Distance between the participants' non-native vowel productions and the target BP vowels reported in Escudero et al. (2009). Given the fact that males and females differ in their vowel formant frequencies, we present male and female data separately.

The Euclidean Distances reported in Table 1 seem to suggest that in most cases, at the initial state of learning, ES monolinguals' production of BP differs from that of native BP speakers. In fact, it appears that the ES speakers produce the non-native BP vowels as more acoustically similar to their own native vowel categories. For instance, the ES participants' non-native productions of /i/ were acoustically closer to their native /i/ vowel than the target BP vowel. This was also the case for BP /o/ and /ɔ/. Furthermore, the male non-native productions of BP /ɛ/ and /a/ were acoustically more similar to their own native /e/ and /a/ vowel categories than the target BP vowels, whereas the female non-native productions of BP /ɛ/ and /a/ were indeed acoustically closer to the target vowels, as observed in Figure 3.

Table 1: The Euclidean Distance (in Hz) between the participants' averaged F1 and F2 non-native vowel tokens and the first and second closest native vowel category as well as the first and second acoustically closest target BP vowels.

	Acoustic Similarity	i	e	ɛ	a	ɔ	o	u
Male	ES	[i] – 14 [e] – 372	[i] – 154 [e] – 203	[e] – 87 [i] – 429	[a] – 44 [o] – 329	[o] – 47 [a] – 132	[o] – 34 [u] – 161	[u] – 37 [o] – 152
	BP	[e] – 90 [ɛ] – 228	[ɛ] – 137 [e] – 239	[a] – 272 [ɛ] – 273	[a] – 182 [ɔ] – 248	[o] – 71 [ɔ] – 132	[o] – 95 [ɔ] – 108	[o] – 34 [u] – 42
Female	ES	[i] – 20 [e] – 553	[i] – 176 [e] – 374	[e] – 65 [i] – 513	[a] – 92 [o] – 547	[o] – 79 [a] – 384	[o] – 119 [u] – 237	[u] – 138 [o] – 242
	BP	[i] – 90 [e] – 254	[ɛ] – 289 [e] – 470	[ɛ] – 32 [e] – 317	[a] – 90 [ɔ] – 527	[ɔ] – 91 [o] – 331	[ɔ] – 108 [o] – 197	[o] – 18 [u] – 150

Additionally, there are a number of cases where it appears that the non-native vowel falls between the closest native and the closest BP vowels. For example, the ES male speakers' production of BP /e/ seems to fall between their native /i/ vowel (ED = 154 Hz) and BP /ɛ/ (ED = 137 Hz). A similar case is observed for the ES females' non-native production of BP /a/ which falls between the target BP /a/ vowel (ED = 90 Hz) and their own native ES /a/ vowel (ED = 92 Hz).

As observed in Figure 3, the female ES /e/ vowel is acoustically close to BP /ɛ/ and it is therefore unsurprising that the ES females produce BP /ɛ/ with similar acoustic properties as the target /ɛ/ vowel (ED = 32 Hz) as well as their own native /e/ (ED = 65 Hz). Similarly, in Figure 2, it is evident that native male BP speakers produce the BP vowels /o/ and /u/ very close together and the ES male /u/ vowel falls in between these two vowels. The production results indicate that ES males do in fact produce BP /u/ with acoustic properties similar to both the target BP /o/ and /u/ (ED = 34 Hz for the former and ED = 42 Hz for the latter) as well as their own native ES /u/ (ED = 37 Hz).

3.2.2. Modelling BP listeners' perception of foreign-accented vowels

In order to determine the expected intelligibility of the ES speakers' non-native production of BP, we ran a cross-language discriminant analysis. Previous studies (e.g., Elvin & Escudero, 2015; Escudero & Vasiliev, 2011; Gilichinskaya & Strange, 2010) have successfully used cross-language discriminant analyses as a means of determining acoustic similarity between the native and target language and to predict real listeners' vowel categorisation patterns. Although it would be ideal to have native BP speakers rate the ES speakers' non-native tokens, the discriminant analysis should provide a good model of how these vowels would be perceived and categorised by native BP listeners in the absence of such data. We trained the model on the target BP vowel tokens and then tested it on the participants' non-native vowel productions, using F1, F2 and F3 (in Bark) as input parameters. The model yielded 91.4 % (males) and 97.1 % (females) correct classification for the trained BP vowels and 54.7 % (males)

and 51.2 % (females) for the non-native vowel tokens. This suggests that, overall, the ES speakers' production of BP would likely be misidentified by native BP listeners as only half of the non-native BP vowel productions were produced with acoustic properties that were similar to the target vowels.

Table 2 shows the percentage of times that each non-native vowel was correctly classified as the target BP token. The vowels that have the highest percentage of correct categorisation are those which are likely to be perceived as the intended vowel. On the other hand, those vowels with a low percentage of correct classifications, but a large percentage of incorrect categorisation to another vowel, are those that the ES speakers fail to produce as separate categories and those which are likely to be misidentified by native BP speakers due to pronunciation errors as a result of their foreign accent.

For instance, it is likely that native BP listeners would perceive the non-native productions of BP /ɛ/ as the intended vowel for the female speakers, but they may have more difficulties correctly identifying the intended vowel for male speakers. As observed in Table 2, the female tokens were correctly categorised 94.7 % of the time, whereas the male tokens were correctly categorised 74.2 % of the time. In the case of BP /ɔ/, the male tokens were categorised better than the female tokens (94.7 % vs. 63.9 %) and therefore, native BP listeners are more likely to accurately perceive the intended vowel produced by male ES speakers rather than female ES speakers.

The intelligibility of the ES speakers' production of BP /a/ may be more inconsistent due to the fact that 77.1 % (female) and 64.7 % (male) of the non-native tokens were correctly categorised but the remaining 22.9 % (female) and 35.3 % (male) of the tokens were incorrectly classified as BP /ɔ/. BP /e/ and /o/ are also less likely to be perceived as the intended vowels. This is because only 37.9 % (female) and 56.3 % (male) of the non-native BP /e/ tokens were correctly classified, with 44.7 % of the non-native female and 40.6 % of the non-native male productions incorrectly classified as BP /ɛ/. In the case of BP /o/, 40 % of the female and 34.3 % of the male non-native tokens were correctly classified with the remaining tokens incorrectly classified as BP /ɔ/.

Table 2: The percentage of non-native male and female vowel tokens classified as the intended BP vowel category. Percentages are rounded to the nearest whole number and the highest classification percentage appears in bold.
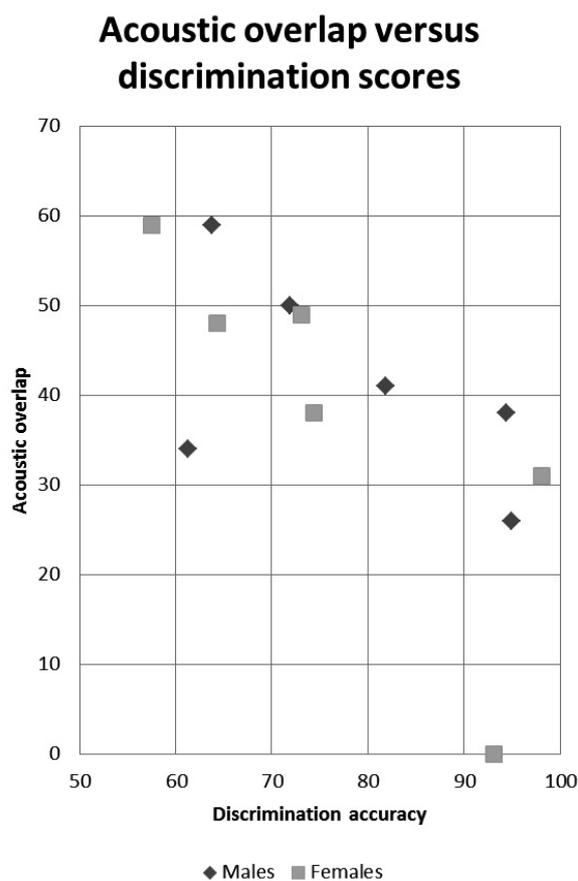
Non-native	M/F	BP vowel category						
		i	e	ɛ	a	o	ɔ	u
i	M	20	**80**					
	F	29	**71**					
e	M	3	**56**	41				
	F		38	**45**	17			
ɛ	M			**74**	10		16	3
	F		3	**95**				
a	M				65		35	
	F				77		23	
ɔ	M				3	3	**95**	
	F			19	8	8	64	
o	M			3		34	**47**	
	F					40	**51**	9
u	M					59		43
	F					83		18

Finally, the non-native BP vowels that would be the least intelligible and which are very likely to be misidentified would be BP /i/ and /u/. As reported in Table 2, only 28.9 % of the female and 20 % of the male BP /i/ tokens were correctly classified as the intended target vowel. Instead, the majority of the BP /i/ tokens were incorrectly classified as BP /e/ (71.1 % of the time for females and 80 % of the time for males). It is likely that BP listeners would perceive the ES females' production of BP /u/ as less accurate than the ES males. That is, 82.5 % of the female tokens were incorrectly classified as BP /o/ with only 17.5 % correctly classified. In the case of the BP males only 42.6 % of the male tokens were correctly classified, whereas 59 % of the tokens were incorrectly classified as BP /o/.

3.2.3. *Relationship between non-native production and perception*

Here we test the hypothesis that there is a monotonic relationship between non-native production accuracy and non-native discrimination. That is, members of contrasts which are easy to discriminate are produced as vowels which are acoustically distinct. Conversely, members of contrasts which are difficult to discriminate are produced as vowels which overlap acoustically. To this end, we calculated acoustic overlap scores from the data in Table 2. For each contrast and gender, the acoustic overlap score is the smaller percentage of classifications when two members of a BP contrast produced by the ES speakers were classified as the same BP vowel (cf. Levy, 2009). For example, the acoustic overlap score of /o-ɔ/ for male

Figure 4: Acoustic overlap scores plotted against discrimination accuracy scores for the six BP contrasts and two genders.

Acoustic overlap versus discrimination scores

ES speakers is 50 %: 34 % of /o/ tokens and 3 % of /ɔ/ tokens were classified as /o/, whereas 47 % of /o/ tokens and 95 % of /ɔ/ tokens were classified as /ɔ/; summing the smaller percentages when both members were classified as the same BP vowel, i.e., 3 % + 47 %, gives 50 %.

Figure 4 plots the acoustic overlap and discrimination accuracy scores for the six contrasts and two genders. Visual inspection does indeed show a relatively strong trend for higher acoustic overlap scores to be associated with lower discrimination accuracy scores, which is confirmed by a Spearman's rank order correlation ($\rho = -0.71$, p = 0.01).

4. DISCUSSION

The present study investigated the interrelation between European Spanish monolinguals' perception and production of Brazilian Portuguese vowels at the initial stage of learning. Testing the interrelation between these two abilities is generally considered problematic due to methodological reasons such as different task demands and not using the same participants across the two tasks. However, we controlled for this by testing the same par-

ticipants on a non-native discrimination task and a non-native repetition task and by using the same stimuli across both tasks. We compared our listeners' non-native production against their own native vowel production.

Our findings indicate an interrelation between non-native perception and production at the initial state of learning. It shall be reminded that the L2LP model posits that individuals at the initial state of learning perceive non-native vowels resourcing to the acoustic properties of their native sound categories and produce non-native vowels using such L1 acoustic properties. Our findings lend support to this claim as the ES participants' non-native vowel productions seem to be heavily influenced by their L1. In particular, less than half of the non-native tokens were produced with similar acoustic properties as the target BP vowels. Furthermore, as observed in Table 1, many of the non-native vowels were acoustically closer to native vowel categories than the target categories.

The L2LP model further states that perception must be in place before development in other speech abilities (e.g., recognition, production) can occur. Our findings provide evidence that perception indeed precedes production as individuals seem to perceive a difference between two vowels before they can produce them as separate categories, while non-native vowels that were difficult to discriminate are produced with properties of a single native vowel. Even with a small sample size, our correlation analysis yielded a strong trend for lower discrimination scores to be associated with a greater amount of acoustic overlap, i.e., both vowels in the contrast being produced as the same vowel(s). Specifically, the BP vowels with the greatest amount of incorrect classifications were BP /i/ and /u/, which were instead predominately produced with similar acoustic properties to BP /e/ and /o/ respectively. It is not surprising that the vowels with the lowest categorisation scores, which would be perceived as heavily accented, are the vowels that correspond to BP vowel contrasts that were perceptually difficult to discriminate, specifically BP /i/-/e/ and /o/-/u/.

The fact that these "heavily accented" vowels were consistently categorised as the other vowel in that contrast indicates that the ES speakers are unable to produce two separate vowel categories for vowel contrasts that they cannot discriminate. On the other hand, the findings from the present study seem to indicate that contrasts that are perceptually easy to discriminate can indeed be produced as two separate vowel categories. Despite the variation in the categorisation of the BP /a/ vowel, the cross-language discriminant analysis model indicated that native listeners are likely to perceive both BP /a/ and /ɛ/ as two separate native vowels.

Interestingly, despite the fact that BP /a/-/ɔ/ was discriminated extremely well, the ES participants still produced these two vowels with some degree of acoustic overlap. That is, a smaller percentage of their non-native BP /a/ tokens were also classified as BP /ɔ/. This finding suggests that even though participants are able to discern a difference between the two vowels in perception, they may not be able to update their production during a short

task, based on their modified perception (Levy & Law, 2010) or further development may be required before they can accurately produce these vowels. As proposed by the L2LP model, learners will need to adjust their category boundaries in perception to then produce similar target language vowels with more native-like acoustic values (Escudero, 2005, 2009).

This finding may also be related to task differences. In the non-native discrimination task the participant is required to make a choice as to which of the two sounds a target sound is more similar, whereas in production they are simply required to repeat the word that they heard. Thus, it may be that there are additional cues that the listener is able to rely on when discerning between two sounds that do not transfer to when only one of these sounds is heard. For instance, some studies suggest that there are allophonic variants of the Spanish mid vowels /e/ and /o/ (Morrison, 2004; Navarro Tomás, 1918). Therefore, it could be that ES speakers use this knowledge in perceptual discrimination as the presentation of several vowels makes them more sensitive to the acoustic cues. However, if the appropriate phonological environments that trigger these allophones in Spanish do not occur in the BP words, their production of these vowels may not be as good. For this reason a Spanish learner of BP would need to learn over time that some of their allophonic variations could be applicable to their production of BP vowels.

It is also interesting to note that we did find that there were some differences between the male and female listeners in the accuracy of the categorisation of the intended non-native vowels. It could be that these gender differences are related to the cross-linguistic acoustic similarity between the native and target language. For example, given the close acoustic similarity between the native female ES /e/ vowel and the female BP /ɛ/, it is not surprising that ES female speakers would make use of their native category to produce BP /ɛ/ and that it would often be categorised as the intended vowel. However, this was not always the case, as the ES males' production of BP /ɔ/ was more often correctly categorised as the intended vowel than the ES female productions.

Male and females have also been shown to differ in their use of acoustic cues when perceiving non-native sounds. Specifically, Wanrooij, Escudero & Raijmakers (2013) found that the males in the "high performers" group (those who are able to use F1 and F2 to perceive the Dutch /ɑ/-/aː/ contrast before training) were more likely to start using F3 after training than females. In both the present study and Wanrooij et al. (2013), males and females only differed in some instances, therefore it would be worthwhile to further explore the meaning of these gender differences in future perception and production studies.

Finally, it is important to note that all of the vowels in the present study were produced in an immediate repetition task and there may have been some repetition effects that influenced our results. For example, it may be that some participants were simply imitating the sounds they heard rather than using the appropriate phonetic representation they have formed for that particular sound. Evidence from

the speech shadowing literature (e.g., Marslen-Wilson, 1985) suggests that in close shadowing, individuals may use the products of on-line speech analysis to drive their articulatory apparatus before they are fully aware of what these products are. We are currently investigating whether or not the ES speakers' production of BP significantly differs when the vowels are produced in an immediate or delayed repetition task. By investigating the data from the delayed repetition task, we will be able to confirm whether or not our speakers' immediate repetitions of the vowels are in fact representative of the categories they have formed for these particular sounds.

In sum, our findings seem to suggest that non-native perception is related to and may precede non-native production. The findings are also in line with the L2LP model because our participants who had no experience with Brazilian Portuguese did indeed produce BP vowels with acoustic properties that were more similar to their own native vowel categories rather than the target vowels. However, we do acknowledge that our cross-linguistic discriminant analysis only shows the acoustic similarity between non-native and target vowels and can therefore only provide a rough idea of how these vowels would be perceived by native BP listeners. Thus, for a more accurate indication of how native-like these vowel productions are, it would be beneficial to have native Brazilian Portuguese categorise and provide a goodness rating for these non-native vowel tokens. Furthermore, as previously mentioned, these vowels were produced in an immediate repetition task and further investigation is needed to determine whether or not their vowel production would differ if the responses were delayed.

ACKNOWLEDGEMENTS

REFERENCES

Baker, W., & Trofimovich, P. (2006). Perceptual paths to accurate production of L2 vowels : The role of individual differences. *IRAL, International Review of Applied Linguistics in Language Teaching*, *44*(3), 231–250. http://doi.org/10.1515/IRAL.2006.010

Best, C. T. (1994). The Emergence of Native-Language Phonological Influences in Infants : A Perceptual Assimilation Model. In J. C. Goodman & H. C. Nusbaum (Eds.), *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words* (pp. 167–224). Cambridge, MA: MIT Press.

Best, C. T. (1995). A direct realist perspective on cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171–204). Timonium, MD: York Press.

Best, C. T., & Tyler, M. D. (2007). Non-native and second-language speech perception: commonalities and complementarities. In O. Bohn & M. J. Munro (Eds.), *Language Experience in Second-Language Speech Learning: In Honor of James Emil Flege* (pp. 13–34). Amsterdam: John Benjamins.

Bohn, O., & Flege, J. E. (1990). Interlingual identification and the role of foreign language experience in L2 vowel perception, *Applied Psycholinguistics*, *11*, 303–328. http://doi.org/10.1017/s0142716400008912

Bongaerts, T., Van Summeren, C., Planken, B., & Schils, E. (1997). Age and ultimate attainment in the pronunciation of a foreign language. *Studies in Second Language Acquisition*, *19*(4), 447–465. http://doi.org/10.1017/s0272263197004026

Colantoni, L., Steele, J., & Escudero, P. (2015). *Second Language Speech*. Cambridge: Cambridge University Press.

Elliott, A. R. (1995). Foreign Language Phonology : Field Independence, Attitude, and the Success of Formal Instruction in Spanish Pronunciation. *The Modern Language Journal*, *79*, 356–371. http://doi.org/10.1111/j.15404781.1995.tb05456.x

Elvin, J. (2016). *The role of the native language in non-native perception and spoken word recognition: English vs. Spanish learners of Portuguese*. PhD Dissertation, Western Sydney University.

Elvin, J., & Escudero, P. (2015). Predicting discrimination accuracy through cross-linguistic acoustic analyses. In The Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK: The University of Glasgow.

Escudero, P. (2005). *Linguistic Perception and Second Language Acquisition*. PhD Dissertation, Utrecht University.

Escudero, P. (2009). The linguistic perception of similar L2 sounds. In P. Boersma & S. Hamann (Eds.) *Phonology in Perception* (pp. 152–190). Berlin-New York: Mouton de Gruyter.

Escudero, P., & Boersma, P. (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition*, *26*(4), 551–585. http://doi.org/10.1017/s0272263104040021

Escudero, P., Boersma, P., Rauber, A. S., & Bion, R. H. (2009). A cross-dialect acoustic description of vowels: Brazilian and European Portuguese. *The Journal of the Acoustical Society of America*, *126*(3), 1379–1393. http://doi.org/10.1121/1.3180321

Escudero, P., & Vasiliev, P. (2011). Cross-language acoustic similarity predicts perceptual assimilation of Canadian English and Canadian French vowels. *The Journal of the Acoustical Society of America*, *130*(5), EL277–EL283. http://dx.doi.org/10.1121/1.3632043

Flege, J. E. (1987). The production of "new" and "similar" phones in a foreign language: evidence for the effect of equivalence classification. *Journal of Phonetics*, *15*(1), 47–65.

Flege, J. E. (1988). Factors affecting degree of perceived foreign accent in English sentences. *The Journal of the Acoustical Society of America*, *84*(1), 70. http://dx.doi.org/10.1121/1.396876

Flege, J. E. (1991). Age of learning affects the authenticity of voice-onset time (VOT) in stop consonants produced in a second language. *The Journal of the Acoustical Society of America*, *89*(1), 395–411.

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. 233–276). Timonium, MD: York Press.

Flege, J. E., Bohn, O.-S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, *25*(4), 437–470. http://dx.doi.org/10.1006/jpho.1997.0052

Flege, J. E., Frieda, E. M., & Nozawa, T. (1997). Amount of native-language (L1) use affects the pronunciation of an L2. *Journal of Phonetics*, *25*(2), 169–186.

Flege, J. E., Mackay, I. R. A., & Meador, D. (1999). Native Italian speakers ' perception and production of English vowels, *The Journal of the Acoustical Society of America*, *106*(5), 2973–2987.

Flege, J. E., Munro, M. J., & Mackay, I. R. A. (1995). Factors affecting strength of perceived foreign accent in a second language. *The Journal of the Acoustical Society of America*, *97*, 3125.

Gilichinskaya, Y. D., & Strange, W. (2010). Perceptual assimilation of American English vowels by inexperienced Russian listeners. *The Journal of the Acoustical Society of America*, *128*(2), EL80–EL85. http://doi.org/10.1121/1.3462988

Johnson, J. S., & Newport, E. L. (1991). Critical period effects on universal properties of language: the status of subjacency in the acquisition of a second language. *Cognition*, *39*(3), 215–258.

Kisler, T., Schiel, F., & Sloetjes, H. (2012). Signal processing via webservices: the use case WebMAUS. In *Proceedings of Digital Humanities* (pp. 30–34). Hamburg.

Levy, E. S. (2009). On the assimilation-discrimination relationship in American adults' French vowel learning. *The Journal of the Acoustical Society of America*, *126*(5), 2670–2682. https://dx.doi.org/10.1121%2F1.3224715

Levy, E. S., & Law, F. F. (2010). Production of French vowels by American-English learners of French: language experience, consonantal context, and the perception-production relationship. *The Journal of the Acoustical Society of America*, *128*(3), 1290–1305. http://doi.org/10.1121/1.3466879

Llisterri, J. (1995). Relationships between speech production and speech perception in a second language. In *Proceedings of the XIIIth International Congress of Phonetic Sciences* (Vol. 4, pp. 92–99).

Marslen-Wilson, W. D. (1985). Speech shadowing and speech comprehension. *Speech Communication*, *4*(1–3), 55–73.

Morrison, G. S. (2003). Perception and Production of Spanish Vowels by English Speakers. In *Proceedings of the 15th International Congress of Phonetic Sciences* (Vol. 2003, pp. 1533–1536). Barcelona.

Morrison, G. S. (2004). An acoustic and statistical analysis of Spanish mid-vowel allophones. *Estudios de Fonética Experimental*, *13*, 12–37.

Morrison, G. S. (2006). *L1 & L2 Production and Perception of English and Spanish Vowels: A Statistical Modelling Approach*. PhD Dissertation, University of Alberta.

Moyer, A. (1999). Ultimate attainment in L2 phonology. *Studies in Second Language Acquisition*, *21*(1), 81–108. http://doi.org/10.1017/S0272263199001035

Navarro Tomás, T. (1918). *Manual de pronunciación española* [1965 (12th ed.)]. Madrid: CSIC.

Piske, T., MacKay, I. R. A., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2 : a review. *Journal of Phonetics*, *29*(2), 191–215. http://doi.org/10.006/jpho.2001.0134

Rallo Fabra, L., & Romero, J. (2012). Native Catalan learners' perception and production of English vowels. *Journal of Phonetics*, *40*(3), 491–508. http://doi.org/10.1016/j.wocn.2012.01.001

Rauber, A. S., Escudero, P., Bion, R. A. H., & Baptista, B. O. (2005). The Interrelation between the Perception and Production of English Vowels by Native Speakers of Brazilian Portuguese Graduate Program in Applied Linguistics. *INTERSPEECH*, *2*, 2913–2916.

Scovel, T. (1969). Foreign accents, language acquisition, and cerebral dominance. *Language Learning*, *19*(3–4), 245–253.

Scovel, T. (2000). A critical review of the critical period research. *Annual Review of Applied Linguistics, 20*, 213–223. http://doi.org/10.1017/S0267190500200135

Sheldon, A. M. Y., & Strange, W. (1982). The acquisition of /r/ and /l/ by Japanese learners of English : Evidence that speech production can precede speech perception. *Applied Psycholinguistics*, *3*(3), 243–261. https://doi.org/10.1017/S0142716400

Smit, U. (2002). The interaction of motivation and achievement in advanced EFL pronunciation learners. *IRAL, 40*(2), 89–116. http://doi.org/10.1515/iral.2002.009

van Leussen, J.-W., & Escudero, P. (2015). Learning to perceive and recognize a second language: the L2LP model revised. *Frontiers in Psychology*, *6*(August), 1–12. http://doi.org/10.3389/fpsyg.2015.01000

Wanrooij, K., Escudero, P., & Raijmakers, M. E. J. (2013). What do listeners learn from exposure to a vowel distribution ? An analysis of listening strategies in distributional learning. *Journal of Phonetics*, *41*(5), 307–319. http://doi.org/10.1016/j.wocn.2013.03.005