PREDICTING VOWEL DISCRIMINANTION ACCURACY THROUGH CROSS-LINGUISTIC ACOUSTIC ANALYSES

Jaydene Elvin and Paola Escudero

The MARCS Institute, University of Western Sydney
J.Elvin@uws.edu.au; Paola.Escudero@uws.edu.au

ABSTRACT

This study compares the acoustic properties of Australian English and Brazilian Portuguese vowels as a means of predicting L2 discrimination difficulty. Euclidean Distances between the vowels of the two languages were computed to quantify acoustic similarity and to predict discrimination difficulty for Australian English learners of Brazilian Portuguese. Results show that Euclidean Distances successfully predict classification patterns in statistical models. We further compared the models' results to those of a previous study reporting Australian English listeners' discrimination of Brazilian Portuguese vowels, showing that real listeners' discrimination difficulty is indeed predicted by the current acoustic analyses.

Keywords: vowel acoustics, discriminant analyses, perceptual assimilation.

1. INTRODUCTION

For many second language (L2) learners, acquiring a new sound system is often a daunting task. Foreign accents and pronunciation difficulty often have a perceptual basis as a result of the influence of the learner's first language (L1) on their L2. Vowels can be particularly difficult for learners to perceive as not all languages share the same type and number of vowels in their vowel inventory.

Models of speech perception such as the Perceptual Assimilation Model [PAM, 1], it's extension to PAM-L2 [2] and the Second Language Linguistic Perception Model [L2LP, 8] account for the difficulties in non-native and L2 perception. They claim that perceptual similarity between native and target language vowel inventories is predictive of non-native and L2 vowel perception.

When two non-native sounds in a contrast are perceived as one native sound, the aforementioned models predict a high degree of discrimination difficulty. This scenario is known as single-category assimilation in PAM and PAM-L2 and as the new scenario in L2LP. This scenario is common for learners with a smaller vowel inventory than that of the target language. However, when two non-native sounds are mapped to two separate native categories, no difficulty is predicted for discrimination. This is known as two-category assimilation in PAM and as the similar scenario in L2LP.

Multiple-category assimilation (MCA, L2LP) or uncategorised assimilation (PAM, PAM-L2) occurs when two or more sounds in a non-native contrast are mapped to two or more vowels in the native language. This is a common scenario for learners of languages whose vowel inventory is smaller than that of their own native language. The problematic nature of this case is still relatively unknown as some studies show difficulty in discrimination in this scenario [e.g., 7, 9], while other studies found no such difficulty [e.g., 15, 16].

Research into the L2 acquisition of languages with larger vowel inventories than the learners' L1 is abundant, as learning new sounds is often difficult for non-native listeners and L2 learners [21]. For example, single-category assimilation has successfully explained Spanish, Portuguese and Russian listeners' difficulties with perceiving the English /i/-/ɪ/ contrast [14-17] because in all of these cases, listeners perceive the English contrast as their single native /i/ category. Studies have also shown that learners' difficulty is not constrained to English vowels, as Spanish listeners' struggle to discriminate the Dutch /a/-/ɑ/ contrast [11, 12].

Much less is known about the acquisition of languages with smaller vowel inventories than the native language. For example, [9] tested Dutch learners' perception of Spanish vowels and found that MCA lead to discrimination difficulty. [5] tested American English learners' perception of Brazilian Portuguese (BP) /e/-/ɛ/ after observing difficulties in their native-like pronunciation of these vowels. Another study [20], tested Californian English listeners' perception of the entire BP vowel inventory and found that learners struggle to perceive BP /i/-/e/, /a/-/ɔ/ and /o/-/u/. More recently, [7] replicated the aforementioned study of BP vowels [20] on Australian English (AusE) listeners and found similar difficulties with the BP /i/-/e/ and /o/-/u/ contrasts.

These two studies [7, 20] attribute these difficulties in perception to MCA. They suggest that discrimination difficulty as a result of MCA only occurs when there is a neutralisation of the L2 contrast. That is, only when two or more of the same native vowels are acoustically close to both target vowels in the L2 contrast, difficulty in discrimination is observed. When MCA does not result in a perceptual overlap, it functions in a similar manner to two-category assimilation, thus resulting in no discrimination difficulty.

The first aim of the present study is to generate more accurate predictions of how AusE listeners will perceive BP vowels by computing Euclidean Distances (EDs) between the acoustic properties of Western Sydney Australian English (WS) and Brazilian Portuguese (BP) vowels that were derived using the same data collection and analysis techniques. These EDs will be used as predictors of classification patterns in a discriminant analysis model. The L2LP model specifically states that for the most accurate predictions, acoustic analyses should be from the same group of listeners intended for testing, using the same method of data collection [8]. Although, [7] also used EDs to predict BP discrimination difficulty for adult AusE listeners from WS, the values in in the ED measurements are taken from [4], whose speakers were adolescent speakers from the Northern Beaches.

As our cross-linguistic acoustic analysis comparing BP and WS vowels was conducted using similar methodologies, predictions are likely to be more accurate. Therefore, the second aim of the present study is to determine whether the acoustic similarity reported from the Euclidean distances and classification model in the present study more accurately predicts previously reported results for AusE listeners of BP vowels [7].

2. METHOD

2.1. Speakers

The first four male and four female monolingual speakers from a corpus of 20 BP speakers reported in [10] were selected for analysis in this study. They were aged between 18 and 30, highly educated and had lived in São Paulo throughout their lives. Participants were considered monolingual if they did not report any knowledge of any foreign language with a proficiency of more than 3 on a scale of 1 to 7 (1 being no experience and 7 being fluent).

To ensure cross-linguistic comparability, we selected four male and four female participants from a new corpus of Western Sydney vowels [6] to match the selected BP speakers. That is, they were highly-educated monolingual speakers of AusE, with AusE speaking parents, born and raised in Western Sydney and aged between 18 and 30.

2.2. Recordings

The recordings in [10] were made in a quiet room at the Escola Superior de Propaganda e Marketing (ESPM) in São Paulo using a Sony MZ-NHF800 minidisk recorder and a Sony ECMMS907 condenser microphone, with a sample rate of 22 kHz and 16-bit quantization. The target vowels, /i, e, ɛ, a, ɔ, o, u/, were orthographically presented on a computer screen embedded in a sentence in five consonantal contexts, namely, /p, t, k, f, s/. The target vowels were always the first vowel produced in a disyllabic CVCV sequence (C = consonant, V = vowel), in which the two consonants were two identical voiceless stops or fricatives yielding nonce words such as /pepe/ and/saso/ (pêpe and sasso).

The WS speakers were recorded in a sound proof booth at the University of Western Sydney using a Shure SM10A-CN headset microphone and an Edirol Quad-Capture UA-55 sound card at 44.1 kHz. Following [10], participants read the target vowels, /iː, ɪ, e, eː, ɜː, ɐ, ɐː, æ oː, ɔ, ʊ, ʉː/, in isolated words and sentences from a computer screen. Words were presented in the same five contexts as presented in [10], namely fVf, sVs, tVt, pVp and kVk and produced in a similar carrier sentence.

For the present study, we chose to focus on the fVf context in both languages, as the studies testing AusE listeners' perception of BP vowels [7, 20] also used vowels extracted from this single context. In this context, there were 224 vowel tokens for BP (4 tokens x 7 vowels x 8 speakers), and 384 in for WS (4 tokens x 12 vowels x 8 speakers).

2.3. Data analysis

We followed the same vowel formant and duration analysis as in [10, 22]. Duration was calculated manually by placing boundaries at the start and end point of each vowel using the Praat [3] program. The "optimal formant ceiling" technique [10, 22] was used to determine F1, F2 and F3 values at 50% of the duration of each vowel. That is, for each vowel of each speaker, the "optimal ceiling" was chosen as the one that yields the least amount of variation for the first and second formant within the set number of annotated tokens for the vowel. Formant ceilings ranged between 4500 and 6500 Hz for females and 4000 and 6000 for males.

3. RESULTS

3.1. Cross-linguistic acoustic analysis

Figure 1 shows the acoustic similarity between BP and WS. To provide a quantitative measure of the acoustic similarity observed in Figure 1, we measured the Euclidean Distance[1] (ED) between the target vowels in a contrast and native vowels. These EDs were calculated by converting the extracted F1 and F2 values from Hz to the Bark auditory scale following the formula in [18][2].

Figure 1: Average male and female F1 and F2 values for BP (black with circles) and WS (grey).

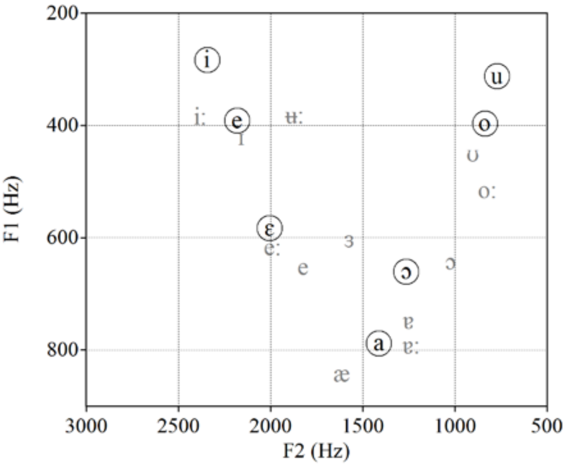


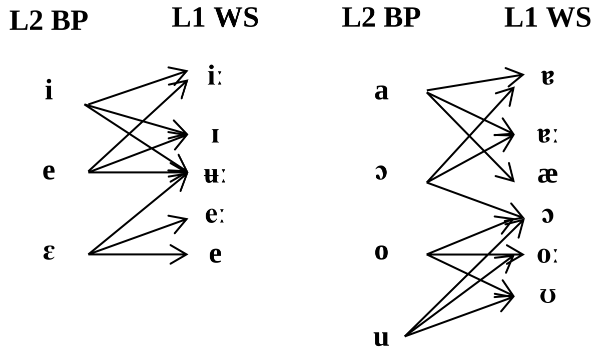
Table 1: The ED between each BP vowel and the three closest WS vowels

BP vowel	Closest AusE vowel 1st	ED	2nd	ED	3rd	ED
i	iː	1.02	ɪ	1.47	ʉː	1.81
e	ɪ	0.30	iː	0.58	ʉː	1.04
ɛ	eː	0.28	e	0.84	ɪ	1.48
a	ɐ	0.82	ɐː	0.87	æ	0.96
ɔ	ɐ	0.66	ɐː	0.98	ɔ	1.33
o	ʊ	0.67	oː	1.09	ɔ	2.41
u	ʊ	1.60	oː	1.94	ɔ	3.35

We observe in both Figure 1 and Table 1 that the closest native vowels to each vowel in the BP contrast often overlap, resulting in a total or partial acoustic overlap. For example, a total acoustic overlap occurs when the three closest native vowels are the same for both vowels in the BP contrast. This is indeed the case for BP /i/-/e/ and BP /o/-/u/. The closest native vowels to both BP /i/ and BP /e/ are WS /iː/, /ɪ/ and / uː/. For BP /o/ and BP /u/ the three closet native vowels are /ʊ/, /oː/ /ɔ/ were the same for each vowel in the contrast.

A partial acoustic overlap occurs when only one or two of the closest native vowels are acoustically close to both vowels in the BP contrast. This is the case for BP /a/-/ɔ/ as WS /ɐ/ was the first and /ɐː/ the second closest vowel for both BP /a/ and /ɔ/. We thus use the EDs in Table 1 as predictors of the likely perceptual assimilation patterns to be observed in the discriminant analysis model as shown in Figure 1 below:

Figure 2: Classification patterns as predicted by EDs.



3.2. Discriminant Analysis

The L2LP model [8] explicitly states that both perceptual assimilation patterns and discrimination difficulty can be acoustically predicted before testing. Previous studies [e.g., 13] have used discriminant analyses as a means of testing whether acoustic values are predictive of listeners' vowel classifications. This method was not used in [7] given the nature of the data in [4]. That is, the data in [4] was collected from a participant group of a different age and using a different method of data collection and analysis to that of the BP data in [10]. As our participant selection and method of data collection and analysis is similar to that of [10], we were able to conduct a discriminant analysis (DA) using formant and duration values to model WS listeners' likely classification patterns and compare them against the ED predictions.

We first conducted a separate linear DA model for each vowel corpus using the F1, F2, F3 (in Bark, measured at its midpoint, i.e., 50%) and duration values of the 224 BP and 384 WS vowel tokens described in the method. On the basis of formant and duration values, the model for BP yielded 84.5% correct classification for the training vowels and 87.5% correct classification for the cross validation set. The model for WS yielded 77.2% correct classification for trained tokens and 72.1% correct classification in the cross validation set.

A cross-language DA was conducted to determine how the WS model classifies BP vowels. The results in Table 2 demonstrate that many BP vowels are assimilated to more than one WS vowel category. The model classifications are very similar to the closest vowels reported in Table 1. In the case of BP /i/ the model classified the closest WS vowel,

/iː/, 100% of the time. A similar case was found for BP /u/ which was classified as /ʊ/ 94% of the time. For the remaining vowels, the model classified the tokens across a range of WS vowels as predicted by the EDs. For example, BP /e/ was classified as WS /iː/, /ɪ/ and / uː/ and BP /a/ was classified across WS /ɐ/, /ɐː/ and /æ/. For BP /ɔ/ the EDs successfully predicted the model classification of WS /ɐː/, /ɔ/, but the EDs did not successfully predict the models' classifications of WS /oː/ and /eː/.

Table 2: Percentage BP vowel tokens classified as a WS vowel. Only values above 10% were included.

BP	WS											
	iː	**ɪ**	**eː**	**e**	**ɜː**	**ɐː**	**ɐ**	**æ**	**oː**	**ɔ**	**ʊ**	**ʉː**
i	100											
e	50	22										29
ɛ			72	13								13
a						13	28	47				
ɔ			16			25			19	25		
o									16		78	
u											94	

Based on the EDs and model classifications, we would predict that the BP contrast /i/-/e/ is difficult to discriminate, as both BP /i/ and /e/ are classified as WS /iː/. Also single-category assimilation is likely to cause difficulty for BP /o/-/u/, as both vowels were predominantly classified as /ʊ/. Finally, despite the MCA patterns for BP /a/when heard in the contrast /a/-/ɛ/, this is unlikely to result in difficulty as there is no overlap between the classified WS vowels across the contrast.

Table 3: Comparison of ED and DA predictions with real listeners' discrimination accuracy.

BP Contrast	BP vowel	Closest Vowels ED	DA results	Discrim. Acc. [7]
/a-ɔ/	a	ɐ, ɐː, æ	æ, ɐ, ɐː	75.63
	ɔ	ɐ, ɐː, ɔ	ɔ, ɐː, oː	
/a-ɛ/	a	ɐ, ɐː, æ	æ, ɐ, ɐː	92.19
	ɛ	eː, e, ɪ	eː, e, ʉː	
/e-i/	e	ɪ, iː, ʉː	ɪ, iː, ʉː	66.25
	i	iː,ɪ, ʉː	iː	
/o-u/	o	ʊ, oː, ɔ	ʊ, oː	65.94
	u	ʊ, oː, ɔ	ʊ,	
/e-ɛ/	e	ɪ, iː, ʉː	ɪ, iː, ʉː	82.81
	ɛ	eː, e, ɪ	eː, e, ʉː	
/o-ɔ/	o	ʊ, oː, ɔ	ʊ, oː	80.31
	ɔ	ɐ, ɐː, ɔ	ɔ, ɐː, oː	

3.3. Comparison with real listeners

Table 3 shows the acoustic predictions for the EDs and DA from the present study and the results from [7]. Our acoustic predictions are in line with the findings from [7]. BP /o/-/u/ had the lowest accuracy as predicted. BP /i/-/e/ also had lower accuracy scores, which is likely due to the acoustic predictions of a partial overlap as a result of MCA. Finally, as predicted BP /a/-/ɛ/ was the easiest contrast to discriminate despite MCA as there was no perceptual or acoustic overlap.

4. DISCUSSION

The present study successfully used EDs as a quantitative measure of predicting classification patterns of BP to WS vowels in a discriminant analysis model. Furthermore, the discrimination difficulty reported in [7] was successfully predicted by the current acoustic analysis.

The DA model yielded similar findings as [7]. In particular, the results suggest that MCA is only difficult when there is a partial or complete acoustic or perceptual overlap. However, lower discrimination accuracy for BP /o/-/u/ was previously attributed to a complete or acoustic overlap in [7], yet our acoustic analysis suggests that this finding is likely a result of single category assimilation to the WS vowel /ʊ/.

In sum, the present study provides further support to the L2LP model claim that acoustics can successfully predict L2 difficulty before testing occurs. The next step in this research would be to confirm the reliability of these acoustic predictions in the EDs and DA by comparing the classification patterns from the present study with perceptual assimilation results from real WS listeners. Future research is also required to determine the role of acoustic similarity in predicting L2 difficulty in word recognition and production.

7. REFERENCES

[1] Best, C.T. 1995. A direct realist perspective on cross-language speech perception. In: Strange, W. (ed), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Timonium, MD: York Press, 171-204.

[2] Best, C.T., Tyler, M. 2007. Non-native and second-language speech perception: commonalities and complementarities. In: Bohn, O. –S., Munro M. (eds), *Language Experience in Second-Language Speech Learning: In Honour of James Emil Flege.* Amsterdam: John Benjamins, 13-34.

[3] Boersma, P., Weenink, D. 1992-2013. Praat: doing phonetics by computer (versions 5.1.3 [ch2], 5.2.26

[ch3], 5.2.40 [ch4]). [Computer program], available from http://www.praat.org/.

[4] Cox, F. 2006. The Acoustic characteristics of /hVd/ vowels in the speech of some Australian teenagers. Aust. J. Linguist. 26, 147-179

[5] Díaz Granado, M. 2011. L2 and L3 Acquisition of the Portuguese Stressed Vowel Inventory by Native Speakers of English. PhD Thesis, University of Arizona, Tuscon, AZ.

[6] Elvin, J. An acoustic analysis of Australian English vowels by speakers from Western Sydney. In preparation.

[7] Elvin, J., Escudero, P., Vasiliev, P. 2014. Spanish is better than English for discriminating Portuguese vowels: acoustic similarity versus vowel inventory size. Front. Psychol. 5:1188

[8] Escudero, P. 2005. Linguistic Perception and Second Language Acquisition: Explaining the Attainment of Optimal Phonological Categorization. PhD thesis, LOT Dissertation Series 113. Utrecht: Utrecht University.

[9] Escudero, P., Boersma, P. 2002. The subset problem in L2 perceptual development: multiple-category assimilation by Dutch learners of Spanish. In: Skarabel, B., Fish, S., Do, A. (eds), Proc. 26th Annual Boston University Conference on Language Development. Somerville, MA: Cascadilla, 208-219.

[10] Escudero, P., Boersma, P., Rauber, A.S, Bion, R.A. 2009. A cross-dialect acoustic description of vowels: Brazilian and European Portuguese. J. Acoust. Soc. Am. 126, 1379-1393.

[11] Escudero, P., Williams, D. 2011. Perceptual assimilation of Dutch vowels by Peruvian Spanish listeners. J. Acoust. Soc. Am. 129, EL1-EL7.

[12] Escudero, P., Williams, D. 2012. Native dialect influences second-language vowel perception: Peruvian versus Iberian Spanish learners of Dutch. *J. Acoust. Soc. Am.* 131, EL406-EL412.

[13] Escudero, P., Vasiliev, P. 2011. Cross-language acoustic similarity predicts perceptual assimilation of Canadian English and Canadian French. . *J. Acoust. Soc. Am.* 130, EL277–EL283

[14] Kondaurova, M., Francis, A. 2008. The relationship between native allophonic experience with vowel duration and perception of the English tense/lax vowel contrast by Spanish and Russian listeners. J. Acoust. Soc. Am. 124, 3959-3971.

[15] Morrison, G. S. 2003. Perception and production of Spanish vowels by English speakers. Proc. 15th ICPhS Barcelona, 203-206.

[16] Morrison, G.S. 2009. L1 Spanish speakers' acquisition of the English /i/-/ɪ/ contrast II: perception of vowel inherent spectral change. Lang. Speech. 52, 436-462.

[17] 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. Proc. InterSpeech Lisbon, 2913-2916.

[18] Traunmuller, H. 1997. Auditory scales of frequency representation. Retrieved 29th January, from http://www.ling.su.se/staff/hartmut/bark.htm

[19] van Leussen, J.W., Williams, D., Escudero, P. 2011. Acoustic properties of Dutch steady-state vowels: contextual effects and a comparison with previous studies. Proc. 17th ICPhS Hong Kong, 1194-1197.

[20] Vasiliev, P. 2013. The Initial State for Californian English Learners of Spanish and Portuguese Vowels. PhD Thesis. University of California, Los Angeles.

[21] Vasiliev, P., Escudero, P. 2013. Speech perception in second language Spanish. The handbook of Spanish second language acquisition, 130-145

[22] Williams, D.P. 2013. Cross-language acoustic and perceptual similarity of vowels: The role of listeners' native accents. PhD thesis, University of Sheffield.

[1] We used the following equation to measure the distance in Bark between the two vowels: d (p,q)= √((p1-q1)^2+(〖p2-q2)〗^2) or d (TV,L1v)= √((TVF1-L1v1)^2+(〖TVF2-L1vF2)〗^2) , where d stands for Euclidean distance, TV for target vowel, L1v for native vowel, and F1 and F2 for this vowel's average F1 and F2 values.

[2] Traunmuller's formula: = (26.81/(1+1960/K5))-0.53