INTERSPEECH 2015

Confusability in L2 vowels: analyzing the role of different features

Mátyás Jani [1], *Catia Cucchiarini* [2], *Roeland van Hout* [3], *Helmer Strik* [2,3]

[1] Faculty of Information Technology and Bionics, Pázmány Péter Catholic University, Hungary
[2] Center for Language and Speech Technology, Radboud University Nijmegen, The Netherlands
[3] Department of Linguistics, Radboud University Nijmegen, The Netherlands

jani.matyas@itk.ppke.hu, {c.cucchiarini, r.vanhout, w.strik}@let.ru.nl

Abstract

The production and perception of L2 vowels are influenced by the L1 vowel system. Most studies on L2 vowel production evaluate the learners' pronunciation using subjective listening tests. In this study we present a novel objective method for investigating learner vowel confusability based on acoustic measurements. Monosyllabic words uttered by Spanish learners of Dutch are analyzed, and basic acoustic features – formant frequencies and duration – are extracted. Native Dutch speakers' measurements are used to obtain models for the Dutch vowels, which are employed to compute likelihood ratios and similarity distributions of the Spanish realizations in comparison to the Dutch target vowels. The likelihood ratios are presented in a matrix format similar to a confusion matrix crossing the target vowels by the vowels as classified. Results based on spectral features alone confirm the existence of an attractor effect of L1 vowels on L2 vowels. Overall, including duration in the analyses decreases the number of confusions. Comparing the confusion values on different feature sets helps analyzing the impact of the specific features. The results of the present study suggest that although the Spanish learners' use of duration is not native-like, it does help reduce confusability among Dutch vowels.

Index Terms: L2 vowel acquisition, vowel confusability, likelihood ratios, pronunciation assessment

1. Introduction

Nativelike acquisition of the vowels of a second language is known to be problematic for many L2 learners [1], [2]. This has been shown for different L1-L2 combinations. Many of the difficulties L2 learners encounter in processing L2 vowels appear to be related to the influence of their L1 [1], [2]. Numerous studies on L2 vowel acquisition have addressed vowel perception while vowel production has received less attention. In particular, there have been relatively fewer studies on L2 vowel production that have employed acoustic measurements (but see [3]–[5]). Acoustic analyses of L2 vowel production can reveal differences that may not be easily perceived by listeners [6]. In addition, acoustic studies may provide insight into how L2 learners deal with the acoustic dimensions involved in L2 vowel production and may in turn provide relevant information for L2 pronunciation training.

In our own research we have investigated Dutch L2 vowel production by Spanish learners through phonological annotation [7] and acoustic measurements [8]. Research based on annotations of speech recordings indicated that Dutch L2 vowel production by Spanish learners is influenced by their L1 vowel system and, more specifically, that the Spanish vowels appear to function as attractors for the Dutch vowels [7]. To complement these studies and gain more detailed information about how Spanish learners realize Dutch vowels we conducted acoustic analyses of F1, F2 and duration. These revealed that Spanish learners produce Dutch vowels with spectral and durational properties that differ considerably from those of native Dutch vowels [8]. The results of these analyses seemed to confirm the attractor effect of the Spanish vowels.

An interesting question at this point is how problematic or confusable these Spanish L1 Dutch vowels are. A common way of establishing this is by having native listeners evaluate L2 vowel quality in perception experiments, as in [9]. However, this procedure draws heavily on subjective evaluations and as such it is error-prone and likely to be influenced by a complex combination of factors. The method can be made more reliable by increasing the number of subjects rating the speech samples, but obtaining large numbers of raters is not always feasible and the influence of many factors cannot be excluded (lexical competitors etc). To overcome the problems related to subjectivity, we propose a novel, more objective approach that makes use of likelihood ratios and allows to investigate the effect of the separate acoustic dimensions involved. In the next section we report on studies on Dutch L2 vowel production by Spanish learners in which such an objective method is employed and investigated.

2. Spanish and Dutch vowels

The Spanish and the Dutch vowels systems are considerably different from each other [7], [10], [11]. In addition, Spanish and Dutch also differ with respect to orthography [12]. Spanish has a shallow orthography while Dutch has a deeper orthography. L1 orthography can affect non-native speech perception [12], [13] and speech production [14]. In Figure 1 the acoustic differences between the Dutch and the Spanish vowels are illustrated. The formant values of the five Spanish vowels are based on analyses of speech material collected and made available to us by [15]. The formant values of the Dutch vowels are based on measurements of Dutch native speech material collected by [16] and made available to us.

In Figure 1 vowel ellipses are used to visualize the distribution of vowel realizations in the vowel space. Two-dimensional Gaussian distributions are fitted on the normalized first two formants. The area of an ellipse covers the 50% of the Gaussian's volume, thus containing approximately half of the realizations. From this figure it is clear how the Dutch vowels are positioned with respect to the Spanish vowels.

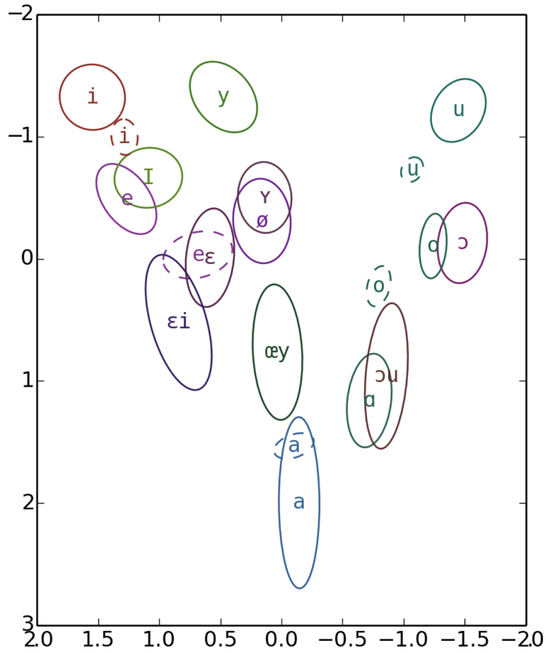

Figure 1: native *Dutch (solid) and native Spanish (dashed) vowel ellipses displaying the normalized (Lobanov) F1 (y-axis) and F2 (x-axis). For more details, see section 2.*

A third, important dimension used to distinguish Dutch vowels, duration, is not visualized in this figure. Short vowels in Dutch are: /i, I, ɛ, y, ʏ, ɑ, u, ɔ/, while the vowels: /a, e, o, ø, ɛi, œy, ɔu/ are long. A previous study based on acoustic measurements [8] revealed that Spanish learners of Dutch use spectral features and duration differently from Dutch native speakers to realize Dutch vowels.

The question at this point is to what extent differences in spectral features and duration in vowels produced by L2 learners may cause confusability between these vowels. We investigate this with respect to Dutch vowels as produced by Spanish learners.

3. Method

3.1. Speech material

In these experiments we used previously collected data by van der Harst [16], and Burgos et al. [8]. For both the Spanish learner and Dutch native speech materials the participants were recorded while reading 29 monosyllabic Dutch words. The selected words contained all 15 Dutch vowels, and ended in /t/ or /s/, as these consonants have less impact on the articulation of the preceding vowels than other consonants [8], [16]. The only missing combination was /y/ with /s/, as there is no Dutch monosyllabic word ending in this consonant, except proper names.

The two sets of speech material were analyzed and measured in similar ways. Duration values were extracted by segmenting the vowels by hand. Formant frequencies F1 and F2 were extracted and hand-checked in both data sets at 25%, 50% and 75% in time. In order to correct for between speaker differences we applied the Lobanov z-score transformation to normalize the formant values [17]. Duration was also normalized using z-score transformation per speaker.

3.2. Likelihood ratio

To find out to what extent the vowels realized by the Spanish learners resemble the target Dutch vowels, we computed likelihood ratios by using a Gaussian statistical classifier. The features we used were duration and the 25%, 50% and 75% F1 and F2 measurements. The likelihood ratios tell us how closely the vowel realizations by the Spanish learners match the target Dutch vowels. They are calculated to obtain the similarity distribution of the Spanish realizations over the Dutch target vowels. Likelihood ratios can vary from 1 (perfect match) to 0 (no match at all).

The likelihood ratio (LR) is obtained by computing the likelihood values of the multivariate Gaussian probability density function (fitted to the training – native – vowels), at the test vowel's feature vector. The LRs were computed as follows:

$$LR_p(x) = \frac{MVG_p(x)}{\sum_q MVG_q(x)} \quad (1)$$

where $MVG_p(x)$ is the likelihood value of the test feature vector x given the Gaussian distribution associated with the native vowel p. The denominator is the sum of the likelihoods of each Gaussian at the test feature vector.

Measures based on likelihood ratios are not uncommon in the field of L2 pronunciation research. In general, they are used for automatic error detection, such as APE (Average Posterior Probability Estimation, [18]) and GOP (Goodness of Pronunciation, [19]) where they are produced by ASR acoustic models (usually Gaussian Mixture Model). Our measure can be interpreted as a special type of APE, with only one Gaussian per phone as acoustic model based on formant and duration values. Instead of averaging over frames we include time as an extra dimension of the Gaussian by using the 25%, 50% and 75% F1 and F2 values of the vowels

The average likelihood ratio values are presented in matrix form to show the confusions between the realizations and the native target phones. The rows in the matrix are associated with the test instances, the columns with the training (native) vowels. Each row contains the average of the LR values of the test instances with the same test vowel. Thus the likelihood ratio matrix (LRM) cells can be defined:

$$LRM_{i,j} = \frac{\sum_{x \in X_i} LR_j(x)}{|X_i|} \quad (2)$$

where X_i is the set of test instances for the i test vowel, j is the target vowel.

4. Results

4.1. Interpretation of the matrices

The matrices defined by equation (2) present the confusions between the test (rows) and the training (columns) vowels. When the learners' vowels are compared to the natives', the test vowels are the learners' and the training vowels are the natives'. In order to understand how the learners' realizations deviate from the natives', we first look at the confusions in the native data because these constitute a kind of benchmark. To obtain a training set and a test set for the native data, 10-fold cross-validation was used, with all samples from a speaker falling in either the training or the test set.

The matrices are presented in Figures 2 to 5. The row and column labels are the phone IPA codes. The last row contains the sum of the columns. High diagonal values mean that the realized test vowels are close to the corresponding training vowels. A high off-diagonal confusion in cell i,j means that test vowel i is often realized acoustically close to the non-target training vowel j. The confusions are not symmetric by default; the similarity of test vowel i to training vowel j does not imply the same similarity of test vowel j to training vowel i. In reality the confusions are often bidirectional.

Diagonal cells with > .5 confusion values are highlighted in green, < .5 values are in red. Off-diagonal values > .1 are also highlighted in red. The average of the diagonal values provides an overall measure of the quality of the fit, and can be found in the figure captions (*DiagAvg*).

4.2. Likelihood ratio matrices

4.2.1. Native matrix

The matrix in Figure 2 shows to what extent vowels produced by native Dutch speakers may be confused with each other based on spectral features (F1 and F2) alone. The average of the diagonals is .86, which denotes a high level of correspondence between the test and the training vowels and therefore low confusion. When we look at the values on the diagonals we see that these differ for the various vowels. The lowest values are found for /ʏ/, /ø/ and /ɔ/. These vowels also exhibit patterns of confusion with /ø/, /ʏ/ and /o/, respectively. We also find a relatively high value for the confusion /o/ - /ɔ/. A fifth remarkable confusion is /e/ being classified as /I/, which also has a counterpart with /I/ being classified as /e/, but in this case the confusion value is lower and below the threshold of .10. So, overall confusion in the native material is low. The confusions that we observe seem to be related to the absence of duration as a dimension of classification. The vowels in the pairs /ʏ/ - /ø/ and /ɔ/ - /o/ and /I/ - /e/ are indeed distinguished mainly by duration and diphthongization.

	i	I	ɛ	y	ʏ	a	u	ɔ	ɑ	e	ø	o	ɛi	œy	ɔu
i	**0.91**	0.04	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00
I	0.04	**0.86**	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.06	0.00	0.00	0.00	0.00	0.00
ɛ	0.00	0.06	**0.81**	0.01	0.08	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.01	0.00	0.00
y	0.00	0.04	0.00	**0.84**	0.04	0.00	0.00	0.00	0.00	0.00	0.08	0.00	0.00	0.00	0.00
ʏ	0.00	0.00	0.05	0.07	**0.71**	0.00	0.00	0.00	0.00	0.00	0.17	0.00	0.00	0.00	0.00
a	0.00	0.00	0.00	0.00	0.00	**0.94**	0.00	0.00	0.05	0.00	0.00	0.00	0.00	0.01	0.00
u	0.00	0.00	0.00	0.00	0.00	0.00	**0.96**	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00
ɔ	0.00	0.00	0.00	0.00	0.00	0.00	0.05	**0.78**	0.00	0.00	0.00	0.17	0.00	0.00	0.00
ɑ	0.00	0.00	0.00	0.00	0.00	0.07	0.00	0.00	**0.90**	0.00	0.00	0.00	0.00	0.00	0.03
e	0.01	0.10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**0.83**	0.00	0.00	0.05	0.00	0.00
ø	0.00	0.00	0.05	0.02	0.18	0.00	0.00	0.00	0.00	0.00	**0.72**	0.00	0.00	0.02	0.00
o	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.18	0.00	0.00	0.00	**0.82**	0.00	0.00	0.00
ɛi	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	**0.96**	0.01	0.00
œy	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.01	0.00	0.02	**0.94**	0.00
ɔu	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.00	0.00	0.00	0.00	0.01	**0.94**
	0.96	*1.10*	*0.93*	*0.97*	*1.01*	*1.03*	*1.01*	*1.00*	*1.00*	*0.96*	*1.02*	*0.99*	*1.04*	*1.00*	*0.96*

Figure 2: *LRM of the native speakers based on normalized formant frequencies. DiagAvg = 0.86*

If we now add duration as a parameter in computing the likelihood ratios we can see how this affects confusability. The results are given in Figure 3.

Including duration increases the diagonal values in general, and decreases the non-diagonal values. Compared to the native matrix without duration, the higher bi-directional confusions, /o/ - /ɔ/ and /ʏ/ - /ø/, disappear. When duration is included we also see that the realizations of /e/ are less similar to /I/.

There is only one decreasing diagonal value, /y/ is reduced from .84 to .73. When duration is included /y/ is more often classified as either /I/, /ʏ/ or /ɛ/, indicating that this vowel is competing with other short vowels, the /ʏ/ in particular. Overall, the native matrix in which duration is included displays no worthwhile confusions between Dutch vowels as realized by Dutch native speakers. In a sense this is also a validation of our method. We now look at what happens in the Spanish learners' data.

	i	I	ɛ	y	ʏ	a	u	ɔ	ɑ	e	ø	o	ɛi	œy	ɔu
i	**0.95**	0.03	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
I	0.06	**0.90**	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
ɛ	0.00	0.01	**0.86**	0.01	0.08	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.00	0.00
y	0.02	0.07	0.07	**0.73**	0.09	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00
ʏ	0.00	0.00	0.06	0.09	**0.83**	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00
a	0.00	0.00	0.00	0.00	0.00	**1.00**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
u	0.00	0.00	0.00	0.00	0.00	0.00	**0.96**	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00
ɔ	0.00	0.00	0.00	0.00	0.00	0.00	0.05	**0.94**	0.00	0.00	0.00	0.01	0.00	0.00	0.00
ɑ	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.02	**0.94**	0.00	0.00	0.00	0.00	0.02	0.00
e	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**0.94**	0.00	0.00	0.06	0.00	0.00
ø	0.00	0.00	0.01	0.00	0.02	0.00	0.00	0.00	0.00	0.00	**0.91**	0.00	0.00	0.06	0.00
o	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	**0.99**	0.00	0.00	0.00
ɛi	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.00	**0.95**	0.02	0.00
œy	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.02	0.00	0.03	**0.94**	0.00
ɔu	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	**0.98**
	1.03	*1.01*	*1.04*	*0.84*	*1.03*	*1.03*	*1.01*	*1.01*	*0.95*	*0.98*	*1.00*	*1.00*	*1.03*	*1.05*	*0.98*

Figure 3: *LRM of native speakers based on normalized formant frequencies and duration. DiagAvg = 0.92*

4.2.2. Learner matrix

The matrix in Figure 4 indicates how the vowels realized by the Spanish learners are related to the Dutch target vowels when only spectral features (F1 and F2) are taken into consideration.

	i	I	ɛ	y	ʏ	a	u	ɔ	ɑ	e	ø	o	ɛi	œy	ɔu
i	**0.61**	0.05	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.26	0.00	0.00	0.05	0.00	0.00
I	0.56	**0.24**	0.05	0.02	0.04	0.00	0.00	0.00	0.00	0.05	0.02	0.00	0.02	0.00	0.00
ɛ	0.00	0.07	**0.72**	0.00	0.02	0.00	0.00	0.00	0.00	0.06	0.02	0.00	0.09	0.03	0.00
y	0.01	0.05	0.00	**0.35**	0.15	0.00	0.30	0.01	0.00	0.00	0.08	0.00	0.03	0.01	0.00
ʏ	0.03	0.03	0.01	0.16	**0.36**	0.00	0.21	0.02	0.00	0.01	0.15	0.00	0.00	0.02	0.00
a	0.00	0.00	0.00	0.00	0.00	**0.69**	0.00	0.00	0.24	0.00	0.00	0.00	0.00	0.05	0.02
u	0.00	0.00	0.00	0.04	0.01	0.00	**0.62**	0.12	0.02	0.00	0.02	0.12	0.00	0.05	0.00
ɔ	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**0.32**	0.30	0.00	0.01	0.30	0.00	0.01	0.06
ɑ	0.00	0.00	0.00	0.00	0.00	0.61	0.00	0.00	**0.21**	0.00	0.00	0.00	0.00	0.17	0.01
e	0.05	0.04	0.10	0.00	0.00	0.00	0.00	0.00	0.00	**0.66**	0.00	0.00	0.15	0.00	0.00
ø	0.00	0.01	0.15	0.04	0.05	0.00	0.00	0.00	0.00	0.02	**0.39**	0.00	0.06	0.26	0.03
o	0.00	0.00	0.00	0.00	0.00	0.00	0.12	0.56	0.03	0.00	0.00	**0.27**	0.00	0.00	0.02
ɛi	0.00	0.00	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.11	0.00	0.00	**0.77**	0.09	0.00
œy	0.00	0.00	0.01	0.03	0.00	0.02	0.00	0.00	0.00	0.00	0.11	0.00	0.16	**0.49**	0.18
ɔu	0.00	0.00	0.00	0.00	0.02	0.00	0.05	0.01	0.04	0.00	0.04	0.17	0.02	0.12	**0.53**
	1.26	*0.49*	*1.08*	*0.67*	*0.65*	*1.32*	*1.30*	*1.03*	*0.84*	*1.16*	*0.85*	*0.87*	*1.34*	*1.29*	*0.85*

Figure 4: *LRM of learners based on normalized formant frequencies. DiagAvg = 0.48*

In this matrix we see that the diagonal values and consequently the average fit (.48) are much lower than in the corresponding native matrix. Some vowels have extremely low diagonal values, the lowest are found for /ɑ/, /I/ and /o/. What is remarkable is that instances of these vowels as realized by Spanish learners are more often classified as another vowel than the target vowel. For instance, Spanish learners' /ɑ/ is more often classified as /a/ (.61) than as /ɑ/ (.21), their /I/ is more often classified as /i/ (.56) than as /I/ (.24) and their /o/ is more often classified as /ɔ/ (.56) than as /o/ (.27). There are also other vowels that obtain low diagonal values and relatively high off-diagonal values. This is the case of /ɔ/ with a diagonal value of .32 and off-diagonal values of .30 for both /ɑ/ and /o/, of /y/ with a diagonal value of .35 and off-diagonal values of .30 for /u/ and .15 for /ʏ/, and /ʏ/ with a diagonal value of .36 and off-diagonal values of .21 for /u/ and .16 for /y/.

Another interesting difference between the native matrix and the learner matrix is that in the former the column sums all have values around 1, while in the latter these values differ considerably. The column sums show how similar the Spanish learners' realizations are to the vowel associated with the given column overall. A number above one suggests that a vowel similar to the column's target vowel is frequently realized. We notice here that for the three extreme vowels in the vowel triangle /i/, /a/, and /u/ and for the diphthongs /ɛi/ and /œy/ the column sums are high. The column sum for /i/ is 1.26. The highest confusion contributing to this sum is that between /I/ and /i/, which means that many of the Spanish learners' realizations of Dutch /I/ are similar to Dutch /i/. As a consequence, the sum for column /I/ is low, which is mainly due to the confusion with /i/. The column sum for /i/ is high in spite of the relatively high classification of Spanish learners' /i/ as target vowel /e/, 0.26.

For Dutch /a/ the column sum is 1.32. A considerable contribution to this high value is the extremely frequent classification of Spanish learners' /ɑ/ as /a/. The target vowel /u/ also exhibits a high column sum value of 1.30. Remarkable contributions to this column sum are the classification of Spanish learners' /y/, /ʏ/ and /o/ as /u/ with the values .30, .21, and .12 respectively.

For the diphthongs /ɛi/ and /œy/ the column sums are 1.34 and 1.29. For /ɛi/ the high value is brought about by a possible diphthongization of /e/ and by a confusion with /œy/, while for /œy/ the high column sum is related to confusions with /ø/, /ɑ/ and /ɔu/.

In this matrix we also see that some vowels have extremely low column sums. This applies to /I/ which has a sum of .49, /ʏ/ with .65 and /y/ with .67. In general, it appears that the vowels that are closest to the five Spanish vowels system (/a, e, i, o, u/) have higher column sums: /a, ɛ, e, i, ɔ, o, u/, which seems to confirm the attractor effect that emerged from annotation data [7]. Two of the diphthongs also get high values, partly due to the diphthongization of the long mid vowels.

	i	I	ɛ	y	ʏ	a	u	ɔ	ɑ	e	ø	o	ɛi	œy	ɔu
i	**0.47**	0.28	0.00	0.05	0.00	0.00	0.00	0.00	0.00	0.14	0.00	0.00	0.05	0.00	0.00
I	0.52	**0.33**	0.08	0.01	0.04	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00
ɛ	0.00	0.06	**0.82**	0.00	0.03	0.00	0.00	0.00	0.00	0.03	0.02	0.00	0.05	0.00	0.00
y	0.00	0.00	0.01	**0.14**	0.04	0.00	0.20	0.04	0.00	0.00	0.44	0.08	0.00	0.05	0.00
ʏ	0.03	0.05	0.01	0.21	**0.44**	0.00	0.17	0.02	0.02	0.00	0.02	0.02	0.00	0.00	0.00
a	0.00	0.00	0.00	0.00	0.00	**0.76**	0.00	0.00	0.11	0.00	0.00	0.00	0.01	0.07	0.05
u	0.00	0.00	0.00	0.02	0.02	0.00	**0.57**	0.19	0.02	0.00	0.02	0.12	0.00	0.05	0.00
ɔ	0.00	0.00	0.00	0.00	0.01	0.00	0.00	**0.53**	0.41	0.00	0.00	0.05	0.00	0.00	0.00
ɑ	0.00	0.00	0.02	0.00	0.00	0.15	0.00	0.00	**0.54**	0.00	0.00	0.00	0.01	0.29	0.00
e	0.01	0.05	0.04	0.00	0.00	0.00	0.00	0.00	0.00	**0.66**	0.00	0.00	0.25	0.00	0.00
ø	0.00	0.00	0.06	0.02	0.03	0.00	0.00	0.00	0.00	0.03	**0.48**	0.00	0.06	0.28	0.05
o	0.00	0.00	0.00	0.00	0.00	0.00	0.07	0.25	0.00	0.00	0.00	**0.63**	0.00	0.00	0.05
ɛi	0.00	0.00	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.11	0.00	0.00	**0.76**	0.09	0.00
œy	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.12	0.00	0.16	**0.54**	0.17
ɔu	0.00	0.00	0.00	0.00	0.02	0.01	0.02	0.00	0.00	0.00	0.04	0.22	0.02	0.13	**0.54**
	1.03	*0.76*	*1.08*	*0.44*	*0.62*	*0.92*	*1.04*	*1.03*	*1.09*	*0.99*	*1.14*	*1.12*	*1.37*	*1.51*	*0.85*

Figure 5: *LRM of learners based on normalized formant frequencies and duration. DiagAvg = 0.55*

We now look at the matrix in Figure 5 with likelihood ratios including normalized duration, which shows how the vowels realized by the Spanish learners are classified when duration is taken into consideration. Also in this case including duration has the effect of increasing the diagonal mean (DiagAvg). However, here the pattern of changes is more complex compared to that in the native data. In this case, some diagonal values increase and other decrease considerably. The diagonal value of /y/ is reduced here as well, but the decrease is more substantial. As was to be expected when duration is included there are fewer confusions between /ɑ/ and /a/, and between /ɔ/ and /o/. The distinction between /I/ and /i/, on the other hand, does not profit so much from the inclusion of duration, as duration is not the distinctive feature for this vowel pair. Furthermore, including duration leads to more confusions between /ɔ/ and /ɑ/ and between /y/ and /ø/.

An interesting result in this matrix is that, in general, the column sums come closer to 1.00, and thus are more similar to those in the native matrix, with the exceptions of the column sums for the diphthongs /ɛi/ and /œy/, which become even higher. In other words, the attractor effect of the L1 vowels is less visible in these data, while two diphthongs function as attractors, /ɛi/ for /e/ and /œy/ for /ɑ/ and /ø/.

5. Discussion and future perspectives

The results presented in the previous sections indicate how the Dutch vowels produced by native speakers and Spanish learners of Dutch L2 are (mis)classified when only spectral information is employed and when duration is added as a classification parameter. In the case of the native vowels, it seems that combining the two sources of information, which are indeed known to be both relevant for vowel production, gives the best results. In the case of the Spanish learners the situation is different. We have seen in previous studies that Spanish learners somehow manage to employ duration to distinguish Dutch vowels, but their use of duration is not nativelike [8]. In some cases they use duration when only spectral properties should be used and in others they employ duration excessively. When duration is included in the likelihood ratio computation, some vowel distinctions improve, but others get worse. The overall effect that can be noticed is, however, that classification is improved and that, notably, the attractor effect that has been observed previously becomes less prominent. In other words, although duration is not realized in a nativelike manner by Spanish learners, the results of the present study suggest that it does help to reduce confusions between Dutch vowels as realized by Spanish learners. The mechanism behind the observations has not been investigated yet. It could be the acoustic similarity between the vowels or the effect of orthography.

With respect to the influence of the L1 vowel system on L2 vowel production, these results suggest that L2 learners in any case try to escape L1 routines and employ new dimensions of the L2, maybe not in a nativelike fashion, but at least in a way that should minimize confusability. Of course we should be careful in drawing conclusions, because these results are based on average data of a group of learners and may not apply to individual learners. For this reason we intend to further explore the potential of the method adopted in this study to see whether it can help shed some light on individual differences in L2 vowel production and confusability. Furthermore, we conducted subjective studies on L2 vowel production, to see how native listeners cope with spectral and durational mismatches in L2 vowel production [20]. By relating the results of such studies to those of the objective method employed in the present investigation, we will be able to gain insight into how the various dimensions are employed by individual L2 speakers and L1 listeners.

6. Acknowledgements

This research was supported by the Centre for Language Studies (CLS) of the Radboud University Nijmegen, The Netherlands and by the Pázmány Péter Catholic University, Budapest. We would like to thank Pepi Burgos and Joop Kerkhoff for their help analyzing of the speech files. Our special thanks go to Sander van der Harst and Paola Escudero for making the acoustic values of the Dutch and the Spanish vowel realizations available to us.

7. References

[1] J. E. Flege, “Second-language Speech Learning: Theory, Findings, and Problems,” in *Speech perception and linguistic experience: issues in cross-language research*, W. Strange, Ed. Timodium, MD: York Press, 1995, pp. 233–272.

[2] C. T. Best, “A direct realist view of speech cross language speech perception,” in *Speech perception and linguistic experience: issues in cross-language research*, W. Strange, Ed. Timodium, MD: York Press, 1995, pp. 171–204.

[3] A. A. N. MacLeod, C. Stoel-Gammon, and A. B. Wassink, “Production of high vowels in Canadian English and Canadian French: A comparison of early bilingual and monolingual speakers,” *Journal of Phonetics*, vol. 37, no. 4, pp. 374–387, Oct. 2009.

[4] G. E. Oh, S. Guion-Anderson, K. Aoyama, J. E. Flege, R. Akahane-Yamada, and T. Yamada, “A one-year longitudinal study of English and Japanese vowel production by Japanese adults and children in an English-speaking setting,” *Journal of Phonetics*, vol. 39, no. 2, pp. 156–157, Apr. 2011.

[5] K. Tsukada, D. Birdsong, E. Bialystok, M. Mack, H. Sung, and J. Flege, “A developmental study of English vowel production and perception by native Korean adults and children,” *Journal of Phonetics*, vol. 33, no. 3, pp. 263–290, Jul. 2005.

[6] A. A. N. MacLeod and C. Stoel-Gammon, “What is the impact of age of second language acquisition on the production of consonants and vowels among childhood bilinguals?,” *International Journal of Bilingualism*, vol. 14, no. 4, pp. 400–421, Dec. 2010.

[7] P. Burgos, C. Cucchiarini, R. van Hout, and H. Strik, “Phonology acquisition in Spanish learners of Dutch: error patterns in pronunciation,” *Language Sciences*, vol. 41, Part B, pp. 129–142, Jan. 2014.

[8] P. Burgos, M. Jani, C. Cucchiarini, R. van Hout, and H. Strik, “Dutch Vowel Production by Spanish Learners: Duration and Spectral Features,” in *Interspeech*, Singapore, 2014, pp. 529–533.

[9] W. Baker, P. Trofimovich, J. E. Flege, M. Mack, and R. Halter, “Child-adult differences in second-language phonological learning: the role of cross-language similarity,” *Language and Speech*, vol. 51, no. Pt 4, pp. 317–342, 2008.

[10] J. I. Hualde, *The Sounds of Spanish*, Pap/Com edition. Cambridge, UK: Cambridge University Press, 2005.

[11] G. Booij, *The Phonology of Dutch*. Oxford, UK: Clarendon Press, 1995.

[12] P. Escudero and K. Wanrooij, “The Effect of L1 Orthography on Non-native Vowel Perception,” *Language and Speech*, vol. 53, no. 3, pp. 343–365, Sep. 2010.

[13] P. Escudero, R. Hayes-Harb, and H. Mitterer, “Novel second-language words and asymmetric lexical access,” *Journal of Phonetics*, vol. 36, no. 2, pp. 345–360, Apr. 2008.

[14] V. D. Erdener and D. K. Burnham, “The Role of Audiovisual Speech and Orthographic Information in Nonnative Speech Production,” *Language Learning*, vol. 55, no. 2, pp. 191–228, Jun. 2005.

[15] K. Chládková and P. Escudero, “Comparing vowel perception and production in Spanish and Portuguese: European versus Latin American dialects,” *The Journal of the Acoustical Society of America*, vol. 131, no. 2, pp. EL119–125, Feb. 2012.

[16] S. van der Harst, “The Vowel Space Paradox: A Sociophonetic Study on Dutch,” Ph.D. Thesis, LOT, 2011.

[17] P. Adank, R. Smits, and R. van Hout, “A comparison of vowel normalization procedures for language variation research,” *The Journal of the Acoustical Society of America*, vol. 116, no. 5, pp. 3099–3107, Nov. 2004.

[18] H. Franco, L. Neumeyer, V. Digalakis, and O. Ronen, “Combination of machine scores for automatic grading of pronunciation quality,” *Speech Communication*, vol. 30, no. 2–3, pp. 121–130, Feb. 2000.

[19] S. M. Witt and S. J. Young, “Phone-level pronunciation scoring and assessment for interactive language learning,” *Speech Communication*, vol. 30, no. 2–3, pp. 95–108, Feb. 2000.

[20] P. Burgos, E. Sanders, C. Cucchiarini, R. van Hout, and H. Strik, “Auris populi: crowdsourced native transcriptions of Dutch vowels spoken by adult Spanish learners,” in *Interspeech*, Dresden, Germany, 2015.