

Research Article

Pepi Burgos*, Roeland van Hout, Brigitte Planken

Matching Acoustical Properties and Native Perceptual Assessments of L2 Speech

<https://doi.org/10.1515/ol-2018-0011>

Received August 24, 2017; accepted February 28, 2018

Abstract: This article analyses the acoustical properties of Dutch vowels produced by adult Spanish learners and investigates how these vowels are perceived by non-expert native Dutch listeners. Statistical vowel classifications obtained from the acoustical properties of the learner vowel realizations were compared to vowel classifications provided by native Dutch listeners. Both types of classifications were affected by the specific set of vowels included as stimuli, an effect caused by the large variability in Spanish learners' vowel realizations. While there were matches between the two types of classifications, shifts were noted within and between production and perception, depending on the vowel and vowel features. We considered the variability between Spanish learners further by investigating individual patterns in the production and perception data, and linking these to the learners' proficiency level and multilingual background. We conclude that integrating production and perception data provides valuable insights into the role of different features in adult L2 learning, and how their properties actively interact in the way L2 speech is perceived. A second conclusion is that adaptive mechanisms, signalled by boundary shifts and useful in coping with variability of non-native vowel stimuli, play a role in both statistical vowel classifications (production) and human vowel recognition (perception).

Keywords: L2 vowel production, acoustical measurements, native transcriptions, segmental deviations, vowel confusions

1 Introduction

Adult learners have difficulties in acquiring the phonology of a second language (L2) (Birdsong & Molis 2001; Long 1990), and many of these difficulties are related to interference from their native language (L1) (Cutler 2012; Flege, Schirru & MacKay 2003). Several models, such as Flege's (1995) Speech Learning Model (SLM), have attempted to explain learners' difficulties in mastering the L2 phonological system in terms of the perceived similarity of segments in the L1 and L2. Flege and others agree that L2 contrasts based on fine-grained phonetic differences, particularly those contrasts that cover areas of the acoustic vowel space in which a single L1 native category is located, influence the perception and production of L2 phones, especially in the case of vowels (Baker & Trofimovich 2005; Best 1995; Bohn 1995; Escudero 2005; Flege, 1995; Flege et al. 2003; Major 2001; McAllister, Flege & Piske 2002).

How can we assess whether L2 vowels are accurately produced? One way to investigate L2 vowel production accuracy is to conduct acoustical analyses (i.e., an objective approach). Comparing the acoustical properties of the target vowels produced by both L2 learners and native speakers can help establish whether the L2 realizations match those of the native speakers, and if not, where and to what

*Corresponding author: Pepi Burgos, Centre for Language Studies, Radboud University Nijmegen, the Netherlands,

E-mail: pepiburgos@gmail.com

Roeland van Hout, Brigitte Planken, Centre for Language Studies, Radboud University Nijmegen, the Netherlands

degree there is a mismatch (e.g., Guion 2003; Iverson & Evans 2007). However, an analysis of L2 vowel production based on acoustics alone does not automatically account for human vowel recognition, that is, how L2 vowels are perceived by native listeners (i.e., a subjective approach). To determine the latter, we can ask native listeners to assess the intelligibility of L2 learners. Munro and Derwing (1995: 76), define intelligibility as “the extent to which a speaker’s message is actually understood by a listener”. Asking non-expert native listeners to orthographically transcribe L2 learners’ speech (Bent & Bradlow 2003; Derwing & Munro 1997) can help us establish whether a passage, sentence or word has been understood. Similarly, the identification of vowels can be assessed by asking native listeners to transcribe isolated, monosyllabic words containing target vowels, produced by L2 learners.

A limited number of studies have investigated the relationship between L2 vowel production and native listeners’ vowel perception. Van Wijngaarden (2001) examined the perception of Dutch vowels embedded in CVC nonsense words produced by native speakers of American English. Dutch vowels which had no equivalent in American English were identified less successfully by the native listeners. Van Wijngaarden’s (2001) findings show that vowels that are difficult for L2 subjects to produce are also difficult for native listeners to recognize. Munro (1993) studied the relationship between acoustical measurements (duration and spectrum) of English vowels produced by native speakers of Arabic and accentedness ratings of those vowel productions by linguistically trained native English listeners. His findings show that the native listeners rated the majority of L2 vowels as accented because they perceived durational and spectral deviances, most of which were attributed to specific characteristics of the Arabic vowel system. The Arabic-English vowels produced by Arabic speakers were found to differ in terms of their temporal properties from vowels produced by native English speakers. For instance, Arabic-English tense-lax pairs exhibited greater duration differences than native English tense-lax pairs. Surprisingly, Arabic speakers’ mean formant frequencies did not differ from those observed for native English speakers, whereas varying degrees of deviance from native English were found in individual productions. In addition, Munro (1993) concluded that degree of accentedness did not correlate with learners’ individual amount of experience, length of residence or daily use of English, suggesting that experience in the L2 does not guarantee success in achieving native-like pronunciation. Nevertheless, several studies have demonstrated that factors such as age of arrival, length of residence, formal instruction, amount of experience, L2 use and motivation play a role in L2 learning, particularly in the acquisition of L2 phones (cf. Moyer 2013; Piske, MacKay & Flege 2001 for a review). These factors may differently affect the performance of individual learners, who in turn may make use of different strategies in acquiring L2 vowels. This can result in large variability in vowel realizations, both within and across learners with the same L1-L2 pairing (cf. Bent, Baesse-Berk, Borrie & McKee 2016).

Variation between learners has been understudied in language acquisition research (cf. Mayr & Escudero 2010; Moyer 2013). It is still unclear what the specific factors are that drive this variability. For decades, studies on second language acquisition have assumed that most limitations in L2 development/acquisition follow from maturational (Critical Period Hypothesis; Lenneberg 1967), muscular (Scovel 1988) and/or cognitive constraints (cf. Moyer 2013). However, recent studies aimed at understanding individual differences in language learning have shown that psychological and social mechanisms also play a role in the accuracy with which a second language is acquired (cf. Larsen-Freeman 2009). The trade-off between these mechanisms may explain why, for instance, learners with the same L1 who are exposed to the same target language at the same age exhibit differences in their L2 development/acquisition. Intrinsic and extrinsic differences can account for this variation in individual performance, particularly when it comes to phonological learning (Moyer 2013). Intrinsic individual differences relate to, for example, differences in aptitude (e.g., mimicry ability: some learners have a special talent to learn languages and to imitate accents), musical talent (cf. Gottfried 2008; Tokuhama-Espinosa 2003), learning styles and strategies (cf. Kolb & Kolb 2005) and gender (cf. Moyer 2013). Extrinsic individual differences may stem from socio-psychological factors, such as identity, motivation and attitudes.

Two other factors that might play a role in individual variability are experience and input. Both are often measured in terms of amount of time (weeks, months, years) of L2 exposure, and in terms of length of residence (cf. Piske *et al.* 2001). However, such measures can lead to misleading assumptions, as amount of

L2 exposure or length of residence does not necessarily account for learners' L2 phonological accuracy (cf. Flege, Frieda & Nozawa 1997). With regard to input, Flege (2012) claims that both the quantity and quality of L2 input are essential to L2 speech learning and need to be taken into account when explaining variability in individual performance.

What the abovementioned studies have in common is that they show that the learners' linguistic performance correlates with non-linguistic factors, and that these factors may account for individual variability. An investigation by Mayr and Escudero (2010) sought to explain individual variation in terms of individual learners' distinct learning paths. The authors investigated – in a perceptual assimilation task and a forced choice identification task – whether individual native English learners of German follow different paths in their perception of six rounded German vowels. Their results showed that learners differed greatly from each other in the way they mapped L2 vowels onto native categories, and that this variation in the observed patterns was highly diverse, but at the same time systematic.

Research has shown that L2 perception (Mayr & Escudero 2010) and production (Munro 1993) differ greatly across individual L2 learners. The variability inherent in L2 learners' production implies that native listeners have to adapt to different pronunciations across learners. Native listeners can rapidly perceive segmental deviations from the norm and can easily detect pronunciation errors by learners that native speakers are not likely to make (Magen 1998). The variability inherent in L2 learners' production implies that native listeners have to adapt to different pronunciations across learners. For example, in the context of L2 vowel realizations, they have to be able to shift their category boundaries to accommodate an ambiguous vowel realization that differs from their usual expectations about phonemic categories (Cutler 2012). These perceptual adaptation processes show that native listeners are able to adjust the L1 boundary between categories to sustain successful language processing (cf. Bradlow & Bent 2008; Cutler 2012).

There is a need for empirical work on individual variation in L2 vowel production and on the perception of these L2 vowel productions by native listeners (cf. Mayr & Escudero 2010; Moyer 2013). The present study seeks to fill this gap. Subsequently, the aim of the present study and the Spanish and Dutch vowel systems will be presented.

1.1 The present study

The aim of the present study is to compare the results of two different approaches aimed at investigating Dutch L2 vowel production accuracy by adult Spanish learners, that is, an objective approach by which the acoustical properties of Dutch vowels produced by adult Spanish learners are investigated and a subjective approach based on the perception of these vowels by non-expert native Dutch listeners. This comparison contributes to answering the question whether Dutch listeners are capable of perceiving deviant L2 vowel realizations, which we know (on the basis of objective measurements) are not produced in a native-like fashion, and to understanding which dimensions or features are important for non-expert native Dutch listeners' perception, as attested from their transcriptions (i.e., subjective measurements).

Previous studies on the acoustical properties of Dutch vowels produced by adult Spanish learners of Dutch (cf. Chapter 3 in Burgos 2018 for the acoustic mapping of the Spanish-Dutch vowels onto the native Dutch vowels; see also Burgos, Jani, Cuccharini, Van Hout & Strik 2014b) and on the perceptual assessments of the same vowels by non-expert native Dutch listeners (cf. Chapter 5 in Burgos 2018; see also Burgos, Sanders, Cuccharini, Van Hout & Strik 2015) have shown that Spanish learners have problems with contrasts in vowel height, vowel length, front rounding and diphthongization. Considering previous findings and given the different types of vowel confusions observed in both production and native perception, we predict that the acoustic cues (vowel height, length, rounding and diphthongization) in the speech signal of the learners' vowel realizations will have a direct impact on both production and perception results. We expect that these acoustic cues may have different weightings in learners' production and, particularly, in native perception, and that they will affect the variability in the production and perception confusion patterns.

An additional aim of the present study is to explain the variability in L2 realizations, and how such variability may be related to individual differences. As well as comparing the acoustical measurements

of the Spanish learners' Dutch vowel productions with corresponding perceptual assessments of those productions by native Dutch listeners, we will examine variability in production and perception in the context of the learners' proficiency level and additional factors that may play a role in L2 vowel accuracy, such as prior linguistic knowledge in multilinguals (De Angelis 2007), length of residence and L2 use (cf. Piske et al. 2001).

1.2 The Spanish and Dutch vowel systems

We now consider the phonological properties of the Spanish and Dutch vowel systems. While Spanish has a straightforward five vowel system (/a, e, i, o, u/; see Figure 1) (Hualde 2005), Dutch has a complex vowel system containing 15 full vowels (monophthongs: /i, y, u, I, ʏ, ɔ, ε, a, a:/; long mid vowels: /e:, ø:, o:/; diphthongs: /ɛi, œy, ɔu/; see Figure 1), next to the reduced vowel /ə/ (Adank, Van Hout & Smits 2004a; Booij 1995). Four features characterize further differences between the two vowel systems (see Table 1). Spanish does not have phonemic vowel length (Hualde 2005; McAllister et al. 2002), whereas Dutch has a strict lax/tense distinction (lax vowels: /I, ε, ɔ, ʏ, a:/; tense vowels: /i, y, u, e:, ø:, o:, a:/), which crosses the short/long distinction (short vowels: /i, y, u, I, ʏ, ɔ, ε, a:/; long vowels: /a:, e:, ø:, o:, ɛi, œy, ɔu/) (Adank et al. 2004a). Spanish does not have the feature of front rounding, as all rounded vowels in Spanish are back vowels (/o, u/) (Hualde 2005), while Dutch has four front rounded vowels (/ʏ, y, ø, œy/). Spanish does not have diphthongs at the phoneme level, that is, single phonemes defined by their trajectory between two vowel positions; instead it has a rich inventory of 14 vowel combinations (Hualde 2005). Dutch, on the other hand, has diphthongs at the phoneme level, such as /ɛi, œy, ɔu/. The Dutch long mid vowels (/e:, ø:, o:/) are not considered to be full diphthongs, but they are slightly diphthongized (cf. Adank et al. 2004a; Van der Harst, Van de Velde & Van Hout 2014). Finally, Spanish is distinguished by three height values (high: /i, u/; mid: /e, o/; low: /a/) (Hualde 2005), whereas Dutch is characterized by four height values (high: /i, y, u/; high mid: /I, ʏ, e:, ø:/; low mid: /ɔ, ε, ɛi, œy, ɔu/; low: /a, a:/) (Booij 1995), as shown in Figure 1.

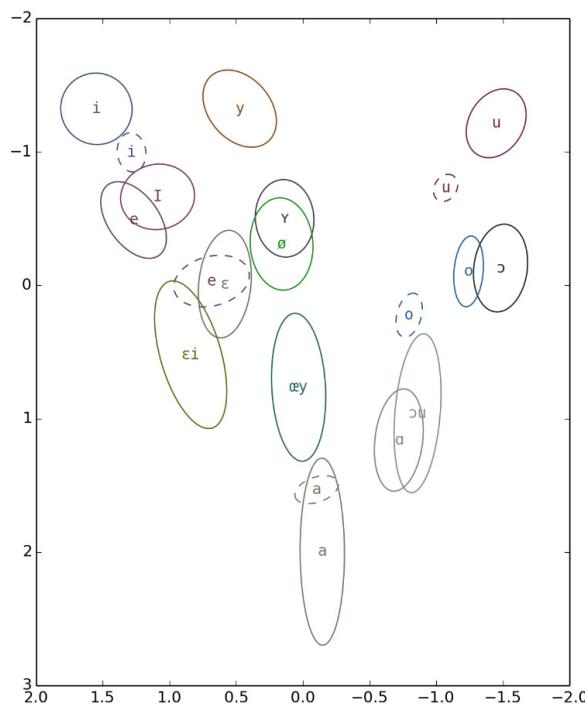


Figure 1. F1 and F2 (normalized) values for the five Spanish vowels (dashed lines), as realized by Spanish L1 speakers and for

the fifteen Dutch vowels, as realized by Dutch L1 speakers; all vowels are measured at 50% of the vowel duration; the mean values are indicated by the vowel symbols; the values for Spanish vowels were drawn from Chládková, Escudero and Boersma (2011); the values for Dutch values were drawn from Van der Harst (2011).

Table 1. Distinctive features of the Spanish and Dutch vowel systems.

	Spanish	Dutch	Comment
Length (duration)	—	+	Dutch has a complex interaction between the distinctions of lax/tense and short/long
Front rounding	—	+	Both languages have back rounded vowels
Diphthong	—	+	Spanish has 14 vowel combinations, but no diphthongs at the phoneme level
Height	3 levels	4 levels	Dutch has an additional height value, as can be observed in the vowel space of Figure 1

2 Method

2.1 Participants

2.1.1 Spanish learners

The speech of 28 adult Spanish learners of Dutch (9 males and 19 females), originating from Spain and a number of Latin American countries (Argentina, Dominican Republic, Guatemala, Mexico and Venezuela), was employed as stimulus material. We are aware of the phonetic differences between varieties of Spanish (Hualde 2005), and their possible influences on Dutch L2 perception (cf. Escudero and Williams 2012 in their investigation on Peruvian and Spanish learners of Dutch), but decided to pool all our Spanish L1 Dutch L2 speech data, as the perceptual differences reported in Escudero and Williams (2012) are negligible and do not bear out investigating these varieties of Spanish separately. In addition, phonetic differences between native speakers of Iberian and Latin American Spanish appear to be few and small when these speakers are highly educated (Navarro Tomás 2004: 7), as the participants in the present study are.

All the learners were living in the Netherlands at the time of the study and had already followed or were taking Dutch courses. All of them reported using Dutch in daily life. As all the learners were familiar with the Common European Framework of Reference for Languages (CEFR), they were asked to assess their own proficiency level in Dutch, and in other foreign languages they spoke, using the CEFR Self-Assessment Grid. They rated themselves in Dutch at one of the following four CEFR levels: A1 (CES (=Cambridge English Scales) 100-119), A2 (CES 120-139), B1 (CES 140-159) and B2 (CES 160-180) (UCLES 2015). Table 2 shows information about the Spanish learners per CEFR language proficiency level (cf. Chapter 3 in Burgos 2018 for more detailed information about the Spanish learners).

Table 2. Scores for speaker variables per CEFR language proficiency level; Age = age at the time of the recording, in years, AoA = age of arrival in the Netherlands, in years, LoR = length of residence in the Netherlands, in years, Use of Dutch = Self-estimated daily use of Dutch, in hours, Origin = where speakers were born and brought up.

CEFR	Age	AoA	LoR	Use of Dutch	Origin	
					Spain	Latin America
A1	32.2 (7.5)	30.2 (6.6)	2.0 (3.1)	6.7 (3.5)	8	2
A2	31.6 (5.3)	30.6 (5.5)	0.9 (0.4)	7.8 (2.5)	6	1
B1	40.2 (1.2)	32.2 (3.3)	8.0 (3.9)	5.0 (2.0)	2	2
B2	41.0 (7.8)	30.8 (7.8)	10.1 (5.9)	13.1 (8.0)	5	2
Overall	35.4 (7.6)	30.8 (6.0)	4.6 (5.3)	8.3 (5.4)	21	7
Range	20–52	19.3–42.0	0.1–20.0	2–24		

2.1.2 Native Dutch listeners

A snowball sampling strategy had been employed to recruit native listeners in an earlier study on the perception of Spanish-Dutch speech (cf. Chapter 5 in Burgos 2018). This sampling technique consists of recruiting subjects from the social networks of a starting set of individuals. In the earlier study, each individual in the starting set was asked to recruit at least five native speaker subjects from his/her networks of family and friends in order to reach a heterogeneous group of native listeners. The individuals themselves could not take part in the experiment. The same sampling technique was used in the present study. The starting set of 25 individuals (7 males and 18 females) in the present study were all undergraduate students of International Business Communication (IBC, Department of Communication and Information Studies) at the Radboud University Nijmegen, in the Netherlands. The native listeners they recruited had to meet the following criteria: 1) at least 18 years old, 2) native Dutch speaker, 3) not linguistically trained, and 4) unfamiliar with Spanish-accented Dutch. A total of 139 native Dutch listeners who met the criteria were recruited. They were all asked to participate in a transcription task, which required them to transcribe the non-native (Spanish-Dutch) stimuli. A total of 132 native Dutch listeners (59 males and 73 females) completed the transcription task. The transcriptions of seven listeners who did not complete the task were discarded. The listeners were heterogeneous in terms of age (range: 18–66 yr old, $M = 32.39$, $SD = 16.26$) and completed education (elementary school ($n = 4$), high school ($n = 73$), vocational training education ($n = 24$), higher professional education ($n = 24$), university degree ($n = 7$)). It should be noted that the transcription task consisted of the non-native tokens interspersed with native tokens of two native Dutch speakers (one male and one female) from the corpus of native speakers of Standard Dutch mentioned earlier. Native stimuli were included to increase the validity of the task, as these native tokens could be used by the transcribers as anchor points (see Chapter 5 in Burgos 2018 for more information about the native Dutch listeners). The transcriptions of both the non-native and native tokens of all 132 listeners were used in subsequent analyses.

2.2 Stimulus materials

The stimulus materials in the present study are from an existing corpus of Spanish L1 Dutch L2 (cf. Burgos et al. 2014b) which includes systematic productions of Dutch sounds that are problematic for Spanish learners. The material we used in the present study comprises a list of Dutch monosyllabic words read out by adult Spanish learners. The same material was used by Van der Harst (2011), and Van der Harst et al. (2014), who obtained recordings of the same list of Dutch words produced by native Dutch speakers. The word list used in the original corpus comprises 278 monosyllabic and disyllabic words representing all the Dutch vowels in different contexts. For our study, we employed a subset of these 278 words, namely 29 Dutch monosyllabic words per speaker. This subset of 29 Dutch words included all 15 Dutch vowels in stressed position followed either by /s/ or /t/, as vowel quality is known to alter only minimally when the vowel is followed by these consonantal sounds (Van der Harst 2011: 146; Van der Harst et al. 2014: 254). Table 3 provides an overview of the 29 Dutch words containing all 15 Dutch vowels, and their corresponding phonological and orthographic representations. No example of a word containing the vowel /y/ followed by /s/ was included, as this combination does not exist in Dutch monosyllabic nouns.

All non-native tokens were recorded in a quiet room at the Linguistics Department of the Radboud University or at the speakers' home, using a headset (Logitech, USB entry DZL-A-0001 4-B) and a laptop (ACER AMD Quad-core Processor A6-3400M with Turbo CORE Technology up to 2.30 GHz). The data were recorded at a sampling frequency of 16 kHz. The Spanish learners read out loud the Dutch words, which were presented on a computer screen one by one, at intervals of three seconds. Each word from the set of 29 words was recorded by each speaker only once, which resulted in 812 word tokens (29 words x 28 speakers). Six speech samples (from six different subjects) were discarded due to erroneous recording. Thus, a total of 806 word tokens from the Spanish learners were subjected to analysis.

We used the same set of 29 words spoken by 20 native speakers of Standard Dutch (10 males and 10 females), collected by Van der Harst (2011) and Van der Harst et al. (2014) to describe the Dutch vowel

system. The Dutch samples were analysed (see below) using similar techniques as those employed for the non-native data.

Table 3. Selected –s and –t words as speech stimuli (Van der Harst 2011); Phon = phonological representation (in IPA), Orth = orthographic representation.

Vowel	Phon	Orth	s-word	t-word	
Monophthongs	/i/	<ie>	<i>kies</i>	/kis/	<i>riet</i>
	/y/	<uu>	-	/y/	<i>fuut</i>
	/u/	<oe>	<i>poes</i>	/pus/	<i>voet</i>
	/ɪ/	<i>	<i>vis</i>	/vɪs/	<i>fit</i>
	/ʏ/	<u>	<i>zus</i>	/zys/	<i>put</i>
	/ɔ/	<o>	<i>vos</i>	/vɔs/	<i>vlot</i>
	/ɛ/	<e>	<i>zes</i>	/zɛs/	<i>vet</i>
	/ɑ/	<a>	<i>gas</i>	/xas/	<i>rat</i>
Long mid vowels	/a:/	<aa>	<i>aas</i>	/a:s/	<i>staat</i>
	/e:/	<ee>	<i>mees</i>	/me:s/	<i>beet</i>
	/ø:/	<eu>	<i>neus</i>	/nø:s/	<i>neut</i>
Diphthongs	/o:/	<oo>	<i>boos</i>	/bo:s/	<i>boot</i>
	/ɛi/	<ij>	<i>ijs</i>	/ɛis/	<i>spijt</i>
	/œy/	<ui>	<i>huis</i>	/hœys/	<i>fluit</i>
	/ɔu/	<ou>	<i>kous</i>	/kɔus/	<i>fout</i>

2.3 Analysis of the speech recordings

The words read by the native Dutch speakers and by the Spanish learners were orthographically transcribed in Praat (Boersma & Weenink 2010), and subsequently segmented following the procedures for vowel segmentation described in Van Son, Binnenpoorte, Van den Heuvel and Pols (2001), and Van der Harst (2011). The segmentation of vowels was done by an experienced transcriber, the first author, who manually segmented each word at the phoneme level. The onset and end of each vowel were determined by looking at the waveform (for example, structure and amplitude). We also looked at information from the spectrogram, formant tracks and auditory cues to determine the beginning and the end of each vowel. The segmentation of vowels was then checked by a native Dutch phonetician, who, where necessary, altered the onset and end of vowels that had already been segmented.

2.4 Acoustical analyses

Acoustical analyses were performed to extract measurements of the first and second formants (F1 and F2) and of the duration of the Dutch vowels produced by adult Spanish learners. It should be noted that measurements of third formant (F3) were not analysed in the current study, as previous research has shown that F3 is not essential in the identification of front rounded vowels in Dutch, and that using F1 and F2 only is sufficient to identify these vowels (Adank 2003; Cohen, Slis & 't Hart 1963 1967; Van der Harst 2011). The first two formants were measured at three equidistant points (i.e., at 25%, 50% and 75% of the vowel duration). This information helps to determine if and how diphthongization is realized by Spanish learners producing all mid vowels and diphthongs in comparison to native speakers, as mid vowels and diphthongs in Dutch are long and show a milder or stronger degree of diphthongization (Adank et al. 2004a; Van der

Harst et al. 2014). All measurements were automatically extracted using an LPC (Linear Predictive Coding) analysis in Praat (Boersma & Weenink 2010). First, every vowel token was assigned a specific number of coefficients, i.e., four, five or six coefficients, based on information from the waveform, spectrogram and formant tracks of the speech signal. Next, an LPC script based on the chosen number of coefficients was run in order to extract F1 and F2 values. The same procedure was repeated for the measurement extractions of duration. All resulting measurements were manually checked by the first author who corrected any errors found. Subsequently, an additional check for outliers, checked at 25%, 50% and 75%, was carried out by the same native Dutch phonetician mentioned earlier, who followed the procedure employed in Van der Harst (2011) and Van der Harst et al. (2014), and corrected any errors found. All vowel realizations were then normalized using Lobanov's (1971) transformation to neutralize formant frequency variations resulting from anatomic differences among informants (cf. Adank, Smits & Van Hout 2004b). Durational values were standardized (*z* scores) for each Spanish learner.

3 Results

Subsection 3.1 presents the results of the acoustic measurements of the non-native and native data. Subsection 3.2 focuses on the native listeners' perceptions of Dutch vowels produced by the adult Spanish learners and the native Dutch speakers. The final subsection, 3.3, compares the results of the production study (acoustic data) with those of the perception study (listener data).

3.1 Acoustic data

The Dutch vowels produced by Spanish learners and native Dutch speakers were analysed using a multinomial logistic regression, which is a statistical classification technique. It is used to predict a vowel classification using a categorically distributed dependent variable, given a set of predictor variables. Based on the acoustic values of the non-native and native speech data, the regression calculates the probabilities of canonical and non-canonical classifications of the vowel realizations. A canonical classification means that the vowel realization is classified as the target vowel, whereas a non-canonical means that the vowel realization is classified under a different vowel (which is not the target vowel). A given vowel realization is classified in the vowel category with the highest probability.

We investigated the non-native and native data using three classification conditions to determine to what degree outcomes depend on the vowel sets to be classified, namely "Total", "Group" and "Individual". In the classification condition "Total", we pooled the non-native and native data, in "Group", the non-native and native data were treated as two independent sets or groups, and in "Individual", the individual Spanish learner data were added, learner by learner, to the data of the native group. The regressions for each classification condition were conducted using F1 and F2 only, and F1, F2 plus vocalic duration. In this way, the analyses could throw light on the extent to which duration plays a role in vowel classification.

3.1.1 Non-native data

We first focus on the results in the three classification conditions. Subsequently, we present the non-native matrix that we obtained in the condition "Group" using F1, F2 and duration.

Table 4 presents the means and standard deviations of canonical classifications for the three conditions "Total", "Group" and "Individual", with F1 and F2 (at 25%, 50%, 75% of the vowel duration) only, and with F1, F2 and duration. We also included the target vowels whose percentages change (Vowel + dur ↑) and do not change (Vowel + dur ↓) after including duration; deviations $\leq 2.5\%$ after including duration are rated as unchanged (Vowel + dur ↔). It can be seen that the average percentage of canonical classifications in the condition "Total" with F1, F2 only, at 61.1%, increases to 72.5% after including duration. A similar increase can be seen in "Group" (61.7% → 74.7%) and "Individual" (72.4% → 89.1%).

However, an increase in the average percentage of canonical classifications does not necessarily mean that each target vowel benefits equally from including duration. Upon closer inspection, we see that in the “Total” condition, almost all target vowels benefit from duration, and adding duration is not detrimental to any target vowel. But three target vowels, the diphthongs *<ij>*, *<ui>* and *<ou>*, are not affected by adding duration; the percentages of canonical classifications for these vowels remain unchanged. Along similar lines, in the “Group” condition, almost all target vowels benefit from duration, but the canonical classification percentages for the target vowels *<e>* and *<ui>* are not affected by duration. Finally, in the “Individual” condition, the canonical transcription percentages for 13 target vowels increase by adding duration, but those for *<ie>* and *<e>* are not affected. In sum, this indicates that with duration, the average percentage of canonical classifications is consistently higher in each classification condition, and for the majority of target vowels. This improvement indicates that duration contributes substantially to an increase in the probability of a canonical classification.

Table 4. Scores for the canonical classifications of the Dutch vowels produced by Spanish learners for the conditions “Total”, “Group” and “Individual”, with both F1 and F2 (25%, 50%, 75%) only, and F1, F2 and duration, and target vowels whose percentages (do not) change after including duration, deviations $\leq 2.5\%$ after including duration are rated as unchanged; Classification C = classification condition, + dur = including duration.

Classification C	F1, F2	F1, F2 + dur	Vowel + dur ↑	Vowel + dur ↓	Vowel + dur ↔
Total	61.1 (15.0)	72.5 (13.8)	<i><ie></i> , <i><uu></i> , <i><oe></i> , <i><i></i> , <i><u></i> , <i><o></i> , <i><e></i> , <i><a></i> , <i><aa></i> , <i><ee></i> , <i><eu></i> , <i><oo></i>	-	<i><ij></i> , <i><ui></i> , <i><ou></i>
Group	61.7 (15.0)	74.7 (13.3)	<i><ie></i> , <i><uu></i> , <i><oe></i> , <i><i></i> , <i><u></i> , <i><o></i> , <i><a></i> , <i><aa></i> , <i><ee></i> , <i><eu></i> , <i><oo></i> , <i><ij></i> , <i><ou></i>	-	<i><e></i> , <i><ui></i>
Individual	72.4 (17.9)	89.1 (13.2)	<i><uu></i> , <i><oe></i> , <i><i></i> , <i><u></i> , <i><o></i> , <i><a></i> , <i><aa></i> , <i><ee></i> , <i><eu></i> , <i><oo></i> , <i><ij></i> , <i><ui></i> , <i><ou></i>	-	<i><ie></i> , <i><e></i>

An improvement (see Table 4) is also observed across the three conditions (“Total” → “Group” → “Individual”). The average percentage of canonical classifications using F1 and F2 slightly improves from “Total” (61.1%) to “Group” (61.7%), and a substantial improvement is observed from “Group” to “Individual” (72.4%). A similar improvement pattern from “Total” (72.5%) to “Group” (74.7%) to “Individual” (89.1%) is found when duration is added (F1, F2 + duration).

The inclusion of native speakers of Dutch in the “Total” condition does not appear to change the canonical scores. The improvement to classifications in the “Group” and “Total” conditions are comparable. The “Individual” condition gives a boost to the canonical scores. In this condition, the vowels of one Spanish learner are classified amongst the vowels of all 20 native speakers. The “Individual” condition can only produce an improvement if the Spanish learners actually produce distinct acoustic properties between the target vowels, however weak. Adding more Spanish learners seems to blur these already vulnerable distinctions, and this effect is probably augmented by the fact that the distinctions are highly variable across and within learners.

Is this improvement in the average percentage of canonical classifications across classification conditions – after including duration – also found for the canonical classifications per target vowel? Table 5 shows the degree of improvement in the percentages of canonical classifications per target vowel and the average percentage of canonical classifications per condition (“Total” → “Group” → “Individual”), using F1 F2 and duration, for all 15 Dutch vowels produced by Spanish learners.

The average percentage of canonical classifications in “Total” (72.5%) is similar to that in “Group” (74.7%), showing only a modest improvement of 2.2%. The differences between the target vowels were only slight (Cohen’s h ; all values were lower than .20; $h \geq .20$ is a small effect). A medium improvement (14.4%; $h \geq .50$) is observed from “Group” (74.7%) to “Individual” (89.1%), with slight to large ($h \geq .80$) differences between the individual target vowels. The rounded target vowels, in particular, seem to benefit (large Δ *<uu>* = 39.3%, medium Δ *<u>* = 12.5%, medium Δ *<eu>* = 20.0%, large Δ *<ui>* = 43.7%), whereas the outcome for *<i>* is the only negative pattern (medium Δ = -26.8%), which may indicate that the distinction *<i>* versus *<ie>* is hard to classify at the level of individual Spanish learners. The striking improvement in the

average percentage of canonical classifications from “Group” to “Individual”, as well as in the percentages of canonical classifications per target vowel (with the exception of <i>) shows that the individual Spanish learners are making distinctions between the Dutch vowels, however weak and variable these may be.

Table 5. Degree of improvement for the percentages of canonical classifications per target vowel and per condition (“Total”(T)→“Group”(G)→“Individual”(I)) for all 15 Dutch vowels produced by Spanish learners (F1, F2 + duration); Vowel = target vowel, Δ = difference between the percentages across conditions, %Can = average percentage of canonical classifications.

Vowel	ie	uu	oe	i	u	o	e	a	aa	ee	eu	oo	ij	ui	ou	%Can
Total	62.5	53.6	76.4	69.6	66.1	76.8	89.1	85.7	78.6	83.6	81.8	85.7	76.8	38.2	63.6	72.5
Group	67.9	53.6	72.7	76.8	71.4	83.9	90.9	83.9	83.9	85.5	78.2	83.9	82.1	41.8	63.6	74.7
Δ T-G	5.4	-	-3.7	7.2	5.3	7.1	1.8	-1.8	5.3	1.9	-3.6	-1.8	5.3	3.6	-	2.2
Individual	69.6	92.9	94.5	50.0	83.9	89.3	96.4	94.6	98.2	96.4	98.2	96.4	94.6	85.5	96.4	89.1
Δ G-I	1.7	39.3	18.1	-26.8	12.5	5.4	5.5	10.7	14.3	10.9	20.0	12.5	12.5	43.7	32.8	14.4

We now focus on the outcomes of the non-native acoustic data in the classification condition “Group”, as this is the condition, unlike “Total”, in which the non-native data is treated as an independent group. Therefore, the outcomes in this condition can best classify the Spanish-accented Dutch vowels based on their own characteristics. We present the results in the “Group” condition employing F1, F2 and duration, as formant frequencies (F1 and F2) and duration are both important properties to be taken into account when characterizing Dutch vowels.

Table 6 presents the matrix representing the classification of the Spanish learners’ Dutch vowel realizations, using a multinomial logistic regression. The columns represent the 15 target vowels corresponding to the nine monophthongs (<ie>, <uu>, <oe>, <i>, <u>, <o>, <e>, <a>, <aa>), the three long mid vowels (<ee>, <eu>, <oo>), and the three diphthongs (<ij>, <ui>, <ou>). The rows show the overall percentages of canonical (marked green) and non-canonical classifications for the 15 target vowels. The column Total shows the average percentage of the sum of all percentages of classified vowels per row.

Table 6. Probability ratio of canonical (indicated in green) and non-canonical classifications for all 15 Dutch vowels produced by Spanish learners in the condition “Group” (F1, F2 + duration); target vowels in the columns, classified vowels in the rows, non-canonical classifications with deviations > 2.5% related to vowel height (in pink), vowel length (in turquoise), rounding (in yellow) and diphthongization (in teal) are also indicated; Vow = target vowel.

Vow	ie	uu	oe	i	u	o	e	a	aa	ee	eu	oo	ij	ui	ou	Total
ie	67.9	0.0	0.0	14.3	1.8	0.0	0.0	0.0	0.0	9.1	0.0	0.0	0.0	1.8	0.0	6.6
uu	0.0	53.6	3.6	0.0	5.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.6	0.0	2.7
oe	0.0	21.4	72.7	0.0	14.3	3.6	0.0	0.0	0.0	0.0	0.0	5.4	0.0	1.8	3.6	7.7
i	21.4	0.0	0.0	76.8	5.4	0.0	1.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	7.3
u	1.8	14.3	9.1	3.6	71.4	3.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.8	6.8
o	0.0	0.0	3.6	0.0	1.8	83.9	0.0	1.8	0.0	0.0	3.6	0.0	0.0	0.0	3.6	6.8
e	0.0	0.0	0.0	5.4	0.0	0.0	90.9	1.8	0.0	3.6	3.6	0.0	3.6	1.8	0.0	7.6
a	0.0	0.0	0.0	0.0	0.0	0.0	0.0	83.9	14.3	0.0	0.0	0.0	5.5	1.8	7.3	
aa	0.0	0.0	0.0	0.0	0.0	0.0	0.0	7.1	83.9	0.0	0.0	0.0	0.0	5.5	6.7	
ee	3.6	0.0	0.0	0.0	0.0	0.0	3.6	0.0	0.0	85.5	0.0	0.0	8.9	0.0	0.0	6.9
eu	0.0	3.6	1.8	0.0	0.0	0.0	1.8	0.0	0.0	78.2	0.0	1.8	12.7	0.0	0.0	6.7
oo	0.0	7.1	7.3	0.0	0.0	8.9	0.0	0.0	0.0	83.9	0.0	0.0	7.3	0.0	7.7	
ij	5.4	0.0	0.0	0.0	0.0	0.0	1.8	1.8	0.0	1.8	0.0	0.0	82.1	9.1	0.0	7.1
ui	0.0	0.0	1.8	0.0	0.0	0.0	0.0	3.6	1.8	0.0	16.4	0.0	3.6	41.8	12.7	5.6
ou	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.8	7.1	0.0	21.8	63.6	6.5	

The results show that 74.7% of all classifications are canonical, whereas 25.3% are non-canonical. The highest percentage of canonical classifications is found for the vowel <e> (90.9%), and the lowest for <ui> (41.8%).

The variability in non-canonical classifications indicates that the Dutch vowels produced by the learners were classified on the basis of their acoustic properties as different vowels than the target vowels. The highest difference was found for the target diphthongs <ui> and <ou>, classified as eight and seven different non-canonical vowel categories respectively, whereas the lowest differentiation was found for the target vowel <aa>, assigned to only two different vowel categories, followed by the target vowels <i>, <o>, <ee>, <eu>, and <oo>, with three categories each.

Upon closer inspection, Table 6 shows non-canonical classifications related to problems with vowel height (e.g., in the vowel confusion <i>-<ie>), vowel length (e.g., in the confusions <a>-<aa> and <o>-<oo>), rounding (e.g., in the confusions <uu>-<u>-<oe> and <ui>-<ou>), and diphthongization (e.g., in the confusions <eu>-<ui> and <oo>-<ou>). Clearly, these problems relate to the four distinctive features listed in Table 1. An examination of the most conspicuous vowel confusions with values higher than 10% brings us to the pair <i>-<ie>. The target vowel <ie> (67.9%) is extensively classified as <i>, as attested by the high percentage of non-canonical classifications, namely 21.4%. In contrast, the target vowel <i> (76.8%) is less frequently classified as <ie> (14.3%). Asymmetry is also found in confusions related to vowel length, as is the case for the vowel pair <a>-<aa>. The target vowel <aa> (83.9%) is often classified as <a> (14.3%), more than <a> is as <aa> (7.1%). A similar situation applies to the target vowel <o> (83.9%), which is frequently classified as non-canonical <oo> (8.9%). Vowel confusions related to rounding are found for the target front rounded vowels <uu> (53.6%) and <u> (71.4%), which are frequently confused with each other, and, especially, with the back rounded vowel <oe>, yielding non-canonical percentages of 21.4% and 14.3% respectively. Rounding is also involved in the <ui>-<ou> confusion. The target front rounded vowel <ui> (41.8%) is often classified as <ou>, and is given the highest percentage of non-canonical classifications in the non-native matrix (21.8%), a higher percentage than for the target back rounded vowel <ou> (63.6%) which is classified as <ui> (12.7%). As to diphthongization, the target long mid vowel <eu> (78.2%) is frequently classified as <ui> (16.4%), more than <ui> is classified as <eu> (12.7%). This indicates that the long mid vowel <eu> is extremely diphthongized by the Spanish learners.

3.1.2 Native data

The Dutch vowels produced by native Dutch speakers were analysed in the classification conditions “Total”, in which the native and non-native data are pooled together, and in the classification condition “Group”, in which the native and non-native data are treated as two independent groups, using F1 and F2 only, and F1, F2 and duration.

Table 7 shows the average percentages of canonical classifications for the two classification conditions “Total” and “Group”, with both F1 and F2 (at 25%, 50%, 75% of the vowel duration) only, and F1, F2 and duration. The target vowels whose percentages change (Vowel + dur ↑) and do not change (Vowel + dur ↓) after including duration are also included; deviations ≤ 2.5% after including duration are rated as unchanged (Vowel + dur ↔).

The average percentage of canonical classifications in “Total” with F1, F2 only, at 84.3%, increases to 91.3% after including duration. A similar increase after adding duration is seen in “Group” (95.2% → 99.2%). In “Total”, ten target vowels benefit from duration, whereas <ie> and <uu> show a decrease. Three target vowels, <oe>, <ij> and <ui> (with deviations ≤ 2.5%), are not affected by adding duration. In “Group”, six target vowels benefit from duration, while the other nine target vowels do not. Overall, including duration leads to consistently higher percentages of canonical classifications per target vowel, although this seems to be less beneficial for the native vowel classifications than it is for the non-native vowel classifications.

The average percentage of canonical classifications of the native data using F1 and F2 only improves from “Total” (84.3%) to “Group” (95.2%), with a greater improvement (10.9%) than that found for “Total” → “Group” in the non-native data (i.e., 2.1%; see Table 5). A similar improvement to that in the non-native data is found

when duration is included (F1, F2 + duration): “Total” (91.3%) → “Group” (99.2%). This substantial improvement seems to suggest that the presence of non-native data in “Total” detrimentally affects the results of the native speech data, indicating that the statistical classifier adapts its classification when non-native data are included.

Table 7. Scores for the canonical classifications of the Dutch vowels produced by native Dutch speakers for the conditions “Total” and “Group”, with both F1 and F2 (25%, 50%, 75%) only, and F1, F2 and duration, and target vowels whose percentages (do not) change after including duration, deviations $\leq 2.5\%$ after including duration are rated as unchanged; Classification C = classification condition, + dur = including duration.

Classification C	F1, F2	F1, F2 + dur	Vowel + dur ↑	Vowel + dur ↓	Vowel + dur ↔
Total	84.3 (11.7)	91.3 (8.3)	<i>,<u>,<o>,<e>,<a>,<aa>,<ie>,<uu> <ee>,<eu>,<oo>,<ou>	<oe>,<ij>,<ui>	
Group	95.2 (6.2)	99.2 (2.2)	<i>,<u>,<o>,<ee>,<eu>,<oo>	-	<ie>,<uu>,<oe>,<e>,<a>,<aa>,<ij>,<ui>,<ou>

To understand this mechanism better, Table 8 shows the native matrix in the classification condition “Total” with F1, F2 and duration. Most non-canonical classifications seem to be related to the vowel confusions observed in the non-native matrix, for example, to the vowel confusions in the pairs <ie>-<i> and <i>-<ie>, associated with vowel height. The <uu>-<u> confusion is related to height as well. Other confusion patterns, associated with rounding, are observed too, such as the non-canonical classifications of the target front rounded <uu> as the front unrounded <i>, or of the target back unrounded <a> as the back rounded <o>. Diphthongal confusions are found for <eu> and <oo>. The feature that does not reflect problems in this matrix is length, indicating that duration does not lead to confusions in the native data because it is a secondary feature in the native-produced vowel distinctions. These results seem to suggest that problematic features found in the statistical classifications of the non-native vowels recur in the classifications of the native vowels when these features are of primary relevance in the native-produced vowel distinctions. The classifier would seem to have adapted to the great variability in the vowel realizations in the non-native data with detrimental results for the native data as a consequence (cf. Berck 2017 who shows that machine learning algorithms are affected by infusing errors in linguistic data).

Table 8. Probability ratio of canonical (indicated in green) and non-canonical classifications for all 15 Dutch vowels produced by native Dutch speakers in the condition “Total” (F1, F2 + duration); target vowels in the columns, classified vowels in the rows, non-canonical classifications with deviations $> 2.5\%$ related to vowel height (in pink), vowel length (in turquoise), rounding (in yellow) and diphthongization (in teal) are also indicated; Vow = target vowel.

Vow	ie	uu	oe	i	u	o	e	a	aa	ee	eu	oo	ij	ui	ou	Total
ie	85.0	0.0	0.0	17.5	0.0	0.0	0.0	0.0	0.0	2.5	0.0	0.0	0.0	0.0	0.0	7.2
uu	2.5	75.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.5	0.0	0.0	0.0	0.0	2.9
oe	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	6.9
i	12.5	5.0	0.0	77.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	6.4
u	0.0	20.0	0.0	0.0	95.0	0.0	7.5	0.0	0.0	0.0	2.5	0.0	0.0	0.0	0.0	7.9
o	0.0	0.0	0.0	0.0	0.0	97.5	0.0	7.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	7.2
e	0.0	0.0	0.0	2.5	5.0	0.0	90.0	0.0	0.0	0.0	2.5	0.0	0.0	0.0	0.0	6.9
a	0.0	0.0	0.0	0.0	0.0	0.0	92.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	6.4
aa	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	2.5	0.0	7.1
ee	0.0	0.0	0.0	2.5	0.0	0.0	0.0	0.0	95.0	0.0	0.0	7.5	0.0	0.0	0.0	7.2
eu	0.0	0.0	0.0	0.0	0.0	2.5	0.0	0.0	0.0	80.0	0.0	2.5	0.0	0.0	0.0	5.9
oo	0.0	0.0	0.0	0.0	0.0	2.5	0.0	0.0	0.0	0.0	95.0	0.0	0.0	0.0	0.0	6.7
ij	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.5	0.0	0.0	90.0	0.0	0.0	0.0	6.4
ui	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	12.5	0.0	0.0	100.0	0.0	0.0	7.8
ou	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	5.0	0.0	0.0	97.5	0.0	7.1

3.2 Listener data

In this subsection, we focus on perception, by examining the native Dutch listeners' transcriptions of the Dutch vowels produced by adult Spanish learners (non-native matrix) and by native Dutch speakers (native matrix).

It should be noted that the non-native and native matrices consist of 15 columns representing the 15 target vowels x 15 rows representing the classified vowels. This was not the case for the non-native and native matrices of the listener data in the earlier perception study (cf. Chapter 5 in Burgos 2018), which contained 15 columns representing the 15 target vowels x 20 rows representing the transcribed vowels. That is, the 20 rows consisted of 15 rows for the 15 Dutch vowels and five additional rows: one row for the frequent non-canonical variant *<ai>* (over diphthongized vowel combination assigned to the target vowel *<ij>*), and four rows in the Rest category (containing transcribed vowels < 5%), including non-canonical variants related to longer duration, diphthongization, other transcriptions and consonants. With a view to comparing the non-native and native matrices of the acoustic data and the non-native and native matrices of the listener data, we decided to alter the number of rows in the original matrices of the listener data. We subsumed the *<ai>* transcriptions under *<ij>*, and, subsequently, distributed the percentages of the four rows in the Rest category throughout the remaining 15 rows representing the transcribed vowels. This allowed us to compare the two 15x15 matrices for the non-native (see Table 9) and native data (see Table 10).

3.2.1 Non-native data

Table 9 presents the non-expert native Dutch listeners' transcriptions of the Dutch vowels produced by the 28 adult Spanish learners. The rows show the overall percentages of canonical (indicated in green) and non-canonical transcriptions of the 15 target vowels.

Our results show that 65.4% of all classifications are canonical, whereas 34.6% are non-canonical. The highest percentage of canonical transcriptions was for the vowel *<e>* (87.3%), while the lowest was for the *<uu>* (33.8%). The last two vowels in the list, namely, the front rounded *<u>* and *<uu>*, do not occur in Spanish and can be considered new for Spanish learners (cf. Flege 1995).

The most striking difference in comparing Table 9 to Table 6 (the acoustic data), is that the non-native matrix of the listener data is much more distributed, in that it shows more variability in the vowel confusions. This indicates that the listeners perceived the vowels spoken by the Spanish learners in various ways. The highest variability was found for *<ui>* and *<eu>*, assigned to 14 and 13 different non-canonical vowel categories respectively. The degree of variability here might be related to the fact that *<ui>* and *<eu>* are front rounded vowels. The lowest variability is found for the target vowel *<i>*, even though it is assigned to as many as six different non-canonical vowel categories.

The non-canonical transcriptions in Table 9 clearly show what the vowel confusions are, and therefore, which features were perceived by the native listeners to be (erroneously) employed by the Spanish learners when producing the Dutch target vowels. The most outspoken vowel confusion, related to vowel height, is observed for the target vowel *<i>* which is assigned to *<ie>* (44.2%). Non-canonical transcriptions of the target vowel *<ie>* as *<i>* (13.7%) were also found, but at a substantially lower percentage, indicating that there is an asymmetrical confusion between these vowels. Other confusions related to vowel height were observed in the target vowel *<u>*, which was frequently transcribed as *<uu>* (10.2%), more than *<uu>* as *<u>* (4.0%). The target diphthong *<ou>* was often perceived as *<oo>* (19.0%), although *<oo>* was seldom transcribed as *<ou>* (1.8%). The most conspicuous confusions related to vowel length were found in the short monophthongs *<uu>*, *<i>*, *<o>* and *<a>*, which were often perceived as having longer duration, and in the long vowel *<aa>* and the long mid vowels *<ee>*, *<eu>* and *<oo>*, which were perceived as monophthongs with shorter duration. In other words, the target short monophthongs were often perceived as long vowels, and the target long vowels as short vowels. An asymmetrical confusion related to vowel length can be seen in the vowels in the pair *<a>*-*<aa>*, which are hard to distinguish, with *<a>* more often perceived as *<aa>* (30.8%) than *<aa>* as *<a>* (14.2%).

With respect to the front rounded vowels, Table 9 shows that the front rounded monophthongs <uu> and <u> are perceived as the back rounded vowel <oe> (30.3% and 37.0% respectively). A similar pattern was found for the front rounded diphthong <ui> which was often transcribed as the back rounded diphthong <ou> (14.6%). However, the front rounded long mid vowel <eu> was perceived differently: either as a front vowel (<ij>, 3.9%) or as a back rounded vowel (<oo>, 9.6%). We found fewer vowel confusions related to diphthongization than to vowel height, vowel length and rounding. The target long mid vowels <ee> and <eu> were often perceived as <ij> and <ui> respectively, indicating that they were over diphthongized. It should be noted that the results reported in Burgos (2018, Chapter 5) also show that native Dutch listeners perceived the extreme diphthongization with which some Dutch vowels were produced by the Spanish learners, especially in the case of the long mid vowels and diphthongs. Finally, the frequent assignment of the non-canonical variant <ai> and the various non-standard transcriptions included in the Rest category in the original non-native matrix indicate that Spanish learners do have problems with diphthongization, even though the non-native matrix reflects fewer confusions (see Table 9).

Table 9. Most frequent canonical (indicated in green) and non-canonical transcriptions of all 15 Dutch vowels produced by Spanish learners, as given by non-expert native Dutch listeners; target vowels in the columns, classified vowels in the rows, non-canonical transcriptions with deviations > 2.5% related to vowel height (in pink), vowel length (in turquoise), rounding (in yellow) and diphthongization (in teal) are also indicated; Vow = target vowel.

Vow	ie	uu	oe	i	u	o	e	a	aa	ee	eu	oo	ij	ui	ou	Total
ie	68.0	0.8	0.0	44.2	0.6	0.0	0.0	0.0	0.0	1.2	0.8	0.0	0.0	0.1	0.1	7.7
uu	0.4	33.8	2.0	1.0	10.2	0.0	0.0	0.0	0.0	0.1	0.5	0.1	0.0	0.7	0.5	3.3
oe	0.1	30.3	72.4	0.1	37.0	1.6	0.3	0.0	0.0	0.0	0.3	5.0	0.0	2.0	4.0	10.2
i	13.7	1.6	0.1	45.0	3.8	0.0	5.6	0.0	0.0	3.1	1.7	0.0	0.3	0.3	0.1	5.0
u	0.5	4.0	2.3	3.9	39.6	3.0	1.7	0.1	0.1	0.0	3.6	0.1	0.0	0.1	2.9	4.1
o	0.0	0.0	4.2	0.0	2.3	81.8	0.0	0.6	0.0	0.0	0.5	20.6	0.0	0.1	1.4	7.4
e	2.1	0.0	0.1	1.6	2.0	0.0	87.3	1.9	0.8	17.4	5.1	0.0	3.7	1.0	0.3	8.2
a	0.1	0.0	0.4	0.0	1.3	5.6	1.2	64.8	14.2	0.3	0.0	2.0	0.0	1.0	2.0	6.2
aa	0.0	0.0	0.0	0.0	0.0	0.0	0.0	30.8	82.9	0.1	0.1	1.3	0.1	0.4	1.0	7.8
ee	12.3	0.3	0.0	4.1	0.0	0.0	1.4	0.0	0.0	58.5	1.1	0.0	6.3	0.1	0.0	5.6
eu	0.0	12.0	0.1	0.0	1.3	0.2	0.3	0.0	0.1	0.3	66.5	0.0	0.0	4.5	3.7	6.0
oo	0.0	16.2	16.2	0.0	1.1	7.3	0.0	0.0	0.0	0.0	9.6	68.9	0.0	4.8	19.0	9.5
ij	2.1	0.3	0.0	0.0	0.1	0.0	2.1	0.3	1.3	18.6	3.9	0.0	85.8	1.3	0.0	7.7
ui	0.5	0.5	0.4	0.0	0.2	0.2	0.3	1.3	0.4	0.3	4.6	0.0	3.5	68.8	8.2	5.9
ou	0.0	0.3	1.7	0.0	0.5	0.2	0.0	0.1	0.1	0.0	1.6	1.8	0.2	14.6	56.8	5.2

3.2.2 Native data

Table 10 shows how the Dutch vowels produced by two Dutch native speakers were transcribed by the native listeners (see section 2.5). The columns present the 15 target vowels, while the rows show the transcribed vowels, reflecting overall percentages of canonical (indicated in green) and non-canonical transcriptions.

Our results show that 81.4% of all transcriptions are canonical, whereas 18.6% are non-canonical. Such a low canonical percentage for the native data was unexpected, particularly because the native data from these two speakers was included in earlier studies (cf. Van der Harst 2011; Van der Harst et al. 2014) and no anomalies were reported in their speech in comparison to the speech of the rest of the speakers in the native database used in those studies.

Table 10 shows that the target vowel with the highest percentage of canonical transcriptions is <oe> (98.6%), whereas the target vowel with the lowest percentage of canonical transcriptions is <e> (53.7%). We did not expect such a low percentage for the identification of the target vowel <e>. A thorough examination of the transcriptions of the native data made it clear that the low percentage for the target vowel <e> can

be ascribed to specific word tokens produced by one of the two native speakers (the target vowel “zes”, “six” was often transcribed as “zus”, ‘sister’). The majority of the non-canonical transcriptions seems to be related to the non-native confusion patterns also found in the non-native matrix of the listener data (see Table 9). For example, the target vowel <i> is transcribed as <ie> (17.1%) (confusion related to vowel height), the target <ee> as <e> (6.0%) (vowel length) and the target <uu> as <ui> (3.8%) (diphthongization). Confusion patterns associated with rounding seem to be distributed over several vowels. Front unrounded vowels were perceived as front rounded vowels: <e> as <u> (44.4%), <i> as <uu> (10.5%), and <ij> as <ui> (21.10%). Rounding is also involved in perceiving front rounded vowels as back rounded vowels, i.e. <uu> as <oe> (23.1%) and in perceiving <a> as <o> (22.2%). The latter distinction also involves height. In sum, it seems that problematic features in perceiving non-native vowels recur in perceiving native vowels, when the latter are mixed with non-native data.

Table 10. Most frequent canonical (indicated in green) and non-canonical transcriptions of all 15 Dutch vowels produced by native Dutch speakers, as given by non-expert native Dutch listeners; target vowels in the columns, classified vowels in the rows, non-canonical transcriptions with deviations > 2.5% related to vowel height (in pink), vowel length (in turquoise), rounding (in yellow) and diphthongization (in teal) are also indicated; Vow = target vowel.

Vow	ie	uu	oe	i	u	o	e	a	aa	ee	eu	oo	ij	ui	ou	Total
ie	92.5	0.0	1.4	17.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	7.9	
uu	3.7	65.4	0.0	10.5	15.6	0.0	0.0	0.0	0.0	3.0	1.5	0.0	0.0	0.0	5.4	
oe	1.2	23.1	98.6	2.6	2.6	19.8	0.0	0.0	0.0	1.5	1.5	7.0	0.0	0.0	10.5	
i	2.5	0.0	0.0	57.9	0.0	0.9	1.8	0.0	0.0	4.5	0.0	0.0	0.0	0.0	4.6	
u	0.0	5.8	0.0	6.6	80.5	1.9	44.4	0.0	1.22	3.0	1.5	0.0	0.0	0.0	9.0	
o	0.0	0.0	0.0	0.0	77.4	0.0	22.2	0.0	0.0	0.0	2.8	0.0	0.0	0.0	9.0	
e	0.0	0.0	0.0	0.0	0.0	53.7	0.0	0.0	6.0	0.0	0.0	1.1	0.0	0.0	3.1	
a	0.0	0.0	0.0	0.0	0.0	0.00	77.8	1.22	0.0	0.0	0.0	0.0	1.1	1.7	5.3	
aa	0.0	0.0	0.0	0.0	0.0	0.0	0.0	96.3	0.0	0.0	0.0	0.0	0.0	0.0	7.1	
ee	0.0	0.0	0.0	2.6	0.0	0.0	0.0	0.0	75.8	0.0	0.0	1.1	0.0	0.0	4.8	
eu	0.0	1.9	0.0	1.3	1.3	0.0	0.0	0.0	0.0	4.5	94.0	0.0	0.0	0.0	6.2	
oo	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.5	90.1	0.0	1.1	10.5	6.5	
ij	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.5	0.0	0.0	76.7	0.0	0.0	6.3	
ui	0.0	3.8	0.0	1.3	0.0	0.0	0.0	0.0	0.0	0.0	21.1	96.7	0.0	0.0	9.9	
ou	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.22	0.0	0.0	0.0	0.0	1.1	87.7	4.2	

3.3 Comparison acoustic and listener data

In this subsection we focus on the non-native data and compare the results of the acoustic data presented in the current study with the outcomes of an earlier perception study (cf. Chapter 5 in Burgos 2018). We first investigate the outcomes of the Spanish learners as a group and then examine individual differences across learners.

3.3.1 Spanish learners as a group

In Table 11, we compare the percentages of canonical classifications per target vowel of the acoustic data in the classification condition “Group” using F1, F2 and duration (see Table 6), with the percentages of the native listeners’ canonical transcriptions per target vowel (see Table 9).

A paired-samples t-test showed a significant difference between the acoustic data and the listener data ($t(14) = 2.31, p = .037$). The average canonical percentages for the acoustic data ($M = 74.7, SD = 13.3$) were higher than for the listener data ($M = 65.4, SD = 16.4$).

Remarkably, for five Dutch target vowels, namely <ie>, <oe>, <o>, <e> and <aa> (counterparts of the five Spanish core vowels /i, u, o, e, a/ respectively), the difference between the percentages of canonical classifications/transcriptions for the acoustic and listener data was fairly small (see Table 11), which means that the results from the statistical classifier and the native listeners coincide to a large extent in classifying/ perceiving the learner vowel realizations of these target vowels as canonical.

At the same time, there are medium-sized differences (Cohens $h \geq .50$; see Table 11 for the percentages) between the acoustic and listener data for four target vowels, namely, for the short monophthongs <i> and <u>, for the long mid vowel <ee>, and for the diphthong <ui>. For the target vowels <i>, <u> and <ee>, the percentages of canonical classifications for the acoustic data are much higher than those for the perception data, indicating that many of the learners' vowel realizations of these three vowels were automatically classified as canonical on the basis of their acoustic measurements. These discrepancies show that the statistical classifier classified the learner realizations of these vowels on the basis of acoustic properties that were decoded differently by native listeners.

The opposite applies to the target front rounded vowel <ui> (see Table 11), which received a much higher percentage of canonical transcriptions for the listener data (68.8%) than for the acoustic data (41.8%). This indicates that human vowel recognition was more accurate than the statistical classifier in perceiving the target vowel <ui>.

Table 11. Percentages of canonical classifications/transcriptions per target vowel and most frequent vowel confusions, resulting from the acoustic and the perception study, confusions related to vowel height (H, in pink), vowel length (L, in turquoise), rounding (R, in yellow) and diphthongization (D, in teal) are also indicated; Vowel = target vowel, AS = acoustic study, PS = perception study, Δ = difference between the percentages of canonical classifications/transcriptions of both studies, Conf = most frequent vowel confusions per study, %Can = average percentage of canonical classifications/transcriptions.

Vowel	ie	uu	oe	i	u	o	e	a	aa	ee	eu	oo	ij	ui	ou	%Can
AS	67.9	53.6	72.7	76.8	71.4	83.9	90.0	83.9	83.9	85.5	78.2	83.9	82.1	41.8	63.6	74.7
PS	68.0	33.8	72.4	45.0	39.6	81.8	87.3	64.8	82.9	58.5	66.5	68.9	85.8	68.8	56.8	65.4
Δ	-0.1	19.8	0.3	31.8	31.8	2.7	2.7	19.1	1.0	27.0	11.7	15.0	-3.7	-27.0	6.8	9.3
Conf-AS	i	oe	u	ie	oe	oo	ee	aa	a	ie	ui	ou	ee	ou	ui	
	21.4	21.4	9.1	14.3	14.3	8.9	3.6	7.1	14.3	9.1	16.4	7.1	8.9	21.8	12.7	
	H	R	R	H	R	L	L	L	L	H	D	D	H	R	R	
Conf-PS	i	oe	oo	ie	oe	oo	l	aa	a	ij	oo	o	ee	ou	oo	
	13.7	30.3	16.2	44.2	37.0	7.3	5.6	30.8	14.2	18.6	9.6	20.6	6.3	14.6	19.0	
	H	R	L	H	R	L	H	L	L	D	R	L	H	R	H	

Table 11 also shows the most frequent vowel confusions per target vowel and the features which were (erroneously) employed by the Spanish learners, namely, vowel height, vowel length, rounding and diphthongization. The rows Conf-AS and Conf-PS in Table 11 indicate that 9 of the 15 target vowels produce the same frequent vowel confusions in the acoustic and listener data (i.e., the target vowels <ie>, <uu>, <i>, <u>, <o>, <a>, <aa>, <ij> and <ui>), whereas this is not the case for six target vowels, namely, <oe>, <e>, <ee>, <eu>, <oo> and <ou>. The target back rounded vowel <oe> was frequently classified as the front rounded vowel <u> in the acoustic data, whereas it was perceived as the back rounded long mid vowel <oo> by the native listeners, which means that the discrepancy between the acoustic and the listener data relies on a difference in weighing front rounding and vowel length. The target front unrounded vowel <e> was often classified as the front unrounded long mid vowel <ee> in the acoustic data, but classified as the front unrounded <i> by the listeners, which indicates a difference in weight assigned to vowel length and vowel height. A similar interpretation applies to the target long mid vowel <ee>, which was frequently classified as the high front vowel <ie> in the acoustic data, but as the front unrounded diphthong <ij> in the listener data. This indicates that the difference here relates to vowel height and diphthongization.

The front rounded long mid vowel <eu> was usually classified as the front rounded diphthong <ui> in

the acoustic data, but often perceived by the listeners as the back rounded long mid vowel <oo>, reflecting a disparity related to diphthongization and rounding. Similarly, the back rounded long mid vowel <oo> was frequently classified as the back rounded diphthong <ou> in the acoustic data, but perceived as the back rounded monophthong <o> by the listeners, indicating that the disparity here relates to diphthongization and vowel length. Finally, the target back rounded diphthong <ou> was frequently classified as the front rounded diphthong <ui> in the acoustic data, but as the long mid vowel <oo> by the listeners, suggesting, different weightings for the features of rounding and vowel height. In sum, larger discrepancies seem to be caused by differently weighing the competing features involved. While the four distinctive features (see Table 1) are clearly involved, their individual impact varies, as exemplified in Table 11.

3.3.2 Individual differences across learners

This subsection discusses the individual patterns found in the acoustic and listener data obtained from the classifications/transcriptions of the Dutch vowels produced by 28 Spanish learners, and the way their individual performance is related to their background characteristics, including their CEFR level, length of residence and daily use of Dutch. We first consider individual differences across learners in the acoustic data and in the listener data separately. Subsequently, we compare the outcomes of the acoustic data and the listener data for each individual learner.

3.3.2.1 Acoustic data

The dissimilarities among the Spanish learners were computed by using a matrix of 15 columns by 15 rows, giving a vector of 225 cells per learner. The analysis resulted in consistent clustering into four groups, irrespective of the clustering method used. We applied the R package *pvclust* (Suzuki & Shinodaira 2006) for a hierarchical cluster analysis with multiscale bootstrapping ($n = 1000$), using Euclidean distances and Ward's method. Two types of probability values are available: approximately unbiased (AU) p -value and bootstrap probability (BP) value, AU is a better approximation. High p -values indicate strong, certain clusters. The values vary between 0 and 100. The result of the hierarchical cluster analysis for the acoustic data is shown in Figure 2.

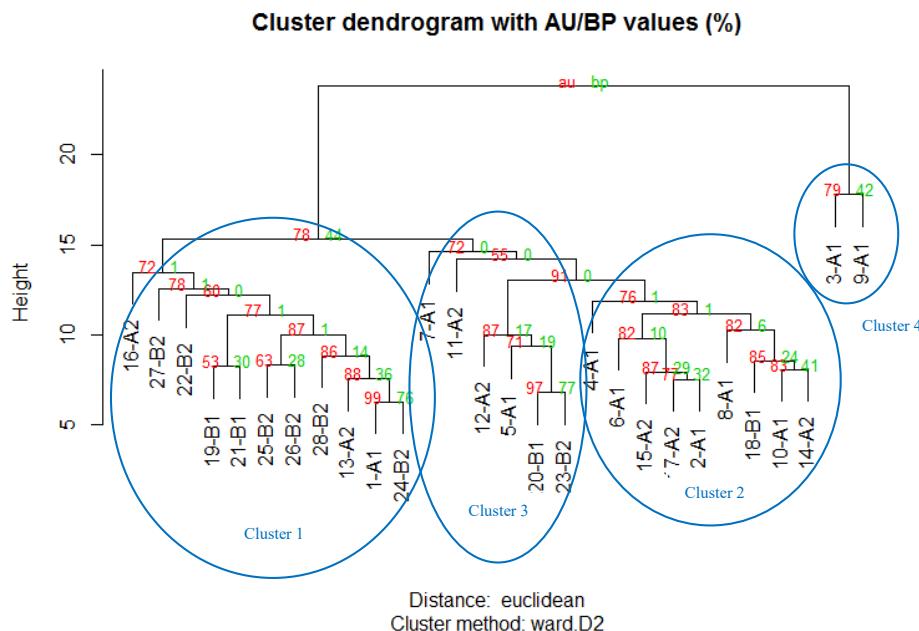


Figure 2. Cluster analysis of the 28 Spanish learners (including their corresponding CEFR language proficiency level) based on the percentages of canonical and non-canonical classifications per target vowel obtained from the acoustic data.

The AU values (in red) in Figure 2 show that three of the four clusters are not entirely separate, whereas the fourth cluster is clearly separated from the rest. There are similarities between three clusters (clusters 1, 2 and 3) and therefore between individual learners. But what are the differences between clusters? Figure 2 shows that the main division is between the three lower clusters (clusters 1, 2 and 3) and the fourth, higher, cluster (cluster 4). Also, a subdivision can be noted between cluster 1 and clusters 2 and 3, and between cluster 3 and cluster 2. Does this clustering result from proficiency differences? Learners with higher proficiency are likely to show a greater consistency in the production of target L2 segments, while learners with lower proficiency will tend to show more variability in their production confusion patterns (Cutler 2012). Table 12 presents the means and ranges of the percentages of canonical classifications for the Dutch vowels produced by the Spanish learners and their CEFR levels, in each of the four clusters.

Table 12. Means and standard deviations (SD) for the percentages of canonical classifications per cluster for the acoustic data, including number of Spanish learners per CEFR language proficiency level; %Can = percentages of canonical classifications.

Cluster	Mean %Can	Range %Can	CEFR levels				Overall
			A1	A2	B1	B2	
Cluster 1	80.5 (8.1)	62.0–90.0	1	2	2	6	11
Cluster 2	73.5 (8.2)	59.0–86.0	5	3	1	0	9
Cluster 3	71.6 (5.2)	66.0–79.0	2	2	1	1	6
Cluster 4	66.6 (1.7)	65.0–68.0	2	0	0	0	2

Table 12 shows that the highest mean percentage of canonical classifications is observed in cluster 1. The range in canonical percentages in cluster 1 overlaps with that of clusters 2 and 3, and even with that of cluster 4. This shows that it is not only the percentages of canonical classifications that determine clustering but also the distributions and percentages of the non-canonical classifications. Most of the Spanish learners are in cluster 1. When proficiency level is considered, we see that cluster 1 contains learners at all four levels, and the highest number of learners with a B2 level. Cluster 2 contains the majority of A1 learners, but also three A2 and one B1 learner. Cluster 3 does not have as many learners as clusters 1 and 2, but it has learners at all four levels. Finally, cluster 4 contains only two learners, both at A1 level.

We next consider the problems faced by learners in each of the clusters. Cluster 1 is characterized by learners with no serious problems associated with vowel length and front rounding, and with few difficulties related to vowel height and diphthongization. Cluster 2 contains learners with problems related to height, diphthongization and front rounding, particularly in the *<uu>-<u>* contrast. Some of these learners also have difficulties with length; their vowel realizations are too long. The difficulties with Dutch vowels for learners in cluster 3 are similar to those observed in cluster 2, but more salient. There is great variability in the vowel confusion patterns associated with the learners in this cluster. The great majority have problems with all four distinctive features: height, length, rounding and diphthongization. They all appear to apply Spanish-like diphthongization (i.e., combining two full vowels) when realizing the long mid vowels and diphthongs.

One of the learners in cluster 1 (learner 1) rather surprisingly has an A1 proficiency level. However, this female learner received the highest percentage of canonical classifications of all the Spanish learners. An explanation for this outcome may be found in this learner's language background. She is a Spanish/Catalan bilingual who had been living in the Netherlands for six months and used Dutch daily (eight hours approximately). She was an MA student of Translation and Interpreting Studies who also spoke English and French at B2 level, and Arabic at A2 level. The fact that she spoke French is perhaps relevant, as French has front rounded vowels. Learner 28, also in cluster 1, is a male post-doctoral researcher who had been living in the Netherlands for three years. Surprisingly perhaps, his use of Dutch was rather limited (approximately four hours a day), especially if we take into account his B2 proficiency level and the fact that, of all the learners in cluster 1, he received one of the highest average percentages of canonical classifications. He was fluent in English (C2 level), French (C2 level) and German (C1 level). French and German have front rounded

vowels, which may account for the phonological accuracy of this learner's Dutch vowel productions. Interestingly, for both of these exceptional learners in cluster 1 (learners 1 and 28), Dutch was their L3 or additional language (La). It is said that prior linguistic knowledge in multilinguals can be useful in the acquisition of an La (De Angelis 2007: 130). Our results suggest that speaking French and German, languages that have front rounding, may help learners to master front rounded vowels in an La, for example, in Dutch.

One B1 learner in cluster 2 seems to provide evidence for the suggestion that phonology acquisition does not always progress along with foreign language proficiency (cf. Burgos, Cucchiari, Van Hout & Strik 2014a). Further examination of this female learner's background (learner 18) does not offer an explanation for the low average percentages of canonical classifications she receives. She is a professional in human ecology who spoke English (C2 level) and French (C1 level). Her length of residence in the Netherlands was ten years, and her use of Dutch was low, namely four hours daily. She appeared to have problems with vowel length, vowel height and diphthongization and, perhaps most strikingly, showed an overreliance on front rounding which led to numerous vowel confusions. Her prior knowledge of French did not seem to help her produce native-like Dutch vowels. Her strategy may have been to apply front rounding for most Dutch vowels, that is, also where this was not appropriate.

The learner with the highest average percentage of canonical classifications in cluster 2 is an A1 learner (learner 2). She had been living in the Netherlands for three years. Her self-estimated use of Dutch on a daily basis was approximately 14 hours. She was a translator and fluent in English (C1 level), German (C2 level) and Italian (C1 level), which may explain her phonological skill when producing Dutch vowels.

The only B2 learner in cluster 3 represents an exceptional case. Learner 23, whose length of residence was 12 years and who used Dutch on a daily basis (eight hours approximately), is a female B2 learner who had a low-intermediate level in English (B1 level). She had severe problems with extreme diphthongization, vowel height and vowel length, and difficulties with the front rounded vowels, in particular <uu> and <u>. This learner reported to the first author that she was fired because customers could not understand her Dutch.

Cluster 4 contains only two learners. Learner 9 is a female MA student of media studies who had been living in the Netherlands for one month. She reported using Dutch daily (6 hours approximately). She spoke English (C1 level), Portuguese (B1 level) and Catalan (B1 level). She had problems with vowel height and her long mid vowels and diphthongs were extremely diphthongized. The average percentage of canonical classifications for learner 3 was lower than for learner 9. The other learner in this cluster (learner 3) is a male university employee who had been living in the Netherlands for ten years, and did not use Dutch very much (two hours a day). He was fluent in English (C2 level) and German (C1 level). He had difficulties associated with diphthongization, vowel height and front rounding. His knowledge of German, which contains front rounded vowels, did not seem to help when producing the Dutch front rounded vowels, as attested by an evident <uu>-<u>-<oe> confusion.

Prior linguistic knowledge of other languages, especially languages with front rounding, seem to contribute to being able to produce Dutch vowels (more) accurately. In this respect, it should be noted that the B2 learner in cluster 3 (learner 23) did not speak any other foreign language (than Dutch) which has front rounded vowels.

3.3.2.2 Listener data

We computed dissimilarities among the speakers by using the original matrix of 15 columns by 20 rows (cf. Chapter 5 in Burgos 2018), giving a vector of 300 cells per speaker. A consistent clustering in three groups was found, regardless of the clustering method used. To exclude the noisy impact of the many cells with rather low frequencies, we excluded those cells in the matrix whose average across the informants was less than 5% of the classifications. The result was a set of 42 cells, a number that obviously is higher than the 15 cells with canonical transcriptions. We again applied the R package *pvclust* (Suzuki & Shinodaira 2006) with multiscale bootstrapping ($n = 1000$), using Euclidean distances and Ward's method. The result of the hierarchical cluster analysis for the listener data is displayed in Figure 3.

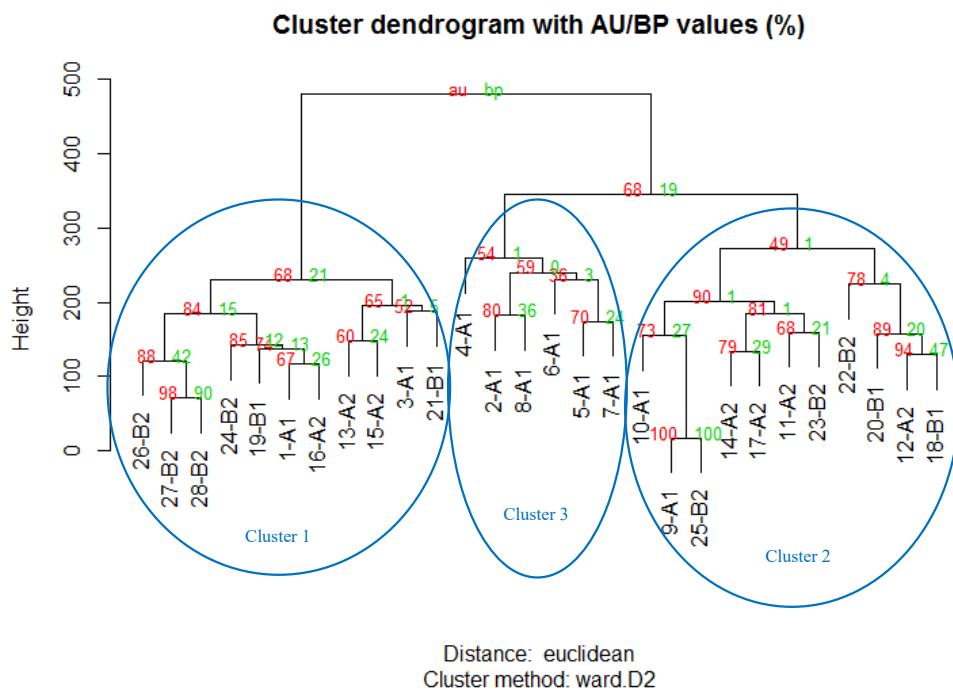


Figure 3. Cluster analysis of the 28 Spanish learners (including their corresponding CEFR language proficiency level) based on the percentages of canonical and non-canonical transcriptions per target vowel obtained from the listener data.

Figure 3 shows a different clustering than that observed in Figure 2. The AU values (in red) in Figure 3 show that the three clusters are not perfectly distinctive, pointing out that there are similarities between the clusters. But what are the differences between the three clusters? Figure 3 shows that the main division is between cluster 1, and clusters 2 and 3. Can this clustering be explained by proficiency differences? Again, L2 learners with higher proficiency can be presumed to have a greater consistency in the realization of phonemic target phones, resulting in higher intelligibility, whereas learners with lower proficiency will likely produce more variable input, resulting in less intelligible realizations (Cutler 2012: 386). Table 13 presents the means and ranges of the percentages of canonical transcriptions and the CEFR language proficiency levels of the Spanish learners in each of the three clusters.

Table 13. Means and standard deviations (SD) for the percentages of canonical transcriptions per cluster for the listener data, including number of Spanish learners per CEFR language proficiency level; %Can = percentages of canonical transcriptions.

Cluster	Mean %Can	Range %Can	CEFR levels				Overall
			A1	A2	B1	B2	
Cluster 1	71.8 (9.0)	56.8–84.8	2	3	2	4	11
Cluster 2	55.5 (5.0)	47.9–64.3	2	4	2	3	11
Cluster 3	43.8 (8.1)	34.0–54.5	6	0	0	0	6

Cluster 1 is associated with the highest average percentage of canonical transcriptions, as shown in Table 13. The range in percentages of canonical transcriptions overlaps to some extent with the two other clusters. Clusters 2 and 3 clearly overlap in this respect. These results exemplify that it is not only the percentages of canonical transcriptions that matter, but also the percentages of non-canonical transcriptions. When it comes to proficiency, cluster 1 has the highest number of learners with a B2 level, but also two A1 learners. Cluster 2 has learners at all four levels. Cluster 3 contains the majority of A1 learners, but also one B2 learner, which suggests that L2 phonology acquisition does not always progress along with foreign language proficiency.

How are the clusters related to the transcriptions? Cluster 1 learners evoke higher percentages of canonical transcriptions, showing an overall better performance, particularly on the front rounded vowels <u> and <uu> and on the long mid vowels. This cluster is characterized by learners with no major difficulties with vowel length, front rounding and diphthongization, and probably on the verge of dealing with problems related to vowel height.

The distinction between clusters 2 and 3 is harder to define. Cluster 2 comprises learners who have difficulties with vowel height and with front rounding. Our results indicate that these learners often realize the vowels <ie>, <uu>, <aa> and <oo> with longer duration. The learners' difficulties with Dutch vowels in cluster 3 are similar to those found in cluster 2, but much more salient. That is, problems with vowel height and particularly with front rounding are more severe for most learners in cluster 3. The duration of <ie>, <uu>, <oe>, <e>, <aa>, <ee> and <oo> are longer. And most importantly, all learners from cluster 3 appear to resort to extreme diphthongization when producing long mid vowels and diphthongs.

Two learners in cluster 1 have an A1 proficiency level, namely learner 1 and learner 3. The background information of these learners has already been commented on in the discussion of the cluster analysis of the acoustic data. Here again, our outcomes seem to suggest that speaking French and/or German, languages that have front rounding, helps in mastering front rounded vowels in an L2, such as Dutch.

The three B2 learners in cluster 2 show that L2 phonology acquisition does not always reflect the level of foreign language proficiency. Learner 25 spoke English (B2 level) and her length of residence in the Netherlands was three years. Her use of Dutch was high, namely 10 hours on a daily basis. She was learning Dutch pronunciation with the help of a Dutch speech therapist at the time of the recording because she had problems being understood by native Dutch listeners. Learner 25 in cluster 2 appeared to have extreme diphthongization. She also had severe problems with front rounded vowels, particularly <u> and <uu>. A similar situation applies to learner 22 in cluster 2. She was a female B2 learner, fluent in English (C2 level), and had been living in the Netherlands for ten years. She used Dutch for an average of six hours a day. She had difficulties with the front rounded vowels, especially <u> and <uu>, and with vowel height and length.

Cluster 3 contains only A1 learners. Learner 4 is associated with the lowest average percentage of canonical transcriptions of all 28 Spanish learners, followed by learner 5, also included in cluster 3. Learner 4 is a female nurse. She had been living in the Netherlands for seven months and used Dutch on a daily basis (11 hours approximately). She was fluent in English (B2 level). She had severe problems with vowel height, vowel length, front rounding and diphthongization. Learner 5 is a male research technologist whose length of residence in the Netherlands was 7 months. He used Dutch for an average of six hours a day. He was fluent in English (B2 level). He also had severe problems with vowel height, vowel length, front rounding and diphthongization, but to a lesser extent than those observed for learner 4.

It should be noted that both B2 learners in cluster 2 and both A1 learners in cluster 3 did not speak any other foreign languages with front rounded vowels, like French or German. This means they could not benefit from existing linguistic knowledge to help their acquisition of Dutch front rounded vowels.

We can conclude that the primary distinction among the three clusters can be related to the front rounded vowels. Our outcomes clearly show that the new feature of front rounding is affected by the L1 feature of back rounding, which leads Spanish learners to produce the front rounded vowels <u> and <uu> as the back rounded <oe>. Recurring pairwise confusions for <a>-<aa> and <i>-<ie> are detected in all three clusters, although learners in cluster 1 appear to perform considerably better when making these vowel distinctions. Although diphthongization does not cause serious difficulties and seems to compensate for problems with vowel length, extreme diphthongization, nevertheless, can lead to intelligibility problems.

3.3.2.3 Comparison of acoustic and listener data

This subsection compares the outcomes of the acoustic and listener data per individual Spanish learner. Table 14 presents the mismatch between the outcomes, based on the clustering analyses, and on the average percentages of canonical classifications/transcriptions per individual learner.

Table 14. Mismatch between the acoustic and listener results based on clustering, and on the average percentages of canonical classifications/transcriptions per Spanish learner with their corresponding CEFR language proficiency level, match (in green), mismatch of one cluster (in orange) and mismatch of two clusters (in red) are indicated; SL = Spanish learners, %Can = percentages of canonical classifications/transcriptions, AD = acoustic data, LD = listener data, Δ = difference between the average percentages of canonical classifications/transcriptions of the acoustic and listener data.

SL	CEFR	Acoustic data	Listener data	%Can	AD	%Can	LD	SL	CEFR	Acoustic data	Listener data	%Can	AD	%Can	LD	Δ
1	A1	cluster 1	cluster 1	90.0	74.1	15.9		15	A2	cluster 2	cluster 1	79.0	61.2	17.8		
2	A1	cluster 2	cluster 3	86.0	54.6	31.4		16	A2	cluster 1	cluster 1	79.0	75.7	3.3		
3	A1	cluster 4	cluster 1	65.0	56.8	8.2		17	A2	cluster 2	cluster 2	83.0	57.1	25.9		
4	A1	cluster 2	cluster 3	59.0	34.0	25.0		18	B1	cluster 2	cluster 2	69.0	57.5	11.5		
5	A1	cluster 3	cluster 3	66.0	34.3	31.7		19	B1	cluster 1	cluster 1	79.0	77.5	1.5		
6	A1	cluster 2	cluster 3	72.0	46.2	31.7		20	B1	cluster 3	cluster 2	72.0	51.5	20.5		
7	A1	cluster 3	cluster 3	69.0	45.0	24.0		21	B1	cluster 1	cluster 1	76.0	66.3	9.7		
8	A1	cluster 2	cluster 3	72.0	48.6	23.4		22	B2	cluster 1	cluster 2	89.0	59.4	29.6		
9	A1	cluster 4	cluster 2	68.0	47.9	20.1		23	B2	cluster 3	cluster 2	76.0	59.2	16.8		
10	A1	cluster 2	cluster 2	69.0	51.7	17.3		24	B2	cluster 1	cluster 1	86.0	74.4	11.6		
11	A2	cluster 3	cluster 2	68.0	58.5	9.5		25	B2	cluster 1	cluster 2	76.0	64.3	11.7		
12	A2	cluster 3	cluster 2	79.0	49.2	36.8		26	B2	cluster 1	cluster 1	90.0	73.4	16.6		
13	A2	cluster 1	cluster 1	62.0	62.5	0.5		27	B2	cluster 1	cluster 1	79.0	83.3	-4.3		
14	A2	cluster 2	cluster 2	72.0	54.0	18.0		28	B2	cluster 1	cluster 1	79.0	84.8	-5.8		

Table 14 shows that while there is a match between two clusters (acoustic data and listener data) for 15 learners, a mismatch of two clusters was found for only two learners. A mismatch of one cluster was observed for 11 learners. Mismatches were also found in terms of the degree of difference between the average percentages of the canonical classifications/transcriptions of the acoustic and listener data for the individual learners. Table 14 shows the average percentages of the canonical classifications/transcriptions for the acoustic and the listener data, and the difference between the two outcomes. As noted earlier in relation to the comparison between the acoustic and listener data of the learners as a group (see section 3.3.1), the average percentage of canonical classifications for the acoustic data (74.7%) is higher than for the listener data (65.4%). Almost all differences are positive, except for two learners, showing that the acoustic classification was more successful than the listener classification. The highest difference between the average percentage of canonical classifications for the acoustic data and for the listener data was observed for learner 12 ($\Delta = 36.8$), but there are more learners with high difference scores. The correlation between the two sets of percentages ($r(28) = .605$, p (two-tailed) = .001) is significant, but not high.

A closer examination of the individual patterns of learner 12 found in the acoustic and listener data, and of her background characteristics, including her CEFR language proficiency level, can help us to understand what the reasons are for such a striking difference. Learner 12 is a Spanish/Catalan bilingual who had been living in the Netherlands for six months and used Dutch daily (13 hours on average). She was working in the pharmaceutical industry and was fluent in English (C1 level), German (B2 level) and French (A2 level). The statistical classifier appears to have classified many of her Dutch vowel realizations as canonical. Her prior linguistic knowledge of other foreign languages might have contributed to her accurate production of Dutch vowels, and proficiency in German and French might have helped her in producing front rounded Dutch vowels accurately. Conversely, the average percentage of canonical transcriptions she received is rather low, indicating that the native Dutch listeners were not always able to decode the acoustic properties of her Dutch vowel realizations. An inspection of the canonical transcriptions for this speaker reveals that she had severe problems with rounding in some Dutch vowels, namely the front rounded vowels <uu> (0.0%), <u> (6.06%) and <eu> (38.1). Back rounding was also problematic, as attested by low canonical percentages

for <o> (40.63%) and <oo> (48.39%), although her production of <ou> (52.5%) was relatively successful. Difficulties with vowel height (e.g., the target vowel <i> (20.59%)) and with extreme diphthongization (e.g., the target vowel <ee> (9.38%)) were also evident. However, not all vowel realizations produced by learner 12 were inaccurate. Her realizations of the target vowels <e> (95.83%) and <aa> (89.66%), which are similar to the Spanish vowels /e, a/, as well as of the vowel <ui> (90.32%) were excellent. In sum, the outcomes of the listener data indicate that learner 12 shows a great variability in her production of Dutch vowels: the production of some vowels was poor, whereas other vowels were accurately produced, reaching near-native canonical percentages. She also shows great variability in the way she applies acoustic features. For instance, she applies front rounding proficiently when producing the front rounded vowel <ui>, reaching a near-native pronunciation, while she is not able to apply this feature properly when realizing the front rounded vowels <uu> and <u>.

The variability observed in the production patterns of learner 12 is not an exception. Interestingly, such variability seems to be present in the features associated with L2 vowel contrasts learners master predominantly. For instance, additional analyses showed that some learners seem to focus on the feature of vowel length first, which will help them to make the <a>-<aa> contrast (based on vowel height and duration) (e.g., learner 10 (A1 proficiency level), with 88.9% for <a> and 79.3% for <aa>), whereas others focus on the feature of rounding, which is necessary to produce the <ij>-<ui> distinction (e.g., learner 8 (A1 proficiency level), with 92.31% for <ij> and 90.63% for <ui>). As a result, we observe considerable variability *within* learners and *across* learners.

4 Discussion

The present study set out to compare the results of two different approaches aimed at investigating Dutch L2 vowel production accuracy by adult Spanish learners. We compared the acoustic properties of Dutch vowels produced by adult Spanish learners (i.e., an objective approach) with the perception of these vowel productions by a varied and extensive group of non-expert native Dutch listeners (i.e., a subjective approach). To this end, we compared statistical vowel classifications obtained from the acoustic properties of the Dutch vowels produced by Spanish learners with human vowel recognition based on the transcriptions of the same Spanish-Dutch vowel productions by a large and varied group of native Dutch listeners.

An additional aim was to explain individual differences and variability in L2 vowel realizations across Spanish learners by investigating individual patterns at the production and perception levels. To establish these individual patterns, we examined the learners' proficiency level in Dutch, as well as factors that could play a role in L2 phonology acquisition, and particularly in L2 vowel accuracy, such as prior linguistic knowledge in multilinguals, length of residence, and daily use of Dutch.

Our outcomes, presented in the non-native matrix (see Table 6), show high variability in the learners' vowel productions. The variety in, and high percentages of, non-canonical classifications assigned by the listeners indicate that the Dutch vowels produced by the Spanish learners were classified on the basis of their acoustic properties as different vowels than the target vowels. The highest variability in non-canonical classifications was found for the target front rounded vowel /œy/ (<ui>), which does not occur in Spanish (new vowel), whereas the lowest variability was observed for the target vowel /a:/ (<aa>) (similar to the Spanish /a/). The non-canonical classifications are related to vowel height, length, rounding and diphthongization. Conspicuous asymmetrical confusions were noted in the contrasts /I/-/i/ (<i>-<ie>) (based on vowel height) and /a/-/a:/ (<a>-<aa>) (based on vowel height and vowel length), in which the vowels /i/ and /a:/ (similar to the Spanish /i/ and /a/ respectively) are frequently classified by the statistical classifier, more than /I/ and /a/. Vowel confusions related to rounding were reflected in the non-canonical classifications of the target front rounded /y/ (<uu>) and /y/ (<u>), two new vowels which are frequently confused with each other, and particularly with the back rounded vowel /u/ (<oe>) (similar to the Spanish /u/). Similarly, the target front rounded new diphthong /œy/ (<ui>) is often classified as the back rounded diphthong /ɔu/ (<ou>). It should be remembered that Spanish does not have front rounding, as all rounded vowels in Spanish are back vowels (/o, u/) (Hualde 2005). This could explain why the Spanish learners

produce Dutch vowels /y/, /ʏ/ and /œy/ – which are new vowels to them – as back and rounded vowels. As to vowel confusions related to diphthongization, the target long mid vowel /ø:/ (<eu>) is often classified as the diphthong /œy/ (<ui>), showing evidence of extreme diphthongization in the learners' realizations.

According to Flege's (1995) SLM, a new phoneme category may be hard to acquire when it seems similar to an existing L1 category. Adult L2 learners may use a single L1 category for two L2 phones classified as similar. In the context of the present study, the Dutch vowels /i, u, ɔ, ε, a:/ can be regarded as acoustically similar to the Spanish /i, u, o, e, a/ and therefore familiar to Spanish learners, whereas the remaining Dutch vowels (monophthongs: /y, ɪ, ʏ, ɑ/; long mid vowels: /e:/, /ø:/, /o:/; diphthongs: /εi, œy, ɔu/) can be considered new for Spanish learners. While the present study did not set out to test Flege's (1995) SLM, it can be concluded that some of our results are in line with the model. They show that Spanish learners have problems in making the fine-grained vowel contrasts /ɪ/-/i/ and /a/-/a:/ because the L2 phones in each pair are non-contrastive in the L1, as both resemble Spanish /i/ and /a/. The long mid vowels and diphthongs are often produced differently than the monophthongs, namely, by applying Spanish-like diphthongization (i.e., combining two full vowels).

Our findings show that the statistical classifier and human vowel recognition coincide to a large extent in classifying/perceiving the learner vowel realizations of the Dutch target vowels /i, u, ɔ, ε, a:/ (see Table 11). This is in line with the Full Copying hypothesis suggested in Escudero's (2005) L2LP model (see also Van Leussen & Escudero 2015 for a revision of the L2LP model). A central assumption of the Full Copying hypothesis is that L2 learners will initially copy their L1 perception to attune L2 segments to their L1 native categories. Over time, exposure to the L2 will help L2 learners to evade their L1-learning mechanisms and to develop optimal L2 perception. We found evidence that the Dutch vowels that are best classified/perceived, namely, /i, u, ɔ, ε, a:/, are those vowels that are copies of the Spanish /i, u, o, e, a/ (see Table 11).

How do the statistical classifications of the learner vowel productions relate to their corresponding perceptions by native Dutch listeners? We assumed that the features of vowel height, length, rounding and diphthongization would play a pivotal role in perception also, but that their cue weightings might vary in comparison to the weightings used in production. The results supported our assumptions, as we found similarities and disparities between the production and native perception results. As expected, comparable results between the statistical classifier and native listeners were found for the five Dutch target vowels /i, u, ɔ, ε, a:/ (<ie>, <oe>, <oo>, <e>, <aa>), because they match the five Spanish core vowels /i, u, o, e, a/ (cf. Flege 1995) (see Table 11). This indicates that statistical vowel classifications and human vowel recognition concur to a great extent. Disparities between the statistical classifier and native listeners were observed too. We found both slight and substantial differences. Slight differences were found for the target vowels /y, ɪ, ʏ, ɑ, ø:/, /εi, œy, ɔu/ (<uu>, <a>, <eu>, <oo>, <ij>, <ou>), whereas substantial differences were seen for /I, ʏ, e:/, /œy/ (<i>, <u>, <ee>, <ui>). These results indicate that the human ear is able to process a large range of variability, as well as subtle and fine-grained characteristics of the speech signal in non-native speech, and suggest that human perception is not mere statistical classification.

The statistical classifier turned out to be more successful in classifying the learner realizations of the target vowels /I, ʏ, e:/ (<i>, <u>, <ee>) as canonical – based on their acoustic properties – than native Dutch listeners, who could not decode these properties or decoded them differently. It is important to take into account that the circumstances for the statistical classifier and native Dutch listeners were different. The statistical classifier considered all the data simultaneously, as a whole set, computing the solution with the best classification result. In contrast, the native listeners considered one stimulus at a time, at most within the context of previous stimuli, so that their classification can be considered to be more local than that of the statistical classifier. Therefore the native listeners not only had less information at their disposal, but their classifications might have been influenced by previous vowels in the set of stimuli they were presented with, which may have allowed them to adapt – and fine-tune – their perception.

Indeed, we found indications of adaptive mechanisms at work both in the statistical vowel classifications of the acoustic data and in native listener vowel recognition, depending on the vowel sets involved. Patterns of vowel confusions and problematic features found in the statistical vowel classifications of the non-native vowels in the classification condition "Total" (in which non-native and native data were pooled) recur in the classifications of the native vowels (see Table 8). For example, problems related to the feature of vowel

height appear in the non-canonical classifications of the target vowels /ɪ/ (<i>) and /y/ (<uu>) classified as /i/ (<ie>) and /y/ (<u>) respectively, whereas difficulties related to rounding and diphthongization are evident from the non-canonical classifications /ɔ/ (<o>) and /œy/ (<ui>), corresponding to the target vowels /a/ (<a>) and /ø:/ (<eu>) respectively. This pattern of performance in the statistical vowel classifications of the native data appears to indicate that the statistical classifier is data-sensitive and may have adapted or shifted its category boundaries to the ambiguous sounds of the non-native speech samples. This adaptive mechanism in boundary shift could help to understand why native front unrounded vowels (e.g., /a/ (<a>)) were classified as back rounded vowels (e.g., /ɔ/ (<o>)). In addition, the improvement observed across the classification conditions “Total” (i.e., non-native and native data pooled together), “Group” (non-native and native data treated as two independent groups) and “Individual” (individual non-native data mixed with the native data group) indicates that the acoustic data set to be analyzed can alter the results in an individual classification condition. Our results for the three classification conditions seem to suggest that the statistical classifier is context-sensitive as it adapts to the nature of the data (non-native and/or native data) inputted to the system. The input of large amounts of non-native data with a large variability in vowel errors seems to lead to boundary shifts as the statistical classifier has to accommodate error-infused data (non-native data) which differ substantially from the “clean” data consisting of target categories only (native data) (cf. Berck 2017).

Similar adaptive mechanisms in boundary shifts were observed for human vowel recognition. When listening to foreign-accented speech, native listeners seem to attend to phonetic details resulting from transfer from the learners’ L1, to navigate specific types of deviations in the speech signal. Recognizing words with segmental deviations implies that listeners have to cope both with sounds that are distorted versions of the native norms, as well as with sounds that can be mapped onto distinct phoneme categories. Native listeners are required to shift their common boundaries to accommodate ambiguous non-native realizations which differ from their experience with native phoneme categories (cf. Bent et al. 2016; Cutler 2012). In sum, adaptive mechanisms in boundary shifts were observed in both the statistical vowel classifications and in human vowel recognition.

The very high canonical vowel classifications obtained in the individual condition in the statistical multinomial regression analysis provides evidence that the vowels of the individual learners have acoustic distinctions, meaning that most vowels are not mergers. Not all learners make the same distinctions and not all distinctions are made with the same degree of distinctiveness. Our results show that the variability in acquiring L2 phones is intricate. There is a great variability both *within* and *across* learners in their production of Dutch vowels, which leads to distinct patterns of vowel confusions per target vowel (cf. Bent et al. 2016; Mayr & Escudero 2010). More specifically, there is great variability *within* learners both in their segmental deviations (cf. Wade, Jongman, & Sereno 2007) and in the way different features (vowel height, vowel length, rounding and diphthongization) are used. Similarly, there is a wide range of variability *across* learners in their abilities and strategies to successfully produce the Dutch target vowels.

Our findings on individual differences across the 28 adult Spanish learners, both for the acoustic and listener data, seem to indicate that phonology acquisition does not always progress along with foreign language proficiency (see Figure 2, Figure 3 and Table 14) (cf. Burgos et al. 2014a). We have provided evidence that higher proficiency levels in Dutch (i.e., CEFR B2 level) do not guarantee success in achieving a native-like pronunciation in Dutch. Other factors that are related to foreign language proficiency are length of residence and substantial L2 use. Earlier studies have shown that these factors do not appear to have a strong effect on L2 pronunciation accuracy (cf. Flege et al. 1997; Munro 1993; Yeni-Komshian, Flege & Liu 2000). Of course, it is possible that additional factors such as intrinsic individual differences (e.g., mimicry ability, learning strategies), or socio-psychological factors (e.g., motivation to sound native-like, attitudes toward the target language and culture) may have played a role in the individual differences in L2 pronunciation accuracy across the Spanish learners (cf. Moyer 2013 for a review of relevant factors in L2 phonology acquisition).

5 Conclusion

The aim of this article was to compare the acoustic properties of Dutch vowels produced by adult Spanish learners and the perception of these vowel productions by non-expert native Dutch listeners. We predicted that the features of vowel height, length, rounding and diphthongization would play a crucial role in native perception, but that their cue weightings might vary in comparison to the weightings used in production. The results supported our prediction, as we found similarities and disparities between the production and native perception results. As expected, similar results between the statistical classifier and native listeners were found for the five Dutch target vowels /i, u, ɔ, ε, a:/, because they match the five Spanish core vowels /i, u, o, e, a/ (cf. Flege 1995). This indicates that statistical vowel classifications and human vowel recognition concur to a great extent. Disparities between the statistical classifier and native listeners were observed too. We found both slight and substantial differences. Slight differences were found for the target vowels /y, ɑ, ɸ, ɔ:, ei, ɔu/, whereas substantial differences were seen for /I, y, e:, oey/. These results indicate that the native human ear is able to process a large range of variability, as well as subtle and fine-grained characteristics of the speech signal in non-native speech.

An additional finding is that statistical vowel classifications and human vowel recognition processes are context-sensitive: in both contexts, classification processes are adapted to the nature of the data (i.e., non-native and/or native data) involved. Including non-native data (with a large variability in vowel realizations) in the analysis of native data led to different results, suggesting that, with changes in the variability of the vowel stimuli, adaptive mechanisms in boundary shifts come into play in both statistical vowel classifications and human vowel recognition.

Our results on individual differences across the 28 adult Spanish learners, both for the acoustic and listener data, corroborate previous findings by showing that phonology acquisition does not always progress along with foreign language proficiency.

Finally, our findings indicate that variability in L2 phonology acquisition is extremely complex. It occurs at different levels: *within* and *across* learners with respect to segmental deviations per target vowel, and *within* and *across* learners with respect to the features (vowel height, length, rounding and diphthongization) they apply to produce Dutch vowels accurately.

References

Adank, Patti. 2003. *Vowel normalization: A perceptual-acoustic study of Dutch vowels*. Radboud University Nijmegen dissertation.

Adank, Patti, Roeland van Hout & Roel Smits. 2004a. An acoustic description of the vowels of Northern and Southern Standard Dutch. *Journal of the Acoustical Society of America* 116(3). 1729–1738.

Adank, Patti, Roel Smits & Roeland van Hout. 2004b. A comparison of vowel normalization procedures for language variation research. *Journal of the Acoustical Society of America* 116(5). 3099–3107.

Baker, Wendy & Pavel Trofimovich. 2005. Interaction of native- and second-language vowel system(s) in early and late bilinguals. *Language and Speech* 48(1). 1–27.

Bent, Tessa & Ann R. Bradlow. 2003. The interlanguage speech intelligibility benefit. *Journal of the Acoustical Society of America*, 114(3). 1600–1610.

Bent, Tessa, Meslissa Baesse-Berk, Stephanie A. Borrie & Megan McKee. 2016. Individual differences in the perception of regional, nonnative, and disordered speech varieties. *Journal of the Acoustical Society of America* 140(5). 3775–3786.

Berck, Peter. 2017. *Memory-based text correction*. Radboud University Nijmegen dissertation.

Best, Catherine T. 1995. A direct realist view of speech cross language speech perception. In Winifred Strange (ed.), *Speech Perception and Linguistic Experience: Issues in Cross-language Research*, 171–206. Timonium, MD: York Press.

Birdsong, David & Michelle Molis. 2001. On the evidence for maturational constraints in second language acquisition. *Journal of Memory and Language* 44(2). 235–249.

Boersma, Paul & David Weenink. 2010. Praat: Doing phonetics by computer. Version 5.1.32. Retrieved from <http://www.fon.hum.uva.nl/praat/> (accessed 14/02/18).

Bohn, Ocke-Schwen. 1995. Cross language speech perception in adults: First language transfer doesn't tell it all. In Winifred Strange (ed.), *Speech Perception and Linguistic Experience: Issues in Cross-language Research*, 279–304. Timonium, MD: York Press.

Booij, Geert. 1995. *The phonology of Dutch*. Oxford, UK: Clarendon Press.

Bradlow, Ann R. & Tessa Bent. 2008. Perceptual adaptation to non-native speech. *Cognition* 106. 707–729.

Burgos, Pepi. 2018. *Non-native pronunciation: Patterns of learner variation in Spanish-accented Dutch*. Radboud University Nijmegen dissertation.

Burgos, Pepi, Catia Cucchiarini, Roeland van Hout & Helmer Strik. 2014a. Phonology acquisition in Spanish learners of Dutch: Error patterns in pronunciation. *Language Sciences* 41. 129–142.

Burgos, Pepi, Mátýás Jani, Catia Cucchiarini, Roeland van Hout & Helmer Strik. 2014b. Dutch vowel production by Spanish learners: Duration and spectral features. *Proceedings of Interspeech 2014*, Singapore. 529–533.

Burgos, Pepi, Eric Sanders, Catia Cucchiarini, Roeland Van Hout & Helmer Strik. 2015. *Auris Populi*: Crowdsourced native transcriptions of Dutch vowels spoken by adult Spanish learners. *Proceedings of Interspeech 2015*, Dresden, Germany. 2819–2823.

Chládková, Kateřina, Paola Escudero & Paul Boersma. 2011. Context-specific acoustic differences between Peruvian and Iberian vowels. *Journal of the Acoustical Society of America* 130(1). 416–428.

Cohen, A., I. H. Slis & J. 't Hart. 1963. Perceptual tolerance of isolated Dutch vowels. *Phonetica* 9. 65–78.

Cohen, A., I. H. Slis & J. 't Hart. 1967. On tolerance and intolerance in vowel perception. *Phonetica* 16. 65–70.

Council of Europe. Common European Framework of Reference for Languages (CEFR). Retrieved from http://www.coe.int/t/dg4/linguistic/cadre1_en.asp (accessed 14/02/18).

Council of Europe. CEFR Self-Assessment Grid. Retrieved from <https://rm.coe.int/168045bb52> (accessed 14/02/18).

Cutler, Anne. 2012. *Native listening: Language experience and the recognition of spoken words*. Cambridge, MA: MIT Press.

De Angelis, Gessica. 2007. *Third or Additional Language Acquisition*. Clevendon, UK: Multilingual Matters.

Derwing, Tracey M. & Murray J. Munro. 1997. Accent, intelligibility and comprehensibility. Evidence from four L1s. *SSLA: Studies in Second Language Acquisition* 20. 1–16.

Escudero, Paola. 2005. *Linguistic perception and second language acquisition: Explaining the attainment of optimal phonological categorization*. Utrecht University dissertation.

Escudero, Paola & Williams, Daniel. 2012. Native dialect influences second-language vowel perception: Peruvian versus Iberian Spanish learners of Dutch. *Journal of the Acoustical Society of America Express Letters* 131(5). 406–214.

Flege, James E. 1995. Second-language speech learning: Theory, findings, and problems. In Winifred Strange (ed.), *Speech Perception and Linguistic Experience: Issues in Cross-language Research*, 233–277. Timonium, MD: York Press.

Flege, James E. 2012. The role of input in second language (L2) speech learning. Paper presented at the VIth International Conference on Native and Non-native Accents of English, ódż, Poland, 6–8 December. Retrieved from <http://www.jimflege.com/> (accessed 14/02/18).

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

Flege, James E., Carlo Schirru & Ian R. A. MacKay. 2003. Interaction between the native and the second language phonetic subsystems. *Speech Communication* 40. 467–491.

Gottfried, Terry L. 2008. Music and language learning. Effects of musical training on learning L2 speech contrasts. In Ocke-Schwen Bohn and Murray J. Munro (eds.), *Language experience in second language speech learning*, 221–237. Amsterdam: John Benjamins.

Guion, Susan G. 2003. The vowel systems of Quichua-Spanish bilinguals. *Phonetica*, 60, 98–128.

Hualde, José I. 2005. *The sounds of Spanish*. Cambridge, UK: Cambridge University Press.

Iverson, Paul & Bronwen G. Evans. 2007. Learning English vowels with different first-language vowel systems: Perception of formant targets, formant movements, and duration. *Journal of the Acoustical Society of America* 122(5). 2842–2854.

Kolb, Alice Y. & David A. Kolb. 2005. Learning styles and learning spaces: Enhancing experiential learning in higher education. *Academy of Management Learning & Education* 4(2). 193–212.

Larsen-Freeman, Diane. 2009. Adjusting expectations: The study of complexity, accuracy and fluency in second language acquisition. *Applied Linguistics* 30. 579–589.

Lenneberg, Eric H. 1967. *Biological foundations of language*. New York, US: Wiley.

Lobanov, B. M. 1971. Classification of Russian vowels spoken by different speakers. *Journal of the Acoustical Society of America* 49. 606–608.

Long, Michael H. 1990. Maturational constraints on language development. *Studies in Second Language Acquisition* 12. 251–285.

Magen, Harriet S. 1998. The perception of foreign-accented speech. *Journal of Phonetics* 26. 381–400.

Major, Roy C. 2001. *Foreign accent: The ontogeny and phylogeny of second language phonology*. Mahwah, NJ: Lawrence Erlbaum.

Mayr, Robert & Paola Escudero. 2010. Explaining individual variation in L2 perception: Rounded vowels in English learners of German. *Bilingualism: Language and Cognition* 13. 279–297.

McAllister, Robert, James E. Flege & Thorsten Piske. 2002. The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian. *Journal of Phonetics* 30. 229–258.

Moyer, Alene. 2013. *Foreign accent. The phenomenon of non-native speech*. Cambridge, UK: Cambridge University Press.

Munro, Murray J. 1993. Productions of English vowels by native speakers of Arabic: Acoustic measurements and accentedness ratings. *Language and Speech* 36(1). 39–66.

Munro, Murray J. & Tracey M. Derwing. 1995. Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning* 45. 73–97.

Navarro Tomás, Tomás. 2004. *Manual de pronunciación española*. Madrid, Spain: Editorial CSIC-CSIC Press.

Piske, Thorsten, Ian R. A. MacKay & James E. Flege. 2001. Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics* 29. 191–215.

Scovel, Thomas. 1988. *A Time to Speak. A Psycholinguistic Inquiry into the Critical Period for Human Speech*. New York, US: Newbury House.

Suzuki, Ryota & Hidetoshi Shimodaira. 2006. Pvclust: An R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* 22(12). 1540–1542.

Tokuhama-Espinosa, Tracey. 2003. The relationship between musical ability and foreign languages. In Tracey Tokuhama-Espinosa (ed.), *The multilingual mind*, 65–80. Westport, CT: Praeger.

University of Cambridge Local Examinations Syndicate (UCLES). 2015. *Cambridge English. The Cambridge English Scale Explained*. Cambridge English Language Assessment. Retrieved from <http://www.cambridgeenglish.org/images/177867-the-methodology-behind-the-cambridge-english-scale.pdf> (accessed 14/02/18).

Van der Harst, Sander. 2011. *The vowel space paradox: A sociophonetic study on Dutch*. Radboud University Nijmegen dissertation.

Van der Harst, Sander, Hans van de Velde & Roeland van Hout. 2014. Variation in Standard Dutch vowels: The impact of formant measurement methods on identifying the speaker's regional origin. *Language, Variation and Change* 26(2). 247–272.

Van Leussen, Jan-Willem & Paola Escudero. 2015. Learning to perceive and recognize a second language: the L2LP model revised. *Frontiers in Psychology* 6. 1–12.

Van Son, R. J. J. H., Binnenpoorte, D., Van den Heuvel, H., & Pols, L. C. W. 2001. The IFA corpus: a phonemically segmented Dutch open source speech database. *Proceedings of Eurospeech 2001*, Aalborg, Denmark. 2051–2054.

Van Wijngaarden, Sander J. 2001. Intelligibility of native and non-native Dutch speech. *Speech Communication*, 35. 103–113.

Wade, Travis, Allard Jongman & Joan Sereno. 2007. Effects of acoustic variability in the perceptual learning of non-native-accented speech sounds. *Phonetica* 64. 122–144.

Yeni-Komshian, Grace H., James E. Flege & Serena Liu. 2000. Pronunciation proficiency in the first and second languages of Korean-English bilinguals. *Bilingualism: Language and Cognition* 3(2). 131–149.