Auris populi: crowdsourced native transcriptions of Dutch vowels spoken by adult Spanish learners

CONFERENCE PAPER · SEPTEMBER 2015

5 AUTHORS, INCLUDING:

Auris populi: crowdsourced native transcriptions of Dutch vowels spoken by adult Spanish learners

Pepi Burgos[1], Eric Sanders[2], Catia Cucchiarini[2], Roeland van Hout[1], Helmer Strik[1][2]

[1] Center for Language Studies, Radboud University Nijmegen, The Netherlands
[2] Center for Language and Speech Technology, Radboud University Nijmegen, The Netherlands

{j.burgos, e.sanders, c.cucchiarini, r.vanhout, h.strik}@let.ru.nl

Abstract

In this paper we report on a study in which Dutch vowels produced by Spanish adult L2 learners were orthographically transcribed by Dutch lay listeners through crowdsourcing. The aim of the crowdsourcing experiment was to investigate how the *auris populi*, the crowd's ear, would deal with possibly deviant L2 vowel realizations. We present data on the transcriptions of the non-expert listeners for all fifteen Dutch vowels. The results of our study indicate that Dutch vowels pronounced by Spanish learners were transcribed differently from their canonical (target) forms by native listeners. The listeners' transcriptions confirm findings of previous research based on expert annotations of Spanish learners' vowel realizations conducted at our lab, namely, that the five Spanish vowels seem to function as "attractors" for the larger set of the Dutch vowels. In general, the results are also in line with the outcomes of acoustic measurements of the same speech material, but there are some interesting discrepancies. We discuss these results with regard to previous studies on the speech production of adult Spanish learners of Dutch and outline perspectives for future research. Finally, given our results, we formulate some evaluative remarks on the *auris populi* methodology for future L2 speech research.

Index Terms: L2 speech, orthographic transcription, crowdsourcing

1. Introduction

Studies on second language (L2) acquisition have shown that adult learners seldom achieve a native-like pronunciation [1], [2]. Accented speech does not necessarily impede communication as long as the pronunciation of the L2 learners is intelligible and native listeners are able to understand the intended message [3]. How can we determine whether accented speech is intelligible? Many studies relied on evaluations of experts. Another approach is to use native lay listeners to judge non-native speech, sometimes even asking them to evaluate specific phonetic contrasts. These approaches, however relevant, cannot answer the question what native listeners hear and perceive when they listen to accented speech. What brings the crowd's ear, the *auris populi*, when that ear has to listen to accented pronunciations of a series of separate words, spoken by a group of L2 learners?

A self-evident manner of finding out whether a word produced by L2 learners has been perceived or understood is by asking native listeners to orthographically transcribe the words uttered by L2 learners. A strong reason for doing this is that learners do not actually communicate with a limited number of experts, but with a various and extensive group of native listeners. A promising way of reaching this group is by crowdsourcing. In doing so, we will not only obtain a large and various group of native listeners, but at the same time we will be able to collect a variety of transcriptions on the speech of many L2 speakers [4], [5].

The aim of the current study is to investigate how the *auris populi*, the crowd's ear, would deal with possibly deviant L2 vowel realizations. The listeners' judgments revealing the "wisdom of the crowd's ear" [6] will help us understand which features of the learner vowel productions may cause confusions in Dutch lay listeners'perception.

In the remainder of this paper, we first present the research background in Section 2. Section 3 describes the method, the crowdsourcing experiment and the quality control. The results are presented in Section 4 and discussed in Section 5. Finally, we draw the conclusions of our study in Section 6.

2. Research background

There are considerable differences between the Dutch and the Spanish vowel inventories [7], [8], [9], [10], [11], [12]. First, Spanish has five vowels (/a, e, i, o, u/) [13], whereas Dutch has fifteen unreduced vowels (seven tense vowels: /i, y, u, eˑ, øˑ, oˑ, aˑ/; five lax vowels: /ɪ, ɛ, ɔ, ʏ, ɑ/; three diphthongs: /ɛi, œy, ɔu/) and the reduced vowel schwa /ə/ [14]. Second, Dutch has a tense/lax distinction, including vowel length (short vowels: /i, ɪ, ɛ, y, ʏ, ɑ, u, ɔ/, long vowels: /aˑ, eˑ, oˑ, øˑ, ɛi, œy, ɔu/), whereas Spanish does not have contrastive vowel length. Third, Dutch has four front rounded vowels: /ʏ, y, øˑ, œy/, whereas in Spanish all rounded vowels (/o, u/) are back.

Previous research has investigated the speech production of adult Spanish learners of Dutch [7], [8], [9], [10]. Studies conducted by Burgos et al. [7], [8] based on samples of extemporaneous speech showed that vowel errors were more frequent and persistent than consonant mispronunciations. For this reason, follow-up research was conducted on the vowels. Burgos et al. [9], [10] reported on studies in which elicited material containing read speech was employed. The use of read speech containing all speech sounds that are problematic for Spanish learners, was aimed at obtaining sufficient mispronunciations to be acoustically analyzed. Burgos et al. [9] studied the production of three vowel contrasts (/ɑ-aˑ/, /ɪ-i/, /ʏ-øˑ/), and that of the Spanish learners' realizations of all fifteen Dutch vowels [10]. Both studies [9], [10] concentrated on the acoustic analysis of the vowels produced by the Spanish learners in comparison to those produced by Dutch native speakers, and concluded that adult Spanish learners do not

September 6 – 10, 2015, Dresden, Germany

employ duration and spectral properties in a native-like manner. Moreover, in Burgos et al. [7], [8], [9], [10] it was found that the L1 phonology influences L2 vowel production and that the five Spanish vowels appear to function as "attractors" for the larger set of Dutch vowels. Based on the results of the studies mentioned above, we can advance the following predictions. First, we hypothesize that Dutch lay listeners will transcribe the tokens produced by the Spanish L2 learners differently from their canonical forms. Second, we expect to find the "attractor" effect phenomenon in the listeners' transcriptions. Third, we predict that deviant patterns found in the acoustic measurements on the same speech material will be mirrored in the listener's transcriptions.

3. Method

3.1. Speakers

To obtain a representative sample of Spanish L1-Dutch L2 vowel pronunciation errors, speech samples from 28 adult Spanish learners of Dutch (9 males, 19 females) with varying degrees of proficiency (A1, n=10; A2, n=7; B1, n=4; B2, n=7, according to the Common European Framework of Reference for Languages [15]) were used in the current study. These data had previously been analyzed in Burgos et al. [10].

3.2. Speech stimuli

The speech stimuli consisted of isolated words in Dutch read by adult Spanish learners. Every speaker read a set of 29 monosyllabic words in which all fifteen Dutch vowels in stressed position were presented. The same elicitation material was previously used in Van der Harst [16] and Van der Harst et al. [17]. All the words ended either in /s/ or /t/, as it is known that these consonants scarcely alter the quality of the preceding vowel [16], [17].

Table 1. S*elected -s and -t words used as speech stimuli from Van der Harst [16]; Vow=Vowel.*

	Vow	s-word		t-word	
Monothongs	/i/	*Kies*	/kis/	*Riet*	/rit/
	/ɪ/	*Vis*	/vɪs/	*Fit*	/fɪt/
	/ɛ/	*Zes*	/zɛs/	*Vet*	/vɛt/
	/y/	-	/y/	*Fuut*	/fyt/
	/ʏ/	*Zus*	/zʏs/	*Put*	/pʏt/
	/u/	*Poes*	/pus/	*Voet*	/vut/
	/ɔ/	*Vos*	/vɔs/	*Vlot*	/vlɔt/
	/ɑ/	*Gas*	/xɑs/	*Rat*	/rɑt/
Long mid vowels	/aˑ/	*Aas*	/aˑs/	*Staat*	/staˑt/
	/eˑ/	*Mees*	/meˑs/	*Beet*	/beˑt/
	/øˑ/	*Neus*	/nøˑs/	*Neut*	/nøˑt/
	/oˑ/	*Boos*	/boˑs/	*Boot*	/boˑt/
Diphthongs	/ɛi/	*Ijs*	/ɛis/	*Spijt*	/spɛit/
	/œy/	*Huis*	/hœys/	*Fluit*	/flœyt/
	/ɔu/	*Kous*	/kɔus/	*Fout*	/fɔut/

Table 1 shows an overview of all fifteen Dutch vowels and their corresponding orthographic and phonological representation. No example of the vowel /y/ followed by /s/ was included, as this combination does not appear in Dutch monosyllabic words, except proper names.

For this experiment we used a set of 29 words produced by 28 Spanish learners. Six speech samples were left out. During the task transcribers were offered a word they had transcribed earlier every 30th token. This was done to calculate the intra-transcriber agreement. The inclusion of repeated items gave a maximum of 833 speech stimuli used in the transcription task.

3.3. Listeners

Prior to participating in the experiment, listeners read the instruction of the transcription task. They were told that they were going to listen to utterances and that they literally had to transcribe what they heard using orthographic spelling. Listeners were allowed to transcribe foreign and non-existing words which might closely represent the heard utterance. An online questionnaire was administered to obtain background information about the listeners. The number of questions presented in the questionnaire was limited to keep the crowdsourcing experiment as simple and accessible to lay listeners as possible. The online questionnaire contained questions concerning mother tongue, gender, age and completed education. Almost 200 listeners participated in the transcription task. Part of the participants were filtered out, resulting in 159 listeners whose data was included in the current study (see section 3.5). All participants were Dutch non-expert native listeners.

3.4. The crowdsourcing experiment

A web application was developed in Django, in which participants could listen to the stimuli and type what they heard. The application was set up in such a way that it was easy to use and also fun to do. Each participant received a score indicating the percentage of "correct" transcriptions. This score was based on the most frequent transcriptions given to a word by all (previous) transcribers. The idea behind providing a score was to motivate the participants and introduce a game element, as the score could be shared on Facebook. This helped recruiting new participants. Participants transcribed 100 tokens on average. See our companion paper by Sanders et al. [18] for a detailed description of the application.

3.5. Quality control

Several criteria were used to filter the data. Only listeners who had Dutch as a native language were included. Secondly, listeners had to transcribe >10 tokens, to be sure that they really got started to perform the task. The maximum of 833 transcriptions per listener was included (three listeners continued to perform a second round).

We used two additional quality control criteria to ascertain the reliability of the data, a measure of intra-transcriber agreement and a measure of inter-transcriber agreement [4], [5]. The intra-transcriber agreement was based on the transcriptions of the repeated items. The inter-transcriber agreement criterion was based on the percentage of shared common transcriptions (see [18]). Listeners failing to meet both agreement criteria were removed from the database. Filtering our data resulted in a total of 17.534 tokens transcribed and 159 listeners.

4. Results

4.1. Listeners' transcriptions, vowel confusions

The listeners' transcriptions show that both consonants and vowels were given canonical and non-canonical transcriptions. We will now focus on the vowels, although consonants also deserve further investigation. Table 2 displays the most frequent listeners' transcriptions per vowel. The fifteen target Dutch vowels are presented in alphabetic order in the columns, except for the last three vowels, corresponding to the three diphthongs. The rows show the transcribed vowels, including both the canonical transcriptions of the target vowels (indicated by the black squares) and the non-canonical transcription <ai>. The percentages in the cells indicate how often a transcription was given to a target vowel. The column Total shows the sum of all percentages of transcribed vowels per row. Transcriptions containing percentages of less than 1% are aggregated in the Rest category (see last row in Table 2). Overall percentages for canonical and non-canonical transcriptions were calculated. Our results indicate that 67.44% of all transcriptions are canonical and 32.56% non-canonical.

The various non-canonical transcriptions (see rows in Table 2) show that there is variation in the way the vowels were transcribed by the lay listeners. The highest variation was found in the long mid vowel <eu> and the diphthong <ui>. The lowest variation appears in the vowel <aa>.

An interesting confusion pattern is found in the non-canonical transcriptions for the target vowels <u> and <uu>, which are often confused with each other, and especially with the vowel <oe>, as displayed in Table 2.

Table 2 shows that the target long mid vowel <ee>, and the target diphthongs <ij> and <ui> have non-canonical transcriptions, such as <ei>, <ai> and <au>, respectively. These transcriptions seem to point to strong diphthongization, as observed earlier in Burgos et al. [9], [10].

The column Total in Table 2 shows that some vowels were more often transcribed by the listeners, namely, <aa>, <e>, <ie>, <o>, <oo> and <oe>, all of them producing percentages above 100. These vowels seem to resemble the five Spanish vowels <a>, <e>, <i>, <o>, <u>, suggesting the idea of the Spanish vowels functioning as "attractors" for the larger set of Dutch vowels, as previously observed in Burgos et al. [7], [8], [9], [10]. A conspicuous case, which needs to be further examined, is the one of the two Dutch vowels <o> and <oo>, which appear to be attracted both by the Spanish vowel <o>.

In order to better understand how lay listeners cope with transcribing specific vowels in a contrast, we decided to study three Dutch vowel pairs <a>-<aa>, <i>-<ie> and <u>-<eu> in more detail. These vowels, produced by Spanish L2 learners, were acoustically analyzed in Burgos et al. [9]. They differ from each other in the way duration and place of articulation are used to make a contrast. The contrast <a>-<aa> is based on duration and place. The distinction between the vowels in the pair <i>-<ie> hinges on place and not on duration, as both vowels are short in native Dutch. The contrast <u>-<eu> is only based on duration, as both vowels have a similar place of articulation and are both front rounded vowels.

Table 2. *Most frequent orthographic representations of all fifteen Dutch vowels transcribed by Dutch lay native listeners; transcribed vowels <1% are aggregated in the Rest category, >10% in grey, >5% in light grey, canonical transcriptions in black squares, the orthographic representation of the target Dutch vowels in the columns, the transcribed vowels in the rows; Vow=Vowel.*

Vow	a	aa	e	ee	eu	i	ie	o	oo	oe	u	uu	ij	ou	ui	Total
a	72.21	13.80	0.25	0.00	0.08	0.00	0.17	2.22	0.74	0.08	1.17	0.00	0.08	1.71	0.08	92.59
aa	23.04	79.12	0.00	0.00	0.08	0.00	0.17	0.00	1.24	0.00	0.00	0.00	0.08	0.81	0.17	104.71
e	1.58	0.67	89.38	18.71	2.18	3.28	1.51	0.08	0.08	0.17	2.33	0.00	4.23	0.32	0.67	125.19
ee	0.00	0.00	1.40	59.31	1.01	1.34	6.54	0.00	0.00	0.00	0.08	0.17	4.39	0.00	0.00	74.24
eu	0.00	0.00	0.08	0.25	65.86	0.08	0.17	0.00	0.00	0.08	0.42	4.78	0.24	3.57	3.76	79.29
i	0.00	0.00	4.94	2.09	0.42	51.34	12.66	0.16	0.00	0.00	2.25	0.17	0.08	0.08	0.17	72.18
ie	0.00	0.00	0.25	2.42	0.67	37.98	73.09	0.00	0.00	0.00	0.42	0.34	0.00	0.00	0.00	115.17
o	0.08	0.00	0.00	0.07	0.25	0.08	0.00	86.34	16.29	2.27	1.33	0.34	0.08	0.49	0.25	107.87
oo	0.00	0.00	0.00	0.00	5.95	0.00	0.08	4.94	72.70	11.09	1.50	5.29	0.00	15.84	1.25	118.64
oe	0.08	0.00	0.00	0.08	0.25	0.34	0.00	2.39	4.38	78.24	34.89	40.96	0.00	3.33	1.75	166.69
u	0.50	0.17	1.23	0.00	2.85	3.36	0.17	0.91	0.00	1.93	44.96	4.61	0.00	1.22	0.50	62.41
uu	0.00	0.00	0.00	0.00	0.50	1.26	0.17	0.00	0.33	1.66	9.16	41.10	0.08	0.08	1.09	55.43
au	0.08	0.00	0.00	0.00	0.00	0.00	0.00	0.16	0.08	0.00	0.00	0.00	0.00	1.95	5.10	7.37
ai	0.08	0.42	0.00	0.08	0.59	0.00	0.25	0.00	0.00	0.00	0.00	0.00	9.36	0.00	0.08	10.86
ei	0.00	0.00	1.07	6.68	0.25	0.08	1.17	0.00	0.00	0.00	0.00	0.00	3.17	0.00	0.17	12.59
ij	0.00	0.42	0.33	6.35	0.34	0.00	0.42	0.00	0.00	0.21	0.00	0.00	66.48	0.00	0.50	75.05
ou	0.17	0.51	0.00	0.00	0.84	0.00	0.00	0.33	0.58	2.18	0.08	0.17	0.08	59.46	3.59	67.99
ui	0.92	0.42	0.08	0.17	3.27	0.00	0.92	0.74	0.08	0.17	0.17	0.51	3.66	7.23	72.10	90.44
Rest	1.26	4.47	0.99	3.79	14.61	0.86	2.51	1.73	3.50	1.92	1.24	1.56	7.99	3.91	8.77	59.11

4.2. Listeners' transcriptions of three vowel pairs

The listener's transcriptions of the vowels in the pairs <a>-<aa>, <i>-<ie> and <u>-<eu> present different patterns (see Table 2). Regarding the pair <a>-<aa>, it appears that the target vowel <a> was more often transcribed as <aa> than the target vowel <aa> as <a>, pointing to an asymmetrical confusion. A similar asymmetrical confusion is found in the transcriptions of the target vowels in the <i>-<ie> contrast. Table 2 shows that the target vowel <i> is frequently transcribed as <ie>, more often than <ie> as <i>. Concerning the pair <u>-<eu>, the <u> seems to be overwhelmingly transcribed as <oe> by the listeners, although it also shows other minimal confusions, as displayed in Table 2. On the other hand, the target vowel <eu> does not have a clear competitor, and has the highest variation of transcribed vowels of all target vowels (see rows in Table 2).

5. Discussion

In general, the results of the listener's transcriptions for the pairs <a>-<aa>, <i>-<ie> and <u>-<eu> are in line with the outcomes of the acoustic measurements of the same speech material presented in Burgos et al. [9]. Asymmetrical confusions between the vowels in the pairs <a>-<aa> and <i>-<ie> have been found in both the listener's transcriptions and in the acoustic data. These are probably due to the existence of Spanish counterparts for the Dutch vowels <a> and <i>. The front rounded vowels in the pair <u>-<eu>, which "are unfamiliar and fall in an empty portion of the native vowel space" [19, p. 123], exhibit a high degree of variation which can be ascribed to difficulties with front rounding, a phenomenon which does not occur in Spanish [9]. An interesting discrepancy between the listeners' transcriptions and the acoustic data was found with respect to duration. The transcriptions do not indicate longer durations of the vowels in question, while objective measurements showed that the learners' vowels were longer than the corresponding ones produced by native speakers [9]. This may suggest that native listeners "somehow" normalize duration in learner speech with little consequences for word recognition and intelligibility. However, it is possible that deviant duration values do have consequences for the degree of foreign accent [3] that is noticeable in the speech of Spanish learners of Dutch. This is definitely a topic that deserves attention in future research.

The listeners' transcriptions show that some vowels were more often transcribed by the non-expert native listeners, as in the case of <aa>, <e>, <ie>, <o>, <oo> and <oe>. These Dutch vowels resemble the five Spanish vowels <a>, <e>, <i>, <o> and <u>, which confirms the idea of the Spanish vowels functioning as "attractors" for some of the Dutch vowels, as advanced in Burgos et al. [7], [8], [9], [10]. In these studies we already discussed the influence of the L1 orthography on L2 vowel production, which appears to play an important role in the vowel mispronunciations and on the activation of such an "attractor" mechanism.

The influence of the native language is also noticeable in the listeners' transcriptions of the long mid vowel <e> and the Dutch diphthongs <ui> and <eu>. Non-canonical transcriptions of the target vowel <ee> as <ei>, which bears a likeness with the Spanish diphthong <ei>, "exhibiting both a similar degree of formant movement" [20, p. 3], point to diphthongization. Previous studies [10] already indicated that Spanish L2 learners tend to diphthongize long vowels more

than Dutch native speakers. The target vowel <ij> was often transcribed by the lay listeners as <ai>. The vowel combination <ai> does not correspond to any vocalic phoneme in native Dutch, but does exist in Spanish and corresponds to the diphthong <ai>. A similar situation applies to the target vowel <ui>, often transcribed as <au>, which is also a diphthong in Spanish. These findings seem to indicate that the "attractor" effect of the Spanish phonology also involves diphthongs, which are combinations of two vowels in Spanish [13]. The idea of the Spanish diphthongs <ei>, <ai> and <au> functioning as "attractors" for the Dutch long mid vowel <ee> and the diphthongs <ij> and <ui> did not appear from previous studies and can be considered an additional finding brought up by the *auris populi*, the crowd's ear.

As to the value of collecting speech transcriptions through crowdsourcing, we did notice that some of the listeners tended to transcribe what they thought the token was, i.e., the canonical transcription, instead of literally transcribing what they heard, because in this way they could get a high score to share it on Facebook. Possibly, this bias made some listeners not entirely perform the task as we wanted, namely, literally transcribing the tokens spoken by the Spanish L2 learners. However, using only canonical transcriptions would not return a 100% score, as for some words the common transcription was non-canonical. Correctedness scores never were higher than 90%, making the game of transcription hard enough to prevent a single strategy to be successful.

The existence of possible biases is an issue that certainly deserves further examination when dealing with crowdsourced native transcriptions (cf. [4], [5]), but the transcriptions we got seem to reflect quite accurately the phonetic variation in the stimuli. Despite the potential drawback of the *auris populi* methodology, we found crowdsourcing to be a valuable tool to collect a large amount of L2 speech transcriptions from an extensive and diverse group of native non-expert listeners.

6. Conclusions

The aim of the current study was to investigate how the *auris populi*, the crowd's ear, would deal with possibly deviant L2 vowel realizations. The transcriptions delivered by lay listeners appear to provide a majority of common transcription plus relevant information on variation and details in the way Spanish L2 learners' pronunciation is perceived by lay Dutch listeners. The listeners' transcriptions mirror the vowel problems and the "attractor" effect found in previous studies conducted at our lab and based on both expert annotations [7], [8] and acoustic measurements [9], [10]. The findings of our study confirm that the native human ear is able to perceive deviations in L2 accented vowel realizations. An additional advantage of crowdsourcing is that considerable amounts of speech material from many L2 speakers can be transcribed or rated when many lay listeners are willing to participate. We would like to draw the following conclusions. First, the results of our study indicate that Dutch vowels pronounced by Spanish learners were often transcribed differently from the canonical forms by Dutch non-expert native listeners. Second, this study revealed that not only the five Spanish vowels, but also three Spanish diphthongs function as "attractors" for the larger set of Dutch vowels. Third, the listeners' transcriptions of three vowel pairs are in line with results of acoustic measurements of the same speech material, but point to a possibly different role of duration deviations. Four, the *auris populi* methodology has proven to be a practical and valuable tool for future L2 speech research.

7. References

[1] M. Long, "Maturational constraints on language development," *Studies in Second Language Acquisition*, vol. 12, pp. 251–285, 1990.

[2] D. Birdsong, and M. Molis, "On the evidence for maturational constraints in second language acquisition," *Journal of Memory and Language*, vol. 44, no. 2, pp. 235–249, 2011.

[3] T. M. Derwing, and M. J. Munro, "Second language accent and pronunciation teaching. A research based approach," *TESOL Quarterly*, vol. 39, no. 3, pp. 379–397, 2005.

[4] G. Parent, and M. Eskenazi, "Speaking to the Crowd: looking at past achievements in using crowdsourcing for speech and predicting future challenges," *Proceedings of Interspeech 2011*, Florence, Italy, pp. 3037–3040, 2011.

[5] M. Eskenazi, G. A. Levow, H. Meng, G. Parent, and D. Suendermann, *Crowdsourcing for Speech Processing: Applications to Data Collection, Transcription and Assessment*. Wiley Publishing, 2013.

[6] S. A. Kunath, and S. H. Weinberger, "The wisdom of the crowd's ear: speech accent rating and annotation with Amazon Mechanical Turk," *Proceedings of CSLDAMT '10, Association for Computational Linguistics*, pp. 168–170, 2010.

[7] P. Burgos, C. Cucchiarini, R. Van Hout, and H. Strik, "Pronunciation errors by Spanish learners of Dutch: A data-driven study for ASR- based pronunciation training," *Proceedings of Interspeech 2013*, Lyon, France, pp. 2385–2389, 2013.

[8] P. Burgos, C. Cucchiarini, R. Van Hout, and H. Strik, "Phonology acquisition in Spanish learners of Dutch: error patterns in pronunciation," *Language Sciences*, vol. 41, pp. 129–142, 2014.

[9] P. Burgos, M. Jani, C. Cucchiarini, R. Van Hout, and H. Strik "Dutch vowel production by Spanish learners of Dutch: duration and spectral features," *Proceedings of Interspeech 2014,* Singapore, pp. 529–533, 2014.

[10] P. Burgos, M. Jani, C. Cucchiarini, R. Van Hout, and H. Strik, "Duration and spectral features of Dutch vowels produced by adult Spanish learners," *Journal of the Acoustical Society of America*, (submitted).

[11] P. Escudero, T. Benders, and S. Lipski, "Native, non-native and L2 perceptual cue weighting for Dutch vowels: The case of Dutch, German, and Spanish listeners," *Journal of Phonetics*, vol. 37, pp. 452–465, 2009.

[12] P. Escudero, and K. Wanrooij, "The effect of L1 orthography on non-native vowel perception," *Language and Speech*, vol. 53, no. 3, pp. 343–365, 2010.

[13] J. I. Hualde, *The sounds of Spanish*, Cambridge: Cambridge University Press, 2005.

[14] G. Booij, *The phonology of Dutch*, Oxford: Clarendon Press, 1995.

[15] CEFR (Common European Framework of Reference for Languages: Learning, teaching, assessment). Online: http://www.coe.int/t/dg4/linguistic/cadre1_en.asp, accessed on 27 Mar 2015.

[16] S. Van der Harst, "The vowel space paradox. A sociophonetic study on Dutch," Utrecht: LOT Dissertation, The Netherlands, 2011.

[17] S. Van der Harst, H. Van de Velde, and R. Van Hout, "Variation in Standard Dutch vowels: The impact of formant measurement methods on identifying the speaker's regional origin," *Language Variation and Change*, vol. 26, no. 2, pp. 247–272, 2014.

[18] E. Sanders, P. Burgos, C. Cucchiarini, and R. Van Hout, "*Palabras*: crowdsourcing transcriptions of L2 speech," *Proceedings of Interspeech 2015,* Dresden, Germany, (submitted).

[19] M. Goudbeek, A. Cutler, and R. Smits, "Supervised and unsupervised learning of multidimensionally varying non-native speech categories," *Speech Communication*, vol. 50, pp. 109–125, 2008.

[20] P. Escudero, and D. Williams, "Perceptual Assimilation of Dutch vowels by Peruvian Spanish listeners," *Journal of the Acoustical Society of America*, vol. 129, no. 1, pp. 1–7, 2011.