INTERSPEECH 2013

On the role of L1 speech production in L2 perception: Evidence from Spanish learners of French

Natalia Kartushina [1], *Ulrich Hans Frauenfelder*[1]

[1] Laboratoire de la Psycholinguistique Expérimentale, Université de Genève, Suisse

Natalia.Kartushina@unige.ch, Ulrich.Frauenfelder@unige.ch

Abstract

Spanish (L2) speakers have difficulty in perceiving the French vowel height contrast /e/-/ɛ/. L2 perception models attribute this perceptual difficulty to native (L1) phonology. For example, the Spanish /e/ vowel, which is perceptually similar to both French vowels, is assumed to influence their perception through an assimilation process. This pattern of assimilation depends upon the degree of the similarity between the vowels. This research aims to shed light on the relevant aspects of L1 phonology and its relation to L2 phonology that can be operationalized to predict L2 perception. Traditionally, measures of perceptual similarity have involved comparing the acoustic or gestural properties of productions by groups of 'referenced' native speakers of the two languages of interest. Here, we rather compute acoustic distances (based on F1 and F2 values) between the native productions by Spanish adolescents learning French (N=14) and a group of native French speakers. In addition to this distance measure, we also introduced a novel measure, the within-speaker variability or compactness in productions of a L1 vowel to predict L2 perception. The results revealed that individual productions of Spanish /e/ and their variability (or compactness) predicted L2 perception accuracy of the French /e/-/ɛ/ vowels.

Index Terms: L1 transfer, L2 perception, interphonology

1. Introduction

Second language (L2) learning takes place at different linguistic levels: phonological, morphological, lexical, syntactic, etc. Although L2 learners can generally achieve a decent overall mastery of the second language (L2), the acquisition of L2 phonology often remains elusive [1, 2, 3, 4]. Despite years of L2 experience, L2 speakers fail to perceive and produce non-native contrasts, and particularly, those that do not exist in their L1. For instance, if we take English as the L2 target, we note that L1 Spanish (SP) speakers have difficulties in discriminating the /i/-/ɪ/ vowel contrast [3], whereas L1 Korean speakers do not distinguish the /e/-/æ/ contrast [4]; both of these L2 contrasts are absent in their respective native phonologies.

The dominant theoretical perspectives attribute these L2 perceptual difficulties to the influence of the native phonological system [5, 6, 7]. However, it remains unclear precisely which aspects of L1 phonology are relevant and how these are best operationalized in predicting L2 success. Traditionally, acoustic and articulatory descriptions of productions by groups of native speakers of the two languages of interest are analyzed to make predictions about the difficulty of L2 perception.

This study proposes alternative ways of operationalizing the similarities and differences between L1 and L2 sounds. More specifically, we take acoustic distances computed for each individual participant rather than group averages found in the literature. In addition, we analyze the productions of the native FR participants who recorded the stimuli as the L2 reference. Finally, we collect a new measure, the distribution (compactness) of the individual speakers' productions. These measures were used to predict perception accuracy of non-native sounds in SP students learning L2 French (FR) in a SP middle-school.

The Full Transfer (FT) model claims that L1 constitutes the initial state of L2 acquisition [8, 9]. That means that at the beginning of L2 learning, the L1 system (i.e., phonological space) is transferred into the L2 "virgin" system. Therefore, at this stage it is assumed that native and non-native languages form a single interlanguage phonological space [10].

During this transfer, non-native sounds are perceptually mapped onto native categories based on their perceived similarity [11]. According to the Perceptual Assimilation Model (PAM, [5, 12]), developed for the perception of L2 sound contrasts, these contrasts are assimilated to native phonemes based on their respective gestural similarity. Three patterns of L2 contrast assimilation are distinguished: Two-Category (TC) assimilation, the two L2 sounds map onto two different L1 categories; Category-Goodness (CG) assimilation, both L2 sounds map onto one L1 category, but one L2 sound is a better exemplar of this L1 category than the other; and Single-Category (SC) assimilation, both L2 sounds are equally good exemplars. Because in SC and CG assimilations both L2 sounds assimilate to one native category, these L2 contrasts are claimed to be more difficult to acquire than the L2 contrasts that show the TC assimilation pattern. According to the SLM framework [6], developed for the perception of individual L2 sounds, these sounds are mapped onto L1 categories on the basis of their perceived acoustic similarity to their closest L1 sound. L2 sounds are classified either as identical or as similar to the L1 sounds, or as new, that is, dissimilar from the L1 sounds. The ease of acquisition of a L2 sound is assumed to depend upon its degree of similarity to the closest native sound category. New categories are created easily for new sounds, are difficult for similar sounds and are unnecessary for identical sounds.

The notion of the perceptual similarity between the transferred L1 sounds and the target L2 sounds is at the core of L2 perception models. Although it remains unclear which properties of the native phonological system actually transfer to L2, the L2 perception models seem to favor the transfer of native phonetic features [9]. These features have generally been conceptualized as L1 prototypes (ideal examples) defined as averages across groups of native speakers and their distributions in the L1 space. They can be measured, for example, in terms of averaged acoustic values (first three

25 – 29 August 2013, Lyon, France

formants: F1, F2, F3) and their standard deviations. For instance, Escudero and Vasiliev [13] found that cross-language acoustical similarities between the formant values of L1 and L2 speech sounds predict L2 phonemic assimilation patterns in perception. They showed that while Canadian FR (CF) /ɛ/ and /æ/ vowels are identified as SP /e/ and /a/ vowels respectively by Peruvian speakers, Canadian English (CE) /ɛ/ and /æ/ vowels are identified as one /e/ SP vowel. This work suggests that the acoustic properties of the L1 vowels found in a group of L1 speakers (corpus analysis) predict the L2 vowel assimilation patterns for a priori any speaker of this linguistic community.

In L1 SP learners of L2 FR, the FR /e-ɛ/ contrast CG assimilates to the SP /e/ phoneme where FR /e/ is generally taken to be close to L1 /e/ and FR /ɛ/ is farther away [5, 14, 15]. Therefore, the contrast is predicted to be well discriminated since FR /e/ is categorized as identical to the SP /e/ and FR /ɛ/ is a discrepant example of the native /e/. However, more recent corpus analysis data [16, 17] have suggested that FR /ɛ/ is acoustically closer to SP /e/ than FR /e/. Accordingly, we could expect CG assimilation for the FR /e-ɛ/ contrast, but in which FR /ɛ/ is a better example of SP /e/ and is assimilated but not the FR /e/ which becomes the discrepant example [5].

As we see here, the predictions change as function of the L1 and L2 production references that are adopted. The situation is even more complicated if we consider that within any given group of native speakers, speech production varies significantly. So it is quite likely that SP /e/ for some Spanish speakers is closer to FR /e/ than to FR /ɛ/ and for others, the opposite is true. It is important then in making predictions to consider the individual L1 productions and their distance from the L2 targets.

It is also important to note that there is not only variation across L1-L2 groups and across individuals within a group, but also within individuals. More specifically, the amount of variation (compactness) observed for the production of the same sound by the same speaker is not constant. Can it be that these individual differences in articulation in L1 transfer into interphonological space and predict individual differences in learning to perceive L2 phonemes? This seems quite likely since the distribution of the vowel exemplars in the L1 acoustic space, and more specifically, the articulatory variability in L1 vowel realizations is related to the perceptual accuracy of this vowel. Perkell and his colleagues have shown that among a group of L1 speakers, some speakers have compact realizations (a compact acoustic space), whereas others do not (a distributed space) [18]. Importantly, the articulation accuracy is tightly connected to vowel discrimination accuracy: individuals with higher discrimination scores pronounced contrasts more distinctly. This finding is compatible with the claim of the DIVA model [19] in which the articulatory movements for the vowels are planned in terms of auditory goal regions in auditory-temporal space. L1 speakers who are able to perceive fine acoustic differences will learn articulatory goals that are smaller, more compact.

The SLM postulates that "the mechanisms and the processes used in L1 remain intact over the life span, and can be applied to L2 learning" [6, p. 239]. If we assume that L1-like perception mechanism is used to perceive L2 sounds and that L1 perception is tightly related to L1 production, we expect to find a relationship between the L1 production compactness and L2 perception accuracy.

To sum up, we propose that the distribution of the phonetic realizations for a given sound transfers into the L2 space of this individual. If this is so, the distance between individual prototypes and L2 sounds as well as the compactness of the L1 vowel space should have effect on the perceptual accuracy of similar L2 vowels. Thus, the smaller the distance between the L1 and L2 sounds is, the more this L2 sound is assimilated and hence identified as this L1 sound. As for the L1 compactness, we hypothesize that speakers with a compact phonological space should perceive L2 vowels better than speakers with broadly distributed phonological spaces.

2. Method

2.1 Participants

Participants were fourteen female SP pupils (mean age 16) studying L2 FR at a SP middle school in Plasencia, Spain. They were all monolingual and had on average 4 years of learning FR. They all came from the same region in Spain and had never lived in a FR speaking country.

2.2 Procedure

L2 perception task: To assess L2 phonological representations in perception we used a five-forced-choice identification task for the FR /e-ɛ/ contrast. The participants heard the isolated vowels and were presented with picture labels (to avoid orthographical bias, [20]) to give their identification response to the FR vowels. Table 1 shows the monosyllabic, non-cognate, concrete nouns containing the vowels used in the task.

Table 1. *Vowels/names of pictures used in the identification task.*

Vowels	i	ɛ	e	u	y
Pictures	lit	verre	clef	louche	jupe

The task was composed of the two phases. During the first, familiarization phase, the participants had to learn the association between the picture label and the vowel it represents. These labels were then used in the testing phase (cf. Figure 1) to identify the vowels. Each of the two target vowels was presented 10 times in random order along with 3 other vowels included in the experiment.

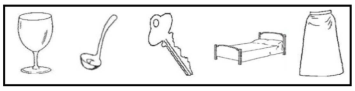

Figure 1: *Pictures used in the testing phase (in order from left to right: verre /vɛʁ/- glass, louche /luʃ/ - ladle, clef /kle/ - key, lit /li/ - bed, jupe /ʒyp/ - skirt).*

L1 production task: To test the participants' L1 production a reading task was used. The participants read a passage from a chapter of a SP translation of "The Godson" by Leo Tolstoy. Participants were tested individually.

2.3 Materials

Five native FR speakers were recorded. They were all from the same region (Paris, Ile-de-France, IDF) to minimize the effect of regional dialect. Multiple speakers were used to increase the variability across vowels and thus to better approximate ecological L2 learning and communication conditions and to avoid the one-speaker bias [21]. They were recorded while reading three lists of the ten FR sentences which contained one isolated vowel, as in “Je prononce /**ɛ**/ comme dans l**ai**t” ([ʒe pronõs **ɛ** kom dã lɛ], “I pronounce /ɛ/ as in milk”). The sentences in the three lists were identical except for the final word that began either with /l/ (lait), /s/ (sait) or /p/ (paix) (milk, know, peace respectively). Isolated vowels were extracted from the sentences; normalized for intensity and controlled for length (mean 210 ms). The vowels were then rated by 5 French speech therapists; the vowels that were accepted as good prototypes of FR by at least four therapists were kept. In total, 20 /e-ɛ/ vowels were selected: 2 vowels * 5 different speakers * 2 tokens per speaker.

2.4 Acoustic analyses

To obtain our participants’ phonological space for L1 SP /e/, we analyzed their SP reading data using Praat [22]. Vowels were selected, manually adjusted, and extracted. On average 47 production exemplars of SP /e/ were obtained per participant. As the target L2 contrast is essentially distinguished by height and to a lesser degree frontness, the first two formants were considered for the analyses. The F1 and F2 of the vowels were calculated. To construct the target L2 FR phonological spaces for FR /e/ and /ɛ/ the productions of the native FR speakers were analyzed in similar fashion.

Furthermore, the SP participants’ native reading productions were also analysed for vowel compactness in L1 acoustic space, using the following formula for the area of the ellipse:

$$A = xy\pi \qquad (1)$$

where x is the F1 standard deviation and y is the F2 standard deviation.

To obtain the distance between SP /e/ and FR /e/, for each SP participant the difference between the average formant values (F1 and F2) for the /e/ vowel and the formant values for the ten FR /e/ vowels pronounced by five native female FR speakers were calculated, giving for each participant 10 distance values for each comparison. For instance, for the SP/e/-FR/e/ height distance, the following formula was used:

$$\text{Distance 1} = F1_{SP/e/} - 1F1_{FR/e/} \qquad (2)$$

where, $F1_{SP/e/}$ is the mean individual value for SP /e/ and $1F1_{FR/e/}$ is the first F1 value for FR /e/ among the ten productions that we have recorded.

Therefore, for each SP participant ten values representing the height (F1) distance and ten values representing the frontness (F2) distance between SP /e/ and FR /e/ vowels were obtained. Identical subtractions (differences between the L1-L2 formant values) were done for SP /e/ and FR /ɛ/ vowels.

3. Results

3.1 Acoustic analysis of L1-L2 productions

L1 compactness: The A value served as the individual L1 compactness indicator. It ranged from 16277 Hz to 57689 Hz.

L1/e/-L2/e/ and L1/e/-L2/ɛ/ height and frontness distances: For every participant 10 values were obtained for each comparison. The range of distances is indicated in the Table 2.

L1-L2 phonemic spaces comparison: Our group acoustic analyses for the native SP and FR productions revealed that both FR vowels are similar to SP /e/, however, FR /ɛ/ is slightly closer acoustically to the SP /e/ than FR /e/ (cf. Figure 2). This finding is compatible with recent data on SP and FR languages [15, 16 respectively]. However, the analysis of individual participant shows considerable interspeaker heterogeneity concerning the acoustically closer L2 vowel: for some participants FR /ɛ/ is almost acoustically identical to SP /e/ and for others, FR /e/ is the closest L2 sound (Figure 2).

Table 2. *The range of height and frontness distances for SP/e/ and FR /e/, /ɛ/ vowels.*

L1-L2 comparison	Height distance, Hz		Frontness distance, Hz	
	minimal	maximal	minimal	maximal
SP/e/-FR/e/	13	212	1	-793
SP/e/-FR/ɛ/	4	-179	2	369

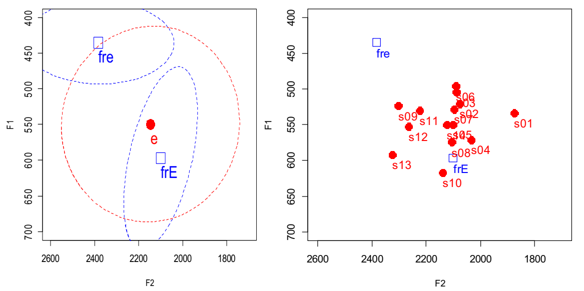

Figure 2: *Group (on the left) and individual (on the right) mean formant values for SP /e/ and FR /e, ɛ/ vowels.*

Based on the group data, two contrast-assimilation patterns can be envisaged. According to the first, FR /e-ɛ/ will follow SC and both vowels will have poor perception scores and will be misperceived as SP /e/. According to the second pattern, FR /e-ɛ/ will show CG assimilation, where FR /ɛ/ will be a better prototype of SP /e/ than FR /e/, therefore this contrast should be perceived better (may vary from moderate to very good, [5]). On the basis of the individual L1 productions, we expect that the assimilation pattern and perception performance vary as function of the individual L1-L2 distances.

3.2 L2 perception task

Student T-test analysis was run on L2 perception scores. The possible perfect score was ten. The results (cf. Figure 3) revealed that FR vowel /e/ was identified better than FR /ɛ/ vowel ($p<.05$), with identification scores of 5.9 and 4.3 respectively (chance level is at 2). FR /ɛ/ was misidentified more often as FR /e/ 53% than the opposite 36%.

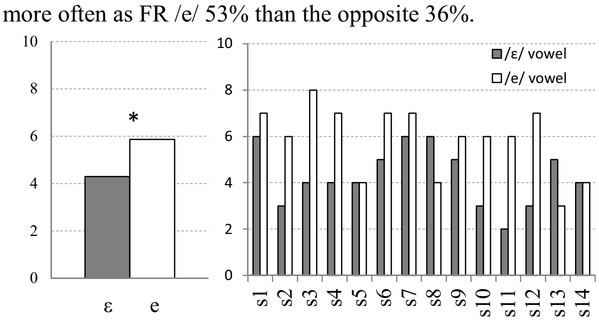

Figure 3: *Group (on the left) and individual (on the right) identification scores for the FR /e/ and /ɛ/ vowels.*

3.3 L1 production and perception of FR /e/

Generalised linear regression analyses were run on the /e/ vowel identification data using the R suite [23], rms package [24]. Three predictors were considered: height and frontness distances between SP /e/ and FR /e/ and the L1 compactness for the SP /e/ vowel.

Among the three predictors both height and frontness distances reached the significant level (p<.001 and p<.05), where the height distance is the best predictor. Once height and frontness distances are taken into account, L1 compactness did not add any predictive power. The post-hoc analyses of the two predictors showed a significant negative correlation (cf. Figure 4) between the height distance and the identification score (r = -.51, p<.05) and also a smaller but significant correlation between the frontness distance and the identification score (r = -.21, p<.05).

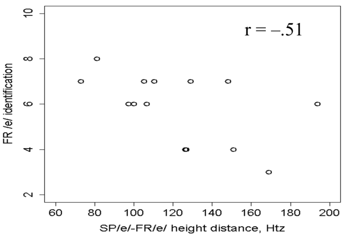

Figure 4: *Scatterplot of the height distance SP/e/-FR/e/ and the FR /e/ vowel identification score.*

3.4 L1 production and perception of the FR /ɛ/

Generalised linear regression analyses were run on the /ɛ/ vowel identification data. Three predictors were considered: height and frontness distances between SP /e/ and FR /ɛ/ vowels and the L1 compactness score for the SP /e/ vowel.

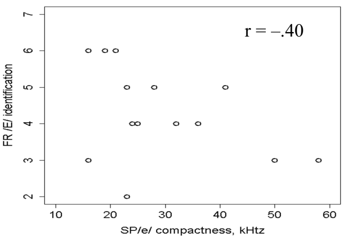

Figure 5: *Scatterplot of the L1 compactness and FR /ɛ/ identification score.*

The results showed that only L1 compactness was a significant predictor of the /ɛ/ vowel identification performance (p<.001). The post-hoc analyses showed a significant negative correlation between the L1 compactness score and the identification score (r = -.40, p<.005) as in Figure 5.

4. Discussion

4.1 L2 /e/ - /ɛ/ contrast perception

The results confirmed the SC assimilation prediction: FR /e/ and /ɛ/ vowels were both identified poorly, more than a half of FR /ɛ/ has been misidentified as FR /e/. In accordance with the predictions of the PAM model for SC contrasts, the /e/ - /ɛ/ contrast is very difficult for the participants to perceive. Moreover, the predictions based on our participants' productions predict their perception results better than those based on the production of averaged SP and FR populations [14, 15, 16, 17], showing the relevance of using the former measures.

4.2 L1 /e/ production and L2 /e/ perception

The perception accuracy score for FR /e/ varied considerably across the L2 speakers (ranging from poor (three) to very good (eight)). Our results revealed that the individual height and frontness distances between the native SP /e/ and FR /e/ vowel predict this variability in L2 perception. In line with the SLM predictions [6], the closer the SP /e/ is to the FR /e/, the better this FR /e/ sound is identified and hence classified as L1 /e/. Those L2 speakers, whose L1 /e/ is acoustically closer to the L2 /e/ phonemic space, identify this L2 /e/ vowel better (hence creating the L2 category according to the SLM) than those, whose L1 /e/ is acoustically distant from the L2 /e/. These results are in line with the recent work [13] showing that cross-language acoustical similarities of L1 and L2 speech sounds predict L2 phonetic perception assimilation patterns. The stronger correlation between the perception performance and the distance measure based on F1 than on F2 is consistent with a recent work on the role of the F1 in perception of the vowel height [25]. Taken together, these data suggest that the individual productions transfer into interphonological space and play a role in perception of similar L2 sounds.

4.3 L1 /e/ production and L2 /ɛ/ perception

The results showed that the overall identification performance on L2 /ɛ/ was poor and varied considerably across participants (from very poor (two) to good (six)). This inter-speaker variability in L2 perception is related to the compactness of their L1 productions. Indeed, we observed a correlation revealing that the more compact L1 /e/ space is, the more the SP speakers perceive the difference between FR /ɛ/ and SP /e/ and therefore correctly identify FR /ɛ/ vowel. These results suggest that the distribution of individual native vowel productions transfers into the interphonological space and affects the perception of the similar L2 sounds. This finding is in agreement with Perkell's general conclusion on the tight relationship between production and perception [18]. Further, these results are consistent with the SLM claim that the mechanisms and processes used to learn L1 are applied when learning to perceive L2 sounds [6].

5. Conclusions

This study examined the perceptual accuracy of Spanish learners on a French "SC assimilated" L2 vowel contrast. We evaluated how well the individual L1 vowel productions of Spanish speakers (distance separating L1 and L2 targets and L1 compactness) predicted their perception of these French vowels. We found that the L1-L2 acoustic distance predicts the perception performance on the acoustically similar L2 vowel (FR /e/), whereas the L1 compactness predicts the L2 perception of the FR acoustically even closer /ɛ/ vowel. From a methodological perspective, this study underscores the importance of using individual rather than group production data to characterize L1 phonological space. From a theoretical perspective, our results suggest that individual articulatory realisations transfer into the interphonological space and influence perception of non-native sounds in L2 speakers. How well these L1 production data can predict the quality of L2 productions is an important open question that we are currently working on.

6. Acknowledgments

We are grateful to the students of the Instituto de Educación Secundaria Parque Montfragüe in Plasencia, Spain for their participation and to the L2 FR professors and to the director of the school for making this study possible. We are grateful to three anonymous reviewers for their comments on an earlier draft of this paper.

7. References

[1] Long, M. H., "Maturational Constraints on Language Development", Studies in Second Language Acquisition, 251-285, 1990.

[2] Iverson, P., Ekanayake, D., Hamann, S., Sennema, A., Evans, B., G., "Category and perceptual interference in second-language phoneme learning: An examination of English /w/-/v/ learning by Sinhala, German, and Dutch speaker", Journal of Experimental Psychology: Human Perception and Performance, 34, 5, 1305–1316, 2008.

[3] Flege, J. E., "Assessing constraints on second-language segmental production and perception", in N. Schiller and A. Meyer [Ed], Phonetics and phonology in language comprehension and production, differences and similarities, 319–355, Mouton De Gruyter, 2003.

[4] Ingram, J. C. L., and Park, S. G., "Cross-language vowel perception and production by Japanese and Korean learners of English", Journal of phonetics, 25, 343–370, 1997.

[5] Best, C., "A direct realist view of cross-language speech perception", in W. Strange [Ed], Speech perception and linguistic experience: Theoretical and methodological issues, 171–204, York Press, 1995.

[6] Flege, J, "Second-language speech learning: Theory, findings, and problems", in W. Strange [Ed], Speech Perception and Linguistic Experience: Issues in Cross-Language Research, 233-277, York Press, 1995.

[7] Piske, T., MacKay, I. R. A., Flege, J.E., "Factors affecting degree of foreign accent in an L2: a review", Journal of Phonetics, 29, 191-215, 2001.

[8] Schwartz, B. and Sprouse, R., "L2 cognitive states and the Full Transfer/Full Access model", Second Language Research, 12: 40-72, 1996.

[9] Major, R., C., "Transfer in second language phonology", in Hansen Edwards, J., G., and Zampini, M., L. [Ed], Phonology and Second Language Acquisition, 63-84, John Benjamins Publishing Company, 2008.

[10] Archibald, J., "Second language phonology, phonetics, and typology", Studies in Second Language Acquisition, 20, 189–211, 1998.

[11] Polivanov, E., "La perception des sons d'une langue étrangère", Travaux du Cercle Linguistique de Prague, 4, 79-96, 1932.

[12] Best, C. T., and Tyler, M., "Nonnative and second-language speech perception: Commonalities and complementaries," in M. Munro and O. Bohn [Ed], Second Language Speech Learning: The Role of Language Experience in Speech Perception and Production, 13–24, John Benjamins, Amsterdam, 2007.

[13] Escudero, P. and Vasiliev, P., "Cross-language acoustic similarity predicts perceptual assimilation of Canadian English and Canadian French vowels", Journal of the Acoustical Society of America, 130 (5), 2011.

[14] Calliope, "La parole et son traitement automatique, collection technique et scientifique des telecommunications", Paris, Milano, Barcelona, Mexico: Masson, 1989.

[15] Bradlow, A., R., "A perceptual comparison of the /i/-/e/ and /u/-/o/ contrasts in English and Spanish: Universal and language-specific aspects", Phonetica, 53, 55-85, 1996.

[16] Chladkova, K., Escudero, P., Boersma, P., "Context-specific acoustic differences between Peruvian and Iberian Spanish vowels", J. Acoust. Soc. Am. 130 (1), 416-428, 2011.

[17] Gendrot, C. and Adda-Decher, M., "Analyses formantiques automatiques de voyelles orales : évidence de la réduction vocalique en langues française et allemande", Proceedings of the Workshop MIDL, 7-12, 2004.

[18] Perkell, J. S., Guenther, F. H., Lane, H., Matthies, L. M., Stockmann, E., Tiede, M., and Zandipour, M., "The distinctness of speakers' productions of vowel contrasts is related to their discrimination of the contrasts", Journal of Acoustical Society of America, Volume 116, Issue 4, 2338-2344, 2004.

[19] Guenther, F., H., "A neural network model of speech acquisition and motor equivalent speech production", Biological Cybernetics, 72, 43-53, 1994.

[20] Bassetti, B., "Orthographic input and second language phonology", in Piske, T. and Young-Scholten, M. [Ed], Input Matters in SLA, 191-206, Clevedon, 2008.

[21] Baker, W., & Trofimovitch, P., "Perceptual paths to accurate production of L2 vowels: The role of individual differences", IRAL 44, 231–250, 2006.

[22] Boersma, P., and Weenink, D., "Praat: doing phonetics by computer (Version 4.3.02) [Computer program]", Retrieved from http://www.praat.org.html, 2005.

[23] R Core Team, "R: A language and environment for statistical computing", R Foundation for Statistical Computing, Vienna, Austria, Retrieved from http://www.R-project.org.html, 2012.

[24] Harrel, F., E., "Rms package", Retrieved from http://cran.r-project.org/web/packages/rms/index.html, 2013.

[25] Chladkova, K. & Escudero, P., "Comparing vowel perception and production in Spanish and Portuguese: European versus Latin American dialects". JASA Express Letter, 131(2), 2012.