Salento Italian listeners' perception of American English vowels

Bianca Sisinni[1], *Paola Escudero*[2] and *Mirko Grimaldi*[1]

[1] Centro di Ricerca Interdisciplinare sul Linguaggio, University of Salento, Lecce, Italy
[2] MARCS Institute, University of Western Sydney, Penrith, Australia

bianca.sisinni@unisalento.it, paola.escudero@uws.edu.au, mirko.grimaldi@unisalento.it

Abstract

The present study investigates Salento Italian (SI) listeners' initial state in the perception of American English (AE) vowels. Results of categorization and discrimination tasks are discussed in terms of the Perceptual Assimilation Model [1] and the Second Language Linguistic Perception model (L2LP; [2]). Further, the categorization results are compared to those of the Peruvian Spanish (PS) listeners in [3] to test the L2LP acoustic hypothesis, according to which the acoustic differences between the five SI and PS vowels will lead to different categorizations of AE vowels. Predictions of differential perceptual development across listener groups are provided.

Index Terms: Classroom-based second language acquisition, vowel identification and discrimination, Salento Italian versus Peruvian Spanish listeners.

1. Introduction

The present study aims at establishing the initial state for Salento Italian (SI) listeners' perception of American English (AE) vowels. The initial state of second language (L2) learning is comparable to that of naïve listeners [4], as described in the Second Language Linguistic Perception model (L2LP) [2] and in the extension of the Perceptual Assimilation Model to L2 acquisition (PAM-L2) [5]. These two models assume that the way naïve listeners assimilate L2 contrasts to native categories predicts how the contrasts will be discriminated and learned during L2 acquisition. If a contrast is assimilated to a single native category, it will be difficult to discriminate and learn, while if it is assimilated to two different native categories, it will be easy. These two perceptual models, however, differ in that the former states that L2 assimilation is determined by the acoustic properties of the native languages, while the latter states that L2 perception is driven by the specific articulatory gestures with which L1 and L2 sounds are produced.

Recent studies have shown that listeners' perception of both native and nonnative sounds match the acoustic properties of their native language [2,4,6] and that when languages or dialects have different productions of the same phonemes, those differences also emerge in nonnative perception [2,7,8] For instance, in [7] it was found that speakers of different varieties of Dutch, i.e., North Holland and East- and West Flemish which have different acoustic realizations of the Dutch vowel /ɛ/, differed in their categorization of the English contrast /ɛ-æ/. Similarly, in [8] it was observed that Peruvian (PS) and Iberian Spanish (IS) speakers discriminated the Dutch contrasts /a-ɑ/ and /i-ɪ/ with different degrees of accuracy, a distinction that is due to the difference in the acoustic properties with which the same PS and IS vowels are produced.

In line with these previous findings, we hypothesize that SI and PS listeners will have different initial states in their perception of AE vowels because although SI and PS have the same five vowel phonemes /i, e, a, o, u/, they are acoustically different. To demonstrate these differences, we recorded the vowel productions of 18 SI native female speakers, who produced SI vowels embedded in a carrier sentence and in the "p-vowel-t-ta" consonantal context. We compared these SI productions to those of the 9 female PS native speakers reported in [9], from which we selected a comparable set of words produced in a carrier sentence and in a similar consonantal context than the one used for the SI vowel productions. Figure 1 shows the average first (F1) and second formant (F2) values of SI and PS vowels.
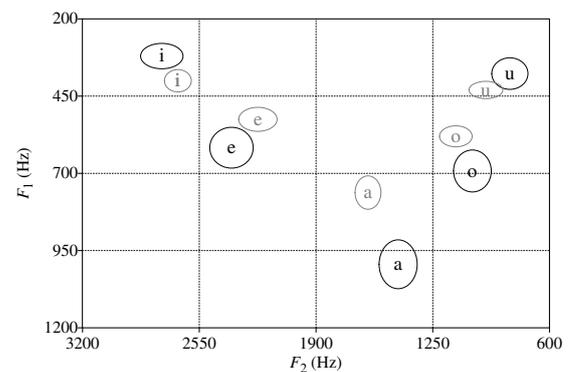

Figure 1. *Average F1 and F2 values of SI (black) and PS (gray) vowels produced by female speakers. Ellipses: one standard deviation from the mean.*

As can be observed in Figure 1, SI /i/ and /u/ have lower F1 values and SI /e/ and /o/ have higher F1 values than the PS counterparts. The F1 value of /a/ is much higher for SI than for PS. With respect to F2, SI vowels are less centralized than PS vowels, that is, SI /a, o, u/ have lower F2 values than the PS counterparts, while SI /i/ and /e/ have higher F2 values.

The present study aims at establishing the initial state in the perception of AE vowels by SI learners using both categorization and discrimination tasks. The results of the two vowel perception tasks are discussed in terms of the PAM and L2LP models. In the discussion section, we compare the SI vowel categorization results to those of female listeners in [3] and make predictions of differential L2 development for the two listener groups.

2. Method

2.1. Participants

Eighteen SI and 10 AE female listeners participated in the study. The SI participants were young monolinguals (mean

25 – 29 August 2013, Lyon, France

age = 20.4 years) from Salento, the southern part of the Apulia region. They were raised by monolingual Italian parents, had started studying English as a foreign language at school at the mean age of 10.3 years and had never been to a foreign country for longer than a month. They were first year students at the University of Salento. Although they studied English since they were children, they can be considered naïve listeners, as suggested by [5], since their L2 exposure was limited to a formal context with L1-accented teachers.

The 10 AE listeners were young listeners (mean age = 21.3) from different cities in the US. They performed the tasks to validate the stimuli and to help determine the SI listeners' level of difficulty with AE vowel contrasts.

2.2. Stimuli

The AE stimuli for both tasks were real words produced by three native female speakers of AE (mean age = 47 years), who were born and raised in the US. They were English lecturers at the University of Salento at the time of recordings and had lived in Italy for a mean of 22.6 years.

Words containing 8 AE vowels, i.e., /iː, ɪ, ɛ, æ, ʌ, ɑː, ʊ, uː/, were recorded. The shape of the words was /pVt/, where V was one of the AE vowels /iː, ɪ, ɛ, æ, ʌ, ɑː/, and /sVt/, where V was one of the vowels /iː, ʊ, uː/. Since /iː-uː/ and /uː-ʊ/ were part of the discrimination task as control and target contrasts, respectively, /iː, ʊ and uː/ needed to be recorded in the same context. Thus, the extra context /s-vowel-t/ was used for these three vowels because there is no English word with /uː/ in the /p-vowel-t/ context. Each word was embedded in the carrier phrase "I say__now" repeated six times. For each vowel, the best out of six tokens, judged by the first author, was edited and normalized in peak amplitude using Praat [10], for a total of 27 tokens (8 different vowels + /i/ produced in an extra context x 3 speakers).

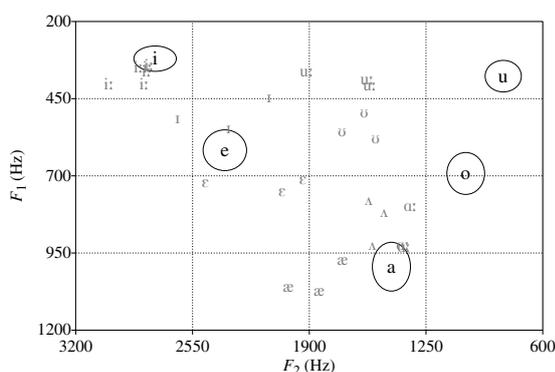Figure 2 shows the average F1 and F2 formants of the five SI vowels and the 27 AE stimuli.

Figure 2. *Average F1 and F2 values of SI (black) and AE (gray) vowels produced by female speakers. The ellipses represent one standard deviation from the mean.*

Segmentation and formant analysis were conducted using Praat [10]. The vowel start (voicing onset) and end (beginning of the subsequent consonant) were manually labeled and the F1 and F2 values were measured over a 25-ms Gaussian window placed around the vowel mid-point. All tokens were visually inspected to manually correct for software mistakes

when extracting formant values. The AE stimuli values compare well with those reported in previous literature [11, 12] except for /æ/, which has higher F1 and lower F2 values than the corresponding vowel in [12], and /uː/, which has higher F2 than the corresponding vowels in [12] and [11]. Based on the acoustic comparison of the SI and the AE vowels in Fig. 2, it is expected that AE /iː, ɪ, ɛ, æ-ɑː-ʌ, uː/ will be classified as native /i, i-e, e, a, u/, respectively. AE /ʊ/ is expected to be classified as SI /o/ and /e/.

2.3. Procedure

All tasks were administered in the listeners' native language by a SI native speaker. Participants were tested individually in a soundproof room listening to the stimuli at a comfortable volume level through headphones. In the categorization task, the 27 AE stimuli were randomly presented one at time. The 18 SI listeners had to identify the incoming AE stimulus with one of the five SI vowels presented on a computer screen, for a total of 1458 responses (18 SI participants X 27 stimuli X 3 speakers).

Subsequently, participants performed 7 discrimination tasks for the AE contrasts /iː-ɪ/, /ɛ-æ/, /æ-ɑː/, /æ-ʌ/, /ɑː-ʌ/, /uː-ʊ/ and /iː-uː/. For each contrast, 8 *change* and 8 *catch trials* were presented. In each trial, they heard three stimuli produced by a different AE native speaker and had to decide which of the three stimuli differed from the others by clicking on a computer screen the number corresponding to the odd item's position in the trial. They could also click on "none", if they thought all three stimuli were the same. Change and catch trials were randomly presented and the interval between the three stimuli in a trial was 300 ms. Despite this short interval, the inclusion of catch trials and the production of the three stimuli by different speakers ensured the categorical perception of the vowel tokens through the activation of a phonetic mode [13]. Importantly, the differences in discrimination accuracy between AE and SI listeners (see Table 2) confirm the language-specific and categorical perception of the AE vowels.

For both tasks, participants were told to guess if unsure and could replay a stimulus or a trial as much as needed. They were given 10 practice trials prior to each task. Responses to these practice trials were not included in the analysis. Participants took 40 minutes to perform both tasks, including a short break between tasks.

3. Results

3.1. SI listeners' categorization of AE vowels

Table 1 shows the mean percentage of times that each AE vowel (columns) was identified as each SI vowel (rows). It can be observed that AE /iː/ and /ɪ/ were both categorized as SI /i/ (100% and 99%), which was not predicted by the acoustic comparison shown in Figure 2. AE /ɛ/ was categorized as SI /e/ (99%), which follows the acoustic comparison, while AE /æ/ was heard as both SI /e/ (58%) and /a/ (42%), which does not follow for the acoustic comparison. AE /ɑː/ and /ʌ/ were heard mostly as SI /a/ (respectively 89% and 67%), and AE /ʌ/ was heard more often as SI /o/ (20%) than AE /ɑː/ (10%), which seems to follow from the acoustic comparison in Figure 2, since /ɑː/ is closer to SI /a/ than AE /ʌ/. AE /ʊ/ was classified more often as SI /u/ (83%) and much less as /o/ (15%), which seems to run contrary to the acoustic

comparison in Figure 2. Finally, AE /uː/ was always heard as SI /u/ (100%). Thus, these results demonstrate that the predictions based on the acoustic comparison of AE and SI vowels were confirmed for only four of the eight AE vowels.

Table 1. *Average vowel categorization of AE vowels as SI vowels. Only percentages above 5% are given.*

	iː	ɪ	ɛ	æ	ɑː	ʌ	ʊ	uː
i	100	99						
e			99	58				
a				42	89	67		
o					10	20	15	
u							83	100

3.2. SI and AE discrimination of AE vowels

A-prime (A') scores were computed on the proportion of "hits" (correct selections of the odd item in change trials) and "false alarms" (incorrect selection of odd items in catch trials), following the formula in [14]. Table 2 shows A' scores for SI and AE listeners on the 7 contrasts tested in discrimination task. An A' score of 1 indicates correct responses in all 16 trials, while a .5 score indicates chance level.

Table 2. *A' scores for the 7 AE vowel contrasts in SI (N = 18) and AE (N = 10) listeners. Standard deviations are in parenthesis.*

	ɑː-ʌ	iː-ɪ	ɛ-æ	æ-ɑː	uː-ʊ	æ-ʌ	iː-uː
SI	.42 (.23)	.64 (.21)	.67 (.23)	.72 (.25)	.77 (.22)	.85 (.15)	.95 (.04)
AE	.80 (.18)	.98 (.02)	.98 (.01)	.95 (.04)	.98 (.02)	.94 (.01)	.99 (.01)

A repeated measures ANOVA on A' scores with Contrast (7 levels) as the within-subject factor and Group (2 levels) as the between-subjects factor was carried out using a Greenhouse-Geisser correction when needed. This analysis yielded significant main effects of Contrast ($F[3.52, 91.60] = 14.97$, $p < .001$) and Group ($F[1, 26] = 52.109$, $p < .001$), as well as a Group * Contrast interaction ($F[6,156] = 4.044$, $p = .001$). To further examine the interaction, we ran nine independent samples t-tests, single-tailed since we expected higher scores for AE than SI listeners, and corrected for multiple comparisons ($\alpha = .05/7 = .00714$). AE listeners had significantly higher scores than SI listener for all 7 contrasts ($ps < .003$).

To test whether contrasts yielded significantly different scores in each group, we ran repeated measures ANOVAs for AE and SI listeners separately. For both groups, there was a main effect of contrast ($F[1.13, 10.19]= 9.66$, $p = .009$ and $F[3.48, 59.31]= 16.07$, $p < .001$, respectively). Bonferroni-corrected post-hoc tests showed no significant difference in accuracy across the different contrasts for AE listeners ($p > .05$), while there was a significant difference for SI listeners. Specifically, /ɑː-ʌ/ and /iː-uː/ yielded the lowest and highest accuracy, respectively ($p < .05$). SI listeners had significantly lower accuracy (p < .05) for /iː-ɪ/, /æ-ɑː/, /uː-ʊ/ and /ɛ-æ/ than for /iː-uː/, and for /ɛ-æ/ than for /æ-ʌ/. The following contrasts

did not have significantly different accuracy scores (all *ps* > .05): /iː-ɪ/ versus /ɑː-ʌ/, /iː-uː/ versus /æ-ʌ/ and /iː-ɪ/, and /æ-ɑː/ versus /uː-ʊ/ and /ɛ-æ/. Thus, SI listeners have high difficulty in discriminating AE /ɑː-ʌ/, moderate to low difficulty for AE /iː-ɪ/, /ɛ-æ/, /æ-ɑː/ and /uː-ʊ/ and, and hardly any difficulty for /æ-ʌ/ and /iː-uː/.

4. Discussion

The aim of the present study was to assess the initial state of AE vowels perception by SI listeners. AE categorization was predictable on the basis of the acoustic similarity between SI and AE for only four vowels, i.e., /iː, ɛ, ɑː, uː/. These findings seem to be in line with previous studies which show that acoustic properties do not always predict nonnative perception [15] and that nonnative categorization is not based solely on acoustic properties but rather also on the articulatory gestures that give rise to them [5]. However, it has been shown that AE vowels have considerable spectral change [16], which may need to be taken into account when predicting perceptual similarity on the basis of acoustic data, as was done in [6].

As predicted by the L2LP and the PAM models, the way SI listeners categorized the AE phonemes predicted in most cases the way they discriminated AE contrasts. Specifically, the AE contrast /iː-uː/ was categorized as two distinct native phonemes and was discriminated with the highest score (.95), while /iː-ɪ/ was categorized as a single native phoneme and received a low-to-moderate score (.64). /uː-ʊ/ was also categorized as a single native phoneme but, surprisingly, received moderate-to-high score (.77), which could be explained by the fact that AE /ʊ/ was partly categorized as /o/ (15%), suggesting that SI listeners are able to hear some differences between AE /uː/, which was always identified as SI /u/ (100%), and /ʊ/.

When the members of the AE contrasts were categorized as more than one SI phoneme, the discrimination varied from moderate to good. Moderate discrimination was obtained for /ɛ-æ/ and /æ-ɑː/, since the members of both contrasts were partially associated with the same SI vowels (AE /æ/ with SI /e-a/, AE /ɛ/ and /ɑː/ with, respectively, SI /e/ and /a/). Good discrimination was in turn obtained for /æ-ʌ/ (.85), probably because /ʌ/ was categorized as both SI /a/ and /o/, which suggests that SI listeners are able to hear some difference between AE /æ/ and /ʌ/. Finally, /ɑː-ʌ/ received the lowest score (.42), which seems to be explained by the very close acoustic proximity of these vowels, as shown in Figure 2.

To further verify the L2LP's acoustic hypothesis, which proposes that differences in native vowel production will lead to different L2 vowel perception, the SI categorization results were compared to those of the 20 *female* PS listeners in [3] and reported in Table 3. SI and PS share the same five vowel phonemes but, in line with the L2LP predictions, their different acoustic realizations and the way they compare differently to AE vowels (see Fig. 1 here and [3]'s Fig. 1) should lead to a differential perceptual assimilation of 5 of the 8 AE vowels, namely /ɪ, æ, ɑː, ʌ, ʊ/.

Despite the fact that SI and PS listeners were exposed to different stimuli, i.e., female natural vs. male synthetic stimuli, the AE vowels /iː, uː, ɛ/ were classified as the same native vowels, i.e., /i, u, e/, and with almost equal percentages (100% vs. 100%, 100% vs. 97%, 99% vs. 95, respectively). This suggests that differences in the categorization of AE vowels by

SI and PS listeners are likely to be due to how their acoustic properties compare to those of AE vowels.

Table 3. *Average categorization of AE vowels as PS vowels by the female listeners in [3]. Only percentages above 5% are given.*

	iː	ɪ	ɛ	æ	ɑː	ʌ	ʊ	uː
i	100	9						
e		91	95	99				
a			5		99	57		
o						42	74	
u							25	97

The groups indeed differed in the categorization of AE /æ, ɑː, ʌ/ since SI listeners categorized /æ/ as /a, e/ (58%, 42%) while PS listeners only as /e/ (99%), /ɑː/ mainly as /a/ (89%) and partly as /o/ (10%), while PS listeners only as /a/ (99%), and /ʌ/ more than /a/ (67% vs. 20%) and less as /o/ (57% vs. 42%). SI listeners categorized /ɪ/ as /i/ (99%), while PS listeners categorized it as /e/ (91%). Finally, SI listeners categorized /ʊ/ more as /u/ (83%) than PS listeners (25%), who categorized it mainly as /o/ (74%), which follows from the acoustic comparison, since PS /o/ is closer to AE /ʊ/.

The different categorization results for SI and PS listeners seem to confirm the L2LP hypothesis according to which dialectal differences in the production of the same vowel phonemes will lead to L2 perception differences [2, 17]. However, it cannot be ruled out that some differences can be due to other reasons. For instance, the difference in PS and SI categorization of AE /æ/ and /ɪ/ may have been due to the considerable spectral change in the natural AE stimuli of the present study, which was not included in the steady-state AE vowels used in [3]. Alternatively, one could think that PS and SI may differ in their use of vowel duration, resulting in different categorization of AE vowels that differ in that dimension. However, duration does not function as a cue to vowel identity in PS or SI, which suggests that neither group should use this cue in the cross-language task used in the present study, where foreign vowels are classified in terms of the listeners' native vowels.

Based on a differential categorization of AE vowels, it can be hypothesized that L2 development will be different for SI and PS listeners. For instance, to accommodate the two contrasts /iː-ɪ/ and /iː-uː/, SI listeners will need to split their native categories /i/ and /u/, while PS will have to shift the boundaries of /e/ and /o/. Conversely, PS listeners need to split their native /e/, while SI listeners need to create a category between /e/-/a/ for successfully perceive the contrasts /ɛ-æ/ and /æ-ɑː/. Both SI and PS need to create a new category between their native /a/ and /o/ to correctly perceive AE /ʌ/ in the contrasts /æ-ʌ/.

In sum, SI listeners' different degree of discrimination accuracy for AE contrasts can be predicted and explained on the basis of both the categorization results and the acoustic relationship between AE and SI phonemes, which is in line with both the PAM and the L2LP models. Further research should show if SI and PS listeners differentially categorize AE vowels in subsequent stages of L2 development.

5. References

[1] Best, C. T., "A direct realist view of cross-language speech perception", in W. Strange [Ed], Speech Perception and Language Experience: Issues in cross-language research, 171-204, Baltimore, MD, York press, 1995.

[2] Escudero, P., "Linguistic perception and second-language acquisition: explaining the attainment of optimal phonological categorization", PhD thesis, Utrecht University, The Netherlands, 2005.

[3] Escudero, P., and Chládková, K., "Spanish listeners' perception of American and Southern British English vowels", J. Acoust. Soc. Am., 128(5): EL254-EL260, 2010.

[4] Escudero, P., and Williams, D., "Perceptual assimilation of Dutch vowels by Peruvian Spanish listeners", J. Acoust. Soc. Am., 129(1): EL1-EL7, 2011.

[5] Best, C. T., and Tyler, M. D., "Non-native and second-language speech perception: Commonalities and complementarities", in O.-S. Bohn and M. Munro [Eds], Language experience in Second Language speech learning: in Honor of James Emil Flege, 13-34, John Benjamins, Amsterdam, 2007.

[6] Escudero, P., and Vasiliev, P., "Cross-language acoustic similarity predicts perceptual assimilation of Canadian English and Canadian French vowels", J. Acoust. Soc. Am., 130(5):, EL277-EL283, 2011.

[7] Escudero, P., Simon, E., and Mitterer, H., "The perception of English front vowels by North Holland and Flemish listeners: Acoustic similarity predicts and explains cross-linguistic and L2 perception", J. Phon., 40: 280-288, 2012.

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

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

[10] Boersma, P., and Weenink, D., "PRAAT: Doing phonetics by computer (Version 4.6.29)", [Computer program], retrieved October 2007. Online: http://www.praat-org/ (University of Amsterdam, Amsterdam) 1992-2010.

[11] Peterson, G.E., and Barney, H.L., "Control methods used in a study of the vowels", J. Acoust. Am. 24: 175-184, 1952.

[12] Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K., "Acoustic characteristics of American English vowels", J. Acoust. Soc. Am., 97: 3099-3111, 1995.

[13] Strange, W. and Shafer, V.L., "Speech perception in second language learners: the re-education of selective perception", in M. Zampini and J. Hansen [Eds], Phonology and Second Language Acquisition, 153-191, Cambridge, Cambridge University Press, 2008.

[14] Sundara, M., Polka, L., Genesee, F., "Language-experience facilitates discrimination of /d-ð/ in monolingual and bilingual acquisition of English", Cognition, 100: 369-388, 2006.

[15] Strange, W., Bohn, O.-S., Trent, S.A., and Nishi, K., "Acoustic and perceptual similarity of North German and American English vowels", J. Acoust. Soc. Am., 115: 1791-1807, 2004.

[16] Hillenbrand, J. M., and Nearey, T. M., "Identification of resynthesized /hVd/ utterances: Effects of formant contour", J. Acoust. Soc. Am., 105: 3509-3523, 1999.

[17] Escudero, P., and Boersma, P, "Bridging the gap between L2 speech perception research and phonological theory", Stud. Second. Lang. Acquis., 26: 551-585, 2004.