Category formation and the role of spectral quality in the perception and production of English front vowels

Ricardo A. H. Bion[1,3], Paola Escudero[2], Andréia S. Rauber[3], Barbara O. Baptista[3]

University of Joensuu[1], Finland; University of Amsterdam[2], The Netherlands &
Federal University of Santa Catarina[3], Brazil

ricardobion@gmail.com, paola.escudero@uva.nl,
asrauber@hotmail.com, b_baptista@terra.com.br

Abstract

This study aimed at comparing the perception and production of English front vowels by 17 proficient Brazilian speakers of English as a second language (L2) and 6 native speakers of American English. Towards this end, three experiments were carried out: (i) a production test measuring the first two formants of the participants' English front vowels, (ii) an oddity discrimination test investigating the formation of vowel categories, and (iii) a discrimination test with synthetic stimuli which assessed the participants' reliance on spectral quality when perceiving English vowels. The results of these experiments suggest a strong relationship between L2 vowel perception and production, since the vowel pairs which were produced with similar formant values by the Brazilian participants were also poorly discriminated in the two perception tests. In addition, the findings suggest that vowel perception might precede vowel production, as high rates on the discrimination of vowel pairs on both perception tests were a prerequisite for differentiating the same two vowels on the production test. Lastly, some Brazilian participants obtained native-like scores on the category formation test without manifesting native-like reliance on spectral quality, indicating that other acoustic cues, such as vowel duration, might be playing a role in their perception of English vowels.

Index Terms: vowel perception, vowel production, English

1. Introduction

The lack of ability of many L2 speakers to produce some vowels accurately is one of the causes of a noticeable foreign accent in their L2 [1], [2], [3]. In addition to this non-native production, L2 speakers also have difficulties in accurately perceiving particular vowels [4], [5], [6], a problem addressed in the literature as perceptual accent [7].

As discussed by Bion, L2 perception has been given considerable attention in recent years [8]. This tendency was initially triggered by perception models such as Flege's [9] Speech Learning Model (SLM) and Best's [10] Perceptual Assimilation Model (PAM), and has been recently revived by Escudero's [11] L2 Linguistic Perception (L2LP) model. In simple terms, these models propose that the ability to perceive nonnative sounds is partially determined by the way new sounds are related to the phonetic categories of the speaker's first language (L1).

The rationales for this common conclusion, however, differ. Flege's SLM sustains that L1 phonetic categories will limit the possibility of L2 category formation because L1 and L2 sounds coexist in a single phonological space, with the L2 sounds being 'filtered' through the learner's L1 sounds. During this process, a mechanism called *equivalence classification* allows the establishment of additional categories for "new" sounds, but not for "similar" sounds. Going against Flege's single phonology, Escudero [11] argues that learners have two separate perceptual grammars for their L1 and their L2, which allow them to adjust existing categories and create new ones by following the same mechanisms as in L1 acquisition.

As for the relationship between L2 perception and production, both Flege [9] and Escudero [11] directly state that many production errors have a perceptual basis. This claim is supported by several studies, which indicate that perception precedes production [1], [12], [13].

When discussing the limitations of previous studies comparing L2 perception and production, Flege emphasizes that most studies focus on a unique phonetic dimension (e.g., VOT), and may not be investigating other important perceptual cues [14]. To solve this problem, Flege suggests that tests investigating L2 category formation should accompany tests focusing on specific acoustic cues.

Thus, the present study compared the perception and production of English front vowels by native speakers of Brazilian Portuguese (BP) by using two perception tests: (i) an oddity discrimination test investigating the formation of vowel categories, and (ii) a discrimination test with synthetic stimuli in order to assess the participants' reliance on spectral quality when perceiving English vowels.

Both approaches for assessing vowel perception have already been adopted in separate studies which investigated the relationship between the perception and production of English vowels by BP speakers. For instance, Rauber et al. investigated vowel category formation with qualitative judgments of vowel production and an odd item perception test [12]. Likewise, Bion et al. investigated vowel production with acoustic measurements of formant values and vowel perception with a discrimination test using synthesized stimuli [15]. These studies found that the English vowel pairs /i/-/ɪ/ and /ɛ/-/æ/ are poorly perceived and produced even by highly proficient BP speakers of English. Expanding on these previous studies, the present study investigated the perception and production of these two vowel contrasts by combining the perception tests of [12] and [15].

Building on these studies, the present study aimed at answering the following research questions: (i) is the ability to distinguish between the members of the English vowel pairs /i/-/ɪ/ and /ɛ/-/æ/ related to their accurate production? If so, which ability is likely to emerge first?; (ii) is a native-like use of spectral quality when perceiving English vowels a prerequisite for reliable category perception?

The hypotheses to answer our questions were based on the discussions of [12] and [15]: (i) there is a relationship between L2 vowel perception and production in that perception precedes production, and (ii) native-like use of spectral quality when perceiving English vowels is not a prerequisite for good perception of new vowel categories, as L2 speakers might use different acoustic cues when perceiving English vowels.

2. Method

2.1. Participants

Seventeen English Language and Literature majors at a Brazilian university, 13 women and 4 men, 18 to 32 years old, participated in the study. The students were proficient English speakers who had American English as their target English variety, as judged in an interview held by a phonetician who is a native speaker of American English. The two perception tests were also given to six female native speakers of English who were 17 to 29 years of age and came from various regions of the United States.

2.2. Experiment 1: L2 vowel production

2.2.1. Procedure

The Brazilian learners of English produced sixteen tokens of each of the 4 American English front vowels (/i/, /ɪ/, /ɛ/, and /æ/), which were embedded in four real words contextualized in final position of sentences (e.g., The past tense of the verb bite is____.). All English target words were monosyllabic, started with one of the following consonants: [p, b, f, s, k, g, h], and ended in [t]. Written versions of the target words were avoided to minimize orthographic influence. Thus, participants were asked to (1) complete each sentence with the word which would best match the sentence semantically; (2) read it a second time with the same word; (3) say the carrier sentence *The last word is* _, completing this with the same word as in the meaningful sentence; and (4) repeat the carrier sentence. The L2 production data were recorded by using a Sony MZ-R70 minidisk with a directional microphone.

2.2.2. Analysis

The data from the minidisk were transferred to a computer and digitized at 10 kHz, 16-bit. The target words were segmented in Praat 4.2 and the vowel formants were measured in Speech Station II. In the second program, the target vowels were visualized in a wideband spectrogram in order to select a 25-millisecond window in the steady-state portion of the vowel before any diphthongal trajectory, in case there was one. Then, the first two formants of the vowels were measured from the LPC filter of the selected window.

Based on the F1 and F2 values of the individual vowels, the Euclidean distance in Hz between adjacent vowels was calculated, as exemplified in Figure 1.

2.3. Experiment 2: Category formation

A Categorial Discrimination Test (CDT), based on Flege et al., was designed to investigate the discrimination rate of the English vowel pairs /i/-/ɪ/ and /ɛ/-/æ/ [5]. The variant of the CDT used for this study was an oddity discrimination test where every trial contained an odd item, or all the three items had the same target vowel. In the former case, the set is called a *change trial* because there is one vowel that differs from the other two, while in the latter the set is called a *catch trial* because all the vowels are the same. In the change trials, the odd item varied in position; i.e., in one third of the trials it was the first item, and in the other two-thirds it was the second and the third respectively. This was done to avoid bias in the answers due to the order of presentation.

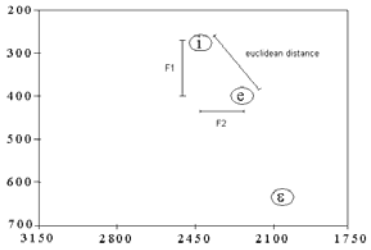

Figure 1 *The Euclidean distance between adjacent vowel pairs*

2.3.1. Stimuli

The CDT contained 32 trials of three items which consisted of eight change and eight catch trials for each of the two vowel contrasts. All sequences were recorded by five native speakers of English (two men and three women) from different U.S. states, i.e., Kentucky, Massachusetts, Michigan, New York and Pennsylvania. The three items of each trial were spoken by three different speakers and were chosen at random from the five native speakers. The sample was recorded in the CSL program, at 10 kHz, with 16-bit accuracy. Following Flege's CDT, the five native speakers produced words that were formed by the insertion of one of the English vowels into a /bVt/ frame [9]. Different from Flege, each word was produced at the end of the carrier sentence "This is a __", to facilitate speaker normalization. The target sentences were presented in a random order by means of the Praat 4.2 program. The interval between the three sentences in each trial was 1.3 seconds and the interval between trials was 2.8 seconds.

2.3.2. Procedure

The participants were given a sheet of paper with four alternatives and were asked to check alternative (a), (b) or (c) to indicate the odd item, or alternative (d) to indicate that all of the items were the same. Feedback was provided in a 5-set practice session before the experiment began.

2.4. Experiment 3: The role of spectral quality

2.4.1. Stimuli

Two edited-speech continua were created: one for the /i/-/ɪ/ and one for the /ɛ/-/æ/ contrast. A phonetician who is also a native speaker of English recorded her reading of the carrier sentences "I said ([bit/ bɪt/ bɛt/ bæt]) now" in a soundproof room in a professional audio studio. Then, using the *Analysis and Synthesis Laboratory* (ASL) from *Kay Elemetrics Corp.,* each glottal pulse of the recorded target vowels was manually marked, measured, and modified. The first two formants of the vowels were modified in continua with nine tokens equally spaced on the auditory-based *mel* scale, but maintaining the

formant frequencies of the original recorded vowels at each end of the continuum. Figure 2 shows the first two formants of the eighteen stimuli created.

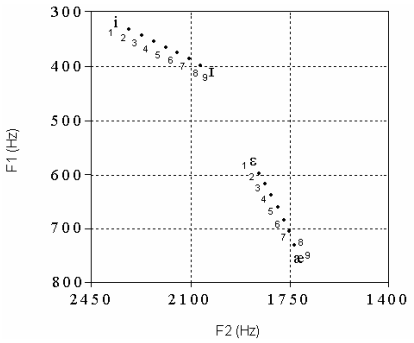

Figure 2 *First two formants of the two continua*

In order to guarantee that perception depended on spectral information alone, the duration of all nine vowels of each continuum was kept constant at a value which was calculated as the mean duration of the two original vowels. Each synthesized [bVt] token was paired with the original word containing [ɪ] or [æ], which are monophthongs that are not present in the Portuguese vowel inventory. The stimuli were recorded on a CD in random order and with an interval of 1.3s between the sentences of each pair and 2.8s between pairs. Each stimuli combination was repeated five times in order to guarantee greater reliability of the results.

2.4.2. Procedure

Participants listened to the stimuli in individual booths with earphones and were asked to decide whether each pair of stimuli was the same or different. A practice test with five pairs of sentences was provided before the main test.

2.4.3. Analysis

The distance necessary for the participants to perceive the vowel contrasts as different was measured based on the point in the continuum in which they started to label the vowel pairs as different in more than 60% of their presentations.

3. Results and discussion

3.1. Experiment 1: L1 and L2 production

The results of the present study are in line with previous studies in that even proficient Brazilian speakers of English had great difficulties in producing a difference between the vowels /ɛ/-/æ/ and /i/-/ɪ/. Our results show that 70% of the Brazilian participants had a Euclidian distance between the vowels /ɛ/ and /æ/ inferior to 60 Hz (mean = 48 Hz, SD = 29 Hz), which significantly contrasts with the mean Euclidian distance of 305 Hz produced by native speakers of English [16]. With respect to the high vowels /i/ and /ɪ/, we found that 70% of the Brazilian participants produced a Euclidian distance smaller than 100 Hz between /i/ and /ɪ/ (mean = 112 Hz, SD = 105 Hz), which was again significantly different from the 357 Hz produced by native speakers [16]. A paired samples t-tests comparing the Euclidian distance for the two contrasts reveals an effect of vowel contrast because the BP participants produced a significantly larger distance between /i/-/ɪ/ than between /ɛ/-/æ/ ($t = 2.607$, $p < .02$, two-tailed). This suggests that, for speakers of Brazilian Portuguese, producing a distinction between the members of the vowel pair /ɛ/-/æ/ is more difficult than between those of /i/-/ɪ/.

3.2. Experiment 2: Category formation

In this perception test, the native speakers of English obtained an average of 96% accurate responses for the /ɛ/-/æ/ vowel pair and an average of 98% accurate responses for the /i/-/ɪ/ pair. In contrast, the BP learners of English obtained an average discrimination rate of 50% for /ɛ/-/æ/, and of 81% for /i/-/ɪ/. It is important to mention that 8 out of the 17 BP participants had a native-like percentage correct (i.e., higher than 90%) categorical discrimination for the /i/-/ɪ/ contrast while none of them had a native-like performance on the /ɛ/-/æ/ contrast. This result suggests that, just like in production, there was an effect of vowel contrast in L2 perception. A paired sample t-test revealed that the difference in percentage of correct discrimination between vowel contrasts was highly significant ($t = 6.815$, $p < .0001$, two-tailed). It is worth mentioning that the same effect was not found in native speakers, who discriminated both contrasts equally well ($t = 1.581$, $p = .175$).

3.3. Experiment 3: The role of spectral quality

In Experiment 3, the native speakers of English started to distinguish both vowel pairs on the fifth member (out of 9) of the vowel continua, after which they recognized the stimuli as different in more than 60% of their presentations.

As for the BP learners, when only spectral information was available for their judgments, seven participants completely failed to discriminate between the vowels /ɛ/-/æ/ because they labeled them as the same vowel in most of their presentations, while four participants failed to distinguish the vowels /i/-/ɪ/. This difference between vowel contrasts is in line with the findings of Experiments 1 and 2 where the mastering of the /ɛ/-/æ/ contrast was shown to be the most difficult. The remaining participants managed to differentiate the two vowel pairs but needed a greater Euclidian distance between the vowels than did the native speakers of English.

3.4. Comparison between tests

Concerning the relationship between perception and production, it was observed that the vowel contrast /ɛ/-/æ/ is the most difficult contrast for BP speakers of English in both perception and production terms.

In order to answer the research question about which ability develops first, one can argue that perception precedes production, , as suggested by Escudero [11] and found in previous studies on L2 vowel acquisition [12], [15]. Considering the results of the production test, it is possible to state that most participants failed to differentiate both vowel contrasts. In contrast, the results of the perception tests demonstrate that some participants had obtained native-like perception, despite their low differentiation rates in production. Thus, it seems that native-like vowel perception is reached before the acquisition of a similar proficiency level in production.

Figure 3 compares the mean Euclidian distances (Hz) in production with the mean scores in the first perception test. This figure serves to illustrate how vowel perception seems to be a

prerequisite for accurate vowel production. A positive correlation was found between these scores (r = .5, p < .01).

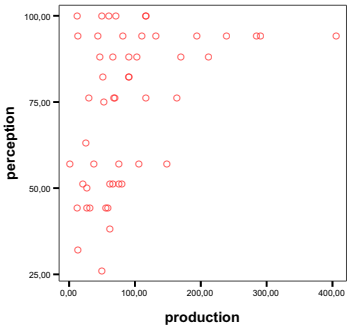

Figure 3 *Comparison of perception and production data.*

As can be observed, many participants obtained good mean discrimination results in the perception tests and only participants with high perception scores (above 75% of accuracy) produced a high Euclidean distance between their vowels; i.e., a difference higher than half of the distance produced by the native speakers. However, it is important to mention that high perception scores do not necessarily mean high differentiation values in production.

Finally, native-like perceptual weighting on spectral quality does not seem necessary for stable category formation, since many participants obtained high scores on the category formation test which involved the categorical discrimination of natural stimuli, but obtained poor scores on the perception test using synthetic stimuli which only varied in spectral properties. As found in many previous studies of L2 vowel perception [1, 4, 5, 11], participants might have relied on different acoustic cues (e.g., duration or diphthongal trajectory) in the natural stimuli to discriminate between pairs of vowels.

4. Conclusion

First, the present study demonstrates that perceiving and producing different L2 sound contrasts pose different degrees of difficulty. For BP speakers, our findings show that the English contrast /ɛ/-/æ/ is more difficult to master than /i/-/ɪ/.

Second, the findings give further evidence both for the hypothesis that L2 vowel perception and production are related, and for the hypothesis that perception precedes production in the acquisition of foreign vowels [11]. Evidence that perception and production are related comes from the fact that the most difficult vowel contrast in the production test was also the most difficult vowel contrast in the two perception tests administered as part of this study. Evidence for the fact that accurate vowel perception might be a prerequisite for accurate vowel production comes from the fact that only participants with high scores in both perception tests (above 75% of accuracy) were able to produce a Euclidian distance higher than 150 Hz between the vowel pairs investigated.

Finally, the fact that BP speakers tended to obtain better results in the perception test investigating category formation than in the test investigating the role of spectral quality raises points for further investigation. Importantly, it indicates that native and non-native speakers might rely on different acoustic cues when perceiving the same sound contrasts.

In addition, it suggests that it is time for a meta-analysis of different studies, focusing on how different perception and production tests assess these two abilities. This meta-analysis might indicate that the apparently conflicting findings of different studies might lie in the different uses of the broad terms 'speech perception and production.'

5. References

[1] Flege, J. E., Bohn, O.-S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics, 25*, 437-470.

[2] Flege, J. E., MacKay, I. R. A., & Meador, D. (1999). Native Italian speakers' perception and production of English vowels. *Journal of the Acoustical Society of America, 106* (5), 1-15.

[3] Munro, M. J. (1993). Production of English vowels by native speakers of Arabic: Acoustic measurements and accentedness ratings. *Language and Speech, 36*, 39-66.

[4] Escudero, P. (2001). The role of the input in the development of L1 and L2 sound contrasts: language-specific cue weighting for vowels, *Proceedings of the 25th Annual Boston University Conference on Language Development* (pp. 250-261).

[5] Flege, J. E., & MacKay, I. R. A. (2004). Perceiving vowels in a second language. *SSLA, 26*, 1-34.

[6] Morrison, G. S. (2002). Perception of English /ɪ/ and /i/ by Japanese and Spanish listeners: Longitudinal results. In *Proceedings NWLC 2002* (pp. 29-48).

[7] Strange, W. (Ed.). (1995). *Speech perception and linguistic experience*. Timonium, MD: York.

[8] Bion, R. (2004). O uso de estímulos sintéticos em testes de percepção de vogais de uma língua estrangeira. *Revista de Estudos da Linguagem, 12* (2), 51-64.

[9] Flege, J. E. (1995). Second language speech learning theory, findings and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233-277). Timonium, MD: York Press.

[10] Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues* (pp. 171-203). Baltimore, MD: York Press.

[11] Escudero, P. (2005). *Linguistic perception and L2 acquisition: Explaining the attainment of optimal phonological categorization*. LOT dissertation series 13, Utrecht University.

[12] Rauber, A. S., Escudero, P., Bion, R., & Baptista, B. O. (2005). The interrelation between the perception and production of English vowels by native speakers of Brazilian Portuguese. *Proceedings of INTERSPEECH 2005* (pp. 2913-2916).

[13] Rochet, B. L. (1995). Perception and production of second-language speech sounds by adults. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 379-411). Timonium, MD: York.

[14] Flege, J. E. (1999). The relation between L2 production and perception. In J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville & A. C. Bailey (Eds.), *Fifth International Congress of Phonetic Sciences* (Vol. 2, pp. 1273-1276). Berkeley, CA.

[15] Bion, R., Baptista, B. O, Escudero, P, & Rauber, A.S.. (2005). *The discrimination and production of English vowels by Brazilian learners.* 1st ASA Workshop on Second Language Speech Learning, Vancouver.

[16] Peterson, G.E. & Barney, H.L. (1952). Control methods used in a study of the vowels. *JASA, 24,* 175-184.