PERCEPTION OF ENGLISH VOWELS AS FIRST AND SECOND LANGUAGE

MARICHE GARCÍA BAYONAS
UNIVERSITY OF NORTH CAROLINA-GREENSBORO

Abstract

This study investigates the *perception* of four English *vowels*: /i/, /ɪ/, /ʊ/, and /u/ by native speakers (NS) and non-native speakers (NNS) of English –Spanish NS. Participants completed one perception task in which they were asked to select their ideal exemplar of the vowels in a method of adjustment (MOA) task. NS (n=17) and NNS (n=17) selected synthetic stimuli with different formant (F1 & F2) values. The results obtained from this study suggest that NNS do not perceive vowels /ɪ/ and /ʊ/ like NS do. However, they can perceive vowels /i/ and /u/ like NS do. For example, the F1 & F2 values of the stimuli selected for /ɪ/ and /ʊ/ by NNS differs from those selected by the NS more than the F1 and F2 values of the synthetic stimuli selected for /i/ and /u/. This finding reveals that learners of English perceive these two vowels /ɪ/ and /ʊ/ differently from the way NS do.

Keywords: *Speech perception; vowel perception; second language acquisition; phonetics; phonology*

1. Introduction

It is necessary to investigate how adults learn to perceive sounds in a second language (L2). These data can help explain the processes underlying L2 phonological acquisition. By investigating how learners perceive L2 sounds we will be able to better comprehend why learners can understand certain words better than others, identify the areas that may pose more difficulty in perception, and determine whether or not the sounds of the first language (L1) interfere in the process of L2 acquisition. By using different types of methodologies and techniques, research on crosslinguistic speech perception in recent years has assembled data that indicate that being able to perceive L2 sounds correctly prevents learners from misunderstandings, incorrect messages, and failures in communication.

A widely known model in L2 phonology is Flege's (1995) Speech Learning Model (SLM), which predicts patterns of cross-language speech perception and production. The model claims that "without accurate perceptual 'targets' to guide sensorimotor learning of sounds, production of the L2 sounds will be inaccurate" (Flege 1995: 238). That is, learners of an L2 may fail to perceive sounds accurately because the L2 sounds may be assimilated to L1 sounds. With regards to the different types of techniques and methods employed in data collection, there are studies that investigated adult L2 vowel perception using discrimination tasks (e.g. Flege, MacKay and Meador 1999; Baker, Trofimovich, Mack and Flege 2002), identification tasks (e.g. Mack 1989, Pallier, Bosch and Sebastián-Gallés 1997), repetition priming paradigm (e.g. Pallier, Colomé, and Sebastián-Gallés 2001), gating paradigm (e.g. Sebastián-Gallés and Soto-Faraco 1999), and the perceptual magnet paradigm (e.g. Kuhl and Iverson 1995; Bosch, Costa and Sebastián-Gallés 2000). A method that was key in the area of vowel perception research was the Method of Adjustment (MOA) task.

Johnson, Flemming and Wright (1993) examined how vowels are perceived by listeners using a MOA task. MOA is a tool that allows researchers to spectrally display the

Revista Electrónica de Lingüística Aplicada (ISSN 1885-9089)
2008, Número 7, páginas 79-89
Recibido: 23/06/2008
Aceptación comunicada: 27/10/2008

mental representation of vowels. Johnson, Flemming and Wright (1993) investigated perception and production of English vowels among NS (n=10). The subjects in the study were first asked to read ten repetitions of a list of English words. Then participants were asked to listen and select vowels from a chart containing 330 synthesized samples. They found that the participants in the perception study chose vowels that were systematically different from the ones they produced, "high vowels were higher, low vowels were lower, front vowels were farther front, and back vowels were farther back" (1993: 505). They concluded that perception and production of vowels are different phenomena even among NS. Johnson et al.'s (1993) MOA spearheaded a new line of research based on the selection of synthesized vowels, which helps determine the mental representation of these vowels.

Other researchers have subsequently utilized the MOA task with NNS like Frieda, Walley, Flege, and Sloane (2000). They utilized the same MOA tasks introduced by Johnson et al. (1993). Frieda et al. (2000) investigated the link between perception and production with respect to the English vowel /i/ by NS of English (n=35). Participants were recorded and then were prompted with some listening tasks. For perception part listeners were presented with a text box containing the same 330 sounds analyzed in the current study. Listeners worked on a computer and were told to locate the stimulus that best represented the vowel /i/. Frieda et al. (2000) found that for perception, listeners preferred /i/s with lower F1 values than their productions and higher F2 values than those of their production speech.

Bayonas (2007) pioneered the application of the MOA to L2 learners, employed it to investigate the perception of ten English vowels and five Spanish vowels. Participants in her study were NS of English and Spanish who were simultaneously advanced learners of Spanish and English. They thus acted both as NS and NNS. Her results yielded vowel-specific discrepancies in the perception of L2 vowels. Put differently, NS and NNS perceived some vowels like NS did (/i/, /u/), whereas all the other vowels analyzed were perceived by NNS differently (located spectrally apart from the results of NS) from NS.

Other types of tasks employed in L2 vowel perception are discrimination and identification, which have been applied to the study of naturally produced speech and to synthetic stimuli that presented learners with pairs of vowels as part of a continuum (i.e. Fox, Flege and Munro 1995; Fox et al. 1995). Among the research done there are several languages whose vowel inventories have been compared (i.e. Bohn 1995; Bohn and Flege,1997; McAllister, Flege and Piske 2002, inter alia)

A more detailed study was needed in order to show how NS and NNS perceive vowels in their L1 or L2 using an MOA task. While other studies (Johnson et al. 1993; Frieda et al. 2000; Bayonas 2007) have previously showed the results of mean values of English and Spanish vowels, an in-depth study on the perception or production of a single L2 vowel, English /i/, and the perception and production of English vowels as L1, the current investigation adds the analyses of three English vowels, /ɪ/, /ʊ/, and /u/, to the research undertaken by Frieda et al. (2000) on the English vowel /i/. The MOA is thus employed in the current study to investigate perception of vowels by both NS and NNS.

2. The Present Study

This research examined the perception of four English vowels: /i/, /ɪ/, /ʊ/, and /u/ by English NS and NNS (NS of Spanish). Spanish has five vowels, /a/, /e/, /i/, /o/, /u/, whereas English has a larger vowel repertoire (ten or eleven by most counts). Therefore, (Spanish-speaking) learners of English have to learn to perceive and produce at least double the number of vowels that exist in Spanish. How do these learners compensate for such a large difference in vowel repertoire? Bayonas (2007) showed that NS and NNS exhibited similar results with regard to vowels /i/ and /u/ however, that was not the case for /ɪ/ and /ʊ/. A more detailed investigation that analyzed the particular case of these four vowels was therefore needed. This study thus, aimed at the differences in selection of synthetic vowels in a MOA task by NS of English and by NS of Spanish learning English. The research question addressed was: are there spectral differences in the selection of /i/, /ɪ/, /ʊ/, and /u/ by NS and NNS?

3. Method

3.1. Participants

The NS of English (n=17 randomly selected from a total of 54 participants) were all students at a Mid-West university. Their ages ranged from 18-34 with a mean of 20 years old. There were six males and eleven females. The NS of Spanish learning English (n= 17) were students at a university in Southern Spain, and graduate students at the aforementioned Mid-West university. Their ages ranged from 23-59 with a mean of 30.3. There were five males and twelve females. No participants reported speech or hearing impediments. Participants that reported being Spanish/English bilingual (having learned both languages from birth) or having learned the L2 in childhood were eliminated from the study. Age was not considered a variable in this particular study. All participants had been studying the L2 from the age of 14 or up. They were all advanced learners of Spanish/English. All participants were phonetically untrained, and volunteered to participate.

3.2. Stimuli

The program "Praat" was used in order to produce 330 steady-state vowel stimuli with varying first and second formants. The first formant (F1) had fifteen possible values, ranging from 200 Hz to 900 Hz, whereas the second formant (F2) had twenty-two possible values, ranging from 800 Hz to 2800 Hz. This set of stimuli mirrors the ones employed by Johnson, Flemming, and Wright (1993) using a different program. Each synthesized vowel had its own wave file which was later hyper-linked to a number on a web-based text table from which participants were to select vowels for the MOA.

3.3. MOA Task

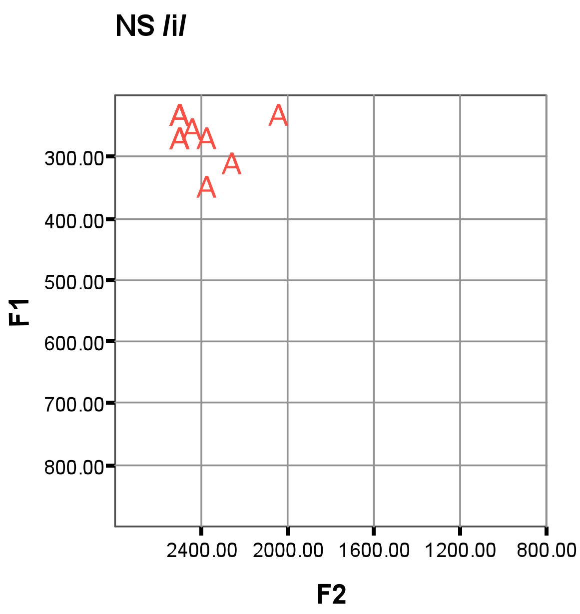
The MOA task was first created and employed by Johnson, Flemming, and Wright (1993). This study was innovative in the incorporation of a larger set of synthesized vowels (330). The synthesized vowels were assigned a wave file and these wave files were each assigned to a number on a chart which was presented to the participants in table form. Clicking on any of the 330 numbers on the chart caused a synthesized vowel sound to play. Sounds placed on higher rows corresponded to high vowels and the sounds on the lower rows corresponded to lower sounds. Sounds on the left corresponded to front vowels and the ones on the right corresponded to back vowels.

3.4. Procedure allotted

Participants were presented with a chart with 330 synthesized vowel sounds on a computer. They were allowed five minutes at the beginning of the test to familiarize themselves with the chart and practice prior to the task. The task contained a set of monosyllabic words which contained the same English phoneme. For example: "heat, see, meat, scene, feed." There were six sets of monosyllabic words, one for each English vowel. Participants were given the following directions to complete the task: "Type in the number of the sound -ONLY ONE-that best depicts the vowel in these groups of words". Participants were expected to select a "prototype" for each vowel sound. There was no time limit for this task. Once they thought they had found the right sound, they had to type the number which played the sound in a box next to the five English words.

4. Results and discussion

The following figures illustrate the answers given by NS (circles) and NNS (squares). Figures 1-4 show results by language group and figure 5 displays the overall mean average results of all participants. Each figure depicts the F1 and F2 values of the synthetic vowels selected as the exemplar. Similar individual results are collapsed in the same circle or square.



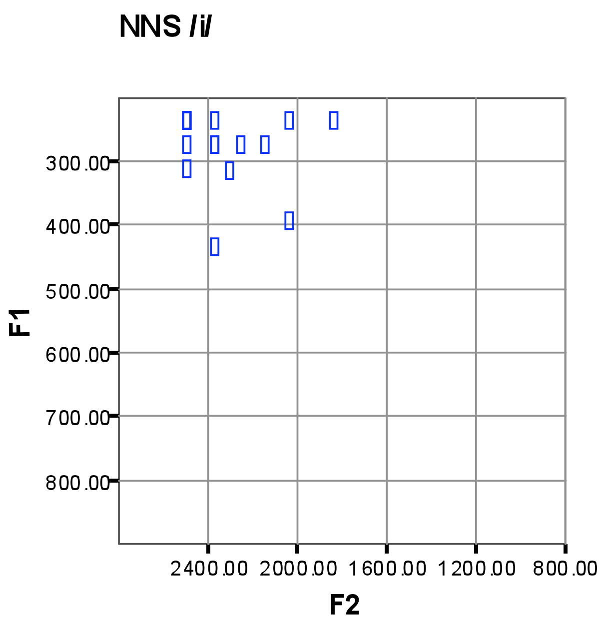


Figure 1. Results of English vowel /i/ as selected by NS and NNS

Figure 1 shows that NS cover a smaller area in the selection of synthetic vowels that resemble English /i/. NNS have a wider area of acceptance for /i/ than NS.

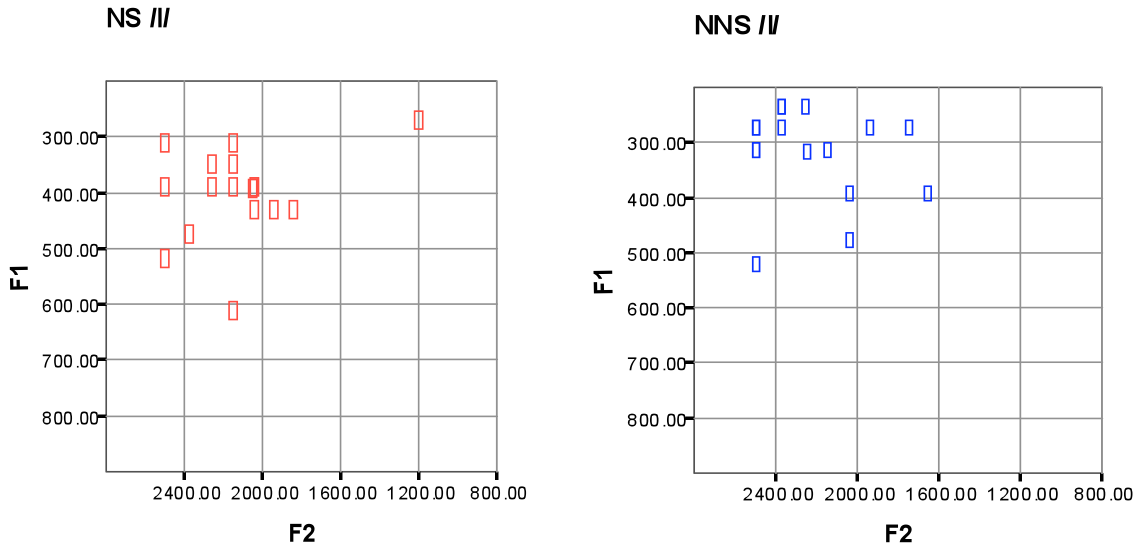


Figure 2. Results of English /ɪ/ NS and NNS

Figure 2 shows a clear discrepancy between NS and NNS. NS selected synthetic vowels with higher F1 than NNS. For NS, there are still many synthetic vowels selected for English /ɪ/ that are identical to the ones selected for English /i/. In that regard, NNS's results differ from those of NS. For example, we can observe the squares in the upper three quadrants in the right-hand side chart of the NNS which means that they selected vowels with higher F1 than the vowels selected by NS. Those three quadrants are empty on the left part of figure 2 that corresponds to NS results. However, some of the results selected by NNS are identical to those of NS.

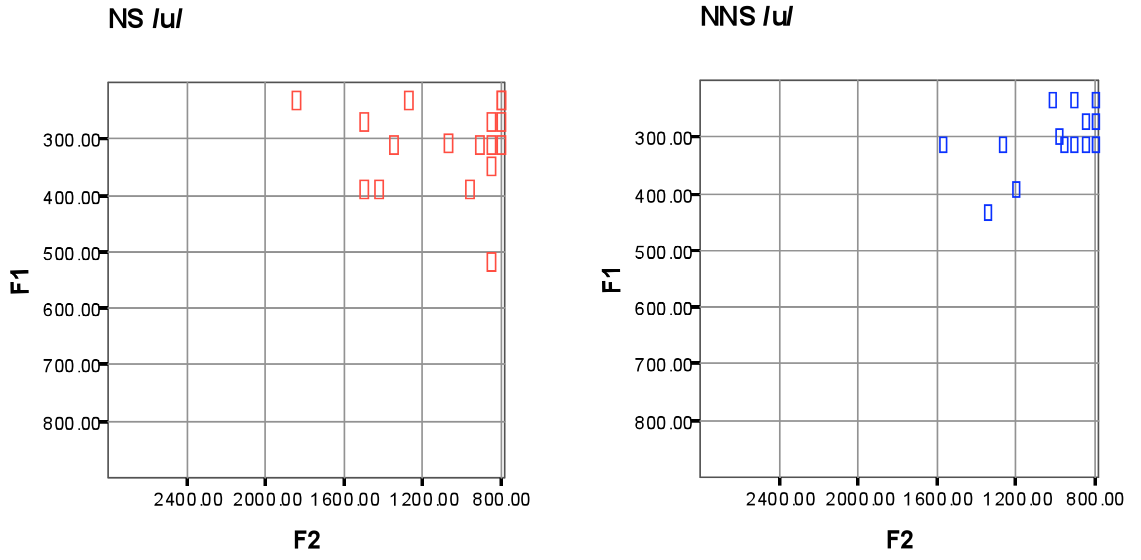
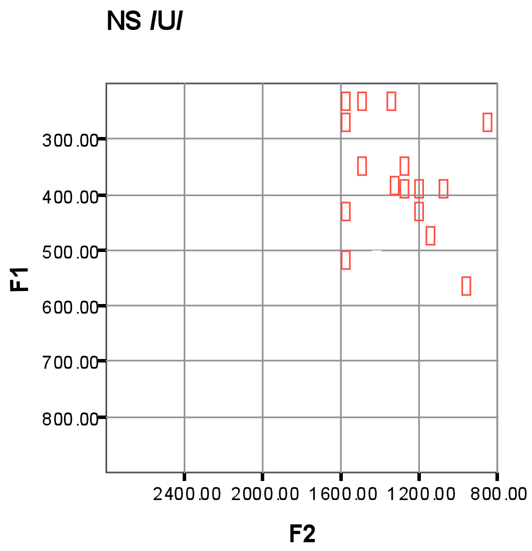


Figure 3. Results of English /u/ NS and NNS

Figure 3 illustrates that the majority of NS and NNS selected vowels with low F2 values. Accordingly, circles and squares cluster in the upper right corner of both charts. There is, however, more disagreement among NS as to the selection of the exemplar /u/ synthetic vowel.



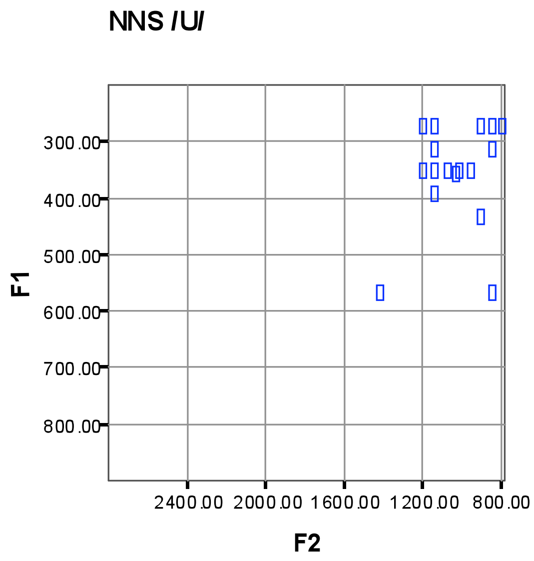


Figure 4. Results of English /ʊ/ NS and NNS

The results represented in figure 4 indicate that overall, NNS and NS selected synthetic vowels with different F1 and F2 values. NS selected vowels within a wider area compared to NNS, and NNS have a preference for synthetic vowels with F1 values not as high as NS and lower F2 compared to NS. When comparing these results to those in figure 3, it can be observed that NNS selected synthetic vowels as exemplars with similar values for /u/ and /ʊ/.

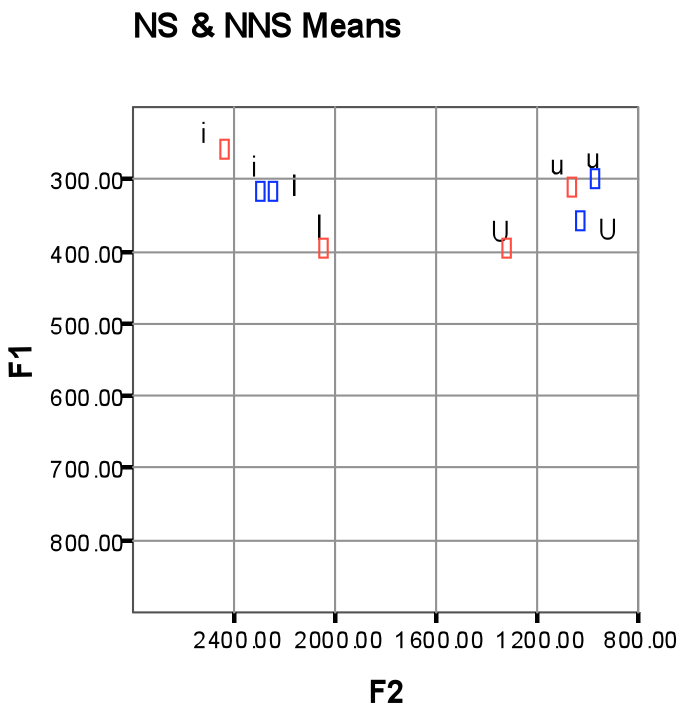


Figure 5. Mean results for all English Vowels by NS-circles and NNS-squares

Figure 5 shows circles (NS) and squares (NNS) that represent the average results of all NS and NNS. As it can be observed, there are obvious differences in vowel perception. For example, /i/ and /u/ have somewhat similar values for both NS and NNS, considering the variability among the groups themselves. However, the values of /ɪ/ and /ʊ/ are for the NNS much closer to their /i/ and /u/ counterpart, than they are for the NS. NNS still do not resemble NS with the selection of values of /ɪ/ and /ʊ/. NNS selected the following F1 and F2 mean values for those vowels: /i/ (F1: 330 Hz, F2: 2305 Hz); /ɪ/ (F1: 331 Hz, F2: 2248Hz); /u/ (F1: 313 Hz, F2: 980 Hz); and /ʊ/ (F1: 371 Hz, F2: 1037Hz). The values selected for /i/ and /ɪ/ are strikingly closer than the ones selected for /u/and /ʊ/.

With regard to the theoretical implications of the current study, part of the SLM is met. Flege's (1995) model establishes that: "Category formation for an L2 sound may be blocked by the mechanism of equivalence classification. When this happens, a single phonetic category will be used to process perceptually linked L1 and L2 sounds (diaphones)." (Flege 1995: 239). In other words, two vowels may be perceived as belonging to one single phonetic category.

In Figure 5, NNS selected formant values for English vowels which cluster around the two pairs of vowels. These pairs of vowels would represent a single vowel category in Spanish. This finding suggests that NNS perceive two L2 vowels based on one category, thus supporting Flege's SLM. This model predicts that, for example, NNS will perceive English /i/ & /ɪ/ based on the formant values of only one vowel: /i/. In this MOA task we find that in Figure 5 by NNS the pairs of vowels: /i/ & /ɪ/; and /u/ & /ʊ/ are closer together than the means obtained by NS. These pairs of vowels selected by NNS seem to be clustered around their vowel counterparts /i u/. It seems as if /i u/ have perceptually linked the pairs of English vowels to one another. One caveat to this remark is that these pairs of vowels are perceived as closer together by NNS than they are by NS, yet they are not identical, which indicates that NNS distinguish between the two, perhaps not at the level that NS do, but there is an indication of learning. The results obtained in this study suggest that there are, indeed, spectral differences in the selection of /i/, /ɪ/, /ʊ/, and /u/ by NS and NNS. As can be seen on Table 1, for vowel /i/ NS selected synthetic vowels with F1 of 250-366 Hz with a mean of 272.5 Hz and F2 values of 2042-2500 Hz with a mean of 2444 Hz, whereas the NNS selected synthetic vowels with F1 of 250-900 Hz with a mean of 330 Hz and F2 values of 1843-2500 Hz with a mean of 2305 Hz. For vowel /ɪ/NS selected F1 values of 288-630 Hz with a mean of 412 Hz and F2 values of 800-2500 Hz with a mean of 2053 Hz, whereas NNS selected F1 values of 250-536 Hz with a mean of 331 Hz and F2 of 1661-2500 Hz with a mean of 2248 Hz. For vowel /u/ NS selected synthetic vowels with F1 values of 250-536 Hz with a mean of 326 and F2 values of 800-1843 Hz with a mean of 1067 Hz, whereas NNS selected F1 values of 250-407 Hz with a mean of 313 Hz and F2 values of 800-1576 Hz with a mean of 980 Hz. Finally, for vowel /ʊ/ NS selected synthetic vowels with F1 of 250-583 Hz with a mean of 400 Hz and F2 values of 851-1576 Hz with a mean of 1325 Hz, whereas NNS selected F1 values of 288-583 Hz with a mean of 371 Hz and F2 of 851-1418 Hz with a mean of 1037 Hz.

	/i/		/ɪ/		/u/		/ʊ/	
	F1	F2	F1	F2	F1	F2	F1	F2
NS Mean	250-366 272.5	2042-2500 2444	288-630 412	800-2500 2053	250-536 326	800-1843 1067	250-583 400	851-1576 1325
NNS Mean	250-900 330	1843-2500 2305	250-536 331	1661-2500 2248	250-407 313	800-1576 980	288-583 371	851-1418 1037

Table 1. NS and NNS selection of synthetic vowel F1 and F2 values.

5. Conclusion

Vowel perception is part of the learning process in foreign language acquisition. Difficulties in L2 perception may lead to difficulty in word recognition (Bradlow and Pisoni 1999), and it has been hypothesized that incorrect perception of L2 sounds may lead to incorrect production of these sounds (Rochet 1995). This study aimed to provide a deeper understanding of how vowels are perceived by L2 listeners as compared to L1 listeners. In view of the results, two trends can be observed: first, this study partly supports Flege's (1995) SLM. Overall, NNS are not able to establish phonetic categories for new or similar L2 vowel sounds. Second, Spanish NS learning English identify English vowels in a manner different from NS in this study.

6. Further Research

One of the limitations in this study is that this study included only advanced learners. A future study should include learners at different proficiency levels in order to explore different possible stages in the perception of L2 vowels to determine whether there are tendencies or patterns that can be predicted based on the learner's (proficiency) level.

Second, this study examined only the perception of vowels. A future study could also include participants' production of those same vowels, which would be measured in terms of their frequencies and then compared to the ones perceived by the same participants. This future study would replicate the one by Johnson, Wright and Fleming (1993) in most aspects and would have NNS in addition to the NS's perception and production.

References

Baker, W., P. Trofimovich, M. Mack, and J. Flege. 2002. The effect of perceived phonetic similarity on non-native sound learning by children and adults. In B. Skarabela, S. Fish and A. Do (eds.), *Proceedings of the 26th Annual Boston University Conference on Language Development*. 36-47. Sommerville, MA: Cascadilla Press.

Bayonas, M. G. 2007. *The Acquisition of Vowels in Spanish and English as Second Language*. Munich: Lincom Studies in Language Acquisition.

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

Bohn, O. S. and J. Flege. 1997. Perception and production of a new vowel category by adult second language learners. In A. James and J. Leather (eds.), *Second Language Speech: Structure and Process*. 53-73. Berlin, Germany: Mouton de Gruyter.

Bosch, L., A. Costa, and N. Sebastián-Gallés. 2000. First and second language vowel perception in early bilinguals. *European Journal of Cognitive Psychology* 12: 189-221.

Bradlow, A. R. and D. Pisoni. 1999. Recognition of spoken words by native and non-native listeners: Talker-, listener- and item-related factors. *Journal of the Acoustical Society of America* 106: 2074-2085.

Flege, J. E. 1995. Second language speech learning: Theory, findings, and problems. In W. Strange (ed.), *Speech Perception and Linguistic Experience*. 233-277. Timonium, MD: York Press.

Flege, J. E., I. R. A. MacKay, and D. Meador. 1999. Native Italian speakers' perception and production of English vowels. *Journal of the Acoustical Society of America* 106: 2973-2987.

Fox, R. A., J. E. Flege, and M. J. Munro. 1995. The perception of English and Spanish vowels by native English and Spanish listeners: A multidimensional scaling analysis. *Journal of the Acoustical Society of America* 97: 2540-2551.

Frieda, E. M., A. C. Walley, J. E. Flege and M. E. Sloane. 2000. Adults' perception and production of the English vowel /i/. *Journal of Speech, Language and Hearing Research* 43: 129-143.

Johnson, K., E. Flemming, and R. Wright. 1993. The hyperspace effect: Phonetic targets are hyperarticulated. *Language and Speech* 69: 505-528.

Kuhl, P. K. and P. Iverson. 1995. Linguistic experience and the "perceptual magnet effect". In W. Strange (ed.), *Speech Perception and Linguistic Experience*. 121-154. Timonium, MD: York Press.

Mack, M. 1989. Consonant and vowel perception and production: Early English-French and French-English bilinguals. *Perception & Psychophysics* 46: 187-200.

McAllister, R., J. E. Flege, and T. Piske. 2002. The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian. *Journal of Phonetics* 30: 229-258.

Pallier, C., Bosch, L. and N. Sebastián-Gallés. 1997. A limit on behavioral plasticity in speech perception. *Cognition* 64: B9-B17.

Pallier, C., A. Colomé, and N. Sebastián-Gallés. 2001. The influence of native-language phonology on lexical access: Exemplar-based versus abstract lexical entries. *Psychological Science* 12: 445-449.

Rochet, B. 1995. Perception and production of second-language speech sounds by adults. In W. Strange (ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Speech Research*. 379-410. Timonium, MD: York Press.

Sebastián-Gallés, N. and S. Soto-Faraco. 1999. On-line processing of native and non-native phonemic contrasts in early bilinguals. *Cognition* 72: 111-123.

Revista Electrónica de Lingüística Aplicada (ISSN 1885-9089)
2008, Número 7, páginas 79-89
Recibido: 23/06/2008
Aceptación comunicada: 27/10/2008