



# Validating a second language perception model for classroom context. A longitudinal study within the Perceptual Assimilation Model

Bianca Sisinni, Mirko Grimaldi

Interdisciplinary Speech and Language Research Centre (CRIL), University of Salento, Lecce, Italy  
bianca.sisinni@unisalento.it, mirko.grimaldi@unisalento.it

## Abstract

The present study verified whether adult listeners retain the ability to improve non-native speech perception and if it can be significantly enhanced in the formal context, a very impoverished context with respect to the natural one. We tested (i) whether perceptual learning is possible for adults in a classroom context during focused phonetic lessons, and (ii) whether it follows the pattern predicted for natural acquisition by the PAM-L2 [1]. The results showed that adult listeners are still able to improve foreign sound perception and this ability seems to occur also in formal contexts in line with the PAM-L2 predictions.

**Index Terms:** non-native phone perception, foreign language acquisition, PAM.

## 1. Introduction

The dominant theories in second language (L2) phoneme acquisition argue that perceptual similarity/dissimilarity between the sounds of L2 and native language (L1) governs their assimilation or non-assimilation and dictates their learnability in adulthood. According to the Perceptual Assimilation Model (PAM) [2], subjects with limited L2 instruction, especially that typified by classroom-only education with instructors with a strong L1 accent, can be considered as naïve listeners, i.e., functional monolinguals not actively learning or using an L2.

The PAM assumes that the way naïve listeners assimilate non-native phones to native phonemes is determined by the detection of commonalities between them [3]. The PAM predicts that the non-native phones can be Categorized or not consistently categorised (i.e., Uncategorised) as exemplars of native phonemes, falling between two or more L1 phonemes. Finally, non-native phones cannot be categorised at all as speech sounds. If two Categorized non-native phones are perceived as acceptable exemplars of two different native phonemes, a very good/excellent discrimination is predicted (Two Category assimilation, TC). Conversely, poor discrimination is expected for Single Category (SC) assimilation, where two non-native phones are perceived as equally good or poor exemplars of a single native phoneme. If two non-native phones are both perceived as a single native phoneme but differ in rating, intermediate degree of discrimination is predicted (Category Goodness assimilation; CG). For the Uncategorised phones, if one non-native phone is perceived as a native phoneme and the other is perceived as an uncategorised speech sound, the predicted degree of discrimination is good (Uncategorised-Categorised assimilation, UC). Two non-native phones assimilated to partially-similar native phonemes will be discriminated from poor to moderate (Uncategorised-Uncategorised assimilation; UU). Finally, the predicted discrimination is good/excellent for the Non assimilable typology (NA) since two non-native phones are not perceived as any speech sound and are easily

distinguishable from each other. A recent extension of this model, i.e., the PAM-L2 [1], refers to adult L2 learners in an L2 immersion context, for which a common L1-L2 system (i.e., an interlanguage) is developing. The L2 perceptual learning seems to be determined by the increase in L2 vocabulary size that causes the learners to *re-phonologise* [4]. This re-phonologisation is stronger after 6-12 months than it is during the first 6 [1] and does not necessarily imply the formation of new set of L2 categories but rather that learners stretch their L1 inventory, perhaps forming new L2 categories [4]. This process may be particularly strong when several L2 phones are assimilated into the same L1 category [4]. Perceptual learning is not predicted for all the non-native phones but only for the ones perceived to be phonetically, even if not phonologically, different from L1 phonemes and for those phones which are not phonetically nor phonologically assimilated to a L1 phoneme/s, but always falling into the native phonological space of the L2 listeners. PAM and PAM-L2 focus on natural communicative situations (SLA). Both important differences and striking similarities exist between immigrant L2 learners in a foreign country and students in a foreign language classroom [5].

To our knowledge, fewer studies have focused on learning an L2 in classroom context (FLA), most of them focusing on production or perception of foreign sounds ([6], [7]). In the present research, perceptual learning of American English (AE) vowels was studied in FLA and L2 focused phonetic lessons centred on multimodal methodology was taken into account [8]. Subjects' discrimination accuracy was investigated for various non-native contrasts falling in different PAM typologies for which different degrees of discrimination are predicted. According to the PAM-L2, perceptual learning was expected for some contrasts only.

A group of Salento Italian (SI) undergraduate students participated. Crucially, SI, the variety of Standard Italian spoken in Southern Apulia, presents only five stressed vowels: /i/, /e/, /a/, /o/, /u/ [9], contrary to the ten monophthongal AE vowels: /i:/, /ɪ/, /e/, /æ/, /ʌ/, /ɑ:/, /ɜ:/, /ɔ:/, /ʊ/, /u:/ [10]. This difference may have implications for L2 vowel learning in the PAM perspective. Actually, it is more likely that individuals with a small L1 vowel system assimilate multiple L2 vowels into the same L1 category [11].

## 2. Method

### 2.1. Participants and phonetic lessons procedure

Nine female native speakers of SI (mean age: 19.1 years), attending the first year at the Faculty of Foreign Languages and Literatures, participated in the study. In a background questionnaire, they reported that they had started studying English at the mean age of 11 years. Five subjects stated they had had lessons with a native speaker of English during secondary school and reported that they had never been to a

foreign country for a period longer than a month. In addition to the university lessons, the subjects received further focused phonetic lessons for 7 weeks (3 hours per week). Before starting the phonetic lessons, they performed two perceptual tests in a pre-test session (November 2007). Post-test sessions, which were the same as the pre-test one, were executed immediately after (January 2008) and after three months (i.e., 2<sup>nd</sup> post-test, April 2008) by the end of the phonetic lessons in a longitudinal manner. A teacher who was an expert phonetician and a bilingual speaker of AE and SI gave explicit instructions about the articulation of AE vowel phonemes. She taught learners the IPA in order to show the differences between the phonological inventories of the L1 (SI) and the L2 (AE) and the difficulties presented by AE. The AE stimuli consisted of real words, drills, tongue twisters and real sentences containing the target phonemes [8]. Subjects were asked to pay attention to the stimuli and to repeat them individually in order to be corrected and thus to avoid the fossilization of erroneous forms. Classroom instruction was thought to be more natural and less controlled than the training conducted in strict laboratory settings, e.g., highly simplified consonant-vowel sequences, modified speech cues, use of feedback, and relatively short periods of training (see the review in [12]).

## 2.2. Stimuli and procedure

Three adult female native speakers of AE were recorded. They produced consonant-vowel-consonant (CVC) AE real words. More specifically, the vowels /i:/, /ɪ/, /e/, /æ/, /ʌ/, /ɑ:/, /ɜ:/, /ɔ:/, /ʊ/, /u:/ were inserted in the /p\_t/ context, and the vowels /i:/, /u/ and /u:/ were inserted in the /s\_t/ context. This was done because, to our knowledge, the phoneme /u:/ does not exist in the /p\_t/ context and another real context was needed to compare this phoneme with the phonemes /i:/ and /ʊ/. The speakers were recorded in a soundproof room by means of CSL 4500 (sampling rate 22.05 kHz). They read a career phrase “I say CVC now” for each of the AE phoneme and the target words were edited and normalised in peak intensity by means of Praat 4.6.29. A total of 36 stimuli were thus derived (12 vowel phonemes x 3 speakers) and were used in two perceptual tests, i.e., the identification and the oddity discrimination tests. For both perceptual tests, the learners were individually tested in a soundproof room via headphones at a comfortable level in the pre-test, the post-test and the 2<sup>nd</sup> post-test. Before each test, they were given a familiarization training test in the presence of the experimenter, the results of which were not analysed.

## 2.3. Experiment 1

The aim of the identification test is to measure the *perceived phonetic distance* [13] between L1 and L2 phonemes, to understand how L1 native listeners categorize non-native phonemes in terms of their native phonological categories. Moreover, this experiment allows one to detect L2 phoneme pairs which can be used in discrimination tests (Experiment 2). The 36 AE stimuli were randomly presented three times each, for a total of 108 stimuli per subject. Subjects were asked to pay attention to the vowels only and to identify them in terms of Italian phonemes by clicking on the computer screen where they were displayed. In order to finally obtain a goodness of fit [14], they had to rate for degree of similarity the non-native phonemes in terms of the native ones, using a scale ranging from 1 (totally different) to 5 (totally similar).

By multiplying the percentage of identification by the goodness rating of that identification, we obtained the fit index (FI) [14] which allows one to raise those identifications that were actually considered as good tokens of the native categories and to lower the ones that were selected because of the lack of good competitors [14]. In the current study, the results relative to the phonemes /ɑ:/, /ɔ:/ and /ʊ/ were not taken into account.

## 2.4. Experiment 2

The purpose of this experiment is to examine the discrimination level of non-native phoneme pairs, and to test the formation of new L2 phonetic categories, from a PAM/PAM-L2 perspective. It can be seen as a categorical experiment since each L2 phoneme was represented by multiple natural tokens [13]. For each contrast, 8 *change trials* and 8 *catch trials* were submitted to each student, each trial made up of 3 items produced by the three different AE native speakers. In *change trials*, the *odd item*, i.e., the different one students had to discriminate, was placed alternatively in the first, in the second or in the third position of the trial, occurring with near-equal frequency in all these positions. The change trials tested students' ability to discriminate two different English phonological categories. In catch trials, the items were three acoustically different tokens of the same vowel category. Therefore, they tested students' ability to ignore acoustic differences among tokens of the same category. For instance, in order to test the contrast A-B, the *change trials* are A-A-B, A-B-A, B-A-A and B-B-A, B-A-B, A-B-B. Similarly, the *catch trials* are A-A-A and B-B-B. If listeners heard a stimulus as different between the three items, they had to click on the computer screen on “1”, “2”, or “3”, which correspond to the positions of the different item heard in the trial. If they did not hear any different stimulus, they had to click on “none”, that is none of the stimuli was different from each other. A trial could be replayed as often as needed, but responses could not be changed. Trials were randomly presented and the interstimulus interval of each trial was of 300 ms. A total of 1080 trials (40 trials X 9 student X 3 sessions) was obtained. A-prime (A') score was calculated for each of the 5 contrasts on the basis of the proportion of “hits” (i.e., the correct detection of the odd item in change trials) and “false alarm” (i.e., the incorrect detection of the odd item in catch trials) by the formula of [15]. An A' score of 1.0 indicates perfect discrimination of a contrast (i.e., the correct response to all eight change trials and catch trials), while an A' score of 0.5 indicates insensitivity to a contrast [14].

## 3. Results

### 3.1. Experiment 1

A series of analysis of variance (ANOVA) was carried out considering the FI as dependent variable, in order to verify if the categorization of L2 phonemes in terms of native phonemes changed from the first to the third session (factor). This analysis, which was yielded on the first two modal identifications, showed that the assimilation of each L2 phoneme did not change across the sessions (i.e., pre-test, post-test, 2<sup>nd</sup> post-test) [F values (1 to 20) = 0,253 to 3,251 p > 0,05]. Thus, for each L2 phoneme, the FI of the three sessions were merged as well as the percentages of identification (Table 1). These data allowed to detect various contrasts to be tested in Experiment 2, that is /i:/-/u:/, /i:/-/ɪ/, /e/-/ɜ:/, /æ/-/ɑ:/ and /æ/-/ʌ/. Moreover, they allowed one to

establish the PAM/PAM-L2 assimilation typologies (see Table 2 and 4.).

Table 1. Mean FI and percentages of identification (% Id.) of L2 phonemes in terms of L1 phonemes (L2\_L1), quality check of L1 exemplar (L1 ex.) and assimilation consistency (Ass. Con.).

L2_L1	FI	% Id.	L1 ex.	Ass. Con.
/i:/_/i/	3.9	100%	Good	✓
/ɪ/_/ɪ/	3.5	98%	Good	✓
/ɛ/_/E/	3.6	97%	Good	✓
/ɑ:_/a/	2.7	89%	Fair	✓
/u:_/u/	3	100%	Fair	✓
/æ/_/a/	1.4	59%	Poor	×
/æ/_/E/	1.2	41%	Poor	×
/ʌ/_/a/	1.8	70%	Poor	×
/ʌ/_/O/	0.4	17%	Poor	×
/ɜ:/_/E/	1.9	85%	Poor	✓

In order to determine how consistently the AE phonemes were assimilated to the SI ones, we considered the identification test results, both in terms of percentages of identification and FI. Namely, those L2 phonemes whose percentage of identification fell between 100% and 75%, e.g., /ɜ:/, were considered as consistently identified as native phonemes, while the ones whose percentage of identification was less than 75% were considered as not consistently identified and the first two modal identifications were taken into account, e.g., /ʌ/. Moreover, in order to assess which AE phoneme could be considered as poor/good instances of native vowels, we adopted the standard deviation (s.d.) criterion inspired by [16]. The non-native phoneme that received the highest FI can be easily considered as a vowel belonging both to the L1 and the L2 phonological inventories. The AE phoneme with the highest FI was /i:/ (i.e., 3.9) and its s.d. was 0.8. By subtracting the s.d. from the FI, various subclasses of American phonemes were detected. Thus, those phonemes that received a FI falling between 3.9 and 3.1 were considered as good exemplars of the Italian phonemes, those that received a FI between 3 and 2.2 (by subtracting the s.d. from 3) were considered as fair exemplars of the native ones and, finally, those with a FI less than 2.1 were considered as poor exemplars of the Italian vowels to which they have been perceptually related (Table 1).

### 3.2. Experiment 2

Table 2 shows the A' scores of the group obtained in the pre-test, post-test and 2<sup>nd</sup> post-test for the five contrasts previously described. This perceptual test was also given to a group of ten AE speakers (AES) as a control. The level of discrimination accuracy of the L2 contrasts in the three sessions was ascertained by a series of one-way ANOVA where the dependent variable (i.e., A') was analysed by a composite factor, i.e., Group/Tests, made up of the results of the three test sessions and the results of the AES. The contrasts /æ/-/ɑ:/ and /ɛ/-/ɜ:/ showed the same pattern of improvement. The statistical analysis revealed a significant effect of the Group/Tests [F(3,29) = 4.926 to 7.806 p > 0.05]. A Tukey test was carried out for both contrasts and showed a general improvement from the pre-test to the 2<sup>nd</sup> post-test (pre-test vs. post-test: p > 0.05; post-test vs. 2<sup>nd</sup> post-test: p >

0.05; pre-test vs. 2<sup>nd</sup> post-test: p < 0.05). Moreover, a native-like discrimination level was obtained in the 2<sup>nd</sup> post-test, since there were no significant differences between the scores of the AE and SI speakers in the last test (p > 0.05). Differently, the significant effect for the contrast /æ/-/ʌ/ [F(3,29) = 5.348 to 11.007 p < 0.01] was determined by the comparison with the AE speakers. Actually, no differences were detected among the high A' of the three sessions (p > 0.05) but, crucially, the discrimination level was statistically equal to the one of the AES in the 2<sup>nd</sup> post-test only (p > 0.05). For the contrast /i:/-/ɪ/, the composite factor's effect was significant [F(3,29) = 21.682 p < 0.01] and the Tukey test showed that this contrast was not discriminated at a native-like level during all the three tests (p < 0.05) and, moreover, a lower A' score was recorded in the post-test (pre-test vs. post-test: p < 0.05; post-test vs. 2<sup>nd</sup> post-test: p < 0.05). Finally, also for the control contrast /i:/-/u:/ a significant effect was carried out [F(3,29) = 14.973 p < 0.01]. If in the pre-test only there was no difference between the discrimination level of SI and AES groups (p > 0.05; post-test vs. AES: p < 0.05; 2<sup>nd</sup> post-test vs. AES: p < 0.05), the A' scores of the students slightly diminished in the post-test (pre-test vs. post-test: p < 0.05; post-test vs. 2<sup>nd</sup> post-test: p > 0.05; pre-test vs. 2<sup>nd</sup> post-test: p > 0.05).

Table 2. A' of pre-test, post-test and 2<sup>nd</sup> post-test of the SI subjects and of the AES as control.

PAM	Contr.	Pre-t.	Post-t.	2 <sup>nd</sup> post-t.	AE
TC	/i:/-/u:/	0.95	0.86	0.90	0.99
SC	/i:/-/ɪ/	0.62	0.32	0.37	0.99
CG	/ɛ/-/ɜ:/	0.76	0.85	0.94	0.99
UC	/æ/-/ɑ:/	0.64	0.82	0.95	0.95
UU	/æ/-/ʌ/	0.81	0.76	0.93	0.99

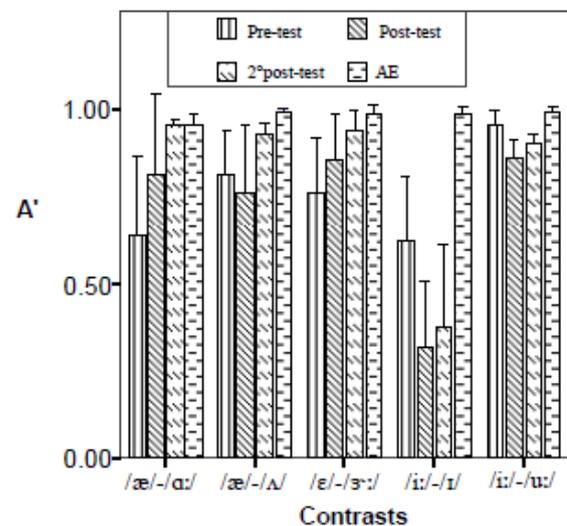


Figure 1. A' scores of the students in the pre-test, post-test and 2<sup>nd</sup> post-test and of the AE speakers.

## 4. Discussion and conclusions

From the PAM perspective, the AE phonemes could be seen as non-native phones with regard to SI speakers. By combining the information derived by the FI and the percentage of identification, the following PAM typologies were detected: /i:/-/u:/ as TC (the two phones totally

assimilated to the native phonemes /i/ and /u/), /i:/-/i/ as SC (both phones assimilated to /i/), /ɛ/-/ɜ:/ as CG (both phones assimilated to /E/ with different degrees), /æ/-/ɑ:/ as UC (the first phone categorised with /a/ and the second perceived as /a/ and /E/) and /æ/-/ʌ/ as UU (the two phones perceived, respectively, as /a/ and /E/ and /a/ and /O/) (Table 2). The expected discrimination for the TC typology is excellent, since the two non-native phones are perceived phonologically and phonetically equivalent to the L1 phonemes. No perceptual learning was expected to occur since the two phones are already present in the SI's phonological system. Actually, the discrimination of the /i:/-/i/ was very high in the pre-test (0.95) and remained high in the following two tests (post-test: 0.86; 2<sup>nd</sup> post-test: 0.90). The predicted discrimination for the SC typology is low and no category formation is expected since the two phones are perceived as equal to each other and to the native phoneme to which they are associated. Actually, the SI group discriminated the contrast /i:/-/i/ with a quite low A' in the pre-test (A': 0.62) and no category formation occurred in the subsequent test sessions (post-test: 0.32; 2<sup>nd</sup> post-test: 0.37). The lowering of the A' of these last two contrasts was due to an increase of false alarms in response to /i:/-/i/. Probably, the students, being aware of the existence of another vowel similar to their native /i/, i.e., /ɪ/, erroneously discriminated these catch trials and this seems to confirm the difficulty of discriminating between the contrasts /i:/-/i/. The CG typology is expected to be well discriminated since the two phones, even though they are assimilated to the same L1 phoneme, are different in "fitting". The formation of a new category is expected just for the phone perceived as more deviant with respect to the native phoneme. For the contrast /ɛ/-/ɜ:/, belonging to the CG typology, the discrimination was good in the first test (A': 0.76) and gradually improved in the following two tests (post-test: 0.85; 2<sup>nd</sup> post-test: 0.94), proving that a perceptual learning for the deviant phone /ɜ:/ seemed to have occurred, especially in the 2<sup>nd</sup> post-test. The contrast /æ/-/ɑ:/ was included in the UC typology. The expected discrimination is good and the formation of a new category is predicted just for the uncategorised phone. The low A' of the pre-test (0.64) was followed by a better A' in the post-test (0.82) and the 2<sup>nd</sup> post-test (0.85). Thus, as for the previous contrast, the phone /æ/ seemed to be perceptually learned especially in the last session. Finally, the prediction for the UU typology was respected. The contrast /æ/-/ʌ/ was highly discriminated (pre-test 0.81) as expected for two non-native phones assimilated to different sets of L1 phonemes. Apart from a small non-significant decrease in the second test (0.76), the discrimination accuracy significantly improved in the third test (0.93). To sum up, this study supports the predictions of PAM and PAM-L2. Our findings are consistent with the PAM since the way that SI subjects discriminated the L2 phones, i.e., AE phonemes, was related to the processes with which they were assimilated to the native categories. The modality of L2 assimilation did not change across the three sessions, while the discrimination ability changed for some contrasts only. We suggest that a longer exposure to the L2 is required to modify L2 categorization abilities, whereas a short period of focused phonetic lessons seems adequate to improve L2 discrimination processes. Moreover, it was proved that L2 perceptual learning takes place for some phones and not for others, as predicted by the PAM-L2.

Additionally, if the rephonologization is predicted to mainly occur in immersion context in 6-12 months [1], here it was observed that it could occur also from 0-6 months of focused phonetic lessons added to university lessons. Actually, these listeners seemed to be able to stretch their five-vowel L1 inventory probably beginning to form new L2 categories. Consequently, the PAM-L2 seems to be a valid perceptual model also for L2 acquisition in formal context.

## 5. References

- [1] Best, C. T., Tyler, M. D., "Nonnative and second language speech perception: Commonalities and complementarities", In M.J. Munro & O.-S. Bohn [Eds], *Second Language speech learning: the role of language experience in speech perception and production*, 13-34, Amsterdam: John Benjamins, 2007.
- [2] Best, C. T., "A direct realist view of cross-language speech perception", in Strange W. [Ed], *Speech perception and linguistic experience: Issues in cross-language research*, 171-204, Timonium, MD: York Press, 1995.
- [3] Best, C. T., McRoberts, G. W., & Goodell, E., "Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system", *JASA*, 109: 775-794, 2001.
- [4] Bundgaard-Nielsen, R., Best, C. T., Tyler, M., "Vocabulary size matters: The Assimilation of Second Language Australian English vowels to first-language Japanese vowel categorise", *Applied Psycholinguistics*, 32(2011): 51-67, 2010.
- [5] Piske, T., "Implications of James E. Flege's research for the foreign language classroom," in M.J. Munro and O.-S. Bohn [Eds] *Second Language speech learning: the role of language experience in speech perception and production*, 301-314, John Benjamins, Amsterdam, 2007.
- [6] Mora, J. C., & Fullana, N., "Production and perception of English /i:/-/i/ and /æ/-/ʌ/ in a formal setting: Investigating the effects of experience and starting age", *ICPhS Proceedings, Saarbrücken, 6-10 August 2007*, 1613-1616, 2007.
- [7] Langeris, A., and Hazan, V., "Cross-language perceptual assimilation and discrimination of Southern British English vowels by Greek and Japanese learners of English," *ICPhS Proceedings, Saarbrücken, 6-10 August 2007*, 1641-1644, 2007.
- [8] Elliot, R. A., "Foreign language phonology: field independence, and the success of formal instruction in Spanish pronunciation", *The Modern Language Journal*, 79: 530-542, 1995.
- [9] Grimaldi, M. (2009). "Acoustic correlates of phonological microvariations." in *Romance Languages and Linguistic Theory 2006*, eds. D. Tock and W. Leo Wetzels (Amsterdam / Philadelphia: John Benjamins), 89-109.
- [10] Ladefoged, P., "Vowels and consonants: An introduction to the sounds of language", Malden, MA, Blackwell, 2001.
- [11] Iverson, P., Evans, B. G., "Learning English vowels with different first-language vowel systems: Perception of formant targets, formant movement, and duration", *JASA* 122 (5): 2842-2854, 2007.
- [12] Ylinen, S., Uther, M., Latvala, A., Vepsäläinen, S., Iverson, P., Akahane-Yamada, R., and Näätänen, R., "Training the Brain to Weight Speech Cues Differently: A Study of Finnish Second-language Users of English", *Journal of Cognitive Neuroscience* 22:6, 1319-1332.
- [13] Flege, J. E., MacKay, I. R. A., "Perceiving vowel in a second language", *Studies in Second Language Acquisition*, 26: 1-34, 2004.
- [14] Mac Kay, I. R. A., Meador, D., Flege, J. E., "The identification of English Consonants by Native Speakers of Italian", *Phonetica*, 58: 103-125, 2001.
- [15] Snodgrass, J., Levy-Berger, G., Haydon, M., "Humans Experimental Psychology", New York, Oxford University Press, 1985.
- [16] Guion, S. G., Flege, J. E., Ahahane-Yamada, R., Pruitt J. C., "An investigation of current models of second language speech perception: the case of Japanese adults' perception of English consonants", *JASA*, 107: 2711-2725, 2000.