PERCEPTION OF FA BY NON-NATIVE LISTENERS IN A STUDY ABROAD CONTEXT[1]

PILAR AVELLO[1]
JOAN CARLES MORA[2]
CARMEN PÉREZ-VIDAL[1]
Universitat Pompeu Fabra[1]
Universitat de Barcelona[2]

Abstract

The present study aims at exploring the under-investigated interface between SA and L2 phonological development by assessing the impact of a 3-month SA programme on the pronunciation of a group of 23 Catalan/Spanish learners of English (NNSs) by means of phonetic measures and perceived FA measures. 6 native speakers (NS) in an exchange programme in Spain provided baseline data for comparison purposes. The participants were recorded performing a reading aloud task before (pre-test) and immediately after (post-test) the SA. Another group of 37 proficient non-native listeners, also bilingual in Catalan/Spanish and trained in English phonetics, assessed the NNS' speech samples for degree of FA. Phonetic measures consisted of pronunciation accuracy scores computed by counting pronunciation errors (phonemic deletions, insertions and substitutions, and stress misplacement). Measures of perceived FA were obtained with two experiments. In experiment 1, the listeners heard a random presentation of the sentences produced by the NSs and by the NNSs at pre-test and post-test and rated them on a 7-point Likert scale for degree of FA (1 = "native" , 7 = "heavy foreign accent"). In experiment 2, they heard paired pre-test/post-test sentences (i.e. produced by the same NNS at pre-test and post-test) and indicated which of the two sounded more native-like. Then, they stated their judgment confidence level on a 7-point scale (1 = "unsure", 7 = "sure"). Results indicated a slight, non-significant improvement in perceived FA after SA. However, a significant decrease was found in pronunciation accuracy scores after SA. Measures of pronunciation accuracy and FA ratings were also found to be strongly correlated. These findings are discussed in light of the often reported mixed results as regards pronunciation improvement during short-term immersion.

1. Introduction

A large body of research into second language (L2) phonological acquisition has analysed the phenomenon of foreign accent (FA), which is the result of perceived differences between the acoustic-phonetic properties of L2 speech and those characterising native speakers' norms: "Listeners hear foreign accents when they detect

---

divergences from English phonetic norms along a wide range of segmental and suprasegmental (i.e., prosodic) dimensions" (Flege 1995). Most FA research has explored the perception of accented speech by native listeners, who have been found to assess accentedness reliably regardless of training or experience (Brennan and Brennan 1981, Flege and Fletcher 1992). These studies have usually been conducted in learning contexts of long-term immersion in the L2 community, and in connection with variables that have been identified as influencing perceived degree of FA, most notably age of onset of L2 learning and L2 experience.

Despite the traditional use of native listeners, a few studies have also analysed the perception of accented speech by non-native listeners. For instance, Flege (1988) found that two groups of Chinese non-native listeners were able to judge degree of FA in Chinese-accented English sentences following the same response pattern observed for native listeners, with judgements from the most experienced Chinese group more closely resembling native listeners' judgements. Similarly, in Mackey et al. (2006), proficient Arabic listeners provided FA judgements of Italian-accented speech in English which were strongly correlated with native listeners' judgements. These findings extended those by Flege 1988, as they suggested that non-native listeners are able to reliably assess accentedness in speech samples from L2 learners even in the absence of a shared L1 background between listeners and learners. More studies have supported the finding that listeners with different L1s may share a similar response to accented speech. Munro et al. (2006) found moderate to high correlations in FA scores, as well as in comprehensibility and intelligibility scores, provided by four different groups of native listeners and non-native listeners with varying L1s who assessed English speech samples with different accents. Derwing and Munro (in press) also obtained high correlations between native listeners' ratings of accented English and ratings from a group of proficient non-native listeners with different L1 backgrounds, concluding that both groups of listeners may be equally reliable to assess L2 learners' speech. The results of these studies, therefore, indicate that non-native listeners who are proficient enough in the L2 they are asked to evaluate can provide reliable FA judgements which closely match those of native listeners. However, these few studies analysing the perception of accented speech by non-native listeners are usually conducted also in long-term immersion contexts, rather than in shorter periods of immersion, such as those typical of Study Abroad learning contexts.

Study Abroad (SA) is a second language learning context which can be defined as a combination of language-based and/or content-based classroom instruction together with out-of-class interaction in the native speech community (Freed 1995). SA programmes have become very popular, for instance, in Europe and America, due to the common sense and long held assumption that immersion in the L2 community results in substantially enhanced L2 knowledge, as such immersion is assumed to offer plenty of opportunities for interaction with native speakers and exposure to a great amount of quality input. Consequently, SA programmes have been encouraged by language instructors and academic administrators and have come to play an important role in governments' L2 learning policies, as a means to promote multilingualism in response to an increasingly globalised international context (see, e.g., Kinginger 2009 and Llanes 2011 for a review of official figures and language learning policies). An increasing body of research has been subsequently devoted to this learning context, in order to account

for the nature of the study abroad experience and empirically assess its impact on L2 learners' linguistic development (see research overviews in DeKeyser 2007; DuFon and Churchill 2006; Freed 1995). For the most part, research has found evidence for a positive effect of the study abroad experience on learners' L2 development, yet actual linguistic gains appear to be related to individual and context variables, such as contact patterns while abroad, L1 and L2 use, L2 exposure, onset level of proficiency, or length of stay, as well as to aspects of programme design (see Pérez-Vidal and Juan-Garau 2011 for a characterisation of SA). A complex picture results of the interaction of all these factors, with findings sometimes providing inconclusive or conflicting evidence, as the benefits of SA are not always clear for all language skills, or the gains reported may fall short of the high expectations arising out of the above-mentioned widespread belief in the substantial effects of study abroad immersion.

Research has analysed the impact of SA on different linguistic domains, and usually in contrast with formal instruction (FI) in at home (AH) institutions. Results have provided consistent evidence of the beneficial role of SA for lexical improvement (Collentine 2004; Llanes and Muñoz 2009), as well as for writing (Pérez-Vidal and Juan-Garau 2009, 2011). Sociolinguistic skills have been the object of considerable research, with studies examining, for instance, communication strategies (Lafford 1995) or pragmatic competence (Barron 2006), and which have also yielded results supporting the positive effect of SA on these areas. However, mixed results have been found for grammar. Results by Collentine 2004 showed a superiority for AH learners over those who went abroad, whereas the opposite was true in Howard (2005). Most SA research has focused on the development of oral skills, traditionally considered to be the linguistic domain most likely to improve as a result of SA, and research findings in general have supported this view. Some studies have analysed the impact of SA on overall L2 speaking proficiency (Brecht et al. 1995, Segalowitz and Freed 2004), and extensive research has also been carried out to analyse gains in L2 learners' fluency (Freed et al. 2004, Juan-Garau and Pérez-Vidal 2007; Trenchs-Parera 2009, Valls-Ferrer 2011). Nevertheless, studies focusing on specific aspects of phonological development in learners' speech production are scarce.

Studies of phonological development during SA generally focus on the differential effects of SA vs FI on production accuracy, and have yielded mixed results. Díaz-Campos (2004) reported a positive effect of both learning contexts on the production of Spanish plosives in two groups of English students of Spanish, although development towards native-like patterns was found to be stronger in the FI group. Contrarily, Díaz-Campos (2006) observed greater gains in the production of Spanish consonants for the SA group as compared with the FI group. Mora (2008) examined the production of VOT in English voiceless plosives by a group of Spanish/Catalan bilingual learners after a two-term FI period at their home university and after a three-month SA term abroad. He found no effect of FI on VOT duration, whereas an increase was observed after SA, although non-significant. However, in a similar study analysing English vowels, significant improvement in production was found after FI, but not after SA (Pérez-Vidal et al. 2011). Højen (2003) found better perceived foreign accent scores after SA as a function of length (average=7.1 months), but production at the segmental level did not improve significantly. Avello (2011) and Avello et al. (in press) reported minor gains in

perceived FA scores and no significant improvement in segmental production accuracy, respectively.

The present study thus explores the under-investigated impact of SA on L2 learners' phonological development by assessing the impact of a 3-month SA programme on the pronunciation of a group of 23 bilingual Catalan/Spanish learners of English by means of both phonetic measures of pronunciation accuracy and perceived FA measures by non-native listeners. The relationship between both types of measures is also explored. Our objectives are thus the following:

- To explore the effect of SA on L2 learners' phonological development (measured by pronunciation accuracy scores and FA scores).
- To explore the relationship between the phonetic properties of L2 learners' speech (objective measures) and perceived degree of FA (subjective measures).

2. Method

2.1. Participants

2.1.1. Speakers

This study is part of a larger, state-funded project called SALA (Study Abroad and Language Acquisition), which aims at uncovering the effects of a short, 3-month SA period on the linguistic development of university level L2 English learners. Data were collected from a group of non-native speakers (NNSs) studying Translation and Interpreting in Barcelona, Spain (*N*=23; 20 females and 3 males). Their age ranged from 17 to 21 (*M*=18.8). At the time of data collection, none of them reported suffering any speech impairment. They all started to learn English as a foreign language (EFL) in AH institutions around the same age (8 years), as established by the curriculum in the Spanish educational system, thus sharing a similar age of onset of L2 learning (AOL). Their acquisition of English took place basically through classroom instruction (i.e., as a FL in their native speech community), sharing also a similar exposure to English of between 700-800 hours.

These learners had to certify an advanced level of proficiency in English (equivalent to a B2 in the Common European Framework of Reference or CEFR) in order to be admitted to the university where they were studying. As part of their Translation and Interpreting degree, they had to specialise in two FLs, English being one of them, and the other language being either French or German. They had thus a similar multilingual profile, since they were all early bilinguals of both Spanish and Catalan, studying English and another FL. They all had a compulsory 3-month study abroad term in an English-speaking country at the beginning of their second academic year.

Speech samples from 6 native speakers (NS) of English served as baseline data to assess the learners' performance. These NSs were also part of the SALA corpus. None of them reported any speech dysfunction. They were young university students enrolled in an exchange programme in Spain (i.e., they were learners of L2 Spanish), with an age

range similar to that of the NNSs. Both groups of speakers had, therefore, a similar profile, and consequently their data were highly comparable.

2.1.2. Listeners

A group of proficient non-native listeners were recruited as judges (NNJs, *N*=37) to assess the NNSs' degree of FA. Their linguistic profile was similar to that of the NNSs, i.e., they were also bilingual speakers of Spanish and Catalan studying English as a FL. They were taking a degree in English Studies in Barcelona, which involved attending Linguistics and Literature content courses taught in English, and by the time of data collection they had completed two courses on English phonetics and phonology. These courses included a comprehensive description of English segmental and suprasegmental properties, phonetic and phonological transcription, and pronunciation training, as well as training in the use of speech analysis software (Praat). The courses were designed to specifically tap on the problems facing L1 speakers of Spanish/Catalan when learning English. They had, therefore, a proficient level of English, a sound knowledge of English phonetics, and were highly familiar with the accented speech they were asked to judge, as they shared the non-native speakers' L1s. They performed two listening experiments (see 2.3.2. below) and completed a questionnaire tapping on their linguistic profile and their degree of familiarity with different native and non-native English accents. They did these tasks for course credit.

2.2. Speech samples

Speech samples were elicited by means of a reading aloud (RA) task in which the participants read the text *The North Wind and the Sun* (NWS, see Appendix 1). This is a standard, 114-word text of which different versions exist in different languages (e.g. French version: Fougeron and Smith 1999; Spanish version: Martínez-Celdrán et al. 2003; RP British English: Roach 2004), and which has been used to document differences characterising English pronunciation in different dialects or by foreign speakers (see Schneider et al. 2004). The fact that the text was the same for all the subjects facilitated contrasting analyses, as the same vowel and consonantal items appeared in all the speech samples, and in the same contexts. In order to assess the effects of the 3-month SA, data were collected prior to the students' departure (pre-test), and immediately after their return (post-test).

The participants were recorded one at a time. They were instructed to read the text first silently on their own, and then aloud at a normal speaking rate to be recorded. They were told that they would be asked a question about the content of the text, which they were to answer as quickly as possible after reading it aloud. This was done so as to draw the participants' attention to the content, in such a way that they were not aware that the focus of interest was pronunciation, and with the aim of obtaining more natural sounding data. The participants read the text out loud, and immediately after finishing, they were asked the following question: *Was the North Wind stronger than the sun?*, which they answered by stating *yes* or *no*.

Data from the NNSs were recorded in sound-attenuated cabins using analogue tape recording technology, and were subsequently digitised in .wav format at 22,050 Hz, with 16-bit resolution. Data from the NSs were digitally recorded in professional sound-proof cabins, using the Pro Tools digital audio platform. The digital files were saved in .wav format at 44,100 Hz (later down sampled to 22,050 Hz), 16-bit resolution.

A sentence from the RA task was selected (see Appendix 2) which presented several segmental and suprasegmental properties that were likely to result in accented pronunciation for our L2 learners (see pronunciation errors in 2.3.1. below). The selected sentence was extracted from each participant's recording, and the resulting files were edited and normalised for intensity at 70.0 dB in order to create the stimuli for the listening experiments. Data manipulation was carried out with Praat 5.1 (Boersma and Weenink 2009).

2.3. Data analyses

2.3.1. Pronunciation accuracy scores

The NNSs' production accuracy was assessed by means of a phonetic analysis (Brennan and Brenna 1981, Trofimovich et al. 2009), which was conducted by the first author on the waveform and corresponding spectrogram of each speech sample. Pronunciation errors were identified and accuracy scores were subsequently computed by counting the total number of mispronunciations in each NNS' pre-test and post-test speech samples. These accuracy scores served as objective, phonetic measures of the NNSs' speech production development, and included mispronunciations affecting segmental articulation (deletions, insertions, and phonological substitutions), as well as stress misplacement. Presented below are some examples of such pronunciation errors:

a) Deletions:
-deletion of [l] in *warm(l)y* (one-segment deletion)
-deletion of final syllable in *travel(er)* (multiple-segment deletion)

b) Insertions:
-insertion of an extra vowel [e] in *immediat[e]ly*
-insertion of a velar consonant at the beginning of *[ɣ]warmly*

c) Substitutions:
-substitution of bilabial approximant [β] for velar fricative [v] in *traveller*
-substitution of dental plosive [d] for dental fricative [ð] in *then*
-substitution of open vowel [a] for close back vowel [ɔ] in *warmly*
-substitution of dental fricative [ð] for alveolar plosive [d] in *immediately*
-substitution of velar fricative [x] for glottal fricative [h] in *his*

d) Stress misplacement:
-stress shift to the penultimate syllable in multisyllabic words: *traˈveller* for *ˈtraveller*, *immeˈdiately* for *iˈmmediately*.

2.3.2. Perceived FA measures

Perceived FA measures consisted of subjective listeners' judgements obtained from the proficient NNJ by means of two listening experiments: a rating task and a paired-comparison task. These experiments provided us with behavioural measures of the perceived degree of FA in the NNSs' pronunciation prior to and immediately after SA. They were self-paced tasks created and run with Praat software (Boersma and Weenink 2009, version 5.1). Both listening experiments were performed during the same session (equivalent to a class activity within the NNJs' course on English phonetics). The rating was conducted first, then the paired-comparison. The whole session lasted around an hour.

a) Experiment 1: Rating
The rating experiment (Munro et al. 2006, Derwing et al. 1998) provided a holistic measurement of perceived FA changes throughout time. The NNJ heard a randomised presentation of the speech samples produced by the NNS (pre-test and post-test) and the NS (baseline). Their task was to rate the degree of FA in the oral samples by means of a 7-point Likert scale, where 1 stood for "native" and 7 stood for "heavy foreign accent". They were instructed to make use of the whole scale. Each stimulus was repeated twice for a total of 104 trials per judge (23 NNSs x 2 times x 2 repetitions + 6 NSs x 2 repetitions), making up a total of 3,848 judgements (104 trials x 37 judges). 10 practice trials were presented before the actual experiment in order to familiarise the listeners with the procedure, allowing them also to check the volume level.

b) Experiment 2: Paired-Comparison
It was expected that the paired-comparison experiment would provide a more fine-grained global assessment of the effect of SA on the NNSs' degree of accentedness, since this methodology consists of directly comparing two items produced by the same speaker at two different testing times. Previous research analysing L2 speech production (Riney and Flege 1998, Bradlow et al. 1999, Højen 2003) has reported it as very sensitive to slight changes in pronunciation of the kind that are most likely to occur after a short SA programme.

First, the NNJ had to decide which sentence was more native-like out of two paired sentences (i.e., produced by the same NNS). Then, they stated their confidence level on a 7-point scale (1="unsure" - 7="sure"). For each NNS, there was a pre-test/post-test trial and a post-test/pre-test trial. The order of presentation was counterbalanced across trials, which were randomised. There were 46 trials per judge (2 orders x 23 NNSs), making up a total of 1,702 judgements (46 trials x 37 judges). As was the case with experiment 1, experiment 2 was also preceded by a few practice trials.

3. Results and discussion

3.1. Pronunciation accuracy scores

Figure 1 below graphically presents the accuracy scores obtained by the NNSs at pre-test (M = 3.95, SD = 2.75) and at post-test (M = 3.30, SD = 2.65). The number of pronunciation errors ranged from 0 to 9 at both testing times, with considerable inter-subject variability, as indicated by the relatively high standard deviation. A paired-samples t-test revealed significant gains in pronunciation accuracy after SA [t(22) = 2.135, p = .044)], i.e., the NNSs produced significantly fewer pronunciation errors after SA than they did before their departure, the eta squared (η^2 = .17) indicating a large effect size. These results suggest that the 3-month SA had a large positive impact on the NNSs' phonological production accuracy, allowing them to significantly improve their segmental articulation and stress production.
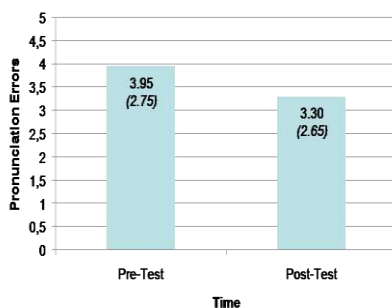

Figure 1: Mean number of pronunciation errors produced by the NNS before and after SA.
SD **in parenthesis.**

3.2. Perceived FA scores

3.2.1. Experiment 1: Rating

The NNJ used a 7-point scale to rate the degree of FA in the speech samples presented to them (1=."native", 7="heavy foreign accent"). Preliminary reliability analyses were conducted to explore consistency in the NNJ' ratings, and they yielded both high intra-rater and inter-rater coefficients. Regarding intra-rater reliability, a strong correlation was found in the judge-based FA scores assigned at each of the two rating repetitions (r = .855, p < .001), which indicates that each judge's first and second repetition ratings were very similar. Inter-rater reliability was examined by means of an intra-class correlation (ICC) analysis which yielded a high Cronbach's Alpha (.996), indicating a high degree of agreement among the judges.

Figure 2 below illustrates the mean FA ratings assigned by the NNJ to the NNSs (pre-test and post-test) and to the NS (baseline). As expected, the ratings for the NS group were very close to 1 ($M = 1.29$, $SD = .17$), indicating that the NNJ successfully identified the native speakers of English, and rated them accordingly. The NNSs' ratings were considerably outside the range of the NSs' ratings both at pre-test and post-test, and significantly differed from them at the two testing times, as shown by independent-samples t-tests ($p < .001$). There was a slight improvement in the NNS' FA scores after SA, since the perceived degree of accentedness decreased from pre-test ($M = 4.88$, $SD = 1.28$) to post-test ($M = 4.68$, $SD = 1.20$). This decrease, however, failed to reach significance [$t(22) = 1.306$, $p > .05$]. These results seem to indicate a positive trend of development towards less accented speech, suggesting that SA might have had some impact on the NNSs' degree of accentedness, although statistically non-significant.
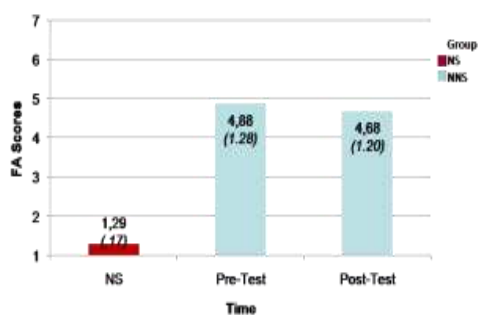

Figure 2: Mean FA ratings (Experiment 1) for NNS (Pre-Test and Post-Test) and NS (baseline). *SD* in parenthesis.

3.2.2. Experiment 2: Paired-Comparison

The paired-comparison experiment complemented the rating experiment, as it was assumed to yield more fine-grained measures of the global degree of accentedness perceived by the NNJ in the learners' speech samples. The combination of the FA scores obtained with both experiments was thus expected to provide us with the necessary information to better evaluate possible changes in the NNSs' speech production.

In the paired-comparison experiment, the NNJ were asked to directly compare the learners' pre-test and post-test speech samples. The NNJ first had to indicate which of the two versions was better (i.e. more native-like), and then used a 7-point scale to state their degree of confidence (1="unsure" - 7="sure"). The data thus obtained were codified as follows: a negative sign was assigned to the selected confidence scale value when the pre-test version was chosen as better, and a positive sign was assigned when the post-test version was preferred. This resulted in scores ranging between -7 and 7, which were further recoded into values between -6 and 6 (see figure 3 below), with positive values indicating that a majority of post-test samples had been preferred as more native-like, and pointing, therefore, to an improvement in speech production after SA.

The mean global FA scores are presented in figure 3. Individual scores ranged between -2.05 and 4.42, indicating large inter-subject variability. 11 out of the 23 learners obtained positive scores (ranging from .29 to 4.42), although in most cases scores were below 2. The positive group mean (.36), although only slightly above 0, can be interpreted as a slight improvement in the NNSs' degree of accentedness, in a similar way to the results of the rating experiment. There was, therefore, a parallelism between the two listening experiments, in the sense that they both seemed to point to a positive, although small effect of SA on the NNSs' perceived degree of FA. These results also matched the gains observed in the analysis of the pronunciation accuracy scores.

Subject	Mean
arri	-,41 (3,68)
begi	-,33 (2,70)
caam	1,53 (3,64)
clfa	-,81 (3,38)
coar	-,22 (3,52)
crga	1,17 (3,60)
esam	-,74 (3,02)
gape	-2,05 (3,77)
gaur	-,31 (3,33)
giro	2,12 (2,95)
gore	,60 (3,21)
goro	1,01 (3,28)
izga	2,91 (3,34)
jigo	4,42 (1,60)
lele	-1,27 (4,05)
miod	-1,38 (3,72)
moll	3,99 (1,41)
paar	-1,63 (4,08)
pare	-1,50 (2,94)
ques	-1,42 (3,66)
rira	1,35 (3,34)
vamo	,91 (3,78)
vipl	,29 (3,01)
Total	**,36**

Scores:

1 to 6 = improvement
-1 to -6 = no improvement
0 = no change.

Data codification: negative value assigned to confidence scale number for a judgement when Pre-Test was chosen as better; positive value assigned when Post-Test was chosen as better.

Values between -7 and 7, recoded to obtain values between -6 and 6:

-7	-6	-5	-4	-3	-2	-1	1	2	3	4	5	6	7
-6	-5	-4	-3	-2	-1	0	0	1	2	3	4	5	6

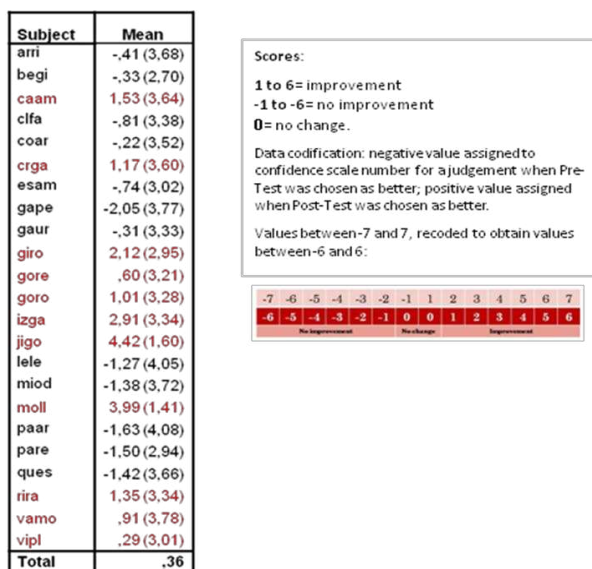No improvement   No change   Improvement

Figure 3: Individual and group mean FA scores for NNS (Experiment 2). 11 subjects (highlighted in red) obtained positive scores, signalling improvement after SA. The group mean (.36) was also positive.

Taken together, these findings suggest that increased experience with the L2 in the context of SA was beneficial for the learners' pronunciation development, as measures of pronunciation accuracy and perceived degree of FA both point towards improved performance after SA. Such an improvement can be explained on the basis of the excellent opportunities for oral practice available while abroad, as the learners take advantage of the exposure to varied and authentic L2 input, and may engage themselves in meaningful interactions in real communicative situations which may lead to useful feedback from native speakers. . The positive albeit moderate effect of SA on pronunciation found in this study is also in accordance with the results of most SA research, which report significant gains in other linguistic skills such as vocabulary (Collentine 2004), writing (Pérez-Vidal and Juan-Garau 2009) and especially oral fluency (Perez-Vidal and Juan-Garau 2007, Trenchs-Parera 2009; Valls-Ferrer 2011).

Interestingly though, when it comes to phonological development, the scant existing research has not provided consistent evidence supporting a large effect of SA on improved pronunciation, despite the positive outcomes shown in other oral abilities, and the fact that oral production is assumed to be one of the most practiced skills while abroad. Hence, our findings regarding improved accuracy in pronunciation contrast with previous research which has mostly focused on the analysis of a limited number of specific vowel and/or consonantal L2 sounds (Avello et al. in press, Díaz-Campos 2004, Pérez-Vidal et al. 2011), and has failed to show a substantial impact of SA on segmental production. These divergences may be attributable to the differences in the selected object of study. Instead of analysing a limited set of discrete units, the present study has targeted a wider range of phonological features, by looking into various phenomena at the segmental and suprasegmental level, including phonemic deletions, insertions and substitutions, as well as stress implementation, which affect not only discrete units, but also syllable structure.

A slight positive impact of SA was found when analysing perceived degree of accentedness, although in this case the improvement was non-significant and suggested no large effect of SA on this domain. This is very much in line with previous FA research within the context of SA. Højen (2003) found a significant improvement in his participants' FA ratings after SA, but when exploring individual differences, he obtained a positive correlation between foreign accent ratings and length of stay (average 7.1 months); improvement was observed for those learners with longer SA (of up to 11 months), whereas learners with 3 to 4 months of SA did not improve significantly. He concluded that length of stay was an important factor for improvement of perceived FA to take place. Similarly, Avello (2011) also failed to find significant improvement in FA scores for a group of participants who had spent a 3-month period abroad. These findings may be explained by the fact that listeners seem to rate speech samples for accentedness holistically (Magen 1998), paying attention not only to aspects of segmental production or stress, but also to other suprasegmental or prosodic properties of speech, e.g. rhythm, intonation, pauses, or connected speech phenomena. In this sense, a 3-month programme may be too short for substantial improvement to accrue in these other areas of pronunciation.

3.3. Relationship accuracy scores/FA ratings

The relationship between the phonetic and FA measures was explored by means of Pearson correlations. A strong correlation was found between the two measures at pre-test (r=.814) and post-test (r=.730), and both correlations were significant at the .01 level (p<.005). This strong correlation points to a relationship between accuracy scores and FA scores, in such a way that the production by the NNS of fewer pronunciation errors resulted in the perception by the NNJ of a lower degree of accentedness, whereas the larger the number of pronunciation errors, the higher the degree of accentedness perceived. These results are in line with previous research which has established a correlation between perceived accentedness in L2 speech samples and the phonetic characteristics of those speech samples in terms of divergences from native-like pronunciation patterns (Brennan and Brennan 1981, Magen 1998). Despite the fact that

improvement in FA scores did not reach significance, it seems that our non-native listeners were nonetheless able to perceive the decrease in pronunciation errors between pre-test and post-test, i.e. they can be considered as "good judges" who correctly performed their task. Given their ability to perceive these differences in pronunciation, and the fact that they were also phonetically trained, it is likely that they also focused on other phonetic-acoustic properties of the speech samples, for instance, at the suprasegmental level mentioned above, which might not have differed substantially after SA, resulting in the differences in significance found for the accuracy scores as compared with the FA ratings.

4. Summary and Conclusions

This study aimed at furthering our understanding of the impact SA may have on L2 learners' pronunciation development. Although the few existing studies suggest that SA does not substantially change learners' pronunciation patterns, our findings indicate that SA may, indeed, result in gains for this specific area, even after a short-term immersion programme of only 3 months.

Phonetic measures of pronunciation accuracy suggest a large impact of short-term SA on production at the level of segmental articulation, as well as at the suprasegmental level of stress implementation, since a significant decrease of pronunciation errors was found in the learners' speech production after SA. However, there is no evidence of a large effect of SA on global FA scores; a positive trend seems to emerge towards less accented speech, but it is not strong and is far from significant.

Despite the differences observed between the two types of scores regarding strength of the SA impact, phonetic accuracy scores and perceived FA ratings are shown to be strongly correlated. This strong correlation points to a relationship between both types of measures, which is interpreted in terms of the proficient non-native listeners' ability to perceive the phonetic characteristics of the speech samples, namely, the decrease in pronunciation errors between pre-test and post-test, assigning worse FA ratings to speech samples containing a larger number of mispronunciations.

To summarise, it seems that SA offers the kind of input and practice that may be conducive to improvement in pronunciation (as is the case in other linguistic areas, specially of oral performance) for those learners who are able to draw on the contact opportunities and the exposure to massive amount of quality input that characterise this learning context. At least our findings regarding pronunciation accuracy seem to indicate so. But these results should be taken with caution, as the learners' FA scores fail to improve significantly or to even approach native-like scores after SA, notwithstanding the significant decrease in pronunciation errors. This may be an indication that substantial improvement is more likely to accrue at the segmental level and regarding stress, but it is possible that other areas of pronunciation not analysed in our study, such as rhythm or intonation, may not be affected by SA, or may require longer periods of immersion to benefit from the SA experience.

References

Avello, P., Lara, A.R., Mora, J.C., and Pérez-Vida, C. In press: The impact of Study Abroad and Length of Stay on Phonological Development in Speech Production. *Proceedings of the 30th Aesla Conference*, Universitat de Lleida.

Avello, P. 2011: Measuring Perceived Pronunciation Gains in Study Abroad: Methodological Issues. *Paper presented at the 29th Aesla Conference*, Universidad de Salamanca.

Barron, A. 2006: Learning to Say 'You' in German: The Acquisition of Sociolinguistic Competence in a Study Abroad Context. In A. DuFon, and E. Churchill (Eds.) 2006: Language Learners in Study Abroad Contexts. Clevedon: Multilingual Matters, pp. 59-90.

Boersma, P., and Weenick, D. 2009. Praat: Doing phonetics by computer, version 5.1. http://www.praat.org.

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., and Tohkura, Y. 1999: Training Japanese Listeners to Identify English /r/ and /l/: Long-Term Retention of Learning in Perception and Production. Perception and Psychophysics, 61(5), 977-985.

Brennan, E. M., and Brennan, J. S. 1981: Measurements of Accent and Attitude Toward Mexican-American Speech. Journal of Psycholinguistic Research, 10(05), 487-501

Brecht, R., Davidson, D., and Ginsberg, R. 1995: Predictors of Foreign Language Gain during Study Abroad. In B. Freed (Ed.), Second Language Acquisition in a Study Abroad Context. Amsterdam: John Benjamins Publishing Company, pp.37-66.

Collentine, J. 2004: The Effects of Learning Contexts on Morphosyntactic and Lexical Development. Studies in Second Language Acquisition, 26(2), 227-248.

Derwing, T. M., and Munro, M. J. In press: The development of L2 oral language skills in two L1 groups: A seven-year study. Language Learning.

Derwing, T. M., Munro, M. J., and Wiebe, G. 1998: Evidence in Favor of a Broad Framework for Pronunciation Instruction. Language Learning, 48(3), 393-410.

Díaz-Campos, M. 2006: The Effect of Style in Second Language Phonology: An Analysis of Segmental Acquisition in Study Abroad and Regular-Classroom Students. In C. A. Klee, and T. L. Face (Eds.), Selected Proceedings of the 7th Conference on the Acquisition of Spanish and Portuguese as First and Second Languages. Sommerville, MA: Cascadilla Proceedings Project, pp.26-39.

Díaz-Campos, M. 2004: Context of Learning in the Acquisition of Spanish Second Language Phonology. Studies in Second Language Acquisition, 26, 249-273.

DuFon, A., and Churchill, E. (Eds.) 2006: Language Learners in Study Abroad Contexts. Clevedon: Multilingual Matters.

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. Timonium, MD: York Press, pp.233-277.

Flege, J. E. 1988: Factors Affecting Degree of Perceived Foreign Accent in English Sentences. Journal of the Acoustical Society of America, 84(1), 70-79.

Flege, J. E., and Fletcher, K. L. 1992: Talker and Listener Effects on Degree of Perceived Foreign Accent. Journal of the Acoustical Society of America, 91(1), 370-389.

Fougeron, C., and Smith, C. L. 1999: French. In IPA (Ed.), Handbook of the International Phonetic Association pp.78-81.

Freed, B. (Ed.) 1995: Second Language Acquisition in a Study Abroad Context. Amsterdam: John Benjamins Publishing Company.

Freed, B. F., Dewey, D. P., Segalowitz, N., and Halter, R. 2004: The Language Contact Profile. Studies in Second Language Acquisition, 26(2), 349-356.

Howard, M. 2005: On the role of context in the development of learner language : Insights from study abroad research. ITL International Journal of Applied Linguistics, 148, 1-20.

Højen, A. D. 2003: Second-language speech perception and production in adult learners before and after short-term immersion. Unpublished doctoral dissertation. University of Aarhus.

Juan-Garau, M., and Pérez-Vidal, C. 2007: The Effect of Context and Contact on Oral Performance in Students Who Go on a Stay Abroad. VIAL, 4, 117-134.

Kinginger, C. 2009: Language Learning and Study Abroad. Basingstoke: Palgrave Macmillan.

Lafford, B. 1995: Getting Into, Through and Out of a Survival Situation: A Comparison of Communicative Strategies Used by Students Studying Spanish Abroad and 'At Home'. In B. Freed (Ed), : Second Language Acquisition in a Study Abroad Context. Amsterdam: John Benjamins Publishing Company, pp. 97-122.

Llanes, À. 2011: The Many Faces of Study Abroad: An Update on the Research on L2 Gains Emerged during a Study Abroad Experience. International Journal of Multilingualism, 8(3), 189-215.

Llanes, À., and Muñoz, C. 2009: A Short Stay Abroad: Does it make a Difference? System, 37(3), 353-365.

MacKay, I. R. A., Flege, J. E., and Imai, S. 2006: Evaluating the Effects of Chronological Age and Sentence Duration on Degree of Perceived Foreign Accent. Applied Psycholinguistics, 27, 157-183.

Magen, H. S. 1998: The Perception of Foreign-Accented Speech. Journal of Phonetics, 26(4), 381-400.

Martínez-Celdrán, E., Fernández Planas, A. M., and Carrera-Sabaté, J. 2003: Castilian Spanish. Journal of the International Phonetic Association, 33(02), 255-259.

Mora, J. C. 2008: Learning Context Effects on the Acquisition of a Second Language Phonology. In C. Pérez-Vidal, M. Juan-Garau and A. Bel (Eds.), A Portrait of the Young in the New Multilingual Spain. Clevendon: Multilingual Matters, pp.241-263.

Munro, M. J., Derwing, T. M., and Morton, S. L. 2006: The Mutual Intelligibility of L2 Speech. Studies in Second Language Acquisition, 28(1), 111-131.

Pérez-Vidal, C., and Juan-Garau, M. 2011: The Effect of Context and Input Conditions on Oral and Written Development: A Study Abroad Perspective. International Review of Applied Linguistics in Language Teaching, 49(2), 157-185.

Pérez-Vidal, C., and Juan-Garau, M. 2009: The Effect of Study Abroad on Written Performance. In L. Roberts, D. Véronique, A. Nilsson and M. Tellier (Eds.), EUROSLA Yearbook Volume 9 (2009). Amsterdam: John Benjamins Publishing Company, pp.269-295.

Pérez-Vidal, C., Juan-Garau, M., and Mora, J. C. 2011: The Effects of Formal Instruction and Study Abroad Contexts on Foreign Language Development: The SALA Project. In C. Sanz, and R. P. Leow (Eds.), Implicit and Explicit Conditions, Processes and Knowledge in SLA and Bilingualism. Washington D. C.: Georgetown University Press, pp.115-138.

Riney, T., and Flege, J. E. 1998: Changes Over Time in Global Foreign Accent and Liquid Identifiability and Accuracy. Studies in Second Language Acquisition, 20, 213-243.

Roach, P. 2004: British English: Received Pronunciation. Journal of the International Phonetic Association, 34(02), 239-245.

Schneider, E. W., Burridge, K., Kortmann, B., Mesthrie, R., and Upton, C. (Eds.) 2004: A Handbook of Varieties of English. Volume 1: Phonology. Berlin: Mouton de Gruyter.

Segalowitz, N., and Freed, B. 2004: Context, Contact and Cognition in Oral Fluency Acquisition: Learning Spanish in at Home and Study Abroad Contexts. Studies in Second Language Acquisition, 26(2), 173-199.

Trenchs-Parera, M. 2009: Effects of Formal Instruction and Stay Abroad on the Acquisition of Native-Like Oral Fluency. The Canadian Modern Language Review, 65(3), 365-393.

Trofimovich, P., Lightbown, P., Halter, R. H., and Song, H. 2009: Comprehension-Based Practice: The Development of L2 Pronunciation in a Listening and Reading Program. Studies in Second Language Acquisition, 31, 609-639.

Valls-Ferrer, M. 2011: The development of oral fluency and rhythm during a study abroad period. Doctoral dissertation. Universitat Pompeu Fabra. Retrieved from http://www.tdx.cat/handle/10803/52064

Appendix I

English version of the *North Wind and the Sun* text used for the Reading Aloud task (from the *Handbook of the International Phonetic Association*, IPA 1999:39), together with the instructions given to the participants.

READING ALOUD TASK
You will be asked a question about the following text.
Read the text twice. First, silently on your own, and then aloud for the examiner to record.
Then, answer the question the examiner will ask you as quickly as possible.

The North Wind and the Sun were disputing which of them was stronger, when a traveller came along wrapped in a warm cloak. They agreed that the one who first succeeded in making the traveller take his cloak off should be considered stronger than the other.
Then the North Wind blew as hard as he could, but the more he blew, the more closely did the traveller fold his cloak around him; and at last the North Wind gave up the attempt.
Then the Sun shone out warmly, and immediately the traveller took off his cloak. And so the North Wind was obliged to confess that the Sun was the stronger of the two.

Appendix II

Sentence used to create the stimuli for the listening experiments together with the corresponding phonetic transcription (standard British English):

Then	the	sun	shone	out	warmly	and	immediately	the	traveller	took	off	his	cloak
ˈðen	ðə	ˈsʌn	ˈʃɒn	aʊt	ˈwɔmli	ən(d)	ɪˈmiːdjətli	ðə	ˈtɹæv(ə)lə	tʊk	ˈɒf	(h)ɪz̥	kl̥əʊk