L2 Phonological Development in Speech Production during Study Abroad

Pilar Avello

TESI DOCTORAL UPF / ANY 2013

DIRECTORS DE LA TESI

Dra. Carmen Pérez Vidal (Departament de Traducció i Ciències del Llenguatge, UPF)

Dr. Joan C. Mora (Departament de Filologia Anglesa i Alemanya, Universitat de Barcelona)

DEPARTAMENT DE TRADUCCIÓ I CIÈNCIES DEL LLENGUATGE

Acknowledgments

First of all I would like to thank my two PhD supervisors, Dr. Carmen Pérez-Vidal (UPF) and Dr. Joan C. Mora (Universitat de Barcelona) for their guidance and advice throughout the process of writing this thesis. My special thanks go to all the researchers in the SALA team and the ALLENCAM research group for many inspiring and insightful discussions on the topics of this thesis, particularly to Dr. Aurora Bel and Dr. Cristina Sanz for their involvement in the first stages of this study, and to all the researchers who took part in data collection and coding duties. Special thanks are also due to all the students who over the years have taken part in the SALA tests, making this research project possible.

I most gratefully acknowledge the funding from the Spanish Ministry of Science and Innovation, which has supported research within the SALA project over the years through several grants. I particularly benefited from the FPI doctoral research fellowship BES-2008-010037, which has made this PhD possible. Within the Ministry's FPI programme I also received funding for a research stay at Université Paris 8, from which this PhD work has greatly benefited. I likewise thank the Department of Translation and Language Sciences for providing support for my studies and for funding the last stage of this PhD.

I am extremely grateful to Dr. Dommergues for welcoming me at Université Paris 8 and for providing all the support I needed for my research stay abroad. I would also like to sincerely thank all the colleagues and friends I have had the privilege to meet and work with at UPF: Itziar, Carmen, Rebecca, Isabel, Margalida, Anna, Victoria, Elena, Juanma, and Estela. It's been a pleasure to work with you and to share the many, many hours we spent in the office.

Pilar Avello

Barcelona, September 2013

Abstract

The present study investigates the impact of a 3-month study abroad (SA) period on second language (L2) phonological development in speech production by means of acoustic-phonetic measures, as well as listeners' assessment of perceived foreign accent (FA). Speech samples were collected from 23 bilingual Spanish/Catalan learners of English before (Pre-test) and after (Post-test) SA. Acoustic-phonetic measures consisted of measurements for voice onset time (VOT) in voiceless plosives and for vowel duration and quality, together with error rate scores resulting from the computation of pronunciation errors. Perceived FA measures were obtained from a group of native listeners (*n*=20) and another group of non-native listeners (*n*=37) who performed a rating task. Results failed to yield a large effect of SA in VOT and vowel measures, although they indicated a slight decrease in perceived FA and a significant improvement in error rate scores after SA. High correlations were found between the acoustic-phonetic measures and the FA ratings.

Resumen

Este estudio investiga el impacto de una estancia de 3 meses en el extranjero (ES) en la producción oral de una segunda lengua (L2) a través de medidas fonético-acústicas y de percepción del acento extranjero. El corpus está constituido por datos orales recogidos de un grupo de 23 aprendices de inglés hablantes nativos de español y catalán. Las muestras de habla fueron recogidas antes (Pre-test) y después (Post-test) de la ES. Las medidas fonético-acústicas incluyen el análisis de la aspiración en oclusivas sordas y de duración y cualidad vocálicas, así como la computación de errores de pronunciación. Las medidas de percepción del acento extranjero fueron proporcionadas por un grupo de oyentes nativos (*n*=20) y otro grupo de oyentes no nativos (*n*=37). Los resultados no arrojan mejoras tras la ES en las medidas de producción vocálica y de aspiración, a la vez que indican una ligera mejora en cuanto a la producción de acento extranjero y un descenso significativo en el número de errores de pronunciación. Se hallaron asimismo correlaciones altas entre las medidas fonético-acústicas y las de percepción del acento extranjero.

Table of contents

List of abbreviations

AH	At home
AIndex	Acoustic index
AOL	Age of onset of learning
B	Bark
B0	F0 in B
B1	F1 in B
B2	F2 in B
CEFR	Common European Framework of Reference
CG	Category-Goodness
CLIL	Content and Language Integrated Learning
CP(H)	Critical Period (Hypothesis)
CV	Consonant/Vowel
dB	Decibel
DD	Duration difference
ED	Euclidian distance
EFL	English as a foreign language
EU	European Union
F0	Fundamental frequency or pitch
F1	First formant
F2	Second formant
FA	Foreign Accent
FI	Formal Instruction
FL	Foreign Language
HO-A	high onset accuracy
Hz	Hertz
IM	Domestic immersion
IP	Input Processing
IPA	International Phonetic Association
JIPA	Journal of the International Phonetic Association
L1	First language
L2	Second Language
LF	Listener factors
LO-A	low onset accuracy
LOR	Length of Residence
LoS	Length of Stay
LREs	Language-related episodes

ms	Millisecond
NA	Non-Assimilable
NL	Native listener
NLM	Native Language Magnet
NNL	Non-native listener
NNS	Non-native speaker
NS	Native speaker
NWS	The North Wind and the Sun
OPI	Oral Proficiency Interview
PAM(-L2)	Perceptual Assimilation Model (L2)
PrEr	Pronunciation errors
RA	Reading Aloud
SA	Study Abroad
SALA	Study Abroad and Language Acquisition
SC	Single-Category
SLA	Second Language Acquisition
SLM	Speech Learning Model
SP	Stimulus properties
SRL	Self-regulated learning
TC	Two-Category
TL	Target language
UC	Uncategorised-Categorised
UIB	Universitat de les Illes Balears
UPF	Universitat Pompeu Fabra
UU	Uncategorised-Uncategorised
VOT	Voice Onset Time

List of figures and tables

Introduction[1]

Study Abroad (SA) programmes have enjoyed an increasing popularity in the last decades worldwide, particularly at the university level. Large numbers of students in areas such as North America, Europe and the Pacific Rim enrol every year in this type of programmes, the duration of which may vary from a few weeks, to a whole term or even a full academic year. The ever-growing popularity of SA is arguably linked to the widespread belief, which has been long held, that an overseas programme has substantial linguistic benefits for students. This belief is based on the assumption that immersion in the target language (TL) community is the best way to acquire the language, due to the opportunities for interaction and the amount and quality of input available in this learning context.

Academic authorities and administrators have played a very active role in the promotion of SA programmes, encouraging their students to go abroad so as to improve their foreign (FL) or second language (L2) proficiency.[2] In fact, a period abroad has been a compulsory requirement for undergraduate language specialists, e.g., in the UK. In these cases, the SA programme is conceived as an integral part of the learners' academic curriculum, and its obligatory nature is a clear indicator of the importance attributed to in-country immersion for L2 development.

Governments have likewise devised foreign language learning policies which aim to facilitate general access to these programmes, such as the

[1] The present study has been funded by grants HUM2007-66053-C02-01/02 and BES-2008-010037 from the Spanish Ministry of Science and Innovation, and conducted within the Competitive research group ALLENCAM (SGR2005-01086/2009-140).

[2] Although scholars have traditionally established a distinction between FL and L2 learning, both terms will be used interchangeably in the present study.

inter-university Erasmus exchange programme within the European context. Created in 1987, the Erasmus programme is part of the EU Lifelong Learning Programme, which has received a budget of nearly €7 billion from 2007 to 2013, and which represents the European commission's commitment to the promotion of multilingualism and mobility across its member states.[3] Hundreds of thousands of students from the different European countries have received an Erasmus grant in order to pursue part of their university studies in a different European country, and Spain, where the present study has been conducted, is one of the countries which have benefited the most from this programme, both in terms of outgoing and incoming students.

In this scenario, the need to empirically assess the actual benefits of SA on learners' L2 development has become evident. An increasing body of research within the field of second language acquisition (SLA) has been thus devoted to this learning context which analyses the effects of SA on the different linguistic skills, as well as the way in which context-specific variables relate to individual variables to account for eventual gains and variation between different SA groups and learners. These studies can provide valuable information in order to inform L2 learning policies and SA programme design. In addition, contributions to this body of research within a European perspective have been particularly called for, as an important part of SA research has been conducted from a North American perspective (Coleman, 1998; Collentine, 2009).

Most likely in contrast with the expectations of the general public, and especially with those of the learners who take part in an SA programme, overviews of the existing SA literature to date do not indicate substantial SA gains for all the different linguistic skills across the board (DeKeyser,

[3] Erasmus mobility scheme website: http://ec.europa.eu/education/lifelong-learning-programme/erasmus_en.htm

2007c; Sanz, forthcoming). Results point to clear benefits in areas such as vocabulary growth or the development of socio-pragmatic skills. Gains have also been found in overall oral proficiency, and especially regarding fluency, which has been one of the most extensively researched areas, and seems to be the one where more progress can be expected as a result of SA.

However, the domain of phonology, which is the focus of the present study, has been the object of relatively little research within the SA literature, and findings so far are inconclusive as to the nature and extent of the changes that can accrue in L2 speech perception and production during a period abroad. This outcome is particularly remarkable, especially if we take into account that one of the main aims of students going abroad is precisely to improve along all aspects of the oral-aural continuum, and with a special emphasis on the development of a more native-like L2 pronunciation, which in FL learners is normally far from native norms as a result of years of exposure to foreign-accented input in a formal instruction (FI) setting at the at home (AH) institutions.

Indeed, research into L2 phonological acquisition in contexts of naturalistic, long-term immersion, has shown that pronunciation is one area of L2 proficiency particularly resistant to change, even in an environment of massive and authentic L2 input exposure. Learners' difficulties in achieving native pronunciation norms are evidenced by a perceptible 'foreign accent' which is the reflection of the learners' L1 phonology. In fact, research into L2 phonological acquisition, which has usually adopted a cross-sectional design, has established that one of the main causes underlying learners' difficulties in acquiring a new L2 phonology is the influence of the already existing L1 phonological system (Best & Tyler, 2007; Flege, 1995).

Taking all this into account, the aim of the present study is to contribute to the under-investigated domain of L2 phonology within the SA literature, in an effort to further our understanding of the type and extent of the benefits that can be expected to accrue during a period abroad. We present the results of a longitudinal, pre-test/post-test design, which assesses the effects of a 3-month SA period on a pool of 23 undergraduate learners of English by means of a multiple-measures approach which differs from previous studies, i.e., we use a set of different measures: acoustic analyses, error rate scores, and listeners judgements of foreign accent.

The dissertation is divided into five main parts or chapters. Chapter 1 presents the theoretical background upon which the present study is built. The first part of the chapter provides a thorough characterisation of the SA learning context and a review of the main findings of SA research, particularly in the domain of L2 phonology. The second part is devoted to the issue of L2 phonological acquisition. First we discuss the acquisition of an L1 phonological system in terms of how it impinges on the subsequent acquisition of an L2 phonology, and we then review some of the most influential models of L2 phonological acquisition which are relevant for the present study.

Once the theoretical background has been established, Chapter 2 provides a brief description of the study and presents the objectives and the research questions which have guided this research. Chapter 3 describes the design of the study, including important information on the context where it took place, and a description of the participants, the data collection procedures and the analyses conducted to address the research questions previously stated. Chapter 4 presents the main findings of our analyses, together with the corresponding discussion in the light of our theoretical background and previous research. The results and discussion

are organised following the research questions. Finally, Chapter 5 outlines the main conclusions and implications derived from our research.

1. Literature review

This chapter provides the theoretical background to the present study. It consists of two main sections. Section 1.1 presents a characterisation and research overview of the study abroad learning context. Section 1.2 gives an account of L2 phonological acquisition. It starts with the acquisition of the native phonology and how it bears on subsequent L2 acquisition, and continues with the presentation of the theories and models of L2 speech learning relevant for the purpose of the present study.

1.1. Study Abroad

There is general agreement amongst researchers that one of the key elements influencing second language acquisition (SLA) is context of learning, an idea which has been particularly emphasised from the field of Study Abroad (Collentine & Freed, 2004; Collentine, 2009; Freed, 1995a). As Collentine notes, "One of the most important variables that affects the nature and the extent to which learners acquire a second language (L2) is the context of learning" (Collentine, 2009: 218), as the interaction between cognitive, sociolinguistic, and pragmatic factors differs across different contexts.

Most SLA research has been devoted to naturalistic or untutored settings of long-term immersion within the L2 or target language (TL) community, and to classroom formal instruction (FI) settings in at home (AH) institutions. The first setting a priori allows for massive exposure to authentic input and limitless opportunities for interaction, whereas the second is characterised by a focus on form and limited quantity and quality of input. Other learning contexts have received also increased

attention during the last decades on the part of SLA researchers. These are the intensive domestic immersion (IM) settings, the Content and Language Integrated Learning (CLIL) programmes, and Study Abroad (SA). Research on SA experienced a boost particularly after Freed's seminal volume *Second Language Acquisition in a Study Abroad Context* (1995a), which provided a state-of-the-art summary of existing research to that date and outlined some of the main theoretical issues to be addressed in future SA research. However, studies addressing the effects of SA seem to be still relatively scarce, especially considering the growing popularity of SA programmes worldwide, and they have traditionally mainly focused on North American settings (Kinginger, 2009; Barquin, 2011).

As defined by Freed (1995a), SA is a learning context which combines "language and/or content learning in a formal classroom setting along with immersion in the native speech community" (p. 5). That is, learners receive classroom instruction which consists mainly of content-based courses taught in the TL (it may include also explicit linguistic training), and at the same time they have access to out-of-class opportunities for practice and interaction in the native-speaking community. This combination is assumed to offer learners rich input and plenty of opportunities for use of the L2, and is presumably conducive to enhanced L2 knowledge.

The interest in the study of SA is due both to its theoretical and socio-pragmatic dimensions. From a theoretical perspective, the analysis of the specific gains, cognitive processes, and individual and context-dependent factors at work in the context of SA constitutes an important contribution to our knowledge of L2 acquisition in general. In this sense, Collentine and Freed (2004) state that: "The study of SLA within and across various contexts of learning forces a broadening of our perspective of the most important variables that affect and impede acquisition in general" (p. 158).

At a socio-pragmatic level, SA programmes have become increasingly popular in the last decades, probably in connection with the importance attached to foreign language (FL) knowledge for mobility in a globalised world and for professional promotion, and due to the widespread assumption that immersion in the TL community has substantial linguistic benefits as a result of the opportunities for interaction and the great amount of quality input provided by this context..

Consequently, SA programmes have been promoted by academic administrators as a means to foster linguistic development, to the extent that a period abroad may be a compulsory requirement for university students with language specialisations, as has been the case, e.g., of the UK (Coleman, 1998). SA programmes have likewise become an important component of governments' FL learning policies (see, e.g., Kinginger, 2009, and Llanes, 2011 for a review of official figures and language programmes). Examples of this type of public-funded programmes are the interprovincial language learning programmes developed in Canada, or the popular European Erasmus exchange programmes which allow European students to study part of their university degree in a different European country, and which are part of the European Union (EU) general policy towards the promotion of multilingualism and mobility amongst its member states. An increasing body of research has been subsequently devoted to empirically assess the actual impact of SA on the different dimensions of learners' L2 linguistic development (see, e.g., Coleman, 1998; DeKeyser, 2007c; DuFon & Churchill, 2006; Sanz, forthcoming, for research overviews; see also section 1.1.2).

As noted by Collentine (2009), this research falls into two main broad categories. Most of the early SA research was conducted with North American learners, whereas an increasing number of studies have focused

on European learners in the last decades, especially within the inter-university Erasmus mobility scheme.[4] Although research in both North America and Europe has tended to analyse university-level learners, American and European SA experiences differ in important ways (Coleman, 1997).

One crucial difference affects learners' L2 pre-departure proficiency level. SA research has analysed American learners with proficiency levels ranging from novice to advanced (Lafford & Collentine, 2006), but most American learners enrolled in SA programmes are usually at the initial stages of acquisition. In contrast, the Erasmus students analysed in European research usually have an upper-intermediate or advanced proficiency level (Collentine, 2009; Barquin, 2011).

Another important difference is related to the design of the SA programme, in the sense that it may influence learners' exposure to TL input and opportunities for interaction. For instance, American programmes are usually sheltered programmes, whereas learners within the Erasmus scheme go abroad in non-sheltered programmes (Kinginger, 2009). American learners thus tend to travel abroad in groups and with tutors who are responsible for planning and supervising their academic programme, and even also part of their out-of-class activities. Erasmus learners, however, usually travel to their host university on their own; they have no tutors supervising their activities while abroad, and they have to take care of finding accommodation and of all the academic and extracurricular activities they engage in. Learners following an American SA programme continue to be mainly within an American educational framework, whereas Erasmus learners are more dependent on local administrative support (Coleman, 1997).

[4] More recently, attention has also been directed to Asia, as Asian learners constitute an increasing percentage of SA learners worldwide (Collentine, 2009).

Nevertheless, despite the differences that may arise across programmes, SA does constitute a learning context with specific characteristics that differ from other contexts, most notably FI, regarding input conditions, interaction, practice, and cognitive and socio-pragmatic factors. The following sections are intended to provide a characterisation of SA, as well as a review of the main SA research findings to date, with a specific focus on the domain of phonology.

1.1.1. Characterisation of Study Abroad

Pérez-Vidal (2011) proposes a characterisation of SA based on three parameters. The author identifies a) the 'macro-level features' of SA, which refer to the external features that define the SA learning context and differentiate it from other contexts, such as FI or CLIL; b) the 'micro-level features' of SA, which consist of the individual learner features that interact with the external features, determining learners' success in their L2 development; and c) the SA 'programme features', or the architecture of the SA programme, i.e., those features that should be taken into consideration when designing a programme abroad. In the following sections these different parameters are described.

1.1.2.1. SA macro-level features

SLA research has identified some specific context-dependent or external factors which are considered central to the characterisation of learning context due to their influence on L2 development, most notably input, interaction, and practice. Macro-level context features can thus be defined on the basis of the particular input conditions of the learning context, the type of interaction the learners can engage in, and the opportunities for practice they can avail themselves of. The following sections present a

discussion of the notions of input, interaction and practice, and how they are related to the SA learning context.

a) Input

Different learning contexts offer different types of exposure to the target language. In this sense, the widely assumed benefits of SA for L2 linguistic development are related to a great extent to the nature of the input available in this learning context. Learners immersed in the target language community may have access to a massive amount of quality input in a large variety of social situations and communicative contexts.

Input is arguably essential for L1 acquisition (see section 1.2.1), and there is general agreement amongst researchers that meaningful input is a very important factor for L2 development. Since the initial works by Krashen (1985) and his notion of 'comprehensible input', the role of input in linguistic progress has been the focus of much research from different approaches (Flege, 2002; Long, 1996; VanPatten & Cadierno, 1993).

Input is a basic concept in one of the most popular current models of L2 speech acquisition, the Speech Learning Model (SLM) by Flege (1995; 2002), who defines L2 speech input as "all L2 vocal utterances the learner has heard and comprehended, including his own", regardless of whether they are well-formed utterances produced by native speakers or utterances produced by non-native speakers of the L2 which might deviate from native norms (Flege, 2009: 175). Researchers have also established a distinction between input and intake. Archibald (2005) notes that input is "the data found in the ambient language", whereas the "subset of this data that is actually taken in and processed by the learner is known as the intake" (p. 292). Bleyhl (2009), defines input and intake along those lines, to further add that "learning then is what the learner makes of intake and

preserves over a certain amount of time for appropriate use in his or her cognitive apparatus" (p.137).

Input is thus widely viewed as the starting point for a set of other processes that should take place in the learner's mind in order for linguistic restructuring and acquisition to accrue. That is, comprehensible input is a necessary condition for improved L2 knowledge, but not a sufficient condition (Larsen-Freeman & Long, 1991; Long, 1996). Input has to be processed in order for L2 acquisition to proceed. Recent trends in research have emphasised the importance of learners' input processing from various perspectives, as noted by VanPatten (2009), with the aim of unravelling how learners process input and which are the cognitive processes that enable them to transform input into intake.

From an input processing (IP) approach (VanPatten & Cadierno, 1993; VanPatten, 2007; VanPatten, 2009), input is language that conveys a meaningful message which is comprehended by the learner. Input processing consists of making use of a series of strategies and cognitive processes that allow learners to comprehend the message by establishing appropriate meaning-to-form connections. IP assumes that L2 learners process input to get meaning first and foremost, a task which would be initially rather effortful, since it would be constrained by the fact that L2 learners have limited working memory and processing capacity as compared with native speakers. This prominence of meaning is encapsulated by the so-called 'Primacy of Content Words Principle', which states that L2 learners will tend to focus on the processing of the content words in the input over any other aspect of form, as these words convey the meaning of the message.

Such content-based processing is very likely to be at play, at least initially, in a communicative and meaning-oriented context such as SA. It is

possible, nonetheless, that "the same input is not processed in the same way by the same learner at different times because the learning process itself is constantly changing" (Verspoor, Lowie, & De Bot, 2009: 63). This would imply that, as learners become more experienced with the L2, they may have more resources freed up in order to attend also to other, more formal aspects in the input to which they are exposed.

Progress in L2 acquisition is also linked to the amount and type of input the learner receives. Since different learning contexts provide different exposure to L2 input, they provide therefore different opportunities for learning. As noted by Moyer (2009), "successful attainment in a second language […] relies on optimal levels of input, quantitatively and qualitatively" (p.168). SLA research has provided evidence of the importance of both input quantity and quality, for instance, on the domain of L2 phonology, which is the focus of the present study (see also section 1.2.2.4).

Input quantity in L2 phonology has been indexed, e.g., as length of residence (LOR) in the target language country, as years of classroom instruction, or as language use patterns (proportion of L1 and L2 use). These measures have been found to have some effect on the development of pronunciation (Díaz-Campos, 2004; Flege & Liu, 2001), but findings are inconclusive (see, e.g., Piske, MacKay, & Flege, 2001 for an overview). This is due to the fact that "in-country residence does not guarantee quality input or interaction," as the learner should "engage in the L2 environment in certain ways, taking advantage of the surrounding input" (Moyer 2009: 162). That is, the quality of the input is also important, and in order to obtain quality input, the learner needs to use the L2 in meaningful communication.

Input quality can thus be understood in connection with actual L2 use and contact with native speakers. Results from some studies indicate that pronunciation is positively affected by regular L2 use with native speakers in informal settings, e.g., to build personal relationships and social networks, and across a multiplicity of domains (Flege & Fletcher, 1992; Moyer, 2004; Purcell & Suter, 1980). Such contact with native speakers across a variety of communicative domains provides access to authentic input and rich opportunities for practice covering different social roles and communicative functions. A learning context such as SA particularly favours these contact patterns and this type of high quality input through the opportunities for interaction of which learners can avail themselves, as discussed in the following section.

b) Interaction

SA is a mainly communicative context which offers learners rich and varied opportunities to engage in meaningful conversational interactions with native speakers in the TL community across multiple social domains. These interactions involve contact with different speakers, using different registers, performing a variety of social roles in different settings also varying in degree of formality and covering a wide variety of topics.

Learners enrolled in an SA programme will need to deal with several academic-related situations, such as attending their content-based lectures, or talking about their academic programme with their tutors and other administrative staff in their host university. But beyond this academic setting, they will also have to deal with all kinds or everyday situations such as looking for accommodation, going to the bank, going shopping, etc., as well as trying to establish a network of personal relationships, all of which is achieved by means of interaction.

Within the field of SLA, it is now widely accepted that there is a link between interaction and L2 learning. The *Interaction Hypothesis* was developed by Long (1983, 1996). It is a combination of the *Input Hypothesis* by Krashen (1985) and the *Output Hypothesis* by Swain (1985; 1993), and it integrates notions such as input, output, interaction or feedback. Interactional approaches posit that interaction leads to L2 acquisition by means of a series of cognitive processes, such as noticing, awareness or attention, which are activated through 'negotiation of meaning' during interaction (Gass & Mackey, 2006, 2007; Gor & Long, 2009; Long, 1983; Long, 1996).

During negotiation for meaning, native speakers (NSs) have been found to deploy a series of input modifications and communicative strategies to address L2 learners. The aim of this modified input, also called 'foreigner register', is to accommodate the language to the needs of the learner, making it more comprehensible and thus facilitating communication. Modifications to the learner-addressed input may consist of simplifications, but linguistic elaboration is more common, and according to Long (1983, 1996) more useful in order to push L2 learners' development forward. Long (1996) describes negotiation for meaning in the following terms:

> "the process in which, in an effort to communicate, learners and competent speakers provide and interpret signals of their own and their interlocutor's perceived comprehension, thus provoking adjustments to linguistic form, conversational structure, message content, or all three, until an acceptable level of understanding is achieved" (p. 418).

As seen in the previous section, input is a crucial element for L2 acquisition, since it provides positive evidence for the well-formed and grammatical structures in the TL. Frequencies of linguistic features, forms and patterns play also an important facilitative role in acquisition (Ellis, N. C., 2002; Gor & Long, 2009). L2 items and patterns which are more

frequent in the input are likely to be more salient and might be therefore more easily noticed and acquired. Exposure to input may likewise result in implicit 'statistical learning' of linguistically relevant regularities (Williams, 2009). This type of learning involves the extraction of statistical regularities present in the input through the activation of implicit learning mechanisms. Humans seem to have powerful mechanisms to detect and extract abstract patterns from linguistic input which can be applied equally to L1 and L2 acquisition. For example, in the domain of phonetics, both infants (Aslin, Saffran, & Newport, 1998) and adults (Saffran, Newport, & Aslin, 1996) have been documented to be able to make use of transitional probabilities between syllables for word recognition.

During interaction learners also receive negative evidence in the form of NSs' feedback to ill-formed or inaccurate structures and forms present in their output, particularly when communication is negatively affected by these deviations from native norms. This feedback may consist of explicit explanations or direct correction of the incorrect form in the learners' output. There are also implicit types of feedback in the form of negotiation strategies including, e.g., confirmation checks, requests for repetition and clarification, or recasts, as well as non-verbal information. Feedback provides thus valuable information regarding the lack of linguistic accuracy in learners' productions and its effects on communication, and plays an important role in drawing learners' attention to problematic aspects in their interlanguage. It is when learners notice the mismatch between the TL input and their own interlanguage representations that restructuring may take place leading to L2 progress.

Output is viewed as an equally important factor for L2 improvement through interaction. As noted by Swain (1985) in the discussion of her Output Hypothesis, output forces learners to go beyond mere

comprehension and semantic processing and pay attention to form in order to produce their own utterances.

There are several functions that output can fulfil to promote acquisition. Production allows learners to notice gaps in their interlanguage systems when they realise mismatches between their output and the input they receive, or when they lack the knowledge to produce the message they need to convey. Following communication breakdowns or misunderstandings signalled through NSs' feedback, output would also serve to formulate and test hypotheses about the L2 system, as learners adjust their language in an effort to produce more native-like output. Production practice is also instrumental in promoting automatisation, which is a necessary factor for L2 proficiency, as explained in the following section. In addition, output can also have a metalinguistic function; within the interactional framework, the construct of 'language-related episodes' (LREs) is used to refer to those situations in which learners consciously discuss and reflect upon their use of language (Swain & Lapkin, 1998). LREs are generally interpreted as a result of a learner's 'noticing the gap' between the L2 system and his interlanguage system.

Notions such as 'noticing' and 'attention' are central to interaction. The importance of noticing for language learning is captured by the *Noticing Hypothesis* posited by Schmidt (1990; 2001), according to which learners need to be aware of the gaps in their language knowledge and to pay attention to the specific aspects of input which can allow them to fill those gaps. As noted by Gass and Mackey (2007), "interaction researchers assume that the cognitive constructs of attention and awareness and the related construct of noticing are part of the interaction-L2 learning process" (p. 187). The processes involved in negotiation for meaning during interactional practice, most notably in the form of NSs' feedback,

are the means whereby learners' attention can be directed to those aspects of language that should be improved, pushing forward L2 progress.

c) Practice

Cognitive approaches to L2 development within recent SLA research, such as skill acquisition theory, have increasingly stressed the important role of practice for progress in L2 acquisition, since it is regarded as a crucial factor for increased automatised language knowledge. From a cognitive perspective, practice is the means by which explicit knowledge can be transformed into implicit, automatised knowledge.

Altarriba and Basnight-Brown (2009) note that "automatic processing […] is considered the cornerstone of the development of proficiency and expertise in an L2" (p. 115). Initial controlled processing, which is required in order to notice the features and patterns of the L2, would develop into automatic processing through extensive practice, repetition and frequency of occurrence and use (Segalowitz & Hulstijn, 2005). This improvement in L2 proficiency would be reflected in a restructuring of the L2 forms which make up the learners' interlanguage system. In addition to these changes at the representational level which might be brought about by input practice in particular, output practice has been acknowledged to be especially important for improvement in specific areas such as pronunciation, which is analysed in the present study (DeKeyser, 2007a; Ellis, R., 1993), since it allows for the automatisation at the motor level of the articulatory gestures involved in L2 production.

Within a skill acquisition theory framework, DeKeyser (2007a) defines practice as "specific activities in the second language, engaged in systematically, deliberately, with the goal of developing knowledge of and skills in the second language" (p.1). Skill acquisition theory traditionally

distinguishes three main stages in skill development: declarative, procedural, and automatic.

In the declarative stage, the L2 learner is presented with explicit information on the structure, rules and patterns of the L2. That is, he acquires 'declarative knowledge' (or 'knowledge that'). In the next step, initial practice allows the learner to act on this declarative knowledge and use it so that it becomes a behavioural routine, or 'procedural knowledge' ('knowledge how'). Once declarative knowledge has turned into procedural knowledge, large amounts of additional practice would result in its automatisation. DeKeyser notes that 'automatisation' is a complex and multi-faceted concept, but in general terms it involves a decrease in "the time required to execute a task (reaction time), the percentage of errors (error rate), and the amount of attention required (and hence interference with/from other tasks)" (DeKeyser, 2007b: 98-99).

This pattern of development in skill acquisition is represented by the 'power law of practice'. This concept refers to the fact documented by skill acquisition researchers that error rate and reaction time in a task both experience a gradual decrease with practice in such a way that the resulting developmental curve follows a mathematical power function, as explained in Newell and Rosenbloom (1981), cited in DeKeyser (2007a). This developmental curve is interpreted as representing an initial and rather quick transition between declarative and procedural knowledge, which is followed by a slower automatisation of the previously acquired procedural knowledge.

DeKeyser argues that SA would be most beneficial for L2 learners if it coincides with the transition between procedural knowledge and automatisation, since this learning context would have the optimal characteristics to provide "the large amount of practice necessary for the

gradual reduction of reaction time, error rate, and interference with other tasks that characterize the automatisation process" (DeKeyser, 2007c: 213).

The notion of 'transfer' is central to this skill-acquisition framework. Transfer of linguistic knowledge and skills is required at two broad levels; transfer from declarative to procedural knowledge, and transfer between different learning contexts. According to DeKeyser, these two types of transfer "are intertwined in study abroad programs" (DeKeyser, 2007a: 9). A minimum declarative knowledge of the L2 system and its initial proceduralisation are necessary for linguistic progress to accrue as a result of SA. In this sense, a period of FI prior to SA would be an ideal setting to provide the learners with this explicit knowledge of language rules and practice for rule proceduralisation. Once an initial functional level of the L2 has been achieved, interactional practice during SA would facilitate the transfer of knowledge and skills acquired in the classroom FI setting to the SA setting in order to complete the proceduralisation process and proceed towards automatisation.

However, there are also limitations to the quantity and quality of actual practice available in an SA context, in such a way that opportunities for L2 practice abroad may not always be as ideal as it is usually assumed. For example, learners may feel tempted to use their L1 in several situations, such as to communicate with other speakers of their L1, because it allows for more effortless interactions, or as a means to cope with cultural shock. Output practice in truly interactional settings may be rather limited, as a great deal of practice may actually be passive and comprehension-based (e.g., attending university lectures, listening to and watching the media, etc.). When learners do engage in interaction, they may not receive appropriate feedback, since native speakers may feel that correcting them will be rude. As a result of this, incorrect interlanguage

forms may become part of learners' speech. In addition, learners' cognitive system may be overloaded due to the pragmatic and social demands of SA. Under the pressure to communicate, it may be difficult for learners to focus on aspects of form and to access their declarative knowledge during on-line processing, and they may end up developing an automatic use of L2 formulas, rather than true automatic use of L2 rules.

Due to all these factors, practice in an SA context may be more limited than initially expected. Therefore, opportunities for optimal practice while abroad will be dependent to a large extent on learners' ability to actively seek for and engage in meaningful interaction with native speakers, which is related to the micro-level learner features discussed in the following section.

1.1.2.2. SA micro-level features

Lafford (2006) emphasises the importance of defining context not only on the basis of the external factors that surround conversational interactions, but also by addressing the learner's "internally-driven factors that affect the individual processes involved in SLA at the micro-level" (pp. 3-4). These internal factors refer to the specific perceptions and expectations attached by individual learners to a particular context of learning, which influence their ability to interact with the environment and take advantage of the type of practice and input available. For example, when engaged in a communicative context, learners would come to have expectations about the roles to be fulfilled by themselves and by their interlocutors, and perceptions regarding the norms of interaction and degree of formality required in each situation, as well as the pragmatic demands of the context.

As noted by Lafford (2006), an SA communicative context has important sociolinguistic and pragmatic implications which lead learners to view language mainly as "a tool to exchange information and participate in important social and interpersonal functions" (p. 4). As already mentioned when discussing the notion of input, in this meaning-oriented context learners are drawn to focus on content in order to comprehend the input they receive and get their own messages across, so as to keep communication going and meet their real world needs. These contextual demands of SA may overload their cognitive system and leave few attentional resources available to focus on form and accuracy. In this situation, a learner's ability to notice L2 forms and integrate them in his developing interlanguage would depend on different individual factors, such as language proficiency level (DeKeyser, 2007c), as well as the learner's "goals and motivations, the learner's perception of the reward system and the definition of successful communication in a given context" (Lafford, 2006: 8).

In their review of the effect of individual differences on L2 acquisition, Dörnyei and Skehan (2003) highlight the fact that learner variables such as aptitude and motivation are good predictors of success in L2 development. According to the authors, motivation is closely linked to what has been termed 'self-regulation' in learning, or 'self-regulated learning' (SRL). Within this paradigm, "researchers attempt to synthesize learner-initiated cognitive, metacognitive, and motivational processes and strategies" (Dörnyei & Skehan, 2003: 612). SRL underscores the importance of the different processes and strategies learners may actively resort to in order to further their L2 development and to maintain sustained motivation for the purpose of improving their linguistic knowledge. These processes and strategies may vary amongst learners,

thus determining the extent to which different learners avail themselves of TL input and contact opportunities in the SA context.

In addition to contextual and individual factors, the architecture of the SA programme may also play a role in the actual gains learners obtain during their experience abroad. Some of the most important features of SA programme design are discussed in the following section.

1.1.2.3. SA programme features

Pérez-Vidal (2011) proposes a tentative classification of some relevant aspects of study abroad that should be carefully considered in the design of an SA academic programme due to their possible influence on SA outcomes. The author identifies eight main variables or programme features, which she divides into two groups, as shown in Table 1.

L2 input and onset level features	**Academic programme features**
Length of Stay (LoS)	Pre-departure preparation
SA living conditions	Point in the curriculum
Opportunities for employment	Academic assignments abroad
Pre-departure language level	Re-entry conditions

Table 1. Features of SA programme design. Adapted from Pérez-Vidal (2011).

One group lists a set of features related to the design of the SA programme as part of the broad curriculum within the learners' academic studies. These academic programme features include a) learners' pre-departure preparation (e.g., by attending a specific preparatory module), which might be particularly beneficial if it is oriented to teach the learners strategies in order to foster contact opportunities; b) the point in the curriculum at which SA takes place (e.g., preceding or following a FI

period); c) academic assignments given to the learners while abroad, which might take the form of interventions to facilitate learners' meaningful interaction with NSs, or to foster the learners' proactive attitude towards their learning process or self-regulated learning (e.g., by writing an academic diary); and d) re-entry conditions at the home institution, in the sense of creating AH conditions that further promote the interactional knowledge developed in the SA context so as to avoid a mid-term or long-term recession in SA gains.

The other group lists a set of features related to L2 input conditions and onset or initial language level. As seen in section 1.1.2.1, input quantity and quality is essential for L2 progress. Input conditions features include length of stay (LoS), which is connected with quantity of input, together with SA living conditions and opportunities for employment, which are both connected with quality of input in terms of opportunities for conversational interaction with NSs in the L2. Pre-departure language level refers to learners' onset level or L2 proficiency before the study abroad programme. For instance, some programmes, such as the European Erasmus mobility scheme, require learners to have achieved a specific level of proficiency in the L2 prior to SA (e.g., upper-intermediate). This is linked to the hypothesis that there may be a threshold level in L2 development at which learners would obtain most gains in an SA context, known as the 'threshold hypothesis'.

Scholars such as DeKeyser posit that learners should reach a specific L2 proficiency level in order to fully benefit from an SA period abroad. He states that learners "should have *functional* knowledge of the grammar that is assumed to be known at an intermediate level" (DeKeyser, 2007c: 217). The rationale for this, from a skill acquisition perspective, is to ensure that learners would have a proficiency level that is high enough to enable them to engage in meaningful and frequent interaction, obtaining

thus the large amount of quality practice that is required so as to proceed in the automatisation of the previously acquired procedural knowledge (see section 1.1.2.1 above).

For instance, Brecht, Davidson, and Ginsberg (1995) reported positive correlations between pre-SA grammar and reading scores and SA gains in speaking, reading and listening. They interpreted this outcome as evidence that formal instruction in grammar at the early stages of learning is a "key element in producing expert language learners who will develop the independent capacity to gather and assimilate information and skills on their own through contact with native speakers" (p. 60), resulting in speaking and listening gains for learners at intermediate and advanced proficiency levels.

Similarly, Segalowitz and Freed (2004) examined the relationship between cognitive processing and oral gains. They found that learners with higher pre-departure measures in lexical access speed and efficiency obtained more SA gains in oral performance. The authors concluded that "oral gains may depend, to some extent, on cognitive readiness to benefit from the learning opportunities available" (Segalowitz & Freed, 2004: 194).

These findings have been interpreted as an indication that higher-proficiency learners may be better able to avail themselves of opportunities for extracurricular practice and interaction while abroad, and also more adept at processing the massive amount of L2 input to which they are exposed.

However, results from a considerable body of research have documented that SA appears to be particularly beneficial for learners with lower onset proficiency levels (Freed, 1995b; Klapper & Rees, 2003; Lapkin, Hart, & Swain, 1995; Llanes & Muñoz, 2009; Towell, 2002). Freed (1995b) found

greater fluency gains after SA for learners with a lower-proficiency level. Likewise, Towell, (2002) reported more progress in fluency scores for lower-onset fluency learners as compared with higher-onset fluency learners, although the lower-level learners still lagged behind the higher-level learners after SA. In a study by Klapper and Rees (2003) comparing grammar gains for lower and higher onset level learners, the authors also reported significantly larger gains for the learners who started out their study abroad with lower grammar scores. In this case, the low-onset learners were actually able to catch up with the learners in the high onset group. Lapkin et al. (1995) reported negative correlations between SA gains in all four linguistic skills and pre-departure scores, which were particularly stronger for listening and oral performance, indicating greater gains for those learners with lower initial level.

As noted by DeKeyser (2007c), it is possible that "the more advanced students are indeed the ones that are learning more in the long run and that the weaker students make the quickest progress at the beginning" (p. 212). It has also been hypothesised that results regarding onset-level effects on SA gains could be influenced by the use of instruments such as the Oral Proficiency Interview (OPI),[5] which may better capture the type of gains made by lower-level learners rather than those obtained by higher-level learners (DeKeyser, 2007c; Llanes, 2011). On his part, Collentine (2009) points out that the notion of different proficiency levels may be particularly convenient from the perspective of academic programme design, and also regarding actual linguistic improvement: "There are most

[5] Developed by the American Council on the Teaching of Foreign Languages (ACTFL), the OPI is a "standardised procedure for the global assessment of functional speaking ability. The OPI takes the form of a carefully structured conversation between a trained and certified interviewer and the person whose speaking proficiency is being assessed." Source: http://www.actfl.org/professional-development/certified-proficiency-testing-program/testing-proficiency.

likely specific domains that require a particular developmental threshold for overall gains to occur" (p. 221). All in all, general findings seem to indicate that even within a minimum, functional knowledge of the L2, lower onset level is associated with larger SA gains.

1.1.2. Study Abroad research

The aim of SA research is to account for the nature of the SA experience and empirically assess its widely assumed positive impact on learners' enhanced L2 proficiency. Research has analysed the effects of SA on different linguistic domains, usually as compared with other learning contexts, such as domestic IM and especially FI, in an effort to establish the differential effects that various learning contexts may have on linguistic gains as learners draw on the different opportunities for practice and interaction provided by each learning environment.

For the most part, SA literature has yielded evidence of a positive effect of the study abroad experience on learners' L2 development. However, actual linguistic gains have been found to be related to the contextual, individual, and programme design factors discussed in the previous section, such as contact patterns while abroad, L1 and L2 use, amount and quality of L2 exposure, or onset proficiency level. Recent sociolinguistic trends of research are also analysing the role of social networks on language development during SA (Mitchell et al., 2013; Mitchell, McManus, & Tracy-Ventura, 2013). The interaction of all these factors results in a rather complex picture, 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 widespread belief in the substantial gains brought about by study abroad immersion. As regards the comparison between SA and FI, results are conflicting and do not

consistently support substantially larger gains for the SA experience over other FI programmes (see research overviews in Coleman, 1998; DeKeyser, 2007; DuFon and Churchill, 2006; Freed, 1995a; Sanz, forthcoming).

As already mentioned, results have provided consistent evidence of the beneficial role of SA for most linguistic domains. Lexical development seems to be one of the areas that improves the most (Collentine, 2004; Ife, Vives Boix, & Meara, 2000; Llanes & Muñoz, 2009; Milton & Meara, 1995; Sanz & Grey, 2012). SA gains have also been reported for listening, which is expected to be widely practiced while abroad (Allen & Herron, 2003; Beattie, 2008; Llanes & Muñoz, 2009). Writing seems likewise to be positively affected by SA (Pérez-Vidal & Juan-Garau, 2009; Pérez-Vidal & Barquin, forthcoming; Sasaki, 2004; Sasaki, 2009), despite traditional claims that scarce practice in this skill while abroad might result in lack of gains. Studies on sociolinguistic skills have also yielded results supporting the positive effect of SA on areas such as the development of communication strategies and native speech norms (Lafford, 1995; Regan, 1995), and pragmatic competence regarding, e.g., the acquisition of L2 politeness patterns (Marriott, 1995), or improvement in the use of the L2 address system (Barron, 2006).

In contrast, grammar is an area for which mixed results have been found. Collentine (2004) reported larger gains for a group of learners receiving FI in an American university over an SA group regarding Spanish grammatical features such as verbal tense and mood, use of copulas and subordination marking. DeKeyser (1991) found similar gains in the use of Spanish copulas *ser-estar* by a group of American learners AH and a group taking part in a six-month SA programme. However, findings in Howard (2005; 2006) suggested that a group of Irish learners of French

who went abroad made more progress in past time marking and agreement than a group receiving FI.

Oral performance has traditionally been considered the linguistic domain most likely to improve as a result of SA, since it is assumed to be the one most widely practiced, and research findings in general have supported this view. General findings indicate that SA is more beneficial than FI for development in overall L2 speaking competence (Collentine & Freed, 2004; Lapkin et al., 1995; Segalowitz & Freed, 2004), as measured mainly by the OPI. Extensive research has also documented gains in L2 learners' oral fluency, which seems to be the domain that develops the most as a result of SA (Freed, 1995b; Freed, Dewey, Segalowitz, & Halter, 2004; Juan-Garau & Pérez-Vidal, 2007; Serrano, Tragant, & Llanes, 2012; Trenchs-Parera, 2009; Valls-Ferrer, 2011). However, studies analysing phonological development during SA are scarce and have provided inconclusive evidence. These studies tend to focus on very specific features of the L2 phonological inventory. Measures are usually limited either to acoustic measurements or phonetic analyses of a small set of segments in the L2 phonological inventory, whereas listeners' perception of degree of accentedness in the learners' L2 speech has been particularly under-investigated.

For instance, Simões (1996) found no consistent gains in the production of vowel quality by a small group of L2 Spanish learners following a short-term programme in Latin-America. Similar results were reported in a preliminary study by Avello (2010a), who failed to find a significant improvement in the production of English vowel quality and duration by a small sample of learners after a 3-month SA programme. In a subsequent study with an analogous group of EFL learners, Avello (2011) also failed to report significant improvement in the learners' speech production regarding perceived degree of accentedness. In contrast, Sanz, Morales-

Front, Nagle and Moorman (2013) reported significant improvement in the production of Spanish plosives for their American learners after a 6-week intensive programme in Spain.

Some studies have compared the impact of SA and FI in AH institutions on learners' gains. Díaz-Campos (2004) found a positive effect of both learning contexts on the production of Spanish word-initial voiced plosives and word-final laterals by two groups of English learners, although development towards native-like patterns was found to be stronger in the FI group. Conversely, in a later study examining possible effects of style on the pronunciation of the same segments (Díaz-Campos, 2006), the author found a superiority of SA learners over FI learners when the register was conversational and informal. Højen (2003) found that the utterances produced by a group of Danish learners of English after SA presented a significantly lower degree of foreign accent as a function of LoS (average=7.1 months), whereas no change was experienced by a similar group of AH learners. However, no difference was observed between the two groups regarding perception and production at the segmental level, as both presented a similar categorisation of English /ʃ/, and neither showed improvement in the production of the English contrast /ɒ-ʌ/.

Other studies have explored the combined effects of FI and SA on learners' phonological development. Mora (2008) examined the production of English voiceless plosives by bilingual Spanish/Catalan learners after a two-term FI period at their home university, and after a subsequent three-month SA term abroad. He found no effect of FI on VOT duration, but a slight, non-significant improvement was observed after SA. This contrasts with the outcome of a study with a similar design, but a different population (Pérez-Vidal, Juan-Garau, & Mora, 2011), in which significant improvement was found in the perception and

production of English vowel and consonant minimal pairs after FI, but not after SA. Lord (2010) analysed the production of Spanish plosives by a group of English native speakers who went on an SA to Mexico following an FI course on phonetics, and compared them with another group participating in the same SA programme, but with no previous FI on phonetics. The author concluded that it was the combination of the two learning contexts that was most beneficial for phonological development, rather than either context in isolation.

As we can see, the picture regarding the acquisition of the L2 phonology during SA is far from clear, especially in comparison with other areas such as fluency, a scenario that calls for further research in this domain. We turn now to the second part of this first chapter, which addresses the issue of L2 phonological acquisition and how it is influenced by the previous acquisition of an L1 phonological system.

1.2. Phonological acquisition

The acquisition of a second language beyond early childhood is always difficult due to the influence of the native language. But perhaps the domain of phonology is the one where the difficulties facing the L2 learner become more obvious, as evidenced by what is normally recognisable as a 'foreign accent' which contains evident traces of the learner's L1 at all the different levels (subphonemic, phonemic, suprasegmental). This section addresses the complexity of the L2 speech learning process, with an emphasis on acquisition at the segmental level, which is the main focus of the present study. Section 1.2.1 offers an overview of L1 phonological acquisition and its crucial implications for the subsequent acquisition of an L2 sound system. Section 1.2.2 provides a description of some of the more influential theories and models which

have informed research in the field of L2 speech learning, and which are relevant for the aims of the present study.

1.2.1. Attunement to the L1 phonetic system

It seems well established that in L1 phonological acquisition speech production follows perception (Ohala, 2008). Before infants' vocal tract and neuro-motor mechanisms involved in speech are developed enough for the production of language, their auditory system already displays an innate and remarkably high sensitivity to the acoustic properties of human speech which is not initially limited to the ambient language, but extends also to unfamiliar languages. Infants are born with a sophisticated perceptual system and are likewise endowed with the mechanisms necessary to identify characteristic patterns of speech and decode the component parts of the speech signal, from words to syllables and down to individual sounds.

Studies have provided evidence that newborns are able to attend to prosodic features of speech in order to differentiate their mother's voice from the voices of other speakers (DeCasper & Fifer, 1980), as well as to discriminate between their native language and other languages (Mehler et al., 1988). Infants are also apt to discriminate segmental speech sounds on the basis of fine-grained acoustic information, such as differences in voicing (Eimas, Siqueland, Jusczyk, & Vigorito, 1971), place of articulation (Eimas, 1974; Morse, 1972), or differences in vowel formant structure and transitional cues (Trehub, 1973) and in third formant transitions for liquids /r-l/ (Eimas, 1975). More importantly, in doing so they exhibit adult-like categorical perception, i.e., they are able to discriminate pairs of stimuli that cross boundaries between categories, but fail to discriminate those that fall within the same category For example,

Eimas et al. (Eimas et al., 1971) found that infants as young as 1 and 4 months are sensitive to the phonetic differences in voicing that cue the phonological contrast between /p/ and /b/, and that peaks in discrimination coincided with adult phonemic boundaries. Categorical perception has been demonstrated by adults and infants also in non-speech stimuli (Cutting & Rosner, 1974; Jusczyk, Rossner, Cutting, Foard, & Smith, 1977), and is likewise exhibited by other species, such as monkeys (Kuhl, 1991). Kuhl and Iverson (1995) note that categorical perceptual ability is a general auditory processing mechanism which has played an important role in the development of the human language sound systems.

1.2.1.1. Language-specific effects in perception

A large body of research indicates that infants are born with a language-general capacity to discriminate different phonetic categories of human speech varying across every acoustic dimension, regardless of whether they are used in their native language or not (Aslin, Pisoni, Hennessy, & Perey, 1981; Best, McRoberts, & Sithole, 1988). In contrast, adults display language-specific perception, which means that their perception of speech is affected by their native language (Lisker & Abramson, 1970), and they have great difficulty in discriminating non-native phonetic distinctions, particularly in cases of similarity between native and non-native categories (see section 1.2.2). This suggests that the initial perceptual acuity which allows infants to discriminate non-native speech contrasts decreases as a function of native language experience. Researchers have explored when and how this L1-induced decrease in perceptual sensitivity takes place throughout human linguistic development.

Werker and Tees (Werker & Tees, 1984a) explored English infants' ability to discriminate Hindi retroflex-dental /ʈa/-/ta/ and Salish velar-uvular /k̀i/-/q̀i/ consonant contrasts by analysing three groups of infants aged 6-8 months, 8-10 months, and 10-12 months. Results indicated a gradual decline in perceptual ability across the three age groups. Whereas infants aged 6-8 months were able to correctly discriminate the non-native sounds, infants of up to 10-12 months already showed a clear decline in their sensitivity to non-native phonetic contrasts. This pattern of results was replicated in a study using a longitudinal design (Werker & Tees, 1984a). Similarly, Polka and Werker (1994) examined the perception of German long and short front-back rounded vowel contrasts /y/-/u/ and /y:/-/u:/ by English infants aged 4 months, 6-8 months, and 10-12 months. According to the results, the 4-month-old infants showed high sensitivity to the non-native vowel contrasts. Sensitivity was reduced for the infants aged 6-8 months, but still significantly better than for the infants aged 10-12 months, for whom a perceptual decline was observed in their discrimination of the non-native vowels.

Findings from these and similar studies were interpreted as revealing a decline in perceptual capabilities within the first year of life, which signals a shift from language-general to language-specific perception due to L1 experience, in such a way that a rather robust L1-attunement can be observed at around 12 months. This L1-attunement involves that perception is adjusted early in life to the contrastive features in the native phonology, whereas those acoustic features which are not relevant in the native system would not be attended to.

Results from other studies are consistent with the view that the observed decline in infants' perceptual ability is attributable to attunement to their L1 phonetic system, rather than to a deterioration of the human auditory system or sensorineural loss. Adults can improve their discrimination of

non-native contrasts with proper training or practice (Logan, Lively, & Pisoni, 1991; Tees & Werker, 1984), and they have been found to perceive non-native phonetic distinctions in specific testing conditions with demands that differ from those characteristic of natural language processing.

For example, Werker and Tees (1984b) found that adult native speakers of English were able to discriminate non-native Hindi and Salish consonant contrasts when presented with stimuli which did not sound like speech, but not when the stimuli were full, speech-like syllables. This suggested to the authors that the decrease in sensitivity to non-native contrasts may be related to the way in which listeners process the speech signal. The authors posited the existence of a 'phonetic' processing level, which would correspond to "natural phonetic boundaries", versus a 'phonemic' level, "corresponding to native language boundaries" (Werker & Tees, 1984b: 1876). When a phonemic processing mode is induced, listeners would only access the acoustic information which is phonologically relevant in their L1. However, when adopting a 'phonetic' mode, they may have access to acoustic information with no distinctive function in their L1, and may thus be able to discriminate non-native contrasts.

The Native Language Magnet (NLM) model developed by Kuhl and colleagues offers an explanation of how L1 experience affects the perception of speech throughout the first year of life, as well as its subsequent production (Iverson & Kuhl, 1995; Kuhl & Iverson, 1995). According to this model, infants are born with innate and universal psychophysical boundaries that allow them to perceive distinctions between phonetic categories in a language-general manner. Exposure to their native language would result in the creation of language-specific category representations due to changes in the perceived distances

between the original phonetic categories. The mechanism that triggers these changes is the 'perceptual magnet effect'.

An L1 prototypical sound is a good instance of a category, and it acts as a 'perceptual magnet' for the other instances of the category, *attracting* them toward itself. This magnet effect causes a shrinking or warping of perceived distances within the acoustic space, so that discrimination around the prototype is reduced. In contrast, poor instances of the category, or nonprototypes, do not act as attractors, and discrimination around them is fairly good. Research has revealed native-language magnet effects as a result of specific L1 experience in both adults and infants as young as 6 months (Iverson & Kuhl, 1995; Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Kuhl, 1994). The magnet effect thus "alters the mechanisms underlying speech perception", and this subsequently affects "both the perception of spoken language and its production" (Kuhl & Iverson, 1995: 121-122). Yet as noted above, these changes do not seem to reflect alterations to general auditory or sensorineural ability, but seem to occur rather "at a higher level, one that involves memory and/or attention" (Kuhl & Iverson, 1995: 142).

The acquisition of L1 magnet effects is predicted to have considerable implications both for L1 and L2 acquisition, as it brings about dramatic and permanent changes to the way in which language is processed. The magnet effect could therefore provide an explanation for the difficulties documented in adults' categorisation of non-native sounds, particularly regarding the problems facing learners of a second language (Aslin et al., 1981; Best et al., 1988).

Learners are thus assumed to perceive the L2 through an L1-altered perceptual space, in such a way that the categorisation of the same non-native sounds differs amongst speakers from different L1 backgrounds as

the native magnets interfere in their discrimination of the L2 phonetic distinctions. Success in the categorisation of L2 sounds depends on their proximity to the L1 prototypes. L2 sounds which are close to an L1 prototype will be difficult to discriminate since the L1 prototype will act as an attractor, whereas sounds which are far from the prototype will be less likely to be attracted to it and could thus be discriminated more easily. This accounts for the fact that, in L2 phonological acquisition, those sounds that are "similar to a category in the adult's native language are more difficult to perceive as different from the native sound, sounds not similar to a native-language category are relatively easy" (Kuhl & Iverson, 1995: 142).

1.2.1.2. Language-specific effects in production

With the development of the articulators and other motor mechanisms involved in speech production, and as infants begin to gain increased control over their speech apparatus, babbling emerges. At around 6-8 months, infants are able to produce adult-like syllables with recognisable prosodic patterns, first in reduplicated consonant-vowel (CV) sequences, and later on in alternating CV sequences (Moyer, 2004; Ohala, 2008). The first stages of acquisition may also present onsetless syllables depending on the frequency of models in the native language, such as the case of languages like Spanish (Morales-Front, 2006). This roughly coincides with the period that signals the beginning of L1-attunement in perception. In fact, changes in speech production seem to run parallel to changes in perception. Infants begin to produce their first words at around 12 months, which again coincides with the period at which evidence of robust attunement to the L1 system can be observed. The first words produced by infants usually follow a simple CV syllable structure, and the different

categories in the infant's native phonological inventory gradually emerge as acquisition proceeds.

Similarly to the development in perception, speech production during these early stages of acquisition appears to reflect both universal, language-general patterns, as well as ambient language influence (de Boysson-Bardies, Halle, Sagart, & Durand, 1989; Lee, Davis, & MacNeilage, 2010; Oller & Eilers, 1982). Infants' output during babbling seems to follow similar tendencies across languages, probably as a result of constraints "based on the structure of the vocal tract as well as immature speech motor control" (Lee et al., 2010: 294). The general frequency patterns which have been observed in cross-linguistic babbling indicate higher percentages of occurrence for plosives and nasals in consonant production, and a predominance of vowels produced in the lower left quadrant of the vowel space (Lee et al., 2010).

Developmental patterns of segmental acquisition are also similar across different languages, and seem to have an articulatory basis in terms of the relative difficulty involved in the production of the different sounds. Some sounds appear to be particularly easy to articulate and are thus acquired at the early stages. This would be the case of plosives and nasals /p, t, k, b, d, g, m, n/, which also happen to be amongst the most common sounds in human language, although different languages may vary, e.g., in the way they implement the voicing feature in plosives. Other sounds, however, such as the fricatives /Θ, ð/, or the different realisations of the liquid /r/, seem to require more complex articulatory gestures, and are acquired at later stages.

Despite these cross-linguistic similarities, the influence of the native language can also be observed in infants' early speech productions. Results from de Boysson-Bardies, Sagart, and Durand (1984) showed that

competent adult French listeners were able to distinguish the babbling of French infants from the babbling of Chinese and Arabic infants at 6 and 8 months, which the authors attributed to the prosodic characteristics of the babbling samples. Consonant and vowel productions in infants as young as 10 months already reflect the characteristic features of the ambient language, as shown, e.g., through the analysis of vowel formants (de Boysson-Bardies et al., 1989). Similarly, L1 token frequency seems to affect the order and rate of acquisition of individual sounds and their presence and frequency of occurrence in infants' speech (Lee et al., 2010), and the same has been observed for syllable structure frequencies (Morales-Front, 2006).

Phonological acquisition proceeds as infants continue to extract regularities from the input they are exposed to, so that knowledge of the L1 phonological system at all levels of structure is apparent both in perception and production at around 18-20 months. Progress in L1 phonological competence continues with increased attunement to the phonetic and phonological features of the L1 throughout the third and fourth years. Around this time, characteristic errors in production also tend to decrease with the maturation of the child's vocal apparatus and the automatisation of the articulatory gestures involved in L1 production. Adult-like competence in the production of the whole L1 inventory and in the perception of native and non-native sounds is achieved by about 8 years, and it is further facilitated by the acquisition of reading and the enhancement of phonological awareness that results from the orthography-onto-sound mapping. The result after this whole process is a robust perceptual attunement to the L1 and a strong automatisation of the gestural movements required for its production, both of which will have an important influence on the later acquisition of a second or foreign language.

1.2.2. L2 speech development

One clear characteristic of L2 speech is that it tends to be noticeably foreign-accented. The fact that a learner's L1 can usually be identified through his accented speech has led researchers to emphasise the role of L1 influence or transfer in the domain of L2 pronunciation.

This is reflected, for instance, in the notion of the L1 acting like a "sieve" in the categorisation of L2 sounds (Trubetskoy, 1969). That is, the L1 is hypothesised to act like a filter through which the L2 is perceived, in such a way that L2 sounds would be associated with the already existing and well-established sounds of the L1 phonetic system based on phonetic similarity, resulting in wrong categorisation of the elements that make up the L2 system (in line with the NLM effect described in section 1.2.1.1).

Different models have been developed in order to address how L1 experience reduces listeners' discrimination sensitivity to non-L1 phonological contrasts, impacting their categorisation of new, non-native sounds, and in what sense the interaction between learners' native and non-native phonetic systems bears on their L2 phonological development. Two of the most influential models have been Best's Perceptual Assimilation Model (PAM), and Flege's Speech Learning Model (SLM). Both models are presented in the following sections.

1.2.2.1. Perceptual Assimilation Model and L2 learning

Best's (1994, 1995) Perceptual Assimilation Model was initially developed as a model of cross-linguistic perception in order to assess naïve listeners perception of non-native or unfamiliar speech sounds. In its recent version, PAM-L2 (Best & Tyler, 2007), the model has been modified so as to extend the principles underpinning the perception of

non-native speech sounds in monolingual subjects to the acquisition of an L2 phonological system by bilingual subjects.

The model was built upon the well-established observation that listeners have great difficulty in the categorisation of non-native phonological contrasts which are based on phonetic features not used contrastively in their L1 (Polka, 1992; Polka & Werker, 1994; Strange, Akahane-Yamada, Kubo, Trent, & Nishi, 2001). However, success in the discrimination of non-native contrasts seems to vary depending on the phonetic characteristics of the native and non-native sound system inventories, specifically regarding the degree of similarity or dissimilarity between native and non-native phones. Within the PAM framework, it is posited that non-native categories tend to be assimilated to the existing native categories in such a way that specific predictions can be made about the categorisation of non-native contrasts on the basis of the type of assimilation taking place between the native and non-native sounds.

In line with the postulates of Articulatory Phonology (Goldstein & Fowler, 2003), PAM claims that the units serving as the primitives for language phonology are the articulatory gestures involved in the production of speech, instead of, e.g., more abstract representations encoding acoustic information present in the speech signal (as is the case of Flege's SLM). Non-native sounds are thus perceived in connection with and assimilated to the native sounds which are more similar articulatorily. When the non-native sound is perceived as an exemplar of a native category, it is *categorised*, although its status in terms of category goodness may vary from being considered a *good* exemplar to a *poor* or *deviant* exemplar of the native category. The non-native sound may also be *uncategorised* if it is not assimilated to any specific native category because it fails to be perceived as an exemplar of any of them, while still being considered a speech sound. There are even cases when a non-native

sound is *non-assimilated* to any native category at all, as it is perceived as a non-linguistic sound, i.e., it is not considered a speech event, but rather some sort of noise. The following specific patterns of assimilation are postulated for non-native contrasts according to how each non-native sound is mapped onto the L1 system:

Two-Category (TC): each individual non-native sound is assimilated to two different native categories. Discrimination is thus predicted to be very good or excellent.

Category-Goodness (CG): the two non-native sounds in a contrast are perceived as exemplars of the same native category, but they differ regarding goodness of fit to that single category. Discrimination is expected to be intermediate.

Single-Category (SC): the two non-native sounds are assimilated to a single native category of which they are considered equally good or deviant exemplars, as opposed to the CG assimilation type. Discrimination in this case is predicted to be rather poor.

Uncategorised-Uncategorised (UU): neither non-native category is assimilated to any native category, even though both are perceived as speech sounds. Expected discrimination varies from poor to moderately good, or even very good, depending on the phonetic differences between the two contrasting sounds, and with regards to the more similar native categories.

Uncategorised-Categorised (UC): one non-native sound is perceived as and assimilated to a native category, whereas the other fails to be categorised. These contrasting sounds are expected to be discriminated very well since they are not considered exemplars of a single native category.

Non-Assimilable (NA): neither non-native category is perceived as a speech sound, so none of them is assimilated to a native category. In these cases, discrimination is predicted to be good or excellent, since the non-native sounds are not encoded phonetically, allowing the listener to attend to the acoustic properties which characterise each sound from a purely auditory processing level.

PAM has proven to be a useful tool for the study of cross-linguistic perception, and findings from several studies are consistent with its predictions and postulates (Best, 1990; Best & Strange, 1992; Halle, Best, & Levitt, 1999). Naïve or non-native listeners have typically been conceived of as monolingual speakers, or 'functional monolinguals'. This is taken to imply that they are not familiar with the language for which they are tested, and that they are not in the process of actively acquiring an L2. Best and Tyler (2007) note that this does not discard "passive exposure to a language other than the L1", particularly in the form of "classroom-only instruction with instructors who have a strong L1 accent" (p. 34). In contrast, L2 learners would be those who are in the process of acquiring an L2 in a mainly communicative setting, as further detailed below.

Best (1995) points out that the approach of PAM "assumes that perceptual learning continues into adulthood" (p. 198) as a result of adjustments to listeners' perception caused by increased experience with the non-native system. In this sense, Best and Tyler (2007), in the revision of the model in order to address its implications for L2 perceptual learning (PAM-L2), stress the idea that it is conceived to account for the effects of the L1 system on non-native perception both at the phonetic and phonological levels.

The interaction between these two levels is particularly relevant from the perspective of L2 learning since, as noted by Best and Tyler (2007), "the phonological level is central to the perception of L2 speech by SL learners […] in a way that it cannot be for L2-naïve listeners perceiving unfamiliar non-native speech" (p. 23). The authors go on to explain how in order for perceptual learning to accrue, L2 learners need to distinguish the L2 phonetic features that are phonologised, i.e., they need to acquire the higher-order, phonological distinctions of the L2 system which are mapped onto the functional forms of the language and which allow, therefore, for further L2 development through the establishment of form-meaning connections. Naïve listeners perceiving a non-native language, in contrast, are only cognizant of the phonological differences which constitute their native phonemic inventory, but they cannot relate the phonetic and phonological levels in an unfamiliar language.

PAM-L2 thus extends the assimilation types established for PAM to the L1-L2 interaction that takes place at the two levels (phonetic and phonological), in order to predict perceptual learning of L2 contrastive categories. For example, for cases when two distinct L2 categories are considered to be good exemplars of the same L1 category (SC L2 assimilation), it is expected that learners will initially assimilate the two L2 categories to the single L1 category at the phonetic and phonological levels, so that differences between L2 contrasting lexical items would not be perceived. In order for new separate phonological categories to be created successfully, learners would need first to develop phonetic categories for the L2 sounds, which could be possible if they are able to distinguish the phonetic properties of at least one of the L2 contrasting categories as a result of increased L2 input.

Best and Tyler (2007) point out that perceptual L2 learning in cases of SC assimilation would be influenced by communicative factors bearing on

learners' ability to distinguish the contrasting lexical items or minimal pairs differentiated through the L2 phonological opposition. If the minimal pairs are high-frequency words, and the phonological opposition is productive in the L2, giving rise to a rather large number of contrasting lexical items which form a dense phonological neighbourhood, communicative needs may push learners to acquire that contrast as they may perceive it as necessary in order to achieve fluent conversational interaction. Conversely, learners might consider that homophony would not be particularly detrimental for communication if the minimal pairs are low-frequency words, or if they come from a sparse phonological neighbourhood. In this sense, studies have found that vocabulary size may be linked to L2 segmental perception (Bundgaard-Nielsen, Best, & Tyler, 2011; Fullana, Miralpeix, & MacKay, 2012).

A different scenario would result from those cases when only one of the contrasting L2 categories is assimilated to an L1 category (UC L2 assimilation). Learners are then predicted to successfully distinguish L2 contrasting lexical items. If the assimilated L2 sound is considered a good exemplar of the L1 category, learners are expected to equate the L1 and L2 categories not only phonologically but also at the phonetic level. The development of a separate phonetic category with better-adjusted L2 values would still be possible in conditions of increased L2 exposure, although rather unlikely due to the high degree of L1-L2 perceived phonetic similarity. It could also be the case that the assimilated L2 sound is perceived as a poor L1 exemplar, even though the L1 and L2 categories might still be equated phonologically. Best and Tyler (2007) give the example of the phoneme /r/ in French and English. L1 English learners of French usually identify English /r/ with French /r/ phonologically, since the two phonemes present similar patterns of distribution and fulfil the same linguistic function in English and French, together with the fact that

they have the same orthographic representation in the two languages. At the phonetic level, however, English [ɹ] (frictionless continuant) and French [ʁ] (uvular fricative) are rather different from each other, and they are likely to be differentiated as two separate phonetic categories. Once learners are aware of the different phonetic properties of each of them, they may develop a new French [ʁ] phonetic category. English [ɹ] and French [ʁ] would thus constitute two different phonetic realisations of the single phonological category /r/.

The principles of PAM can thus be extended to L2 learning within the PAM-L2 framework, and the generated predictions and hypotheses can likewise be empirically tested in order to complement the findings on L2 speech learning that result from research conducted within other approaches such as Flege's SLM (see section 1.2.2.2). Up to now, general findings on L2 perceptual learning seem to be in line with results reported for monolingual non-native perception regarding, for instance, learners' problems in the discrimination of L2 contrasts which do not exist in their L1, or which are difficult to categorise due to similarities with the L1 system.

For example, many studies have documented poor discrimination of the English /r-l/ consonant contrast by L1 Korean and Japanese learners, whose native languages do not contain this phonological opposition (Yamada, 1995). Studies on the perception of L2 vowels have likewise provided evidence of learners' difficulties for L2 vowel categorisation. Rochet (1995) found that English and Brazilian Portuguese learners differed in their categorisation of French /y/, a front rounded vowel which does not exist in either language, and which English learners tend to identify with English /u/ whereas Brazilian learners tend to identify it with Portuguese /i/. Studies have also shown that L1 Spanish/Catalan learners have great difficulty in distinguishing the English vowel contrasts /i:-ɪ/

and /æ-ʌ/ (Cebrián, 2006; Cebrián, 2007; Flege, Bohn, & Jang, 1997), as their L1 systems have only one vowel in the phonological space covered by the two contrasting L2 vowels.

PAM-L2 views L2 learners mainly as late learners, or learners who start to learn an L2 once they have acquired their L1, displaying thus robust and adult-like language-specific attunement to the phonetic and phonological categories of their native system, similarly to monolinguals' patterns of native attunement. But in contrast with the description for 'functional monolinguals' provided above, Best and Tyler (2007) characterise L2 learners as "people who are in the process of *actively learning* an L2 to achieve functional, communicative goals" (p. 16). That is, L2 learning is assumed to take place in a naturalistic immersion context and during meaningful conversational interaction with native speakers of the L2, who are expected to provide correct models for the L2 categories, as opposed to the sometimes impoverished and even incorrect models which might be available in a formal instruction setting, where input is limited and usually foreign-accented. L2 learning in a communicative setting has also been the focus of research within the SLM approach discussed in section 1.2.2.2 below.

L2 learners are considered to differ from naïve non-native listeners along other factors. They have experience or exposure to L2 input, and are therefore familiar with the syllabic, phonotactic and phonological patterns of the language they are learning. They are expected to have contact with native speakers of the L2 through conversational interaction. This would allow them to have experience both in the perception and production of the L2 distinctive contrasts. Besides, face-to-face interaction would also allow them to have access not only to the auditory properties of the L2 sounds, but also to visual information regarding the motor routines and gestural constellations involved in their articulation (which are considered

the primitives of phonology in PAM). Finally, and as touched upon above, L2 lexical acquisition might have an influence on phonological learning, since it could facilitate learners' phonologisation of contrastive features distinguishing minimal pairs, a process which could be pushed forward as a result of learners' need for efficient communication

1.2.2.2. The Speech Learning Model

One of the most popular models of L2 phonological acquisition is the Speech Learning Model by Flege (Flege, 1995; Flege, 2002) The model posits a series of postulates and hypotheses in order to account for the difficulties facing L2 learners' perception and production of L2 sounds, and the processes that influence ultimate attainment in L2 phonological acquisition. SLM focuses specifically on perception and production at the segmental level, and has proven to be very useful, and therefore widely favoured, in informing and guiding L2 speech research, as it allows for the generation of testable predictions and can be used to discuss and interpret large and varied sets of empirical data.

A considerable body of research has analysed the factors affecting individual variance in L2 speech, and has revealed that age is arguably the factor which has the greatest impact on L2 ultimate attainment (see also section 1.2.2.4). In general, it seems that for L2 native or near native-like competence to be achieved in the different linguistic domains, and particularly in phonology, "earlier is better" (Larsen-Freeman & Long, 1991). Late bilinguals (those who come into contact with the L2 around puberty or later) have been consistently found to speak their L2 with stronger foreign accents than early bilinguals (those who come into contact with the L2 within early childhood), and to differ from L2 monolingual native speakers in their production and perception of L2

sounds more than early bilinguals do (Flege, Munro, & Mackay, 1995; MacKay, Flege, Piske, & Schirru, 2001; MacKay, Meador, & Flege, 2001; Munro et al., 1996; Munro, Flege, & Mackay, 1996; Piske, Flege, MacKay, & Meador, 2002).

A common explanation for these oft-observed age effects on L2 speech is that L2 acquisition is constrained by a biologically-based critical period (first posited for L2 acquisition by Lenneberg, 1967). According to this critical period hypothesis (CPH), the neurological mechanisms which make L1 acquisition possible would be less efficient, or even non-operative, after the end of the critical period, in such a way that native-like L2 speech learning would be no longer possible due to the loss of brain plasticity caused by age-related neurological maturation. Different ages have been hypothesised for the end of this critical period, traditionally coinciding with puberty, e.g., at around 12 years (Scovel, 1988), or 15 years (Patkowski, 1990). Authors such as DeKeyser (2000) have proposed that the effects of neurological maturation on L2 acquisition can be observed as early as around 6-7 years of age.

However, Flege (1995; 2002) claims that there are reasons to be cautious in accepting an explanation to age effects in L2 phonological acquisition based solely on the existence of a putative critical period associated with neurological maturation. As noted above, there is no clear consensus amongst researchers on the beginning and end of the critical period. There is evidence that some late bilinguals who start to learn an L2 after the end of the hypothesised critical period may be able to perform within native-like production and perception patterns (Birdsong, 2007; Bongaerts, Planken, & Schils, 1995; Bongaerts, van Summeren, Planken, & Schils, 1997). Conversely, early bilinguals who start to learn an L2 in early childhood have often been found to present divergences in phonological performance from monolingual native speakers of the TL (Flege,

MacKay, & Meador, 1999; Guion, Flege, & Loftin, 2000; MacKay, Meador et al., 2001; Piske et al., 2001).

Hence, the SLM developed by Flege calls into question the existence of a critical period for L2 acquisition, and offers an alternative account for the age-related difficulties underlying the phonological acquisition of an L2 on the basis of four postulates and seven hypotheses (see Flege, 1995: 239). These postulates and hypotheses have been derived from empirical data found in experimental studies carried out by Flege and colleagues (Bohn & Flege, 1990; Flege & Eefting, 1987; Flege, 1993; Flege, Schirru, & MacKay, 2003). According to this alternative account, the differences between native and L2 speech production would be due to the interaction between the learners' L1 and L2 systems, and to input conditions.

Contrary to the CPH, SLM posits that the mechanisms and processes used for the acquisition of the L1 remain accessible over the life span and can also be used in the acquisition of an L2 (or any other language learned after the mother tongue). The distinctive features that characterise the sounds of the L1 and L2 inventories consist of acoustic-phonetic cues extracted from the speech signal input, and they form abstract, long-term memory representations called 'phonetic categories'. These phonetic categories act as perceptual targets that guide learners' production, so that incorrect categorisation of an L2 sound would result in inaccurate production. In this sense, the model contends that L2 production errors have mostly a perceptual basis, as inaccurate perception leads to incorrect representations of the L2 sounds. However, it is not ruled out that some L2 production errors should have other causes, e.g., related to output constraints at the motor level.

An L2 phonetic category may differ from the native category if learners fail to perceive an acoustic feature which has contrastive relevance in the

L2 but not in the L1, or which exists in both languages but with a different status or allophonic distribution. Incorrect categorisation may also result from the fact that learners may weigh acoustic cues differently from native speakers of the TL. All these problems affect L2 perception since learners tend to perceive L2 sounds through the 'filter' or 'sieve' of their L1 phonetic system (Archibald, 2005; Best, 1995; Trubetskoy, 1969).

Learners' ability to create a new L2 category depends on the perceived degree of dissimilarity between the L2 and the phonetically closest L1 sounds. That is, L2 learners must perceive at least some phonetic distance between two similar L1 and L2 sounds for a new L2 category to be created. SLM claims that increased experience in the L2, in terms of input quantity and quality, may allow learners to distinguish the differences between L1 and L2 sounds, leading to L2 category formation. Once new L2 categories have been established, L2 segmental production accuracy can be improved. Improvement in perception would lead thus to improvement in production. If learners fail to perceive the phonetic differences between the L1 and L2 sounds, then new category formation will be blocked.

The model predicts that the greater the perceived phonetic dissimilarity between the L1 and L2 sounds is, the more likely it is to be discerned, and to result therefore in new L2 category formation. This implies that the creation of new L2 categories will be more difficult in those cases in which there is great phonetic similarity between the L1 and L2 sounds, whereas phonetic dissimilarity between L1 and L2 sounds will facilitate L2 category formation. The likelihood to discern these cross-language phonetic differences is assumed to decrease as age increases (see section 1.2.1).

According to the model, therefore, success in the creation of new L2 categories will be influenced by perceived phonetic similarity and learners' age. In addition to these two main factors, exposure to L2 input and patterns of L1-L2 use are also hypothesised to play a role. For example, when an L2 sound is similar to an already existing L1 sound, the likelihood of category formation will increase if L2 learning begins in early childhood and there is large exposure to quality input, together with a high use of the L2, but it will decrease exponentially as age of L2 learning increases, particularly if exposure to L2 input and L2 use are low. Conversely, for an L2 sound which is clearly dissimilar to any sound in the L1 phonetic system, category formation could be possible even if L2 learning takes place in adulthood, provided that there is a sufficient amount of L2 exposure and a low L1 use.

The L1 and L2 categories are assumed to coexist in a common phonological space. SLM posits that the subsystems of the two languages will interact and influence each other within this shared space. Unlike the earlier view of 'cross-language interference' as the one-directional influence of the L1 on the L2 (Lado, 1957), and in line with Grosjean (1989), interaction here is considered to be bi-directional. That is, not only will the L1 subsystem influence the L2, but the L2 subsystem will also affect the L1. The extent to which each language will influence the other is related to language dominance, in a way that "there might be a stronger influence of the L1 on the L2 for late bilinguals, but a stronger influence of the L2 on the L1 for early bilinguals" (Flege, 2002: 222).

According to the model, the phonetic categories in the L1 and L2 subsystems interact by means of the mechanisms of 'category assimilation' and 'category dissimilation', depending on whether a new L2 category has been created or not. Foreign-accented L2 production

would be the result of this interaction between the subsets of L1 and L2 categories.

Results from some studies indicate that the L1 system continues to develop during childhood and becomes more stable as the speaker grows into adolescence (Hazan & Barrett, 1999; Nittrouer, 2002), exerting a greater influence on the perception of non-native sounds. Thus, as the L1 categories become stronger, L2 sounds are more likely to be linked or equated to the closest similar-sounding L1 sounds through the process of 'equivalence classification' or category assimilation (Best, 1995). That is, the L2 vowel or consonant is considered to be a realisation of an already-established L1 category. When this category assimilation mechanism operates, the creation of new L2 categories is blocked and production of L2 sounds is then expected to differ from L2 monolinguals' norms.

Category assimilation would take place when the L1 and L2 sounds are very close to each other within the phonetic space. Yet if the two sounds are not identical phonetically, learners may be sensitive to the sub-categorical differences between them. In this case, SLM predicts that L2 learning would still be possible, even in the absence of new L2 category formation, through the development of 'merged' L1-L2 categories, or 'diaphones'. This happens when an existing abstract representation is gradually modified to represent the phonetic characteristics of the L1 and L2 sounds which have been perceptually linked, resulting in values that differ from typical monolingual L1 and L2 phonetic category representation. The merged category will be subsequently used in the perception and production of the L1 and L2 sounds which are subsumed under it.

When the learner is able to create a new category for an L2 sound, either because it is distant enough from any other L1 sound, or because it does

not exist in the L1, category dissimilation may occur as the L1 and L2 categories in the shared L1-L2 phonetic space tend to be dispersed in order to maintain the contrast between the two phonetic subsystems. L1 and L2 categories which are relatively close may thus deflect away from each other, in which case both of them are expected to differ from native-like categories. A learner's newly established L2 category may likewise differ from native norms if it is the result of weighing phonetic cues differently from L2 native speakers.

Flege (1995, 2002) stresses the role of the input conditions learners are exposed to in triggering the changes that affect category organisation and representation in their phonetic systems. He points out that the age effects observed in L2 speech acquisition might be due to the fact that age seems to be usually confounded with other variables such as L2 exposure, or amount of L1 and L2 use (see section 1.2.2.4). In this sense, the common differences between early and late L2 learners could be related to age-dependent variation in quantity and quality of L2 input, rather than to maturational issues, as social and demographic factors might allow early learners to receive richer and more adequate input than late learners. What underlies this explanation is the basic tenet of the SLM regarding the malleability of the learners' phonetic system given the right input conditions. That is, L2 speech acquisition may be possible regardless of age provided that a massive amount of authentic TL input is available so as to prompt category creation and modification.

Input conditions, in terms of amount and quality, can be defined in relation with different learning contexts, such as FI and SA. FI in AH settings is typically characterised by limited L2 input exposure, which is usually restricted to the classroom and often foreign-accented. SA, on the other hand, offers the possibility of massive exposure to native and rich L2 input for L2 acquisition. SA could thus be expected to have a positive

effect on learners' ability to distinguish differences between similar L1 and L2 sounds which could have been previously equated, resulting in modifications to their phonetic system which might lead to the creation of new L2 phonetic categories as a previous step for subsequent improvement in production accuracy.

Several studies conducted within the SLM framework have supported the main postulates and hypotheses of the model. For example, findings in Flege (1987) have provided evidence of category assimilation and the development of merged L1-L2 categories for plosives. Flege analysed the production of the voiceless plosive /t/ by two groups of English/French and French/English bilinguals. Voiceless plosives in English and French differ mainly regarding 'voice onset time' (VOT), which is the time lapse between the release burst of the plosive and the beginning of vocal fold vibration. VOT is long-lag (aspirated) in English, and short-lag (unaspirated) in French.

Results from this study indicated that the learners in the two groups produced VOT values in their L2 which failed to adjust both to the L1 and the L2 norms, but were rather intermediate between the two. Interestingly, when producing VOT in their L1, learners also failed to reflect monolinguals' norms, producing instead similarly intermediate L1-L2 values. These results were interpreted as indicating that the existing L1 category had evolved as a result of the dual source input in order to reflect the properties of the /t/ tokens in French and English. This type of cross-language influence has been commonly found in other studies analysing VOT production (Flege, Frieda, Walley, & Randazza, 1998; Mora, 2008; Wrembel, 2011; 2013).

Studies for vowels have similarly produced evidence of L1-L2 phonetic interaction. Flege, Schirru and MacKay (2003) examined L1 Italian early

and late learners' production of the English diphthong /eɪ/, which is usually equated with the Italian monophthong /e/ by native Italian speakers. Results showed a tendency for the early bilinguals to produce English /eɪ/ with significantly more tongue movement than monolingual native speakers of English, whereas the group of late bilinguals produced /eɪ/ with less tongue movement than the English native speakers. The longer-than-native average movement of the early learners was taken as a signal of dissimilation in their effort to maintain the contrast between the phonetic category created for the L2 sound /eɪ/ and the L1 category /e/. Conversely, the short movement produced by the late learners was attributed to the development of a merged category between English /eɪ/ and Italian /e/ as a result of their inability to create a new L2 category for the L2 vowel.

To sum up, SLM challenges the view that L2 acquisition is constrained by a critical period for language learning which would be particularly noticeable in the domain of speech perception and production. Instead, SLM posits that the mechanisms underlying language learning remain accessible for L2 learning over the life span, and proposes therefore two alternative explanations for the commonly found divergences in L2 perception and production from native norms. One explanation is related to the interaction between the L1 and L2 phonetic systems, with the L1 exerting a stronger influence as L1 categories become stronger throughout childhood and adolescence. The other explanation is related to L2 input conditions, which are considered crucial in promoting L2 category formation. Amount and quality of L2 input, nonetheless, may be subjected to age-related variation (with early child learners receiving more authentic input than adult L2 learners), and may also vary between different contexts such as FI (where L2 input is rather limited) and SA (characterised by immersion in a rich L2 input setting).

1.2.2.3. The role of L1 transfer and universals

Results from the research reviewed so far suggest that restructuring in L2 learners' phonological system as they progress in the acquisition of difficult L2 phonological contrasts may be possible under conditions of increased L2 experience. An important issue addressed by researchers is the role of L1 transfer and universals in the acquisition of L2 phonology, particularly regarding to what extent L2 learners may be able to make use of L2 features which are not exploited contrastively in their L1. It has been established that individuals develop language-specific perception early during L1 acquisition, in such a way that they learn to attend to and weigh those features which have a distinctive function in their L1 system through a process of L1 perceptual attunement (see section 1.2.1). In order to be able to make use of new L2 features absent in their L1, learners would thus need to re-attune their phonetic system to the L2-specific features.

Different authors have different views concerning the difficulty involved in this re-attunement process based on the importance attributed to the role of L1 transfer and of universal or language-independent factors. Some hypotheses which are of relevance for the current study are reviewed below.

a) The Feature Hypothesis

Based on the hypotheses and postulates of Flege's SLM (see section 1.2.2.2), McAllister, Flege, and Piske (2002) stated the so-called 'feature hypothesis', which underscores the role of the L1 phonetic system in learners' acquisition of new L2-specific features not exploited in their L1. This hypothesis states that "L2 features not used to signal phonological contrast in L1 will be difficult to perceive for the L2 learner and this

difficulty will be reflected in the learner's production of the contrast based on this feature" (McAllister et al. 2002: 230). Successful acquisition of an L2 contrast would thus depend on the existence of the feature upon which it is built in the L1 system. If a contrastive L2 feature is likewise used to signal L1 phonological distinctions, this would facilitate the creation of new L2 categories conducive to the acquisition of the contrast. Conversely, if the L2 feature has little relevance in the L1 phonetic structure, or is absent from its phonological patterns, the acquisition of the L2 contrast would be more difficult, as the creation of new L2 phonetic categories would be eventually blocked.

Therefore, in the light of this hypothesis, learners would not be expected to become readily sensitised to new L2-specific features not exploited in their L1. It is this mismatch between L1 and L2 distinctive features that would negatively affect the creation of L2 phonetic categories, reducing learners' ability to correctly perceive and produce L2 phonological contrasts. The feature hypothesis represents a softer position regarding the implications of the L1 system for phonological learning as compared, e.g., with the 'deficit hypothesis' (Brown, 2000), which posits that a feature that is absent in the L1 will not be apt to be acquired by the L2 learner. In order to test the feature hypothesis, McAllister et al. (2002) explored the acquisition of quantity distinctions in Swedish mid /ø:-ø/, /ɛ:-ɛ/ and non-mid /ʉ:-ʉ/, /a:-a/ vowel contrasts by native speakers of Estonian, English and Spanish, which are three languages that vary as regards the use and relevance of durational features to signal phonological oppositions.

Similarly to Swedish, duration is a prominent feature in the phonology of Estonian, affecting both vowels and consonants. In English, some vowel contrasts present duration differences which are concomitant with the spectral properties of the vowels, but these differences in vowel length are regarded as phonologically irrelevant, since vowel quality is considered

the main distinctive feature in these phonological oppositions. In the case of Spanish, duration does not have any distinctive function, and seems to play no role as a cue for segmental categorisation in its phonological system. Quantity distinctions can thus be argued to be more salient or relevant in Estonian, they seem to play an intermediate role in English, and are least relevant in Spanish. In accordance with the feature hypothesis, McAllister et al. (2002) predicted that the acquisition of Swedish vowel duration contrasts would be highly successful for the native speakers of Estonian, who are sensitised to phonological duration differences in their L1, whereas acquisition would be difficult for the native speakers of English and Spanish, since duration is not exploited phonologically in either language.

Results indicated that the L1 English learners were better at acquiring the Swedish vowel quantity contrasts than the L1 Spanish learners, and the L1 Estonian learners were better than the English learners. This suggested to the authors the possibility to refine the feature hypothesis and reformulate it in the form of a 'feature prominence hypothesis' according to which "the relative importance of a feature in the L1 will determine the extent to which the feature is successfully used in producing and perceiving phonological contrasts in the L2" (McAllister et al. 2002: 254). That is, learners' access to specific L2 phonetic cues will be mediated by previous degree of L1 experience with them.

b) The Desensitization Hypothesis

In contrast with the position adopted by proponents of the feature hypothesis, Bohn (1995) takes a different stance that de-emphasises the role of the L1 phonological system in the observed systematic differences between learners' and native speakers' speech perception and production. He stresses the fact that, while the influence of L1-attunement is well-

documented in cross-language speech perception and L2 phonological acquisition, research has likewise revealed that L1 transfer does not always explain the deviating patterns characteristic of non-native speech. As some authors have noted (Eckman, 2008; Major, 2001; Wode, 1981), these deviating patterns reflect also the influence of 'universals' or general, language-independent perception and production strategies.

For example, Wode (1981) discusses the existence of universal processing and production strategies underlying both L1 and L2 acquisition. Major (2001) posits three main sources of influence on L2 speech learning, namely: L1 factors, L2 factors, and universals (those phenomena which cannot be traced back neither to the L1 nor the L2). In his Ontogeny and Phylogeny model, Major (2001) tries to account for the relative influence of each of them throughout the L2 learning process. He contends that L1 factors play a more important role in the early stages of acquisition, whereas the influence of universals gradually increases as acquisition proceeds, and then decreases as knowledge of the L2 increases. Eckman (2008) develops the notion of 'markedness' or typological universals in L2 phonology, which is based on the principle that particular segments, syllable structures or phonotactic patterns are more common across natural languages than others, forming a markedness hierarchy. Those forms which are more common are considered to be less marked and easier to be acquired, and those which are less common would be more marked and more difficult to acquire.

In keeping with these approaches, Bohn (1995) proposes that, in order to categorise particular non-L1 vowel contrasts, non-native speakers would resort to general, language-independent auditory strategies, showing a tendency to make use of specific acoustic cues which may be particularly salient or have enhanced auditory impact, regardless of their experience with them in their L1. That is, under this view, it is posited that certain

cues, rather than others, will be "easy to access whether or not listeners have had specific L1 experience with them" (Bohn, 1995: 280).

In a series of experiments on the categorisation of vowel contrasts, Bohn and colleagues (Bohn, 1995; Bohn & Flege, 1990) explored the perception of English vowels by native speakers of German, Mandarin and Spanish. While vowel duration in German is a more important cue to signal phonological distinctions than it is in English (Bohn & Flege, 1992; Strange, 2007), vowel contrasts in Spanish are only signalled through spectral differences, but not duration (Quilis, 1993). As for Mandarin, duration seems to be used to differentiate tonal contrasts, but it is not a cue to segmental contrasts either (Bohn, 1995). In order to conduct these experiments, two vowel continua were created ranging from /i/ to /ɪ/ and from /ɛ/ to /æ/ which varied in duration and formant frequency values. Results indicated that, whereas native English speakers attended mainly to spectral cues for the identification of the English vowels, the learners of German, Mandarin and Spanish, all relied more on duration than on spectral differences, irrespective of the differences between the three languages regarding the use of duration as a distinctive feature.

These findings were not consistent with an explanation based on L1 transfer, and they led Bohn to formulate the 'desensitization hypothesis'. This hypothesis is based on a general, language-independent principle, which is stated as follows:

> "whenever spectral differences are insufficient to differentiate vowel contrasts because previous linguistic experience did not sensitize listeners to these spectral differences, duration differences will be used to differentiate the non-native vowel contrast" (Bohn, 1995: 294-295).

That is, when non-native speakers are not sensitised to specific L2 spectral features due to the fact that they are not exploited contrastively in their native vowel system, they will resort to differences in duration, even if

their L1 does not make use of duration either to signal segmental contrasts. For example, certain vowels in the English vowel system are located rather close within specific areas of the vowel space and are thus acoustically very similar, so that native speakers of English have become sensitised to fine-grained spectral differences in those areas. That is the case, for instance, of vowels /i/ and /ɪ/, which occupy the high-front portion of the vowel space. In contrast, the vowel systems of Mandarin and Spanish present only one vowel in the area occupied by English /i/ and /ɪ/. Therefore, native speakers of Mandarin and Spanish did not need to develop linguistic sensitivity to spectral differences in those areas.

In the absence of readily available spectral cues that might enable them to differentiate these vowels, both Mandarin and Spanish learners of English seem to be able to attend to duration as a distinctive cue, even though neither Mandarin nor Spanish make use of duration as a segmental distinctive feature. Hence, such reliance on the non-L1 feature of duration suggests the use of some language-independent auditory strategy for speech perception, based most likely on the particular saliency of duration as an acoustic cue, rather than the transference of a specific native perceptual mechanism.

The desensitization hypothesis is therefore built on the principle that some phonetic features of the speech signal are more salient auditorily than others. This would be in line with the postulates represented, e.g., by the 'redeployment hypothesis', proposed by Archibald (2005; 2009) in response to the deficit hypothesis mentioned above (Brown, 200), and which underscores the importance of cue saliency and robustness for L2 phonological acquisition based on Wright's work on phonetic cue robustness (Wright, 2004). According to the redeployment hypothesis, the knowledge and features of the L1 phonological system can be redeployed and re-assembled for the acquisition of the L2 phonology, and at the same

time learners would be able to acquire L2 features which do not exist in the L1 as long as they are encoded by robust acoustic cues.

1.2.2.4. Foreign accent

An important body of research into L2 phonological acquisition has examined the phenomenon of *foreign accent* (FA, also referred to as *accentedness* in the literature). These studies have been conducted mainly with immigrant populations in learning contexts of long-term immersion in the TL community, and they have usually analysed the contribution of different individual and context-dependent variables to the perception of FA in L2 learners' speech, most notably age of onset of L2 learning and L2 experience.

As pointed out by Munro (2008), interest in the study of foreign-accented speech has been motivated by theoretical as well as pragmatic reasons. From a theoretical perspective, research into the phenomenon of FA is of relevance regarding general issues in second language acquisition, such as the interplay between the processes involved in L2 perception and production (Best, 1995; Flege, 1995), the nature of age-related constraints on L2 acquisition in connection with a hypothesised critical period and regarding issues related to ultimate L2 attainment (Long, 1990; Patkowski, 1990), or the role of universals and cross-linguistic influence in L2 phonology (Eckman, 2008; Major, 2001).

From a pragmatic perspective, a better understanding of which specific features of L2 speech contribute more to a foreign accent may inform useful actions to be taken for efficient approaches in the teaching of L2 pronunciation (Piske et al., 2001). Piske (2007) further notes that, despite the differences between immigrant learner populations and learners in a foreign language classroom setting, both learner groups present also

similarities, in such a way that "factors that have a significant influence on immigrants' success in learning an L2 might also provide important indications as to how the effectiveness of foreign language teaching could be increased" (Piske, 2007: 302). A foreign accent may also be relevant from a social perspective, as it may have negative social consequences and affect communication (see section c) below).

a) Characterisation of a foreign accent

Foreign accent has been described, for instance, as "the extent to which an L2 learner's speech is perceived to differ from native speaker (NS) norms" (Munro & Derwing, 1998: 160). It has also been characterised as "non-pathological speech produced by second language learners which differs in partially systematic ways from the speech characteristic of native speakers of a given dialect" (Munro, 1998: 139). In his seminal work providing a full account of his SLM for L2 phonological acquisition, Flege (1995) wrote about foreign accent that "Listeners hear foreign accents when they detect divergences from English phonetic norms along a wide range of segmental and suprasegmental (i.e., prosodic) dimensions" (p. 233).

These different descriptions stress the nature of FA as a perceptual phenomenon related to listeners' processing of L2 speech productions. In this sense, a foreign accent is the result of perceived differences between specific properties of L2 speech and the patterns that characterise native speakers' norms; i.e., it is the perceptual correlate of objective, acoustic-phonetic features present in L2 learners' pronunciation. As pointed out by Flege (1995), these L2-specific features can take place both at the segmental level (e.g., divergences from the range of native-like acoustic values, or number and severity of pronunciation errors), and at the suprasegmental level (in terms of stress, rhythm and intonation patterns

which differ from native norms), as it has been evidenced through findings from accent detection studies.

Listeners have been found to be highly sensitive to divergences from native patterns in L2 oral productions even on the basis of very short speech samples, down to individual segments or just parts of a segment (Flege, 1984). For instance, several studies have found significant correlations between listeners' assessments of foreign accent and segmental error counts consisting of phonemic substitutions, deletions and insertions, and such error counts are often found to be good predictors of degree of accentedness (Avello, Mora, & Pérez-Vidal, 2012; Brennan & Brennan, 1981; Cunningham-Andersson & Engstrand, 1989; Munro & Derwing, 1995a). These findings provide evidence of listeners' ability to recognise differences from native phonological norms in L2 learners' speech at the distinctive, phonemic level.

Moreover, some studies have shown that native listeners are likewise sensitive to non-distinctive phonetic differences in L2 learners' realisations of TL phonemes. Flege (1984) presented native English listeners with speech samples from native speakers and native French speakers of L2 English which differed in duration, from whole phrases to just parts of segments corresponding to /t/, /u/, and /i/ tokens. The listeners were found to correctly identify the non-native speakers regardless of stimuli duration, even when presented with just short, 30-ms samples of syllable onset /t/ tokens. Similarly, Flege and Hammond (1982) found that, in order to mimic Spanish-accented English, native speakers of English not only produced phonological substitutions, but also altered phonetic features such as VOT duration in voiceless plosives and syllable-final lengthening. Taken together, the results reported in these studies have shown that segmental divergences from native-like norms, both at the level of distinctive, phonemic differences, as well as at the level of

non-distinctive, subphonemic differences, are salient to native listeners and contribute to the perception of a foreign accent.

Research has also explored the role of suprasegmental phenomena in the perception of L2 accentedness. For example, Anderson-Hsieh, Johnson and Kohler (1992) and Munro and Derwing (1999) identified variables representing measures of prosodic features, such as non-L1 rhythm or stress patterns, which presented strong correlations and high predictive power regarding degree of foreign accent. Munro (1995) reported that native English listeners were able to identify Mandarin speakers when presented with low-pass filtered English speech samples in which segmental information was almost non-existent, but which contained prosodic information that could have been used for accent detection in the absence of more salient segmental cues. In a study similarly designed to analyse prosody and accent detection, Van Els and DeBot (1987) manipulated native and L2 Dutch speech samples by removing pitch variations. The authors found that the loss of these prosodic properties resulted in a lower correct identification of native versus non-native speakers by Dutch listeners. All these findings thus provide evidence that a foreign accent can also be perceived through non-native prosody.

b) Factors affecting foreign accent

Results from the literature examining the phenomenon of foreign accent to date indicate that perceived accentedness is influenced by a number of factors which appear to differ in terms of their impact on L2 pronunciation, and which interact in a complex manner, to the extent that more than often they have been confounded. Researchers have tried therefore to assess the specific contribution to overall degree of FA of these different factors, which include age of L2 learning, L2 exposure, input quality and quantity, and patterns of L1 and L2 language use.

Age of onset of learning (AOL) has been the most examined factor in the FA literature. It normally refers to learners' age of first massive exposure to the target language in a predominantly L2 speaking context, and has been traditionally operationalised as age of arrival in an L2 country. This variable has been strongly connected to research examining the possible existence of the hypothesised 'critical' or 'sensitive' period for the acquisition of a second language, since the domain of pronunciation seems to be particularly subject to incomplete acquisition both for adult and adolescent learners who start learning an L2 beyond early infancy (Long, 1990; Oyama, 1976). Even as these learners progress in their L2, a foreign accent is still normally perceptible which is considered a characteristic feature of their L2 speech, as shown by results from many FA studies which have found that higher AOL is associated with a higher degree of perceived accentedness (Asher & García, 1969; Flege, 1988; Flege & Fletcher, 1992; Oyama, 1976).

As touched upon in section 1.2.2.2 dealing with Flege's SLM, this oft-observed lack of native-like performance in L2 learners' pronunciation has been taken as evidence supporting the CPH which some authors postulate for language acquisition (Lenneberg, 1967; Patkowski, 1990; Scovel, 1988). Other authors prefer to talk instead of a 'sensitive period' for the successful mastery of L2 pronunciation (Long, 1990; Oyama, 1976), after which acquisition would be imperfect and irregular. Proponents of these hypotheses posit biological and maturational constraints on L2 acquisition, in such a way that native-like L2 phonological performance beyond the hypothesised critical or sensitive period for language acquisition, which is generally considered to end around puberty, would no longer be possible due to these age-related constraints, leading to the emergence of a clearly perceptible foreign accent in the L2 learner's speech.

However, results from some studies have shown that adult learners may indeed be able to acquire native-like pronunciation (Birdsong, 2007; Bongaerts et al., 1995; Bongaerts et al., 1997). Conversely, studies such as Flege, Frieda and Nozawa, (1997), or Guion, Flege and Loftin (2000) have also shown that an early age of L2 acquisition (as low as 3.2 years in the former, and around 6 years in the latter) does not guarantee accent-free pronunciation. These results tend to be, nonetheless, an exception. In general, research has revealed a gradual increase in L2 learners' foreign accent as age of acquisition increases, AOL being usually an important predictor of degree of accentedness (Flege, 1988; Flege & Fletcher, 1992; Flege, Munro, & Mackay, 1995). This finding points towards a linear relationship between age and degree of foreign accent, instead of the sharp discontinuity around adolescence that the CPH would imply.

Despite the clear influence of AOL on L2 phonological development, findings reported in the FA literature as a whole indicate that, even though it appears that 'the earlier the better' for L2 pronunciation, early L2 acquisition may not be a sufficient condition for total native-like mastery of the L2. This has led authors such as Flege to call into question the CPH and to posit factors other than biological age-related constraints as the cause of foreign accent, most notably social and context-dependent factors regarding type of L2 exposure (in terms of amount and quality of input), or patterns of L1 and L2 use.

L2 experience or amount of L2 exposure has been the second most studied factor considered to influence degree of accentedness. Since most studies on foreign-accented L2 speech have analysed long-term immersion contexts, this variable has been typically indexed as length of residence (LOR) in the L2 country. Research assessing the importance of LOR for L2 pronunciation has yielded somehow mixed results. Studies such as Asher and García (1969), Flege and Fletcher (1982), or Flege et al. (1995)

have found that LOR had an impact on foreign accent, although it was not as important a predictor of L2 accentedness as AOL. In contrast, other studies, such as Flege (1988), or Oyama (1976) have failed to report a significant LOR effect.

Piske et al. (2001) interpreted these results on the basis of differences in learners' proficiency level. As noted by Flege (1988), LOR might have an initial and rapid effect on the pronunciation of early, low-level learners (i.e., those who are at an initial learning stage), but in the case of more experienced, high-level learners, further LOR exposure would be unlikely to result in significant FA reduction. Results from other studies (Riney & Flege, 1998) have supported the view that LOR effects on perceived foreign accent depend on learners' being or not at an early stage of L2 acquisition.

Piske (2007) provided another explanation for the contradictory results regarding LOR effects on pronunciation based on a study by Flege and Liu (2001) exploring L2 competence in English as a function of LOR and L2 exposure. In this study, the authors compared the effect of LOR on two groups of Chinese learners of English, a group of non-students (1.7 versus 6.6 years) and a group of students (2.5 versus 7.3 years). The two groups were considered to differ in terms of amount of exposure to authentic input from native speakers, as the students group were assumed to be exposed to a predominantly L2-speaking environment in which they received a substantial amount of native input, whereas the non-students group had a more limited contact with native speakers and were, therefore, exposed to a reduced amount of authentic input. LOR effects were observed for the students group, but not for the non-students group. Flege and Liu (2001) concluded that LOR in itself may not be an accurate index of L2 exposure, and that amount of contact with native speakers should be taken into account as well.

In line with this claim, Højen (2003) found a strong correlation between LOR and FA gains in L2 English during a period abroad, but the correlation was stronger between FA gains and an overall measure of total input which combined LOR and self-reported use of English while abroad. These findings underscore the importance of substantial and authentic, high-quality L2 input for the language learning process. It seems, therefore, that additional time spent in an L2 context would not be a sufficient condition for pronunciation improvement, and that amount and quality of L2 input should also be considered, as stressed by Piske (2007): "progress in learning an L2 is dependent on both the quantity and the quality of the L2 input L2 learners receive" (p. 306).

Finally, several studies have found that language use patterns are also an influential factor on L2 pronunciation, especially regarding learners' use of their L1 while immersed in the L2 learning environment (Flege, Munro, & MacKay, 1995; Flege et al., 1997; Flege, Yeni-Komshian, & Liu, 1999; Piske et al., 2001). In these studies, L1 and L2 use have been normally evaluated by means of self-assessment questionnaires in which learners have to estimate, for instance, the amount of time they spend using their L1 and L2 in different contexts, amount of contact with L2 native speakers, or L1 and L2 proficiency. Results in Flege et al. (1995) revealed that language use patterns constituted a significant predictor of foreign accent ratings for Italian learners of L2 English, explaining 15% of the total variance. In a follow-up study (Flege et al., 1997), the role of L1 use was further explored by creating two groups of early Italian/English bilinguals who were AOL-matched (around 6 years), but who differed in percentage of L1 use (3% and 36%). The authors reported an L1 use effect as the learners with higher L1 use were perceived to have a significantly stronger FA than the learners with lower L1 use. Similar results were obtained by Flege et al. (1999), who examined Korean

learners of L2 English that were also matched for AOL. They found that those learners who presented a pattern of higher L2 use (English) and lower L1 use (Korean) had significantly better FA scores than those who presented the opposite pattern (lower use of English and higher use of Korean). Results in Piske et al. (2001) showed that the L1 use effect observed for early bilinguals was also extended to late Italian/English bilinguals.

All these studies, conducted mainly in naturalistic contexts of immersion in a TL setting, indicate that, although AOL has been found to be the most influential factor in the development of L2 pronunciation and the best predictor of changes in FA, different pronunciation outcomes for AOL-matched L2 learners may arise which can be attributable to differences in factors such as type and amount of input or patterns of language use (see also section 1.2.2.2).

c) Effects of a foreign accent

FA studies have also been conducted in connection with other dimensions of L2 speech, such as speaking rate, fluency, comprehensibility, or intelligibility, in order to clarify the interaction between these different speech dimensions and how each of them affects listeners' processing of non-native speech, as well as to assess the extent of the influence that the L1 may exert on L2 speech production regarding each of these dimensions., These studies aim to shed light on the best strategies that would facilitate the development of L2 learners' fluent and successful communication in the L2 environment, which is usually the ultimate goal of the language learner in a context of immersion in the target language community, as stressed in section 1.1 when discussing the characteristics of SA.

The relationship of foreign accent with intelligibility and comprehensibility has been of particular interest for researchers. Derwing and Munro (1997) defined intelligibility as "the extent to which the native speaker understands the intended message", and comprehensibility as "judgments on a rating scale of how difficult or easy an utterance is to understand" (p. 2) Intelligibility is thus usually assessed by asking listeners to transcribe oral stimuli, whereas in order to measure comprehensibility listeners are typically asked to make scalar judgments by using the same Likert-type scales utilised to measure FA.

Although research findings suggest that accentedness is to some extent a dimension of L2 speech independent from intelligibility and comprehensibility, and that a foreign accent does not necessarily impede effective communication in the L2 context (Derwing & Munro, 1997; Munro & Derwing, 1995a), some authors have nonetheless noted that FA may indeed reduce L2 speech intelligibility (Flege, 1988; Munro, 2008). Munro, for instance, pointed out that "it is widely recognized that L2 users at times have difficulty making themselves understood, sometimes because of pronunciation errors that make their speech unintelligible", and added that "a detailed understanding of the situation in which pronunciation errors lead to communication breakdowns has yet to be developed" (Munro, 2008: 197).

Scholars have also pointed out other negative consequences that a foreign accent may have for L2 learners (Derwing & Munro, 2013; Flege, 1988). For example, native listeners in the L2 community may attach negative connotations or prejudices to particular foreign accents. Some studies have analysed to what extent foreign-accented speech may trigger irritation on the part of listeners (Wrembel, 2010). Such prejudices towards accented speech may be a consequence of stereotypes or socio-economic differences between the learner and the native listener. Negative

attitudes towards accented speech on the part of listeners may also arise as a result of difficulties to process and understand heavily-accented L2 speech as compared with native speech. Results from some studies indicate that L2-accented speech might involve extra-processing costs for listeners (Munro & Derwing, 1995b; Weil, 2003), as signalled by the finding of reaction times which are longer for non-native than for native speech, showing that non-native speech may take longer to be understood.

A foreign accent may thus affect the interaction between native listeners and L2 learners in different ways. It may notably reduce the intelligibility or comprehensibility of L2 speech due to divergences from native patterns. Furthermore, this diminished intelligibility may cause increased processing difficulties for listeners, who may lack the patience or willingness to interact with L2 learners, or develop negative attitudes towards foreign-accented speech.

d) Design and methods in FA studies

Research in the field of FA usually adopts the form of experimental studies that tend to focus on naturalistic settings of long-term immersion, such as the case of immigrants from different backgrounds in their L2 communities (Asher & García, 1969; Flege et al., 1999; Guion et al., 2000; Munro, Derwing, & Morton, 2006; Piske et al., 2001; Purcell & Suter, 1980). Most of these studies have a cross-sectional design in which oral data are elicited at a single point in time. Common data elicitation techniques include controlled tasks such as asking participants to read lists of words, sentences or whole paragraphs, or the repetition of recorded speech samples modelled on native speech.

As noted by Munro (2008: 202), the use of these controlled tasks may result in unnatural or "better-than-normal" speech. Samples of extemporaneous speech may thus also be collected so as to obtain more

natural sounding oral data in a less controlled setting, e.g., by means of picture story telling or personal narrative tasks. However, Munro (2008: 202) also notes that the elicitation of controlled data such as read material has the advantage of controlling for grammatical or lexical errors, which are frequently present in extemporaneous speech and may affect listeners' judgements. Another advantage of collecting read material is that it facilitates analyses involving the comparison of groups of speakers (or of the same speakers at different points in time).

Participants in this type of studies typically include a small group of native speakers of the L2, who provide baseline data, and at least one group of L2 learners who differ along one or more of a series of factors, such as AOL, LOR, or L1 and L2 use. These factors act as independent or predictor variables for degree of FA, which is the dependent variable. FA is conceptualised in terms of listeners' judgements of accentedness as measured through some type of scalar procedure, usually by means of a Likert, equal-appearing interval scale (7-point or 9-point scales are most commonly used). Munro (2008) stresses the importance of using listeners' judgements in order to assess L2 speech: "From the standpoint of communication, there is no useful way to assess accentedness […] except through listener responses of some sort" (p. 200).

Listeners' judgements have been traditionally collected from groups of native listeners, whether expert (linguistically trained) or unsophisticated. Despite the focus of FA research on exploring native listeners' perception of L2 speech, some studies have also analysed the perception of accented speech by non-native listeners. Interest in the perception of L2 learners' speech by different groups of listeners is related to concerns regarding to what extent listeners' perception of speech may vary as a function of properties of the speech signal itself, or as a function of listener differences. Munro (2008) addresses this issue and proposes a model to

conceptualise the different dimensions of L2 speech, based on Gass and Varonis (1984) and Varonis and Gass (1982). This model establishes a distinction between stimulus properties (SP component) and listener factors (LF component). A similar response pattern from different groups of listeners would indicate a strong influence of the SP component, and minimal influence of listener-based characteristics represented by the LF component.

The evidence available from the studies analysing the perception of accented speech by groups of listeners with different L1 backgrounds, including native and non-native listeners, seems to indicate that when assessing L2 speech samples stimulus properties are likely to be more relevant than listener factors. Like the bulk of FA research, these studies are typically conducted also in contexts of long-term naturalistic immersion, and not so much in periods of shorter immersion, such as those characteristic of SA learning contexts. For instance, results in Flege (1988) showed that two groups of Chinese-speaking listeners differing in experience provided judgements for Chinese-accented English sentences which paralleled those obtained from English native listeners, the more experienced Chinese group more closely resembling the native English group. MacKay, Flege, & Imai (2006) extended the findings in Flege (1988), as they found that FA ratings of Italian-accented English samples from proficient Arabic listeners strongly correlated with native English listeners' ratings. These findings pointed to the ability of non-native listeners to reliably assess L2 speech for degree of accentedness, even in cases when listeners and learners do not share the same L1 background.

More studies have provided evidence that reactions to accented speech from listeners with different L1 profiles may follow similar patterns. Munro et al. (2006) found that FA scores provided by four groups of listeners with different L1s yielded moderate to strong correlations as they

assessed English samples produced by L2 learners with those same L1s (the same correlations were obtained for comprehensibility and intelligibility). Similarly, Derwing and Munro (2013) obtained strong correlations between ratings from native English listeners and from proficient non-native listeners differing in their L1 profile, and concluded that both native and non-native listeners provided equally reliable evaluations of L2 learners' speech. Taken as a whole, the outcomes from these studies indicate that proficient non-native listeners are able to accurately and reliably assess degree of accentedness in L2 speech, providing FA ratings that closely match native listeners' ratings.

2. Objectives and research questions

This chapter presents the objectives which have motivated the research carried out in the present study, and the research questions herein addressed.

2.1. Objectives

Building on the research findings and theoretical frameworks presented in the previous chapter, the present study aims at contributing to a better understanding of the under-investigated effects of SA on L2 learners' phonological development in speech production. In order to do so, it analyses the changes, or lack thereof, in the speech production of a group of bilingual Spanish/Catalan learners of English following a 3-month SA programme. Learners' production was assessed by means of a set of measures which have been commonly used in the literature analysing L2 speech production (Munro, 2008; Zampini, 2008) as reviewed in the previous chapter, although rarely combined in the analysis of pronunciation within an SA context so as to provide a better account of the actual impact of SA on this domain.

These measures consist, on the one hand, of objective, acoustic-phonetic analyses, which include acoustic measurements examining changes in segmental production, and measures of pronunciation accuracy involving pronunciation error rate scores at the segmental level (phonemic insertions, deletions and substitutions) and suprasegmental level (stress misplacement). On the other hand, foreign accent ratings were also used, which constitute more subjective measures of listeners' perception of speech. The study also examines the relationship between the acoustic-

phonetic properties of speech and listeners' assessment of foreign accent, and to what extent the acoustic measures and error rate scores can account for variance in the foreign accent ratings.

2.2. Research questions

The present study addresses three research questions, which are presented below:

Research Question 1 (RQ1)

Is there an effect of SA on L2 learners' speech production as assessed through objective, acoustic-phonetic measures?

Research Question 2 (RQ2)

Is there an effect of SA on L2 learners' speech production as assessed through subjective measures of listeners' perceived FA?

Research Question 3 (RQ3)

Are acoustic-phonetic measures related to perceived FA ratings, and to what extent are they good predictors of variance in FA ratings?

3. Method

This chapter presents the experimental design devised for the current study in order to address the research questions stated in section 2.2 above. In the following sections we provide a description of the participants and data collection instruments and procedures, together with an explanation of the different measures and data analysis procedures used.

3.1. Design

This study has been conducted within the large, state-funded project called Study Abroad and Language Acquisition (SALA), based at Universitat Pompeu Fabra (UPF) in Barcelona, and in collaboration with Universitat de les Illes Balears (UIB), in the Balearic Islands (Spain). This project addresses the effects of a 3-month SA programme on L2 linguistic development by upper-intermediate undergraduate learners of English enrolled in Translation and Interpreting studies at the UPF in Barcelona (see, for instance, Pérez-Vidal & Juan-Garau, 2011, or Pérez-Vidal, forthcoming, for a detailed description).

The SALA project has a longitudinal, pre-test/post-test design, whereby data were collected at different points in time. Learners were first tested upon their university entrance (T1). They were tested again after a two-term, 80-hour English FI period (T2), and immediately upon their return from a 3-month SA (T3). A last data collection was conducted 15 months after the learners' return from SA (T4). Since the present study aims to specifically assess the short-term effects of the SA period, we have

focused our analyses on the data collected prior to the students' departure abroad (Pre-test) and immediately after their return (Post-test).

SALA language assessment instruments consisted of a battery of exam-like tests covering different linguistic skills: grammatical knowledge, listening comprehension, writing, and L2 phonological competence both in perception (by means of a test assessing the perception of L2 sounds), and in production (assessed through three tests for the elicitation of oral data). Participants were also asked to complete three questionnaires designed to provide information about their linguistic profile, their attitudes and motivation to learn English, and the conditions and patterns of language use and L2 contact they experienced while abroad.

The students who agreed to take part in the SALA project were paid in order to encourage participation and to ensure an acceptable number of longitudinal subjects, in such a way that robust and reliable data analyses could be conducted. Data were collected from three consecutive cohorts of students at UPF, which make up a large pool of informants: Between 60 and 80 longitudinal subjects and between 155 and 250 cross-sectional subjects, with variation depending on the different tests. A group of native speakers of English was also recruited at UPF and UIB to provide baseline data which served as an index of language proficiency in order to examine learners' performance, as well as their potential development towards native-like patterns as a result of the SA learning context.

3.2. Participants

Participants in this study were drawn from the large SALA corpus. As explained in the design section 3.1 above, the SALA database consists of a main body of data collected from a large group of non-native speakers

of English (NNSs), as well as data from a small group of native speakers (NSs) which is used for comparison purposes.

3.2.1. Non-native speakers (NNSs)

This group was composed of undergraduate students of Translation and Interpreting at the UPF (*n*=23, 20 females and 3 males)[6]. Their mean age at the beginning of the study was 18.8 (range 17-21). None of them reported suffering any speech impairment at the time of data collection. Participants in this group were included in the study based on the sound quality of their recordings, in order to make sure that reliable acoustic analyses could be performed. Due to the logistic difficulty involved in collecting oral data from the large number of SALA participants, some of them had been recorded in sound-attenuated rooms, whereas most of them had been recorded in normal classrooms using digital recorders. Therefore, the former exhibited less background noise than the latter, such as hiss or voices, the frequencies of which may interfere with the frequencies of the voice signal analysed, and were thus considered more appropriate to be included in the study on account of their better sound quality, which also resembled more closely the optimal sound quality of the NSs' recordings (see section 3.3).

All participants had a similar linguistic profile. They reported being bilingual in either Spanish/Catalan (87%) or Spanish/Basque (13%), as they had been regularly exposed to Spanish and one of the other two languages since childhood and considered themselves proficient in both.[7]

[6] Due to the higher number of females taking Translation and Interpreting studies, it was not possible to have a balanced composition of males and females. This also prevented possible gender effects.

[7] Catalan and Basque are official, together with Spanish, in Catalonia and the Basque country, respectively. Therefore, it is normal for students in these territories to grow up with a knowledge of the two official languages.

Participants studied English as a foreign language in AH primary and secondary education institutions, following similar foreign language curricula under the Spanish Educational System *LOGSE*.[8] Under this system, participants completed 12 years of primary and secondary education (from age 6 to age 18). Foreign language instruction was introduced at age 8 (in Grade 3), and students received an average of 2.5 hours of English instruction per week from Grade 3 to Grade 10, and 2 hours per week during Grades 11 and 12, making a total of around 1,100 hours throughout 10 years (Pérez-Vidal, forthcoming).

Besides, students also received an additional FI period which covered the first and second terms of their first academic year at UPF, prior to the SA period, during which they took two English language courses which were included as compulsory subjects in the curriculum for their degree. Each of these subjects represented 40 hours of instruction, for a total of 80 hours. The courses were taught in English and focused on morpho-syntactic aspects of the language, and included also practice in vocabulary-building and the different language skills, in order for the students to improve their overall oral and written competence in English. No specific training was provided in English phonetics or phonology, or in English pronunciation. Therefore, all participants shared a similar exposure to the English language through AH classroom instruction, and a similar AOL of English as a foreign language.

In order to be admitted into the UPF Translation and Interpreting degree program, the students had to sit an entry test in which the minimum English proficiency level required was equivalent to a B2.1 in the Common European Framework of Reference (CEFR), whereas the level required to take part in the SA program was equivalent to a B2.2, both

[8] Source: Spanish Government: http://www.boe.es/buscar/doc.php?id=BOE-A-1990-24172

corresponding to an upper-intermediate level (see Beattie, forthcoming). As part of their Translation and Interpreting degree, they had to specialise in two foreign languages, English being their first foreign language, and the other being either French or German. All the students were exposed to a 3-month study abroad period in an English-speaking country, which constitutes the treatment condition under which they were tested. A description is provided below particularly regarding the amount and quality of English input and the type of practice which were available in this learning context.

Characteristics of the SA context

Each academic year at UPF is divided into three terms, each term lasting three months. During the first term of the second academic year, students were required to take a compulsory SA programme in an English-speaking host university. More than half the students went to the British Isles (69.5%), and the rest to other destinations, mainly in North-America (30.5%). The length of the stay was around 90 days, usually comprising a one-week introductory module for exchange orientation purposes, and 12 weeks of instruction during which the students followed content courses at the host university. The length of the programme thus corresponded with one of the three trimesters that make up a UPF academic year, while also meeting the minimum 90-day length requirement for the accreditation of an exchange abroad within the Erasmus scheme. Under the Erasmus programme students could benefit from small grants to help cover travel and living-abroad expenses.

The SA programme was equivalent to another subject in the participants' degree, and it was evaluated based on the courses they attended at the host institution. Participants were chiefly enrolled in Modern Languages and Humanities departments. The programmes were tailored-made according

to their interests and the host institutions' regulations, so that the number of courses and hours varied among participants. However, a minimum of 4 courses was set as a requirement. These were mainly content-based courses related to the participants' Translation degree and taught through the medium of English. In some cases the programme also included a course on the participant's second foreign language (French or German), or other foreign language. Participants reported receiving an average of 8.5 hours of instruction per week.

Students travelled to their host universities in small groups, together with other UPF classmates who shared the same destination (minimum of 2 people, maximum of 7). Regarding accommodation, more than half of them stayed in single rooms at residence halls (56.5%), while the rest chose different accommodation options, such as staying with a host family or sharing an apartment with native speakers of English and also with both native and non-native speakers. Students reported a fairly high level of interaction with native speakers of the target language. This was assessed by asking them to estimate their degree of contact with native English speakers by means of a 5-point scale (1='never', 2='seldom', 3='sometimes', 4='often', 5='very often'), most students scoring 4 or 5 (M=4). They reported a similarly high use of English with other non-native speakers (M=4.10 on the same scale). Students were also exposed to more native input through the host country media, by means of activities such as listening to the radio or watching TV and films (M=3.30 on the scale), being exposed mainly to British and American standard accents.

Consequently, and in contrast with the FI context, in this SA learning context students had access to a massive amount of exposure to English through both in-class and out-of-class quality input, as well as plenty of opportunities for communicative interaction in real-life, everyday

situations. It seems thus reasonable to believe that a context offering such rich and varied input and numerous practice opportunities, which are the trademark characteristics of SA, would highly benefit L2 competence, particularly regarding oral proficiency. Indeed, as already noted, this is a widespread belief which is one of the main reasons for the high numbers of students enrolling in these types of exchange programmes.

3.2.2. Native speakers (NSs)

Speech samples from 21 native speakers of English (17 females and 4 males), also drawn from the SALA corpus, served as baseline data in order to compare native and non-native oral performance, and to assess learners' potential oral proficiency gains. None of them reported any speech problem at data collection. They were young university students from the US (12) and the British Isles (9) enrolled in an exchange programme in Spain (either at UPF or UIB), with an age range similar to that of the NNSs (*M*=20 years). Data from the two groups were thus highly comparable, as both had similar profiles. Whereas the NNSs were tested at several points in time so as to analyse their development across time, the NSs sat for the SALA tests only once, as their data were used only for comparison purposes.

Although some concerns have been raised against seeing L2 learners as 'failed native speakers' (Cook, 1999: 185), the importance of the native speaker L1 competence in SLA research is also widely acknowledged, since "The native speaker's 'competence' […] or 'knowledge of the language' is a necessary point of reference for the second language proficiency concept" (Stern, 1983: 106, cited in Cook, 1999). Foster and Tavakoli (2009) also note the vital role of native speaker baseline data in order to validate research on L2 acquisition, as the possibility of

comparing L1 and L2 data for the same task and under the same conditions allows the researcher to more reliably attribute any observed proficiency differences to possible L2-specific extra processing costs. Zampini (2008) particularly stresses the need to collect L1 data from native speakers in studies assessing L2 phonology. She points out that the common methodological approach in this field consists in obtaining the relevant measures for the analysed aspect of L2 speech (acoustic-phonetic measures or listeners' responses) and compare the non-native values with the mean values of native speakers.

3.3. Data collection

Speech samples were drawn from the SALA reading aloud (RA) task, which consisted of the rendition by the participants of Aesop's fable "The North Wind and the Sun" (NWS; see Appendix 1 for RA handout with the full text and instructions). This is a short, 114-word text of which different versions exist in different languages, since The International Phonetic Association (IPA) has encouraged its use as a standard oral elicitation resource to illustrate the pronunciation of different languages and language varieties (see IPA, 1999). It has been widely used to document differences characterising English pronunciation by L1 users with different dialects or by L2 users (Schneider, Burridge, Kortmann, Mesthrie, & Upton, 2004).

The RA task was selected over the other two SALA oral tasks (an interview and a role play), which elicited extemporaneous speech by reproducing what was intended as a real-life interactional setting. The use of the RA task was deemed appropriate as it was thought that the speech thus elicited would more accurately represent the L2 learners' actual pronunciation competence in English, by eliminating the additional L2

processing-related demands that are inherent to the production of extemporaneous speech, particularly in tasks of an interactional nature, such as the other two SALA oral tasks. Besides, having all participants read the same text at all the different testing times ensured that the same vowel and consonant items appeared in all the speech samples and in the same phonetic contexts, thus facilitating contrastive analyses between native and non-native segmental production, and analyses assessing changes in NNSs' segmental production across the different testing times.

A member of the research team was present during the recordings to explain how to perform the task and to answer possible doubts or questions. Instructions were also provided on the RA test handout, which participants were asked to read carefully. Participants were recorded individually. They were instructed to read the text twice, first silently on their own in order to become familiar with it, and then aloud to be recorded. The researcher told them that, after reading the text the second time, they would be asked a question about the text which they were to answer by merely stating 'yes' or 'no' as quickly as possible. This was done to draw the participants' attention to the content so that they were not aware that the focus of interest was pronunciation, with the aim of obtaining more natural-sounding data. Immediately after reading the text out loud, participants were asked the following question: "Was the North Wind Stronger than the sun?"

Data from the NNSs were recorded in sound-attenuated cabins at UPF using an analogue tape recorder and were subsequently digitised at 22,050 Hz, in 16-bit, uncompressed wave format, monaural mode. Data from the NSs were recorded in sound-proof cabins at UPF and UIB with the assistance of a professional sound technician, using the Pro Tools digital audio workstation platform for Microsoft Windows. The digital files were

saved in wave format at 44,100 Hz (later downsampled to 22,050 Hz), 16-bit monaural.

3.4. Data analyses

Development in the learners' speech production across time was assessed through both acoustic-phonetic analyses of the speech samples, through which objective measures were obtained of acoustic-phonetic speech properties, as well as listeners' FA ratings, which provided subjective measures of perceived degree of accentedness.

Acoustic measurements were conducted on a series of tokens occurring in the NWS text (section 3.4.1.1). An excerpt extracted from the NWS text served to perform a more in-depth analysis, whereby additional error rate measures of pronunciation accuracy (section 3.4.1.2) and perceived FA ratings (section 3.4.2) were also obtained. This excerpt consisted of a short sentence containing several segmental and suprasegmental properties that were known to be problematic for L1 Spanish/Catalan EFL learners, and were thus likely to result in non-native, accented production patterns (see Appendix 2 for the transcribed sample sentence, and section 3.4.1.2 for some examples of pronunciation errors). The following sections present the different measures used in order to address the different research questions in the study.

3.4.1. Acoustic-phonetic measures

A series of acoustic-phonetic analyses were conducted on the participants' speech samples to obtain objective measures of their speech production development during SA. These analyses are described in the sections below.

3.4.1.1. Acoustic analyses

Acoustic analyses consisted of objective measurements for VOT in voiceless plosives and for vowel duration and quality in the English contrasts /i:-ɪ/ and /æ-ʌ/. The decision to study these English categories was informed by theoretical models of L2 speech acquisition such as Flege's SLM (1995, 2002) or Best & Tyler's PAM-L2 (2007), which allow for the prediction of difficulties in L2 segmental acquisition attributable to the influence of the L1 system (see section 1.2.2), as well as by previous research reporting difficulties in the categorisation of these English sounds for L1 Spanish and Spanish/Catalan learners. Acoustic measurements were conducted with Praat speech analyses software (Boersma & Weenink, 2008) on the data elicited through the NWS text (see 3.3. above, Appendix 1).

a) VOT

A well-documented acoustic difference between voiceless plosives in English and in Romance languages such as Spanish and Catalan is VOT, which is the time interval between the release burst of the stop closure and the beginning of vocal fold vibration for voicing. In English, the release of the plosive is followed by a perceptible burst of noise or plosion, after which air goes through the vocal folds during an interval of time that is called aspiration and which precedes the beginning of vocal fold vibration. In Spanish and Catalan, the plosion after the release burst is weaker than in English, and vocal fold vibration begins shortly after it, or at about the same time. English voiceless stops occurring in stressed syllable onsets have thus long-lag, aspirated VOT duration values,

whereas voiceless stops in Spanish and Catalan are short-lag or unaspirated (Flege et al., 1998; Ladefoged, 2005; Mora, 2008).[9]

Speakers of L1 Spanish and Catalan tend to produce English plosives inaccurately due to the phonetic similarity between the L1 and the L2 categories. As predicted by SLM or PAM-L2, the L2 stops are linked to the similar L1 stops and assimilated to them, and this hinders the creation of new L2 categories with English-adjusted long VOT duration. These learners usually produce English voiceless plosives with VOT values which are intermediate between the L2 long-lag values and the L1 short-lag ones (Flege et al., 1998; Mora, 2006; Mora, 2008).

It could be expected, therefore, that the learners in this study would equate the English voiceless plosives with the Spanish/Catalan plosives, thus producing inaccurate VOT values in English. In accordance with SLM, it could also be expected that increased exposure to rich and authentic TL input during SA would help to sensitise the learners to the acoustic differences between L1 and L2 VOT duration leading to an improvement in L2 production, given the amount of practice and authentic input that the SA context may offer.

[9] Mean VOT values for English /k/ usually range between 50-80 ms, they are slightly less for /t/ and about 30-40 ms for /p/. In Romance languages VOT for /k/ is about 30 ms, it is even less for /t/ and almost 0 for /p/. VOT values usually increase as the constriction is located further back in the mouth.

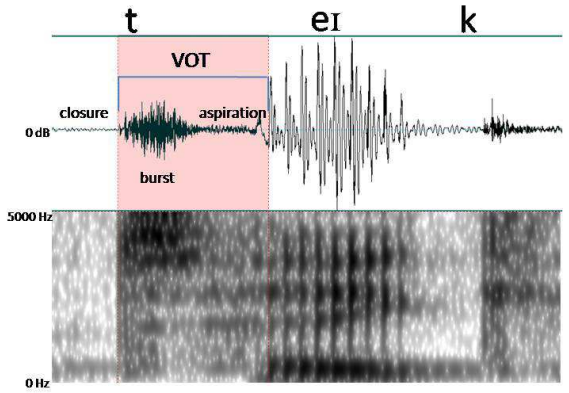

Figure 1. Waveform (upper panel) and spectrogram (bottom panel) for English "take" as produced by a male NS

VOT measures were conducted on the voiceless stops /t, k/ which appear in the NWS text on stressed syllable onsets.[10] There were two tokens of velar /k/ (*came*, *could*), and three tokens of dental /t/ (*take*, *took*, *two*), for a total of 335 VOT measurements (5 plosives x 23 NNSs x 2 times + 5 plosives x 21 NSs). As shown in Figure 1 above, VOT was gauged as the distance in ms between the points on the waveform signalling the onset of the stop release burst as indicated by the first peak of irregular perturbations on the sound wave (shown as a release bar in the spectrogram) and the beginning of vocal fold vibration as indicated by the first peak of periodic energy (corresponding to the first vertical striation in the spectrogram).

b) Vowel quality and duration

The vowels in the English contrasts /i:-ɪ/ and /æ-ʌ/ present distinctive features of duration (vowel length) and quality (spectral differences in vowel height and degree of frontness) which have no distinctive function in either Spanish or Catalan (see Ladefoged, 2005; Spencer, 1996 for

[10] Unfortunately, in this text there is no /p/ token occurring in stressed syllable onsets.

detailed acoustic and articulatory descriptions of these sounds). The main distinctive feature between /i:-ɪ/ is a quality distinction, according to which /i:/ is described as tense and /ɪ/ as lax. There are also inherent duration differences between both vowels which are implemented as a correlate of the main tense-lax distinction.

For the articulation of tense /i:/ the root of the tongue is pulled forward so that the tongue is stretched and raised towards the hard palate, near the roof of the mouth. It is thus a close vowel, articulated near the periphery of the front vowel space within the vocal tract. As a result of the tension in the muscles of the tongue and cheeks during its production, the lips are rather spread. For the production of lax /ɪ/ the root of the tongue moves back to its normal, resting position, so that the tongue is displaced back and down. Consequently, /ɪ/ is centralised and less close than /i:/, the lips being loosely spread or in a more neutral position as the tension in the muscles is reduced.

These articulatory characteristics are acoustically reflected in spectral differences between the two vowels regarding both vowel height (represented by the first vowel formant or F1) and frontness (represented by the second vowel formant or F2). In general terms, F1 values decrease as vowel height increases, and F2 values increase as the vowel is more fronted. Consequently, although /i:/ and /ɪ/ are both high-front vowels, tense /i:/ presents a low F1 and high F2, as it is a peripheral vowel, articulated with the tongue high up in the mouth and fronted. In contrast, lax /ɪ/ has a lower and more centralised point of articulation, corresponding with higher F1 and lower F2 values than its tense counterpart.

As mentioned above, there are also inherent duration differences between both vowels, since /i:/ is typically a long vowel and /ɪ/ a short vowel

(Ladefoged, 2005; Peterson & Lehiste, 1960). These differences in length are, nonetheless, context-dependent, as it is well established that consonantal environment influences the duration of vowel nuclei in English, so that /i:/ and other long vowels are reduced when followed by fortis consonants, whereas /ɪ/ and other short vowels are usually lengthened before lenis consonants (Peterson & Lehiste, 1960; Roach, 2004; Spencer, 1996). Duration is therefore considered a secondary or redundant feature for the categorisation of this vowel contrast, concomitant with the main distinction based on quality (Bohn, 1995; Strange, 2007).

This view is supported by experimental evidence from vowel pattern recognition and perception studies showing that native speakers of English are able to distinguish these two vowels by relying on their spectral features, regardless of the distortion of typical durational patterns (Hillenbrand, Clark, & Houde, 2000; Ylinen et al., 2010). However, it is important for L2 learners to implement both the appropriate differences in quality and also in duration for these vowels, as durational differences may be an important cue to distinguish other English phonetic categories, e.g., word-final obstruents such as bit/bid, or beat/bead (Hogan & Rozsypal, 1980).

The categorisation of the English tense/lax /i:-ɪ/ contrast involves therefore a complex combination of spectral and durational features which L2 learners need to learn for an accurate production of these two vowels. These quality and duration differences do not exist neither in Spanish nor in Catalan, which have only one, rather peripheral high-front vowel /i/. The two contrasting L2 vowels /i:-ɪ/ tend thus to be assimilated to the similar L1 category /i/, which roughly corresponds to the phonological space occupied by the L2 categories.

Regarding the contrast /æ-ʌ/, the main distinctive feature between these two vowels is a front/central opposition that Spanish and Catalan also lack for open vowels. The vowel /æ/ is articulated within the low-front area of the vowel space, so that it is more peripheral and has higher F1 and F2 values. In contrast, the vowel /ʌ/ is centralised, which is reflected in lower values for F1 and F2. There are also inherent duration differences between /æ-ʌ/, as peripheral vowels are typically realised with a tenser articulation and longer duration than centralised vowels, although the difference in length between /æ-ʌ/ is smaller than between /i:-ɪ/ (Hillenbrand et al., 2000), and it is equally affected by context. Again, as was the case with the high-front contrast, Spanish and Catalan also lack a distinction for open vowels. The closest L1 category is a single and rather central vowel /a/, to which the English L2 categories /æ-ʌ/ tend to be assimilated.

Hence, the acquisition of these two English contrasts usually poses considerable difficulty for L1 Spanish/Catalan learners. They have to learn how to integrate two L2-specific distinctive features, vowel duration and quality, which are not used contrastively in their L1s, and they also have to learn that quality is the main distinctive feature, whereas differences in duration are a redundant, secondary phonetic feature which is affected by context. Previous studies have documented the perception and production problems facing L1 Spanish/Catalan learners of English in the acquisition of the contrasts /i:-ɪ/ and /æ-ʌ/ (Avello, 2010a; Cebrián, 2006; Fullana & MacKay, 2003; Mora & Fullana, 2007). These learners usually fail to distinguish the low vowels /æ-ʌ/, and when they are able to differentiate the high-front vowels /i:-ɪ/, they tend to do so only in terms of duration, but not quality, in contrast with native speakers' patterns.

Consequently, the learners in our study could be expected to assimilate the two vowels in the English contrasts /i:-ɪ/ and /æ-ʌ/ to the corresponding L1 vowels /i/ and /a/ due to the phonetic similarity between the L2

categories and the closest L1 categories, and to the learners' lack of sensitivity to the relevant acoustic features that distinguish these English vowels, resulting in inaccurate production which may lead to L2-accented speech. As pointed out in the case of VOT above, exposure to the L2 could be expected to have a positive effect on the learners' categorisation of these sounds conducive to an improvement in their production, as a consequence of the rich amount of authentic L2 input which is likely to be available during SA.

Vowel measures were obtained for all the tokens /i:-ɪ/ and /æ-ʌ/ occurring in the NWS text as the nuclei of stressed syllables (unstressed syllables were not included as they are typically weakened in English, a phenomenon which affects vowel duration and quality). There were three /i:/ tokens (*agreed, succeeded, immediately*), seven /ɪ/ tokens (*wind*x4, *which, considered, did*), six /ʌ/ tokens (*sun*x3, *one, other, up*), and five /æ/ tokens (*traveller*x4, *wrapped*), for a total of 1,407 vowel tokens (21 tokens x 23 NNSs x 2 times + 21 tokens x 21 NSs). Vowel onset and vowel offset for each vocalic token were identified on the Praat window display of the sounds, including vowel transitions (see Figure 2). A Praat script was used (Avello, 2010b; Pérez-Vidal et al., 2011) to measure the duration in ms of the highlighted vowel token along with frequency values in Hertz (Hz) for F0 or pitch, F1 and F2, which were gauged on a 20-ms window of the central part of the vowel, coinciding with its steady state and highest amplitude. As explained above, the measures of F1 and F2 provide a representation of the spectral characteristics of the vowels (height and frontness).

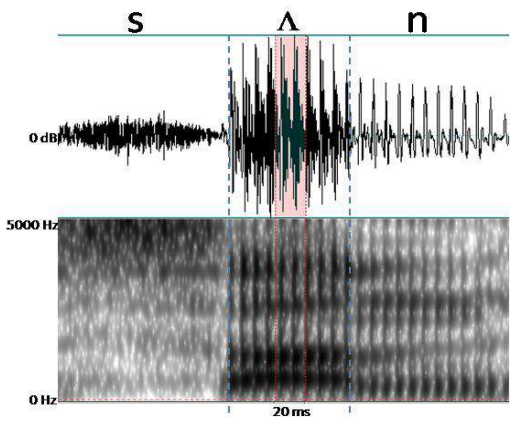

Figure 2. Waveform (upper panel) and spectrogram (bottom panel) of English “sun” as produced by a male NS.

These vowel frequency measures were later converted to Bark (B), which is a psychoacoustic scale that better represents the perceptual characteristics of sounds (F0=B0; F1=B1; F2=B2), and were then subjected to a normalisation procedure in order to reduce inter-speaker variability. This procedure consisted in the computation of B1 minus B0 (B1-B0) as a measure for vowel height, and B2 minus B1 (B2-B1) as a measure for degree of vowel frontness (Syrdal & Gopal, 1986). Euclidean distances (EDs) were then computed by means of the normalised measures in order to gauge the distance between the two vowels in each contrast within the two-dimensional vowel space (i.e., the distance between vowels /i:-ɪ/ and vowels /æ-ʌ/). Duration differences (DDs) were also calculated between the longer vowel and the shorter vowel within each contrast. These ED and DD measures were subsequently used to assess the robustness of the distinctions produced for the two contrasts in terms of quality and duration, respectively (Flege et al., 1997; Mora & Nadeu, 2012).

3.4.1.2. Error rate scores

Error rate scores served as measures of learners' pronunciation accuracy, and consisted of a computation of pronunciation errors (PrEr) obtained by means of a phonetic analysis (Brennan & Brennan, 1981; M. J. Munro & Derwing, 1995a; Trofimovich, Lightbown, Halter, & Song, 2009). As mentioned in section 1.2.2.4, previous studies have found a relationship between such measures of L2 learners' error counts and listeners' perception of L2-accented speech (Avello et al., 2012; Brennan & Brennan, 1981; Cunningham-Andersson & Engstrand, 1989; M. J. Munro & Derwing, 1995a).

These error rate scores served also as objective measures of learners' speech production development. The phonetic analysis was conducted on the same speech samples used to construct the stimuli for the listening experiment developed to measure perceived FA (see section 3.4.2.2). These speech samples presented both segmental and suprasegmental features that could result in learners' mispronunciations, leading to accented speech production (see Appendix 2). Mispronunciations were identified through a perceptual analysis with the assistance of Praat-displayed waveforms and the corresponding spectrograms for each speech sample, and included phoneme substitutions affecting segmental articulation (deletions, insertions, and substitutions), as well as lexical stress misplacement. Below are some examples of different pronunciation errors encountered in the analysed speech samples:

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 *tra<u>v</u>eller*
- Substitution of dental plosive [d] for dental fricative [ð] in *<u>th</u>en*
- Substitution of open vowel [a] for close back vowel [ɔ] in *w<u>a</u>rmly*
- Substitution of dental fricative [ð] for alveolar plosive [d] in *imme<u>d</u>iately*
- Substitution of velar fricative [x] for glottal fricative [h] in *<u>h</u>is*

d) Stress misplacement:

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

3.4.2. Perceived FA measures

Measures of perceived degree of FA were obtained from two different groups of listeners who performed a rating task so as to examine the perception of accentedness in the participants' speech samples by listeners with different profiles. In the following sections a description is provided of the listener groups and the procedures for the rating task.

3.4.2.1. Listeners

Two groups of listeners were recruited differing in terms of their linguistic profile (English native listeners vs. non-native listeners) and their knowledge of linguistics and training in English phonetics and phonology

(naïve or unsophisticated vs. trained or sophisticated). Listeners completed the rating task and also a linguistic profile questionnaire which provided demographic data and information on their linguistic background and their degree of familiarity with different native and non-native accents of English, with specific reference to the type of foreign-accented speech they were asked to evaluate.

a) Native listeners (NLs, *n*=20)

The native listener group consisted of L1 English native speakers (males=5; females=15; mean age=21.85). 15 of them came from the UK and the rest from the USA. They were exchange students of L2 Spanish enrolled in different degree studies at the UPF: Translation and Spanish and English studies (7), International studies and Political sciences (5), Law (5), and Economics (3). They were naïve, unsophisticated listeners, with no specific training in linguistics or phonetics. They reported familiarity with different standard British and American native accents, and were also highly familiar with Spanish/Catalan-accented English.

b) Non-native listeners (NNLs, *n*=37)

The group of non-native listeners had a linguistic profile similar to that of the NNSs in that they were also bilingual L1 Spanish/Catalan speakers studying EFL (males=8; females=29; mean age=22). They had a proficient level of English, since they were studying English Philology at Barcelona. Their academic curriculum included courses on English Literature and Linguistics, and at the time of data collection they were just finishing the second of two courses on English phonetics and phonology. These courses were designed to specifically tap on the problems facing L1 Spanish/Catalan speakers when learning English pronunciation, and they included the analysis of English segmental and suprasegmental characteristics, phonetic and phonological transcription, and training on

perception and production of English pronunciation. They were thus trained or sophisticated listeners, with a sound knowledge of English pronunciation. They also reported familiarity with British and American accents, and were highly familiar with the Spanish/Catalan-accented speech they had to assess, as they shared the NNSs' L1 background.

3.4.2.2. Rating task

This task provided us with a global measurement of perceived foreign accent in the NNSs' speech prior to and immediately after their SA. It was a self-paced task created and run with Praat software (Boersma & Weenink, 2009). The same speech samples used for the accuracy scores analyses were also used to create the stimuli for this task (see Appendix 2). The audio files were normalised for intensity and saved in wave format at 22.050 Hz, with 16-bit resolution and in monaural mode. The rating task was equivalent to a class activity within the NNLs' course on English phonetics and phonology. The NLs were paid in order to encourage their participation. At the beginning of the session, the listeners were given a handout with the description of the task, as well as with the instructions on how to run it with Praat. After completing the rating task, they filled out the linguistic profile questionnaire.

The listeners were presented the speech samples produced by the 23 NNSs (PreTest and PosTest) and by 6 NSs used as baseline. They were instructed to focus on pronunciation and to rate the degree of FA in the speech samples by means of a 7-point, equal-appearing Likert scale, where 1 stood for "native" and 7 stood for "heavy foreign accent" (see Figure 3). They were also instructed to make use of the whole scale.

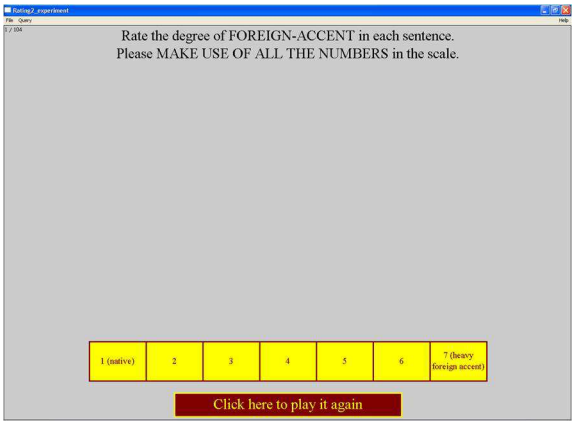

Figure 3. Praat experiment window for the Rating task

This methodology has been widely used in research on L2 speech production analysing such constructs as FA, as well as intelligibility and comprehensibility (see section 1.2.2.4. A 9-point scale has been most commonly used in this type of studies, in which participants usually differ greatly in proficiency level, as well as in AOL and/or L2 exposure. However, a 7-point scale was deemed more appropriate for the data in the present study, taking into account the smaller degree of variability in our oral samples (NNSs with a similar age, AOL, exposure to the TL and proficiency level), as compared to the majority of FA studies.

Each stimulus was repeated twice for a total of 104 trials per listener (23 NNSs x 2 times x 2 repetitions + 6 NSs x 2 repetitions), making up a total of 5,928 ratings (104 trials x 57 listeners). Each listener heard the stimuli in a different randomised order. The listeners could play each trial twice before choosing their answer. After correctly rating a sample, they had to click on a "next" button to listen to the following sample. If a stimulus was wrongly rated, the listeners could click on an error button to listen to it again and correct their answer. A short 10-trial practice session was provided before the experiment so as to allow listeners to become familiar

with the procedure and to self-adjust the volume of the headphones used to present the samples to a comfortable level. In order to avoid fatigue, listeners had the possibility to take a short pause after the first half of the trials had been presented.

In the following chapter, we present and discuss the main findings obtained with the measures and procedures just described.

4. Results and discussion

This chapter presents and discusses the main findings from the analyses described in Chapter 3. It has been organised into three main sections, each dealing with one of the three research questions stated in Chapter 2. Accordingly, section 4.1 addresses RQ1, section 4.2 addresses RQ2, and section 4.3 addressed RQ3. The different research questions are restated at the beginning of each section, followed by the results and discussion for each research question.

4.1. Acoustic-Phonetic analyses

This section addresses RQ1: *Is there an effect of SA on L2 learners' speech production as assessed through objective, acoustic-phonetic measures?* Section 4.1.1 reports the results of acoustic analyses measuring VOT and vowel quality and duration. Section 4.1.2 reports the results of phonetic analyses exploring pronunciation accuracy by means of error rate scores.

4.1.1. Acoustic measures

In this section we present the results for the analyses exploring the effect of SA on learners' production of VOT and vowel quality and duration. Acoustic analyses were performed on the oral data collected from the group of learners before and after SA. Native speaker baseline data were used for comparison purposes.

4.1.1.1. Results

Table 2 and Figure 4 present the mean VOT duration values produced by the NSs (baseline) and by the NNSs at the two testing times.[11] A series of correlations showed that VOT values for /t/ and /k/ were strongly correlated for NSs (r=.58, p<.01) as well as moderately to strongly correlated for NNSs (Pre-test: r=.49; Post-test: r=.51; p<.05), indicating that participants who produced longer VOT duration for /t/ also produced longer VOT duration for /k/, and vice-versa. Therefore, the VOT values obtained separately for /t/ and /k/ were averaged in order to create a composite measure of overall VOT which serves as a global index for participants' VOT production.

Group	Time	VOT /t/	VOT /k/	overall VOT
NS	(baseline)	70.12 [11.7]	65.06 [10.4]	67.59 [9.8]
NNS	Pre-test	53.00 [17.8]	45.34 [13.3]	49.17 [13.4]
	Post-test	51.16 [18.7]	44.73 [15.8]	47.95 [15.0]

Table 2. Mean VOT duration (ms) for NSs' baseline data (n=21) and for NNSs (n=23) as a function of testing time (Pre-test, Post-test) and category (individual and composite measures). *SD*s in brackets.

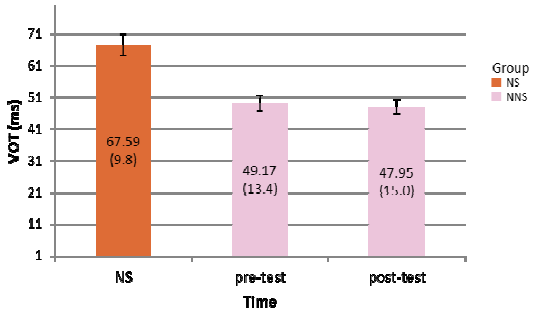

Figure 4. Mean VOT duration for NSs and NNSs averaged across all tokens.

[11] The fact that VOT for /t/ is longer than for /k/, both for NSs and NNSs, may be attributable to phonetic context

An inspection of the VOT descriptives shows that NNSs produced VOTs for /t/ and /k/ which were longer than the short-lag VOTs characteristic of voiceless plosives in Romance languages such as Spanish and Catalan, but they were still significantly shorter than the long-lag VOTs characteristic of English voiceless plosives, as indicated by independent-samples t-tests comparing NSs' and NNSs' values at the two testing times [for all cases $t(42)>3.80$, $p<.001$]. That is, NNSs' VOT values were halfway between typical short L1 values and typical long L2 values. As mentioned in the Methods section, evidence of cross-linguistic influence in the production of VOT duration has been documented in previous studies (Flege et al., 1998; Flege, 2002; Mora, 2008; Wrembel, 2011).

These intermediate L1-L2 values remained very similar both at Pre-tests and Post-test, differences in VOT duration being within a narrow 3-ms range, which suggests that SA did not have a large effect on NNSs' L2 laryngeal timing patterns. This was confirmed by paired-samples t-tests conducted with the individual and composite VOT measures as the dependent variables and *time* (Pre-test, Post-test) as the independent, within-subjects factor, as these analyses did not yield any significant change as a function of time ($p>.05$). This indicates that increased L2 experience, even in a context of massive exposure such as SA, did not impact learners' production of English voiceless plosives.

Table 3 and Figure 5 illustrate the mean Euclidean distances (EDs) and duration differences (DDs) between the vowels in the contrasting pairs /i:-ɪ/ and /æ-ʌ/ as produced by the two speaker groups. As explained in the Methods section, the greater the Euclidean distance values are, the more robust the distinction between two contrasting vowels is in terms of quality, whereas the longer the duration difference values are, the more robust the distinction between the two vowels is regarding duration.

Group	Time	ED /i:-ɪ/	ED /æ-ʌ/	DD /i:-ɪ/	DD /æ-ʌ/
NS	(baseline)	3.05 [0.8]	1.20 [0.7]	44.60 [17.0]	22.30 [14.3]
NNS	Pre-test	0.93 [0.7]	0.50 [0.3]	26.69 [17.0]	15.81 [14.8]
	Post-test	0.65 [0.4]	0.64 [0.3]	23.65 [11.5]	17.40 [15.7]

Table 3. Mean Euclidean distances (B) and duration differences (ms) for NSs' baseline data (n=21) and NNSs (n=23) as a function of vowel contrast and testing time. *SD*s in brackets.

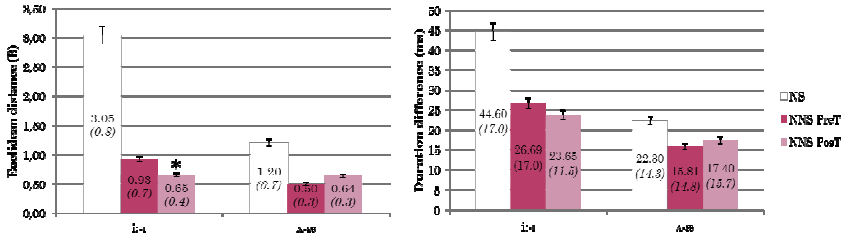

Figure 5. Mean Euclidean distances (left) and duration differences (right) for NSs and. NNSs as a function of vowel contrast and testing time.

Figure 6 is a formant chart plotting the average values of height and frontness for /i:-ɪ/ and /æ-ʌ/ as produced by the NSs (baseline) and the NNSs at Pre-test, to illustrate native and non-native production of these contrasts. As expected, the vowels produced by the NSs in each of the two minimal pairs were very distinct and fairly distant within the acoustic vowel space regarding both height and frontness, indicating a clear, robust quality distinction. Indeed, paired-samples t-tests showed significant differences between the NSs' two high-front vowels /i:-ɪ/ and two low vowels /æ-ʌ/ both in terms of height [in both cases $t(20)>5.93$, $p<.001$, $\eta^2>.637$] and frontness [in both cases $t(20)>-3.98$, $p<.001$, $\eta^2>.441$]. In the high-front tense/lax pair, /i:/ presented a rather close and peripheral articulation, whereas /ɪ/ was more open and less fronted, with values located more towards the centre of the vowel space. As for the low front-central contrast, /æ/ was more fronted and open than /ʌ/, and /ʌ/ had rather centralised values.

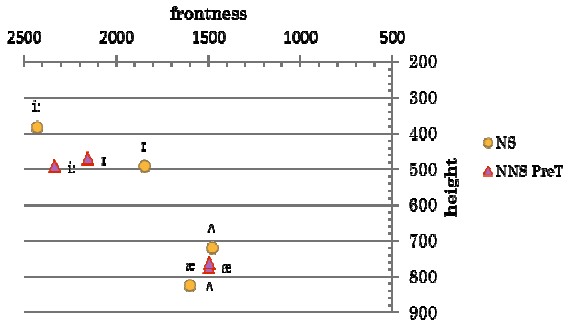

Figure 6. Formant chart for English vowel contrasts /i:-ɪ/ and /æ-ʌ/ as produced by the NSs (baseline) and the NNSs (Pre-test).

In contrast, the two vowels produced by the NNSs for each contrasting pair presented a rather reduced distance within both the height and frontness dimensions. As a result, the non-native high-front vowels /i:-ɪ/ and low vowels /æ-ʌ/ were much closer to each other than the native vowels, seemingly sharing the same acoustic vowel space, which suggests that the learner group did not produce a robust distinction between the two L2 distinctive vowels of each minimal pair in terms of quality. As indicated by paired-samples t-tests, the NNSs did not produce significant differences between tense-lax /i:-ɪ/ and front-central /æ-ʌ/ neither in height [in both cases $t(22)<.45$, $p>.05$] nor in frontness [in both cases $t(22)<.59$, $p>.05$]. The NNSs' high-front vowels were placed halfway between the vowel spaces corresponding to the two native categories /i:/ and /ɪ/, and the low vowels presented an even greater degree of overlap. It is very likely, therefore, that when attempting to pronounce the two distinct vowels in each English contrast, the NNSs were actually producing the same vowel as far as the two-dimensional height-frontness space is concerned.

Regarding duration, however, the values produced by the NNSs followed a pattern that resembled more closely that of the NSs'. The two speaker

groups implemented very similar duration values for the peripheral vowels /i:/ [NSs' *M*=120.88, *SD*=20.8; NNSs' *M*=126.11, *SD*=21.5] and /æ/ [NSs' *M*=122.13, *SD*=23.4; NNSs' *M*=126.81, *SD*=12.9], which were longer than the centralised vowels /ɪ/ [NSs' *M*=76.27, *SD*=15.7; NNSs' *M*=104.11, *SD*=9.9] and /ʌ/ [NSs' *M*=99.83, *SD*=14.9; NNSs' *M*=111.03, *SD*=10.7], although the NSs' centralised vowels were shorter than those produced by the learner group. In fact, paired-samples t-tests showed significant differences in duration between the tense-lax vowels /i:-ɪ/ as realised both by the NSs [$t(20)=12.01$, $p<.001$, $\eta^2=.88$] and the NNSs [$t(22)=4.90$, $p<.001$, $\eta^2=.52$].

These data illustrate NSs' use of quality features in the realisation of the vowel contrasts /i:-ɪ/ and /æ-ʌ/, and NNSs' difficulty in producing similar quality-based distinctions for these contrasts. The data also illustrate how NSs integrated both durational and spectral features in the distinction of the tense-lax contrast /i:-ɪ/, whereas the NNSs seemed to rely on durational features only in order to distinguish these two L2 vowels. As reported in the Methods section above, previous studies have documented how L1 Spanish and bilingual Spanish/Catalan speakers usually fail to implement non-L1 vowel spectral distinctions while at the same time being able to produce temporal distinctions which are equally inexistent in their L1 in the production of English /i:/ and /ɪ/.

As expected, these differences between native and non-native vowel production were evidenced by the fact that the NSs produced larger EDs and also longer DDs than the NNSs between the two vowels of each contrast (illustrated in Table 3). Independent-samples t-tests comparing native and non-native ED and DD measures showed significant differences between NSs and NNSs in ED for the two vowel contrasts and in DD for /i:-ɪ/ [in all cases $t(42)>3.43$, $p<.001$], whereas the differences between NSs and NNSs in the production of duration distinctions for /æ-

ʌ/ did not reach significance at any of the two testing times [$t(42)<1.60$, $p>.05$]. This measure of DD for /æ-ʌ/ was therefore not included in any other analyses.

Phonological development in NNSs' vowel production towards more native-like patterns would thus imply an increase in the measures of ED and DD after the SA period. As shown in Table 3, this was the case of ED for /æ-ʌ/, however the opposite was true regarding ED for /i:-ɪ/, whereas DD for /i:-ɪ/ was very similar both at Pre-test and Post-test. A series of paired-samples t-tests were subsequently conducted comparing Pre-test and Post-test ED and DD measures, so as to better explore the effects of SA on NNSs' production of vowel quality and duration.

Regarding DD for /i:-ɪ/, and in line with the results for VOT reported above, no significant change was found in duration differences between these vowels as a function of time ($p>.05$). In fact, Table 3 and Figure 5 above show that, as was the case with VOT, changes in DD values before and after SA were within a very narrow span.

Results from the analyses of ED measures for the vowel contrasts /i:-ɪ/ and /æ-ʌ/ were mixed. As mentioned above, the pattern of changes during SA differed between the two vowel contrasts (see Figure 7). The contrast /i:-ɪ/ presented a decrease in ED values following SA which actually reached significance [$t(22)=2.53$, $p=.019$, $\eta^2=.22$], suggesting that NNSs produced less of a quality distinction between the two vowels. However, for the contrast /æ-ʌ/ there was an increase in ED during the SA context, and even though it did not reach significance ($p>.05$), the observed gains seem to signal a trend of improvement in the production of quality distinctions between these two vowels after SA, as opposed to the results for the /i:-ɪ/ contrast.

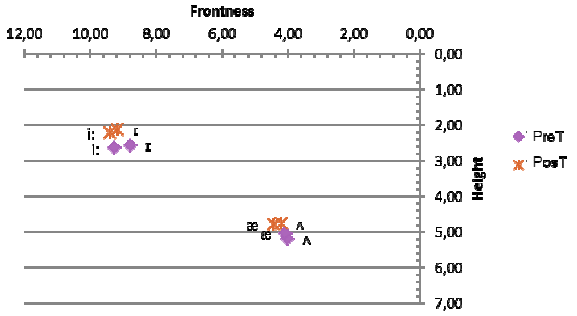

Figure 7. Vowel plot (B) for NNSs' vowel contrasts /i:-ɪ/ and /æ-ʌ/ as a function of testing time (Pre-test, Post-test).

An inspection of the vowel plot in Figure 7 shows a shift up and forwards in the NNSs' vowels within the vowel space between the two testing times. Particularly in the case of the tense-lax contrast, both /i:/ and /ɪ/ seemed to undergo a simultaneous and parallel movement after SA, at the same time that the ED between them decreased, which further suggests that the two vowels were treated by the NNSs as two instances of the same L2 category in terms of quality. In contrast, as reported above, there was no change in DD values for this contrast at Post-test, indicating that the initial significant differences in duration between /i:/ and /ɪ/ were maintained after SA, although such durational differences still fell significantly short of NSs' DD values. It appears, therefore, that SA did not change NNSs' production of this contrast towards more native-like patterns, as duration seemingly continued to be the main distinctive feature used by NNSs in order to produce the /i:-ɪ/ contrast.

4.1.1.2. Discussion

Taken together, the analyses assessing VOT duration in voiceless plosives and the realisation of the vowel contrasts /i:-ɪ/ and /æ-ʌ/ yielded no evidence of significant improvement in the learners' production of these

L2 sounds. This would indicate that the increased L2 exposure experienced by the learners during SA failed to trigger a substantial restructuring in their phonological representations for these L2 categories leading to more target-like production.

Regarding VOT, measures at Pre-test revealed duration values which were intermediate between English long-lag VOTs and Spanish/Catalan short-lag VOTs, in line with findings from previous research on English VOT production by L1 Spanish and Spanish/Catalan learners (Flege et al., 1998; Flege & Eefting, 1987; Mora, 2008). This would indicate that some type of L2 phonetic learning regarding English voiceless plosives had already taken place before SA, probably in terms of the creation of a merger between L1 and L2 VOT values, as further explained below. Such changes in the learners' categorisation of English plosives could be attributable to the previous FI setting they had experienced prior to their stay abroad. These VOT values remained unchanged at Post-test, which suggests that a short immersion period of three months might not be long enough to trigger further changes towards the production of more L2-adjusted, long-lag VOT values, even when learners already display intermediate L1-L2 VOT production patterns.

Findings from some studies indicate that changes in VOT production may nonetheless arise when these intermediate L1-L2 values have not yet been achieved before a period of increased L2 exposure during SA, i.e., in the absence of evidence showing previous phonetic learning. For instance, in a study comparing the effects of a 3-month versus a 6-month SA programme on two groups of L1 Spanish/Catalan learners of English, Avello and Lara (forthcoming) found that the 6-month group experienced a significant increase in their English VOT values for /k/ after their period abroad. Before SA, these learners produced /k/ with short-lag, L1-like values, and after SA, VOT duration reflected intermediate L1-L2 values.

This was the only significant change observed for VOT in Avello and Lara, and the only case in which VOT duration before SA still reflected an L1-like pattern. The rest of pre-SA VOT measures already reflected intermediate L1-L2 duration values for both learner groups, and they remained stable after SA. However, Avello and Lara (forthcoming) failed to find a differential effect of length of stay (three versus six months) on L2 VOT production; i.e., when intermediate VOT values were already displayed by both learner groups before SA, a period of six months abroad did not result in longer VOT duration as compared with a period of three months, VOT values remaining equally unchanged after each SA programme.

Findings of cross-linguistic influence on VOT duration are an example of the posited differences between typical monolingual linguistic competence and the linguistic competence characteristic of bilingual (or multilingual) speakers. Results from different studies have shown that patterns of speech perception and production in bilingual speakers should not be expected to reflect purely monolingual norms (Flege & Eefting, 1987; Flege, 1987; MacKay et al., 2001). As noted by McAllister et al. (2002), L2 learners "are apt to produce and perceive L2 phonetic segments differently than do individuals who are monolingual native speakers of the target L2" (p.230).

As far as VOT laryngeal timing is concerned, typical monolingual production involves, for example, short duration values for voiceless plosives in unaspirated languages such as Spanish and Catalan, and long duration values in aspirated languages such as English. However, as already commented, L2 production by bilingual speakers of languages with different VOT patterns has been commonly found to result in compromise values which are intermediate between typical monolingual values, reflecting the phonetic properties of both the L1 and L2 systems.

This is the result reported in the present study for L1 Spanish/Catalan learners of English, and similar results have been reported for other language pairs, such as Italian/English bilinguals (MacKay et al. 2001), or French/English and English/French bilinguals (Flege & Hillenbrand, 1984; Flege, 1987). Evidence of the production of these compromise VOT values by bilingual speakers has been accounted for as the result of the development of 'merged categories' within the framework of Flege's SLM (1995, 2002).

According to the SLM (see section 1.2.2.2), the influence of the learner's L1 system, with already existing and well-established L1 categories, would hinder the creation of new L2 categories with totally L2-adjusted values in cases of L1 and L2 category similarity, which is the case of plosives in English and Spanish/Catalan. The L1 and L2 categories are hypothesised to share a common phonetic space, and their interaction would lead to a restructuring in this shared phonetic space causing the development of merged categories. These merged categories, or diaphones, have characteristic values which reflect both L1 and L2 patterns, differing therefore from purely L1 and L2 norms, and they would underlie learners' production in the two languages.

The intermediate L1-L2 values found in the present study for VOT in L2 production are thus consistent with the existence of merged categories in the learners' phonetic system, and are in line with previous research documenting similar L1-L2 interaction effects on VOT production. For instance, MacKay et al. (2001) found that L1 Italian learners of English who produced English voiced plosives with more L2-adjusted values, i.e., without prevoicing, would also produce Italian voiced plosives with less prevoicing, reflecting greater influence of English VOT patterns for voiced plosives on their L1 production as their English production was more L2-like. Flege (1987) analysed the production of VOT in voiceless

plosives by L1 French learners of English and by L1 English learners of French, and showed that VOT duration values in English and French for both groups were intermediate between typical native-like norms in monolinguals.

Concerning the production of the English vowel contrasts /i:-ɪ/ and /æ-ʌ/, results failed to yield evidence of significant improvement in the realisation of these English contrasts in terms of more robust distinctions either in vowel quality or duration. These results parallel those for VOT, and could likewise be explained on the basis of the interaction between the L1 and L2 systems.

Regarding vowel quality, the front-central contrast /æ-ʌ/ presented a positive trend towards increased ED values, although non-significant. The opposite pattern was observed for the high-front contrast /i:-ɪ/, as it presented a decrease in ED values following SA which reached significance. Learners' EDs remained relatively short at the two testing times, differing from the significantly longer EDs produced by the NSs, and reflecting rather similar height and frontness values for each pair of contrastive vowels. These reduced EDs suggested that the two vowels in each contrast shared a similar phonological space, pointing to an inability on the part of the learners to correctly implement the differences between these contrasting L2 vowels.

As for duration, DDs remained unchanged between Pre-test and Post-test. DDs for /i:-ɪ/ were larger than for /æ-ʌ/, which is in line with NSs' patterns, since inherent duration differences between /i:/ and /ɪ/ are typically larger than between /æ/ and /ʌ/ (Hillenbrand et al., 2000). There were no significant differences between learners' and NSs' DD values for /æ-ʌ/. DD values for learners' /i:-ɪ/ were significantly shorter than those

produced by NSs', although learners did produce tense /i:/ as significantly longer than lax /ɪ/ at the two testing times.

These results are thus in line with previous research analysing the acquisition of the English contrasting vowels /i:-ɪ/ and /æ-ʌ/ by L1 Spanish and Catalan learners (Avello, 2010a; Bohn, 1995; Cebrián, 2006; Cebrián, Mora, & Aliaga-García, 2011; Fullana, 2005; Mora & Fullana, 2007). Findings from these studies have shown that these learners have great difficulty in making use of spectral information for the differentiation of these contrasting vowels, as well as a tendency to over-rely on duration as a distinctive cue in the distinction of /i:-ɪ/, contrary to native speakers' use of quality as the main distinctive feature.

As pointed out in the Methods section, both Spanish and Catalan lack the spectral and duration features that distinguish the vowels in these English minimal pairs. The English vowel contrasts /i:-ɪ/ and /æ-ʌ/ are located in sections of the vowel space in which there is only a single L1 category, /i/ and /a/ respectively. Therefore, the problems facing these learners in the acquisition of these vowels are generally attributed to the assimilation of the two distinct L2 vowels in each contrast to a single L1 vowel, due to the phonetic similarity between the L1 and L2 sounds. This is consistent with the postulates of Best's PAM (1995), which claims that L2 sounds are initially assimilated to the closest similar-sounding L1 sounds, and Flege's SLM (1995, 2002), which states that L2 learners need to perceive some differences between similar L1 and L2 sounds in order to be able to establish new L2 phonetic categories leading to more native-like production.

Similarly to Flege's SLM, Best's PAM model of cross-linguistic perception (Best, 1995) and its extended version for L2 acquisition, PAM-L2 (Best & Tyler, 2007) posit that the learners' L1 exerts an influence on

the way the L2 is perceived due to the existence of well-established L1 categories (see section 1.2.2.1). Different patterns of assimilation between L1 and L2 phones are postulated, each one corresponding to a pattern of discrimination for L2 phonological contrasts. The case of the analysed contrasts /i:-ɪ/ and /æ-ʌ/ could be an example of SC assimilation, whereby the two distinct L2 phones are perceived as exemplars of the same L1 category (/i/ for the tense-lax contrast, /a/ for the front-central contrast), and are thus assimilated to it. This notion of 'assimilation' in PAM-L2 is comparable to the notion of 'equivalent classification' in SLM. Equivalent classification would lead to the creation of merged categories with mixed values representing the properties of the L1 and L2 phones, as seen with regards to VOT above. This would prevent the creation of new categories for /i:-ɪ/ and /æ-ʌ/ with better-adjusted L2 values, resulting in the observed non-L2 production patterns.

McAllister et al. (2002) noted that from Flege's SLM can be derived a 'feature hypothesis', which they explicitly stated as follows: "L2 features not used to signal phonological contrast in L1 will be difficult to perceive for the L2 learner and this difficulty will be reflected in the learner's production of the contrast based on this feature" (McAllister et al., 2002: 254). This hypothesis stresses the role of L1 transfer as a determinant factor in L2 phonological acquisition, and it implies that L2 learners will not readily make use of features that are not contrastive in their L1.

In contrast, Bohn (1995) claims that the categorisation of L2 sounds is not determined completely by the L1 system, and that other factors related to general auditory strategies should also be taken into account. Based on the results of a number of experiments on L2 vowel categorisation (Bonh, 1995; Bohn & Flege, 1990), he posited the 'desensitization hypothesis', according to which "linguistic desensitization to spectral differences between vowels causes listeners to differentiate foreign vowel pairs on the

basis of duration differences" (Bohn, 1995: 295). This hypothesis followed from the finding that whenever subjects seemed to be unable to use spectral differences to distinguish an L2 or non-native contrast, e.g., because the L2 had two vowels covering an acoustic space where there was only one L1 vowel, they would resort to duration cues regardless of whether duration was used contrastively in their L1. He assumes that this was due to the fact that "duration cues in vowel perception are easy to access whether or not listeners have had specific linguistic experience with them" (Bohn, 1995:294).

The learners in the present study failed to produce the non-L1 spectral features distinguishing /i:-ɪ/ and /æ-ʌ/, but they were able to make use of durational differences between /i:/ and /ɪ/,even though duration is not used contrastively in their L1 vowel systems either. The fact that EDs for /i:-ɪ/ actually decreased after SA, whereas DDs remained unchanged, points towards an overreliance on duration in these learners' implementation of the /i:-ɪ/ contrast. As mentioned above, these results are in line with previous studies analysing the acquisition of these English vowels by L1 Spanish/Catalan learners (Cebrián, 2006; Cebrián, 2007; Fullana, 2005; Mora & Fullana, 2007), and together they provide evidence which is consistent with Bohn's desensitization hypothesis. The trend towards an increased ED for /æ-ʌ/, although non-significant, may be due to the lack of interference of duration, since duration differences between these contrasting vowels are rather irrelevant in comparison with /i:-ɪ/ (Hillenbrand et al., 2000).

It should also be pointed out that the two vowels in both the /i:-ɪ/ and /æ-ʌ/ contrasts seemed to experience a parallel displacement, resulting in closer and more fronted points of articulation. That is, all four vowels were shifted up and forwards within the vowel space. The fact that this displacement affected the four vowels in a similar way could indicate that

there was some other type of restructuring taking place in the learners' phonological system. The observed vowel shift could signal, for example, changes in their vowel system as a whole, or it could be the result of changes affecting consonantal place and manner of articulation.

Taken as a whole, the results in the present study regarding VOT and vowel production contrast with the findings reported in most studies assessing the effect of SA on other domains of L2 acquisition. As pointed out in section 1.1.2, SA has been generally found to have a clear positive impact on learners' linguistic development, particularly when compared with the usually more modest effects of an FI context. Previous findings have provided evidence of SA gains in different linguistic domains, such as lexical development (Collentine, 2004; Llanes and Muñoz, 2009; Milton & Meara, 1995), writing (Barquin, 2012; Pérez-Vidal & Juan-Garau, 2011; Sasaki, 2009) or listening (Allen & Herron, 2003; Beattie, Valls-Ferrer, & Pérez-Vidal, forthcoming). SA has been found to be likewise beneficial for the development of L2 oral competence regarding, for instance, overall oral proficiency, and particularly more native-like fluency (Brecht et al. 1995; Pérez-Vidal & Juan-Garau, 2007, Segalowitz & Freed, 2004; Valls-Ferrer, 2011). This is in keeping with the common expectation that oral production is particularly likely to improve during SA, as it is assumed to be one of the most practiced skills while abroad and to specially benefit from the massive L2 exposure that SA may offer.

However, our results are consistent with the scant research analysing L2 phonological development during SA. These studies have provided inconclusive evidence regarding the potential benefits of SA on improved L2 perception and production at the segmental level (Avello, 2010a; Díaz-Campos, 2004; Højen, 2003; Lord, 2010; Pérez-Vidal et al., 2011; Sanz et al., 2013; Simões, 1996), despite the positive SA outcomes that seem to accrue in most other skills. In addition to the influence of the L1 sound

system explained above, the lack of a more positive impact of SA on the production of the vowel contrasts /i:-ɪ/ and /æ-ʌ/ and on VOT laryngeal timing patterns might have been due to the fact that the relevant phonetic features for the correct categorisation of these L2 categories were not salient enough for the learners in the SA context, probably as a consequence of the mainly communicative nature of study abroad.

As pointed out by Pérez-Vidal (2011) in her description of the SA macro-level features, the type of interaction in which learners engage themselves during an experience abroad is mainly meaning-oriented. This is true for the content-based classroom lectures they have to attend, and especially so for the out-of-class conversational exchanges with native speakers they have to face on a daily basis.

As part of their SA programme, the learners were required to attend courses dealing with similar contents to those they followed in their home university, with the difference that these courses were taught through the medium of English in classes where most of the students were English native speakers. The learners had thus to direct their attention to the content being taught and focus on the teachers' explanations in order to be able to follow those courses and complete their assignments.

Besides, learners had to take part in multiple out-of-class interactions and perform the typical everyday tasks which are inherent to the fact of having to move to a different country and stay there for a sustained period of time. For instance, they would have to look for accommodation, open a bank account, purchase a new mobile phone, go to the supermarket, take care of the paperwork related to their SA programme at their host university, etc. In dealing with all these situations, they would be facing the challenge of "assuming different social roles, within a variety of

human relationships, and in a myriad of social domains" (Pérez-Vidal & Juan-Garau, 2011: 161).

In this setting, where learners are pressed to cope with the needs of everyday life and their main concern is to comprehend the input they receive and get their own message across, it is very likely that L2 processing will be too demanding for learners to pay attention to very specific aspects of form, such as fine-grained acoustic features. From an IP perspective, VanPatten (2007) points out L2 learners' limitations in terms of information storage and processing capacity during conversational interaction, and how these limitations result in a primacy of meaning over form, so that learners will show a tendency to focus their attention on the processing of content words with meaningful information over aspects of form which are not essential for comprehension.

This primacy of meaning over form could underlie the mixed results which have been found regarding SA outcomes in areas such as phonology. Grammar is another L2 domain for which SA outcomes have also been inconclusive, paralleling the mixed results for phonology. For example, findings in Howard (2006) indicated greater gains in French subject-verb agreement for a group of SA learners over a group of AH learners receiving FI. In contrast, DeKeyser (1991) found no evidence of superior gains for SA students over FI students in the acquisition of Spanish copulas *ser-estar* (to be), whereas Collentine (2004) found a superiority of FI over SA in the acquisition of grammatical features which are usually emphasised by the Spanish curriculum (e.g., verb morphology and subordination). A similar superiority of FI over SA has been found in some studies analysing perceptual gains in L2 phonology, such as Pérez-Vidal et al. (2011), or Mora (forthcoming), which reported significant gains in segmental discrimination during FI but no changes during SA.

Such results in phonology and grammar suggest that FI may foster greater progress in the acquisition of discreet formal units of the L2, especially if they are emphasised in the foreign language curriculum, since learners in an SA context are drawn to pay attention to aspects related to content which they may consider more relevant for conversational interaction. Thus, in the absence of explicit instruction on L2 phonology, it may be difficult for SA learners to focus their attention on specific L2 forms (unless somehow enhanced during interaction, see section 4.1.2 below on error rate scores). That could have been the case of the fine-grained acoustic measures of VOT and vowel quality and duration analysed in the present study, as they might not have been perceived by learners as essential for successful communication.

In addition, learners might have realised that a foreign accent may not necessarily be detrimental for the intelligibility of L2 speech (Derwing & Munro, 1997; Munro, 2008), and consequently they could have focused their attention on the development of other L2 skills while abroad. This would be consistent with the results from other studies within the SALA project which do report major SA gains in domains such as listening (Beattie et al., forthcoming), fluency (Trenchs-Parera, 2009; Valls-Ferrer, 2011), writing (Pérez-Vidal & Juan-Garau, 2009, 2011), or intercultural awareness (Merino & Avello, forthcoming).

Another factor which could have hindered greater development towards more native-like pronunciation of the analysed categories is learners' patterns of language use while abroad, particularly concerning the use of their L1. As explained in the Methods section, the learners in the present study travelled to their SA host university together with other UPF students who shared the same destination, and with whom they were normally acquainted as they attended the same Translation and Interpreting classes at UPF. It is therefore not surprising that they reported

spending some amount of time with their UPF peers, specifically an average of 1.5 on a 3-point scale (1='most time', 3='little'). Moreover, learners further reported having frequent contact with their families during their stay abroad. On a scale from 'a' to 'e' ('a'='more than once a day' and 'e'='none'), most learners reported 'b' ('a few times a week'), at the end of the scale indicating frequent contact, whereas none reported 'd' or 'e', at the opposite end of the scale. These L1 contact patterns indicate that learners continued to make frequent use of their L1 even while abroad. Sustained L1 influence during SA could have thus prevented changes towards the formation of new L2 categories with more target-like values, since the observed intermediate L1-L2 values discussed above pointed to the existence of merged categories both before and after SA. Learners reported being also frequently exposed to non-native L2 input as well, measured through a scale from 1="never" to 5="very often" (*M*=4.10). Being Erasmus students, they were very likely to be in contact with other Erasmus students taking part in similar exchange programmes in the English-speaking host university. Hence, exposure to this L2 input containing non-TL features could have further contributed to the lack of significant development towards more native-like production of the analysed L2 sounds.

4.1.2. Error rate scores

This section reports the results from the phonetic analyses conducted on the learners' speech samples used for the rating task (see section 3.4.2.2), so as to further assess the effect of SA on pronunciation accuracy by looking at a wider range of phonological phenomena than those addressed in the acoustic analyses just reported in section 4.1.1 above, both at the segmental level (including vowel and consonant insertions, deletions and phonemic substitutions), as well as at the suprasegmental level of lexical

stress (see data analysis procedure and mispronunciation examples in section 3.4.1.2).

4.1.2.1. Results

Error rate scores were calculated for each L2 English learner by computing the total number of pronunciation errors at Pre-test (prior to SA) and at Post-test (immediately following SA). PrEr ranged between 0 and 9 at both data collection times, with considerable inter-subject variability, as indicated by the relatively high *SD*s shown in Figure 8. This figure also illustrates a decrease in PrEr from Pre-test (*M*=3.95) to Post-test (*M*=3.30). A paired-samples t-test conducted to assess the effect of SA on learners' error rate scores revealed that the decrease in pronunciation errors turned out to be significant [$t(22)=2.135$, $p=.044$], the eta squared statistic ($\eta^2=.17$) suggesting a large effect size.

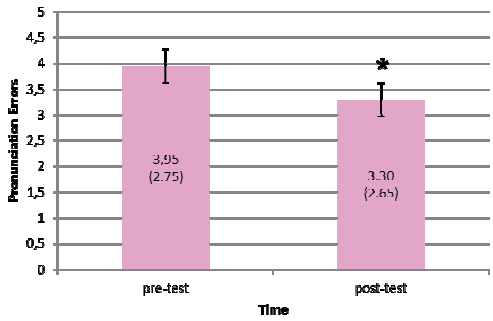

Figure 8. Mean number of PrEr for NNSs at Pre-test and Post-test (range=0-9). *SD*s in parentheses.

This finding seems to indicate that the 3-month SA period had a positive impact on the learners' pronunciation accuracy in terms of a reduction in pronunciation errors at the segmental level and at the level of lexical

stress. Nonetheless, despite the overall significant improvement observed in the group error rate scores, the high *SD*s reported above pointed towards substantial individual variation in the learners' level of pronunciation accuracy, in such a way that some ceiling effects might have existed for learners with better pronunciation accuracy level (it was noted that some learners produced very few PrEr already at Pre-test). Hence, further analyses were conducted in order to explore whether there was an effect of learners' pre-SA or onset level of pronunciation accuracy on their SA outcomes in PrEr scores.

Two groups of learners were created according to their onset level of pronunciation accuracy by means of a median split (Isaacs & Trofimovich, 2010; Isaacs & Trofimovich, 2011), so that both groups differed maximally in their Pre-test PrEr scores (Low versus High). Learners whose pre-SA error rate scores were over the median (those who produced a large number of pronunciation errors at Pre-test) were grouped into the low onset accuracy level group (LO-A, n=12), whereas learners with scores below the median (those producing few pronunciation errors at Pre-test) were assigned to the high onset accuracy level group (HO-A, n=11). Table 4 and Figure 9 illustrate the differences between the two learner groups. The LO-A group experienced a decrease in pronunciation errors between Pre-test and Post-test, whereas the mean PrEr scores for the HO-A group remained unchanged at the two testing times. In terms of SA gains, this represents mean gains of 1.25 (SD=1.4) for the LO-A learners, but 0 gains (SD=1.1) for the HO-A learners, the difference in SA gains between both groups being significant [$t(21)$=-2.22, p=.038, η^2=.190].

Group	Time	Mean	SD
HO-A (*n*=11)	Pre-test	1.64	1.2
	Post-test	1.64	1.3
LO-A (*n*=12)	Pre-test	6.08	1.9
	Post-test	4.83	2.7

Table 4. Descriptives for error rate scores (PrEr) before (Pre-test) and after SA (Post-test) as a function of onset pronunciation accuracy level (Low:LO-A, High: HO-A).

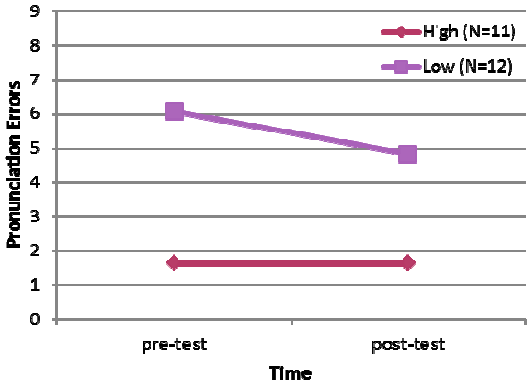

Figure 9. Changes in error rate scores for the LO-A group (Low, *n*=12) versus the HO-A group (High, *n*=11).

A mixed ANOVA was run with *time* (Pre-test, Post-test) as within-subjects factor and *learner group* (LO-A and HO-A) as between-subjects factor. This analysis yielded a significant main effect for *time* [$F(1, 21)=4.92$, $p=.038$, $\eta^2=.190$], a significant main effect for group [$F(1)=26.97$, $p<.001$, $\eta^2=.562$], and a significant *time* x *group* interaction [$F(1, 21)=4.92$, $p=.038$, $\eta^2=.190$]. Paired-samples t-tests indicated that the decrease between the mean pre- and post-SA PrEr scores for the LO-A learners was significant [$t(11)=2.92$, $p=.014$, $\eta^2=.436$]. Yet despite this significant decrease in pronunciation errors for the lower level learners after the SA period, the significant difference in PrEr scores observed

between the two groups at Pre-test [$t(21)$=-6.68, p<.001, η^2=.679] was still maintained at Post-test [$t(21)$=-3.68, p=.002, η^2=.392].

Taken as a whole, these results reveal different SA outcomes in PrEr scores for learners who differed in pronunciation accuracy level prior to their SA, in such a way that the overall improvement observed in the group analysis was due to the improvement obtained by those learners who started out their SA experience with a lower accuracy level. There was no effect of SA on PrEr scores for the high level learners, as they obtained the same mean scores both before and after SA, producing a low number of pronunciation errors at the two testing times. In contrast, the low level learners presented a significant reduction in pronunciation errors as a result of SA. Despite this improvement, however, the low level learners continued to produce a significantly higher number of pronunciation errors than the high level learners after the SA period, which means that their SA gains still did not allow them to catch up with their more accurate counterparts.

4.1.2.2. Discussion

Taken together, these results suggest a positive impact of SA on the development of pronunciation accuracy as measured through error rate scores, since a significant decrease was found in the number of pronunciation errors produced by learners between Pre-test and Post-test. The outcome from the PrEr scores thus contrasts with the outcome from the acoustic measures reported in section 4.1.1 above, which failed to reveal a significant improvement in the production of the vowel and consonant L2 categories analysed, as well as with general SA findings regarding L2 phonology, which have not provided consistent evidence supporting a large effect of SA on improved pronunciation. These

divergences regarding SA outcomes can be attributable to differences in the scope of the selected object of analysis and how it was measured.

While acoustic analyses can provide fairly precise objective measurements for many different aspects of speech at the subphonemic, phonemic and suprasegmental levels, they may also have their drawbacks. The acoustic analyses conducted for the present study focused on a limited set of discrete units or phonetic features selected a priori, i.e., during the design of the study, from the whole L2 phonological inventory. These analyses may fail to capture actual changes in learners' production if such changes have accrued for features other than those which were included in the analysis. Conversely, it is possible that acoustic analyses might capture very subtle acoustic changes or nuances in some features of learners' speech which could nonetheless not be linguistically relevant, as they might not be perceived by listeners.

In this study, however, a significant effect of SA was observed when using an error rate measure covering a wider-reaching object of analysis which included a series of phonological phenomena (phonemic insertions, deletions, substitutions, lexical stress) as they occurred in the learners' actual speech production (see types of errors in section 3.4.1.2). These phenomena might have been more salient to L2 learners or perceived as more important for fluent communication than the other, very specific and fine-grained acoustic features, which might, in turn, have been particularly subjected to influence from the L1 phonological system, as discussed in section 4.1.1 above. Despite the absence of explicit formal instruction which could overtly draw learners' attention to specific features of L2 pronunciation, it seems that the SA context allowed the learners to attend to some aspects of the L2 phonological system resulting in more accurate L2 performance, and making it possible for them to overcome the influence of their L1 system in some cases.

The improvement found in the present study for PrEr scores could be explained on the basis of learners' patterns of L2 use and exposure to native input. We saw in section 4.1.1 that learners reported a relatively high degree of L1 use which was hypothesised as possibly detrimental for improvement in the L2 VOT and vowel quality and duration features analysed due to the influence of the L1 system. However, learners also reported a considerably high degree of L2 use in terms of contact with native speakers and exposure to native input. Regarding contact with native speakers, learners reported a mean of 4 on a 5-point scale, where 1 was "never" and 5 was "very often", most reporting either 4 or 5. Learners also experienced further exposure to native input through the media, as they reported a mean of 3.30 on the same scale. These patterns of L2 contact and exposure to L2 input suggest a predisposition on the part of the learners to engage in interaction with native speakers and to avail themselves of TL input. Learners' readiness to take advantage of the contact opportunities with the TL available in the SA context is a crucial factor for successful L2 linguistic development, as emphasised in section 1.1.2.2 when discussing the SA micro-level features.

As noted in section 1.1.2.1 dealing with the SA macro-level features, the main asset of the SA context from the perspective of the language learning process is precisely the fact that it provides excellent opportunities for oral practice through interaction in meaningful, communicative situations, and massive exposure to rich and high quality TL input. Interactional approaches (Gass & Mackey, 2007; Gor & Long, 2009; Long, 1996) have established a connection between conversational interaction and progress in L2 acquisition through the process of *negotiation for meaning*, during which NSs' input is modified in order to facilitate successful communication with L2 learners, resulting in a type of foreigner register that is mainly well-formed, and constitutes therefore a source of positive

evidence for the correct grammatical structures in the L2 system. During the course of communication in an SA setting, interactional patterns may likewise bring about negative evidence for structures which are ungrammatical and incorrect in the TL, e.g., through overt or implicit corrective feedback from NSs to non-target like forms in learners' output. These interactional processes may lead to enhanced saliency of specific TL features and structures which might be particularly relevant for fluent communication.

Hence, within this interactional framework, the type of meaning-oriented communication characteristic of SA might have favoured the observed decrease in mispronunciations since the interplay between input, output and feedback could have enhanced the contrast between the deviant or non-native like forms in the learners' interlanguage and the TL forms present in the input to which they were exposed. For example, as a consequence of the modifications used by NSs to accommodate the input to the learner, foreigner register is usually characterised by a slower rate of deliverance, repetition, and more careful and clearer articulation. All of this can result in more robust phonetic cues at the segmental and suprasegmental levels, so that the TL forms may become perceptually more salient and easier to process. The saliency of a TL form might likewise be increased through frequency of occurrence in the native input. As revealed in research analysing the role of input frequencies in L2 acquisition, high-frequency linguistic features, items and patterns might be more easily acquired since they are more likely to be noticed (Ellis, N. C., 2002; Gor & Long, 2009). Thus, token frequency of a TL item or form might have drawn learners' attention, allowing them to notice the divergences between the native form and the non-native form in their interlanguage.

Another way in which the divergences between TL and interlanguage forms could have been brought into the learners' attention is through their interlocutors' feedback to a wrong or non-native like form in learners' output, especially in those cases in which the deviation of the interlanguage form from the TL form might have negatively affected fluent communication, which is the primary goal of learners in a communicative setting such as SA. Feedback from native speakers might have taken up different forms, such as requests for repetition or clarification if the non-TL phonological encoding hindered the intended meaning, or negative feedback provided by means of either overt correction, or indirect correction via recasts. In fact, since the learners in the present study were language specialists and generally reported great motivation to improve their knowledge of English, it is highly possible that they took an active involvement in their language learning process during SA which might have facilitated their noticing of linguistic features in the TL which differed from their own interlanguage features.

The processes and input modifications that take place during conversational interactions in SA might thus favour the saliency of language items which could then be more easily noticed and attended to by learners. It is a widely accepted assumption in the SLA field that some degree of attention to form and noticing of linguistic features and rules are important processes for the acquisition of an L2, as postulated by Schmidt through the Noticing Hypothesis (2001), and pointed out in approaches stressing the importance of focus on form (Doughty, 2003). During oral communication, learners need to focus their attention on the relevant pieces of information in order to be able to understand the intended meaning of the message and follow the conversation. Selective attention would allow them to filter out irrelevant acoustic properties of the speech

signal while concentrating on the relevant phonetic features, thus facilitating the encoding of new L2 items in memory.

But even though the acquisition of a new item may begin with attentional focus and controlled processing, progress in L2 acquisition involves also the development of automatised processing, which can be achieved with repetition, frequency and practice (Segalowitz & Hulstijn, 2005; DeKeyser, 2007a). In this sense, the interactional nature of SA can provide learners with the opportunities for practice needed for the automatisation of learning that can lead to L2 improvement, particularly as far as oral competence is concerned, as is the case of the reduction of pronunciation errors. During oral communication learners are able to access both auditory and visual information regarding the phonetic features of a TL form and its correct articulation, and increased practice in oral production specifically facilitates rehearsal and repetition of the articulatory gestures required for accurate production.

From a general skills acquisition theory perspective, DeKeyser (2007c) points out that the study abroad setting is an optimal context for the automatisation of already proceduralised linguistic knowledge, as the massive and varied type of practice that SA can offer would act as a facilitative factor for the transition between rule proceduralisation and rule automatisation. That is, under this view, learners would benefit the most from their experience abroad if they have previously acquired a minimum procedural knowledge of the TL system, equivalent to at least an intermediate level of proficiency (DeKeyser, 2007c), which might endow them with the skills necessary to engage in successful communication while abroad so as to push their L2 learning process further. In this sense, DeKeyser stresses the important role of previous explicit language instruction in an FI context at home, as this learning context is typically characterised by “an explicit instructional treatment of language, and

learning mostly takes place through structural practice" (Pérez-Vidal & Juan-Garau, 2011: 161), and it constitutes therefore an ideal means to foster the acquisition of declarative L2 knowledge and its subsequent proceduralisation.

We should recall that this was precisely the case of the undergraduate learners in our study, who had received around 1,100 hours of foreign language instruction on English during primary and secondary education at home, plus an additional 80-hour FI period consisting of two English language courses during the first year of their degree at UPF prior to their departure abroad. Actually, as noted in section 3.2.1, at the end of the FI period at UPF learners were expected to have an upper-intermediate English proficiency level or B.2 in the CEFR, in order to take part in the SA programme.

The FI context preceding SA thus provided learners with more classroom practice for further proceduralisation of the TL rules and structures so as to acquire a proficiency level that would purportedly place them in a better position for the automatisation of knowledge during their stay abroad. With respect to oral communication and speech development, this FI context might have been particularly relevant to enhance learners' familiarity with the L2 sound system, even though there was no explicit instruction on phonetics or phonology. For example, learners were faced with the challenge of having to follow language courses which were taught exclusively through the medium of English, which is not usually the case in primary and secondary foreign language education settings, where teachers' L1-L2 code-switching is frequent in order to make sure that learners are able to understand the explanations about the L2 grammar. Besides, despite the fact that interaction and amount and variety of L2 input during FI were limited in comparison with the SA context, learners did have opportunities for interaction and exposure to authentic

spoken native input through activities designed for the practice of oral and listening skills, probably to a larger extent that what is usually the case in school and high school, together with the input they received from native speaker teachers.

In this sense, it could be argued that the previous FI learning experience would have played an important role in providing the learners with the necessary language knowledge, skills and resources in order to be able to take full advantage of the possibilities offered by the study abroad period for further L2 development to accrue towards automatisation, and this would be reflected in learners' improved PrEr scores. This is the premise of the SALA project, and as already noted, results from the different SALA studies to date generally support it, since positive effects of SA following the previous FI period have been found for several skills, such as writing (Barquin, 2012; Pérez-Vidal & Juan-Garau, 2011), listening (Beattie et al., forthcoming), or oral fluency (Juan-Garau & Pérez-Vidal, 2007; Valls-Ferrer, 2011).

Hence, through their engagement in conversational interactions, learners might have progressively realised the differences between the increasingly salient TL forms in the native input, and the forms in their interlanguage as they were initially reflected in their output. Increased practice would have then allowed learners to proceed from more attentional controlled processing to gradually more automatic processing or, in terms of skills acquisition theory, to automatise previous procedural knowledge (DeKeyser, 2007a), in line with the idea that "with frequent repetition [...] comes the ability to perform a task much more easily and [...] more quickly" (Altarriba & Basnight-Brown, 2009: 125).

This increased automaticity would have triggered a gradual category restructuring in the learners' interlanguage sound system, reflected in the

observed decrease in mispronunciations. Altarriba and Basnight-Brown (2009), following McLaughlin and Heredia (1996),emphasise that automaticity leads to restructuring in the interlanguage system of advanced L2 learners, as their internal representations and rules are rearranged and reformulated in order to reflect their increased language proficiency. The findings reported in section 4.1.1 above regarding VOT and vowel production, which pointed to the existence of merged categories with intermediate L1-L2 values, would similarly constitute evidence for restructuring in the learners' sound system.

In addition, the nature of the cognitive processes which are mainly at work during SA could have likewise favoured the restructuring in the learners' interlanguage leading to the observed improvement in error rate scores. When discussing the macro-level features of SA, Pérez-Vidal (2011) pointed out that the primarily communicative nature of the SA context, with its focus on meaning, triggers the activation of implicit learning mechanisms. Implicit learning of an aspect of language takes place incidentally when paying attention to a different aspect (e.g., learning of form while attending to content) and without awareness of what is being learnt (Hulstijn, 2003). Implicit learning thus differs from more controlled processes in the sense that it does not involve as much attentional focus and resources. One type of such implicit learning could be 'statistical learning', which consists of "the absorption of statistical regularities in the environment through implicit learning mechanisms" (Williams, 2009: 328). For example, exposure to the great amount of both oral and also written input available in the SA context could facilitate learners' L2 orthographical-phonological encoding, such as the mapping of "th" onto /ð/, or that of "v" onto /v/ in the case of learners or English.

But what would constitute the restructuring in the learners' interlanguage leading to the observed decrease in mispronunciations? According with

the re-deployment hypothesis (Archibald, 2005; 2009), L2 learners would be able to re-deploy the knowledge and structures of their L1 phonology in order to assist the process of L2 phonological acquisition. In some cases, the acquisition of a new L2 form or category may be the result of the re-combination of L1 features which are integrated into a new feature composition for the L2 category. For example, English /v/, as in 'tra**v**eller', involves the integration of the features [labiodental], [+continuant] and [+voice]. The phoneme /v/ thus contrasts with /f/, which has the feature [-voice], a phonological opposition in terms of voicing that Spanish and Catalan both lack, having only the voiceless phoneme /f/ in their phonological inventories.[12]

The features [labiodental], [+continuant] and [+voice] are all nonetheless present in the learners' L1 sound systems of both Spanish/Catalan as part of different configurations for different categories within the two phonemic inventories. The features [labiodental] and [+continuant] are combined in the L1 category /f/. The feature [+voice] appears in quite a number of L1 obstruents, such as the categories /b,d,g/, both in their realisation as plosives [b,d,g] in onset position (following pause and [+stop] obstruents), where voicing is the main distinctive feature with regards their voiceless counterparts /p,t,k/, as well as in their word-medial and coda position [+continuant] allophones [β,ð,ɣ]. The feature [+voice] appears likewise in other [+continuant] L1 obstruents, such as the phonemes /z/ and /ʒ/ in Catalan (e.g., 'casa' /ˈkazə/, 'gel' /ʒɛl/), and also as a result of context-dependent assimilation phenomena, like those affecting sequences where /s/ is followed by a [+voice] obstruent, in which /s/ has a voiced realisation due to coarticulation effects of the following voiced sound (e.g., 'còsmic' [ˈkɔzmik], 'desde' [ˈdezðe]).

[12] There is a phoneme /v/ in some Catalan dialects, but this phoneme is not part of the phonological inventory of the variety spoken by the subjects in this study.

Successful production of /v/ thus involves the integration of the feature [+voice] together with the features [labiodental] and [+continuant], resulting in a voicing opposition between /f/ and /v/.

In other cases, improved production would be the result of learners' acquisition of new L2 distributional patterns for phones which exist both in the L1 and L2, but with a different distribution in each language. This is what happens, for instance, with the [+stop] phone [d] and the [+continuant] phone [ð]. English /d/, as in 'imme**d**iately', and /ð/, as in '**th**en' or '**th**e', are two independent phonemes and they appear in different contrasting words. In Spanish and Catalan, [d] and [ð] are two allophones of the phoneme /d/ in complementary distribution, [d] occurring in syllable onset position preceded by a pause or by another [+stop] obstruent with the same place of articulation, and [ð] occurring in all other positions. Both [d] and [ð] can thus alternate as realisations of /d/ in the same word as a function of variations in phonetic context (e.g., 'día' [dia], 'este día' [ˈesteˈðia]). In order to overcome these L1-constrained patterns, learners need to correctly categorise these two phones as different L2 phonemes. This involves the phonologisation of the feature [+stop] for /d/ and the feature [+continuant] for /ð/, and the subsequent implementation of these features in the production of each category, so that the full stop /d/ is produced in intervocalic position in words like 'immediately' instead of the continuant /ð/, and the continuant /ð/ appears in syllable onset position after a pause in the words 'then' and 'the' instead of the full stop /d/.

Robust phonetic cues play an important role in order to overcome the constraints of the L1 system, as they favour the processing of the L2 input by facilitating the retrieval and coding of segments and segmental sequences from the speech signal (Archibald, 2005). In his review of cue robustness, Wright (2004) refers to these cues as 'information in the

acoustic signal that allows the listener to apprehend the existence of a phonological contrast' (p. 36). Phonetic cues are robust when they are somehow enhanced, e.g., because they are redundant, because they have great auditory impact or show great resistance to environmental masking.

The transmission of the speech signal in normal communication of the type taking place in a natural SA setting is affected by environmental masking, such as background noise, as well as factors related to listeners' reception of the signal, such as occasional distractions that may result from a processing overload, or the very functioning of the auditory system. Wright (2004) notes that robust phonetic cues facilitate the processing of the speech signal in natural conditions as they are more resistant to the signal degradation that is likely to occur in these non-ideal conditions.

Both internal cues (within a segment), such as the formant structure of a vowel, and transitional cues (between adjacent segments), such as vowel formant transitions, provide important information for the phonological encoding of the speech signal (regarding, e.g., voicing, and place and manner of articulation). Transitional cues may be particularly robust due to the redundancy that results from articulatory overlap and to the changes that occur in signal modulation, both of which would increase the auditory saliency of the cues.

As explained by Wright (2004), speech production is characterised by gestural overlap affecting the articulation of adjacent sounds. These coarticulation effects result in cue redundancy, especially in sequences where consonants and vowels are in contact, as information for the categorisation of adjacent segments appears encoded in the internal cues within the segments and in the transitional cues between those segments. For instance, vowel formant transitions contain rich acoustic information,

as they encode cues for vowel quality but also for place and manner or articulation of the adjacent consonants.

Changes in the modulation of the speech signal along the amplitude and frequency dimensions may also be a contributing factor to the increased robustness and auditory impact of acoustic cues, since the auditory system appears to react to these changes in modulation through an increase in activity that favours the processing of the acoustic information (Wright, 2004). Such differences between low and high amplitude and lower and higher frequencies usually coincide also with the transitions between adjacent segments, particularly in sequences where consonants and vowels appear in alternation. This cue enhancement effect could be greater in the type of foreigner register that is frequently used by native speakers in interactions with L2 learners, as the changes in signal modulation occurring at segment transitions could be more marked due to the slower rate and clearer articulation that usually characterises this type of speech.

It should be pointed out that interactional communication may further enhance cue robustness due to its bimodal nature. As already mentioned, during their engagement in conversation, participants have access to auditory cues regarding the acoustic-phonetic characteristics of the speech signal, and also to visual cues providing information for the articulatory gestures involved in the production of that speech signal. Indeed, humans have been found to integrate both auditory and visual information during speech perception, an interaction encapsulated by the so-called 'McGurk' or 'McGurk-MacDonald effect' (McGurk & MacDonald, 1976), which has documented the effects that a mismatch between auditory and visual cues may have on perception. Phonetic training studies addressing L2 pronunciation instruction have likewise assessed the relevance of visual information in the acquisition of L2 sounds through the comparison of

auditory-only and articulatory-based, audiovisual training methods (Aliaga-García, 2013; Cebrián & Carlet, 2012).

Some studies seem to indicate that robust transitional cues may in some cases be more salient and easier to process than internal cues, particularly when an L2 contrast is encoded with non-L1 features (see Archibald, 2005). Lack of robustness and saliency of internal phonetic cues that are not sufficiently enhanced could further explain some of the differences between the significant improvement in the PrEr scores reported in this section and the lack of significant gains observed in the acoustic measures reported in section 4.1.1.

As already explained, the phonological contrasts /iː-ɪ/ and /æ-ʌ/ are distinguished mainly by the differences in quality reflected in the formant structure of the vowels, which is an internal cue. Vowel formant structure better reflects the relative distance between the different vowel formants in the steady-state of the vowel, where there are no movements of the formants. A clear formant structure with relatively long and stable steady-states for vowels could be obtained, for example, in a controlled laboratory setting where production might be characterised by hyper-articulation induced by pronunciation awareness. However, this is not usually the case in spontaneous communication in natural settings, where the focus is on meaning and there tends to be less awareness regarding pronunciation. In interactions between native speakers, the need to quickly and efficiently get the message across may considerably speed up speaking rate, in such a way that formants may be quite unstable and vowels may have a rather short or even no steady-state at all.

Vowel formants could be similarly affected by communication constraints in native/non-native interactions where a foreigner register is used. Even though, as already mentioned, some adjustments are usually made to this

type of learner-addressed speech in terms of slower rate or clearer articulation to facilitate L2 learners' processing, a predominant meaning-oriented focus would still be present which would rarely result in excessive hyper-articulation. Vowels could have slightly longer steady-states than in more fast-paced speech, albeit vowel formants might continue to present a considerable degree of movement. In these conditions, learners could be sensitive to differences in formant frequencies of contrasting vowels that are fairly distant within the acoustic vowel space (e.g., front versus back, high versus low), but internal formant structure might not be a robust enough cue for L2 learners to be able to encode the frequency differences between the contrasts /i:-ɪ/ and /æ-ʌ/ due to the fact that the two vowels in these minimal pairs are rather close and occupy a similar vowel space.

The discrete nature of segmental units results in their frequency of appearance as part of different patterns and in different phonetic contexts. As mentioned above, some of these contexts, for instance those in which vowels and consonants are in contact (CVCV sequences), particularly favour the presence of robust transitional cues, all of which may contribute to learners' successful processing of segmental information and phonological encoding if they engage in repeated interaction while abroad. In fact, some of the difficult L2 segments for which improvement in error rate scores was observed for the English learners in this study occur in words that can be expected to be recurrent in TL discourse, as they are mostly function words and are therefore fairly frequent in English. That is the case, for instance, of initial /ð/ in 'then' and 'the', which also appears in other high frequency words such as 'this/that' (and their plural forms), as well as /h/ in 'his' (which also appears in the feminine 'her').

Ladefoged (2005) stresses the idea that mapping of a specific sound pattern to a specific TL word or form would result from token frequency in the input, as he notes that "the more a learner of a foreign language hears a particular pattern of sound and identifies it as a particular word, the more strongly does that pattern become part of the group of patterns representing the word" (p. 103). At the same time, repeated output practice would allow learners to practice the different articulatory gestures involved in the production of the L2 forms. Correct mapping of sound patterns onto the corresponding L2 words is particularly relevant for accurate production of lexical stress in polysyllabic words, since word stress is not fixed in English (unlike, e.g., Finnish or French, where stress falls on the first and last word syllables, respectively), and therefore learners need to encode and store information about the stress pattern of a word together with its phonemic composition.

English being a stressed-timed language, stressed syllables are very salient auditorily due to the great contrast between stressed and unstressed syllables, as unstressed syllables are considerably reduced regarding length and intensity. Stress is also robustly encoded in English by means of several phonetic cues, such as increased loudness and longer duration, and especially changes in pitch (Ladefoged, 2003; 2005). English stress prominence would facilitate the recognition and correction of wrong lexical stress patterns, for example in those cases in which L1 Spanish/Catalan learners shift stress from left to right, probably due to L1 influence (e.g. *tra'veller* for *'traveller*, or *imme'diately* for *i'mmediately*). Successful sound pattern-onto-L2 word mapping would likewise allow learners to correct possible deviations from TL phonotactic constraints that may result, for instance, from cases of L1-induced epenthesis or deletions (e.g. onset modification [gw] in 'warmly'), as these would be likewise fairly salient phonetically.

A final comment regarding the outcome of the error rate scores should make reference to the onset level effects that seemed to arise as a function of the learners' pre-departure level of pronunciation accuracy. When learners were divided into two groups based on their pre-SA PrEr scores, the lower accuracy learners in the LO-A group were found to obtain significant SA gains in their PrEr scores, whereas scores for the higher accuracy learners in the HO-A group remained unchanged, as they produced a relatively low number of mispronunciations that signalled a fairly high degree of accuracy. The significant decrease in mispronunciations experienced by the LO-A learners did not allow them, nonetheless, to catch up with the HO-A learners.

Onset or pre-departure language level is identified by Pérez-Vidal (2011) as one of the relevant SA programme features that may influence learners' linguistic outcomes while abroad (see section 1.1.2.3). Interest in the analysis of how learners' onset level may impact their SA gains (or lack thereof) is connected with the issue of the possible existence of a threshold level. Some scholars have claimed that learners would need to acquire a minimum proficiency level in the L2 language in order to obtain the greatest benefit from SA, e.g., by being able to automatise previous procedural knowledge (DeKeyser, 2007c). Llanes (2011) notes that there are clear practical implications for this issue, since knowing the ideal proficiency level required for optimal L2 learning to accrue while abroad could inform better programme design policies in order for learners to benefit the most from their SA experience.

However, results in the SA literature show a general tendency for learners with lower proficiency levels to obtain greater gains during SA than higher level learners (Freed 1995b; Klapper & Rees, 2003; Lapkin et al., 1995; Llanes & Muñoz, 2009). Results in the present study are thus consistent with this generally reported larger effect of SA on learners with

lower L2 proficiency. Even though our learners had a rather advanced level of L2 proficiency (they were required to have at least a B.2 level), those learners with lower pronunciation accuracy level before SA were found to significantly improve their PrEr scores after SA, whereas scores for the learners with higher pre-SA pronunciation accuracy did not change. Our results on pronunciation accuracy are in line with findings from other studies within SALA which have also reported larger SA gains for lower level learners, for instance, in listening (Beattie et al., forthcoming), perceptual phonological development (Mora, forthcoming) or fluency (Valls-Ferrer, 2011).

Results in the present study could nonetheless be due to ceiling effects for the higher level learners. The learners in the HO-A group produced very few pronunciation errors already at pre-test, performing not far from native level. In contrast, the significantly higher number of pronunciation errors produced at pre-test by the learners in the LO-A group gave them more room for improvement as a result of their experience abroad. Another possible explanation is that those learners who already had a rather high onset level of pronunciation accuracy before SA could have improved on other linguistic areas or less salient features for which they might have had a processing advantage over the learners with a low onset level.

4.2. Listeners' FA judgements

This section addresses RQ2: *Is there an effect of SA on L2 learners' speech production as assessed through subjective measures of listeners' perceived FA?* FA measures were obtained by means of the rating task described in section 3.4.2.2 above, and consisted of ratings for degree of FA in learners' speech before and after SA collected from two groups of

listeners (see description in section 3.4.2.1): a group of L1 English NLs and a group of L1 bilingual Catalan/Spanish NNLs. The listeners heard the speech samples produced by the NNSs (n=23) and a group of baseline NSs (n=6) and rated them for degree of accentedness on a 7-point scale (1="native", 7="heavy foreign accent").

4.2.1. Results

First of all, rating consistency was checked for the FA ratings obtained from the two groups of listeners by exploring both intra-rater and inter-rater correlation coefficients.

Regarding intra-rater reliability, a strong *Pearson-r* correlation was found for the listener-based scores assigned at each of the two rating repetitions by the NLs (r=.849 p<.001), as well as by the NNLs (r=.855, p<.001), which indicates that each listener's first and second repetition ratings were strongly correlated, i.e., every listener, whether native or non-native, assigned similar ratings at both repetitions.

Inter-rater reliability was next assessed by conducting an intra-class correlation (ICC) analysis for each group of listeners. Results yielded a high Cronbach's Alpha for both the NLs (α=.991) and the NNLs (α=.996), indicating a high degree of agreement among the listeners in each group.

Figure 10 illustrates the mean perceived FA ratings assigned in the rating task by the two groups of listeners to the two speaker groups, the NNSs (Pre-test and Post-test), and the NSs (baseline). As expected, the ratings accorded to the NSs were very close to 1 (NLs' M=1.04, SD=0.06; NNLs' M=1.29, SD=0.17). In contrast, one-sample t-tests showed that the FA ratings accorded to the NNSs were significantly higher than the NSs' ratings at Pre-test [$t(22)$=18.18, p<.001, η^2=.93 for the NLs' scores;

$t(22)=13.51$, $p<001$, $\eta^2=.87$ for the NNLs' scores], and they remained so at Post-test [$t(22)=17.87$, $p<.001$, $\eta^2=.92$ for the NLs' scores; $t(22)=13.39$, $p<.001$, $\eta^2=.86$ for the NNLs' scores]. This indicates that the two groups of listeners, regardless of their L1, were able to successfully identify the native speakers of English and distinguish them from the non-native speakers.

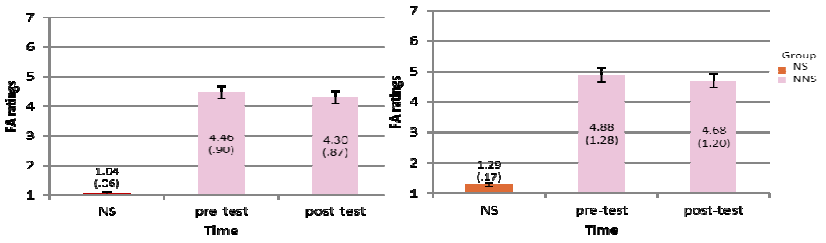

Figure 10. Mean FA ratings for NSs (baseline) and NNSs (Pre-test and Post-test) assigned by the NLs (left) and the NNLs (right). Scores: 1="native", 7="heavy foreign accent".

Both the NLs' and NNLs' ratings pointed towards a slight decrease of perceived FA in the NNSs' oral production between Pre-test (NLs' $M=4.46$, $SD=0.90$; NNLs' $M=4.88$, $SD=1.28$) and Post-test (NLs' $M=4.30$, $SD=0.87$; NNLs' $M=4.68$, $SD=1.20$). However, paired-samples t-tests with testing *time* as the within-subjects factor showed that this decrease was non-significant according both to the NLs' ratings [$t(22)=1.44$, $p>.05$] and the NNLs' ratings [$t(22)=1.31$, $p>.05$]. Interestingly, Figure 10 shows a tendency for the NLs to provide more lenient FA ratings than the NNLs. In order to explore this difference, the FA ratings obtained from the two listener groups were averaged across all participants and compared by means of paired-samples t-tests with *listener group* as the within-subjects factor. This analysis revealed that the NNLs assigned significantly higher FA ratings ($p<.01$) than the NLs, both at Pre-test [NLs' $M=3.76$, $SD=1.62$; NNLs' $M=4.12$, $SD=1.84$; $t(28)=$-

2.88, η^2=.23] and at Post-test [NLs' *M*=3.63, *SD*=1.55; NNLs' *M*=3.99, *SD*=1.76; *t*(28)=-3.39, η^2=.29].

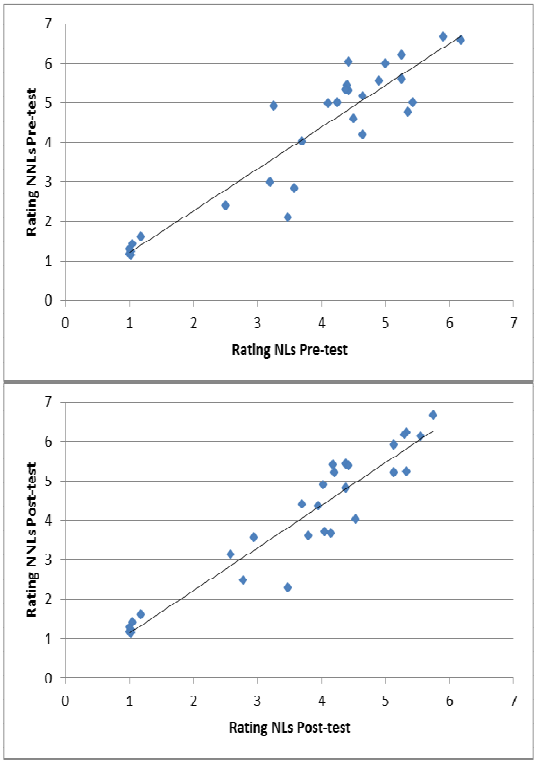

Figure 11. Correlations at Pre-test (upper graph) and at Post-test (lower graph) between the FA ratings assigned to NSs (*n*=6, at the bottom-left of the scatterplots) and to NNSs (*n*=23) by the two listener groups: NLs (horizontal axis) and NNLs (vertical axis).

Nonetheless, *Pearson-r* correlations showed that the FA ratings assigned by the native and non-native listener groups were strongly correlated at the two testing times (*r*=.931 at Pre-test, *r*=.949 at Post-test), the significance level being *p*<.001 in the two cases (see scatterplots in Figure 11). These highly significant and strong correlation coefficients indicate that, despite the fact that the non-native listeners were, overall, more severe listeners than the native listeners, the FA ratings obtained from the two listener groups followed similar patterns that closely resembled each other.

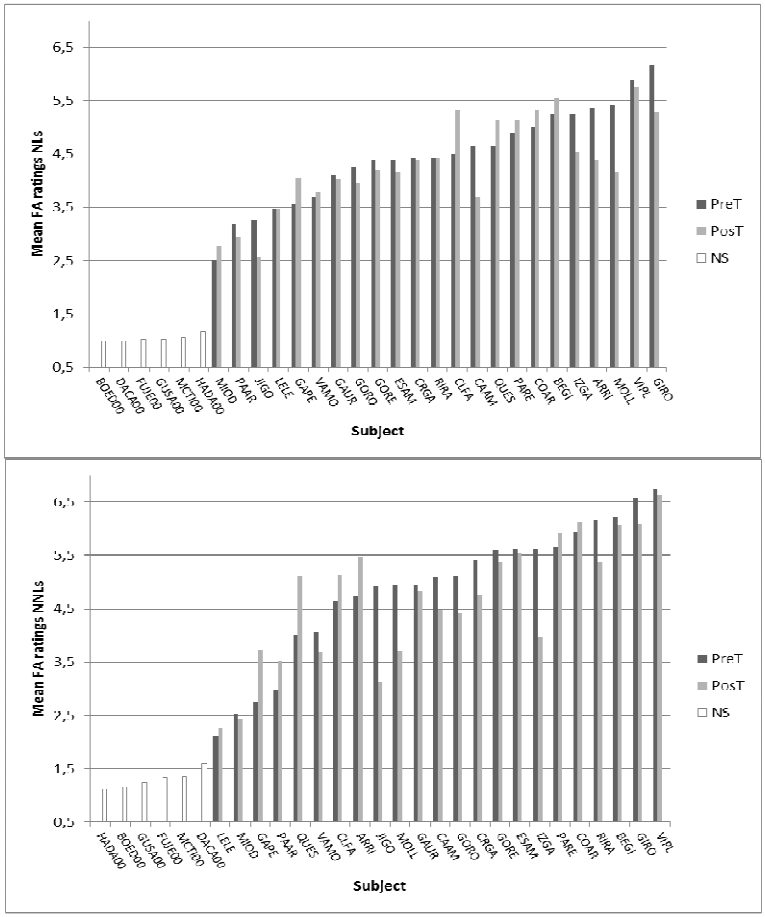

Figure 12: Individual mean FA ratings for NSs (baseline) and NNSs (Pre-test and Post-test) assigned by the NLs (upper graph) and the NNLs (lower graph). Scores: 1="native", 7="heavy foreign accent".

An inspection of the NNSs' individual mean FA ratings reveals rather high inter-subject variability (Figure 12). 13 subjects were found to decrease their degree of accentedness according to the NLs' ratings, and 15 according to the ratings provided by the NNLs. 11 out of all these subjects were found to improve by the two groups of listeners, showing, again, a great parallelism between the FA ratings provided by both listener groups.

These results suggest that SA might have had some positive impact on the NNSs' oral production as measured by global FA ratings, since a positive

trend towards less accented speech appeared to take place according to the mean FA ratings assigned both by the native and the non-native listeners to the NNS group. Yet the effect of SA was not large and turned out to be far from significant. It should also be noted that none of the NNSs managed to perform within native-like range, as there was not any NNS who received native-like ratings from either group of listeners. Nonetheless, it is worth pointing out that the observed changes, even though very small in general, were always in the direction of a slight decrease of accentedness, and that the same pattern of results was obtained from the FA ratings assigned both by the native and non-native listener groups.

4.2.2. Discussion

One possible explanation for the lack of a stronger impact of SA on perceived FA scores could be related to the length of the study abroad programme. LoS is one of the SA programme features identified by Pérez-Vidal (2011) as influencing SA outcomes, in the sense that it determines amount of input received or L2 exposure. LoS would be thus similar to LOR, which is the variable traditionally used as an index of amount of L2 experience in studies exploring the acquisition of L2 speech within long-term immersion contexts.

In an experiment addressing SA changes in overall foreign accent by a group of L1 Danish undergraduate learners of English, Højen (2003) reported significant improvement in his participants' FA scores after their experience abroad. However, the participants in his study presented considerable variation in terms of length of stay in the English-speaking environment. The mean LoS was 7.1 months (range=3-11 months), which is rather long as compared with the short 3-month stay experienced by the

learners in the present study. When analysing these individual differences, Højen found a strong positive correlation between LoS and gains in FA scores between pre-test and post-test (r=.606, p<.05). The learners who benefited the less from SA were those with stays of only 3 to 4 months, which is in line with the results in the present study, whereas the greatest SA gains were obtained by the learners who stayed abroad up to 11 months. Højen interpreted these results as an indication that the length of the stay abroad may be an important factor for SA gains in FA scores to accrue.

Findings regarding the role of LOR on degree of foreign accent have been rather mixed, with some studies reporting an effect of LOR on FA ratings while other studies have failed to do so (see section 1.2.2.4). In general terms, it seems that the impact of LOR depends to a great extent on learners' age and stage of L2 learning. In the literature analysing long-term immersion contexts, an LOR effect has been usually found for early L2 learners, i.e., learners who first came under massive L2 exposure in the L2 country before the end of a hypothesised critical period (around puberty), whereas LOR does not seem to impact late or adult L2 learners after a rapid initial phase of improvement (Flege, 1988). In this sense, it has been claimed that most L2 phonological learning for late or adult learners would take place around the first year of massive exposure in an L2 setting, and pronunciation would then fossilise, resisting further changes after the initial period of gains (Flege, 1988; Selinker, 1972). More inexperienced learners (those at an early stage of L2 learning) could benefit from additional L2 exposure, whereas very experienced learners (those with increased years of immersion) would be unlikely to benefit from further exposure to the L2.

Results from the present study, as well as those in Højen (2003), seem to be consistent with these general findings for LOR. Although the learners

in this study had been learning English as a foreign language since childhood in at home institutions, opportunities for access to authentic English input outside the classroom are rather limited in Spain (unlike in many North-European countries), since there is no English presence in the media (e.g., films and TV series are dubbed), the UPF SA programme allowing them to live in an English-speaking country up to three months for the first time. It can be argued, therefore, that this SA experience provided our learners with their first massive exposure to authentic and rich conversational English in a TL environment. However, three months might have been too short a period of time for learners to significantly improve their FA scores. Taking into account that the tendency was for a decrease in accentedness in the learners' speech, as perceived by our two different groups of listeners, it could have been the case that foreign accent would have continued to gradually decrease with an increase in length of stay, for instance, up to the average 7.1 months or 11 months at which significant improvement arose in the study by Højen (2003).

In addition, the outcome in the learners' FA scores might have been affected by other factors, such as actual amount and quality of L2 input, and learners' patterns of language use. Regarding amount and quality of L2 input, Piske et al. (2001) attribute the inconclusive findings of research on LOR effects partly to the fact that "LOR only provides a rough index of overall L2 experience" (p. 197). Piske (2007) further stresses the importance of considering time of immersion in an L2 setting together with the nature of the L2 input learners receive. In this sense, Højen (2003) created a composite measure which weighted length of stay by self-reported English-language input while abroad, and found that the correlation between this composite measure and gains in FA scores (r=.811, p<.001) was highly significant and stronger than the correlation between length of stay and gains in FA scores reported above. He

concluded that this outcome signalled the importance of quality native input for improved FA scores, in line with previous findings (Flege & Liu, 2001).

As noted in section 4.1, learners in the present study appeared to have a high degree of contact with English native speakers, since they reported a mean of 4 on a 5-point scale (1='never', 5='very often'), but they also reported a similarly high degree of contact with non-native speakers (*M*=4.10 on the same scale). It should be remembered that the learners in the present study were Erasmus students, and they were likely therefore to be in contact with other Erasmus students from whom they could have received foreign-accented English. This would have reduced the quality of the L2 input they were exposed to, a fact which could have contributed to the lack of significant development towards more native-like pronunciation, just in the same way it could have affected the acoustic measures discussed in section 4.1.1.

Learners' patterns of language use during immersion have been likewise found to have an impact on FA scores, particularly as far as amount of L1 use is concerned, since results in previous research indicate that more frequent use of the L1 is associated with higher degree of foreign accent (Flege et al., 1997; Piske et al., 2001). As noted already when discussing the acoustic measures in section 4.1.1 above, learners reported some degree of contact with other UPF students in the same SA destination (1.5 on a 3-point scale where 1='most time', 3='little'). Besides, learners reported keeping contact with their families while abroad by means of a scale from 'a' to 'e' ('a'='more than once a day' and 'e'='none') where most learners reported 'b' ('a few times a week'). Hence, it is possible that these patterns of frequent L1 use should have also prevented the learners' from obtaining greater gains regarding degree of foreign accent,

similarly to the negative effect they might have had on the acoustic measures as well.

Another explanation for the very modest changes in FA ratings could be related to the fact that, when asked to assess speech samples for overall degree of foreign accent, listeners seem to do so holistically (Magen, 1998), which means that they rely on different aspects of pronunciation, both at the segmental level but also at the suprasegmental level, paying attention to prosodic speech properties such as overall prosody, rhythm or intonation. As noted in section 1.2.2.4, research has found that prosodic features of speech, such as intonation and rhythm, contribute to degree of foreign accent (Munro & Derwing, 1999; Anderson-Hsieh et al., 1992).

This could explain the contrast between the lack of significant improvement in the learners' FA ratings and the significant decrease in error rate scores regarding segmental production and stress reported in section 4.1.2. The 3-month SA programme might have been too short to have a similar positive impact leading to greater SA gains in prosodic areas of pronunciation such as speaking rate, connected speech phenomena, rhythm or intonation, resulting in the observed slight changes in foreign accent.

Taking into account that the outcome of the FA ratings parallels the similar lack of clear improvement found in the acoustic measures reported in section 4.1.1, another possible explanation could be that the measures of foreign accent also reflected effects of the processing demands facing the learners in a context of full immersion within the TL environment, as it was hypothesised regarding the acoustic measures, due to the characteristic tendency for learners to direct their attention to meaning rather than form in an SA setting.

Another factor which could have influenced listeners' assessment of degree of foreign accent when rating the non-native speech samples is the rather homogeneous composition of the learner group in terms of proficiency level. Although there were differences in pronunciation between the learners, they all shared a similar and fairly advanced English language level (B.2 or upper-intermediate), which could have made the rating task more difficult for the listeners, as they were faced with the challenge of having to discriminate very subtle changes in accentedness within and across learners. For listeners, the task of rating non-native speech for degree of accentedness would probably be easier with a pool of learners showing a wider range of proficiency levels, form low to advance. This is usually the case in the FA literature (see review in section 1.2.2.4), where differences in FA scores arise as a result of considerable inter-subject variation in terms of L2 proficiency, normally due to differences in AOL, as well as in other variables such as L2 exposure, L2 and L1 use, etc. (Flege, 1988; Derwing & Munro, 2013; Piske et a., 2001.).

Nonetheless, despite the difficulty involved in their task, it should be emphasised that the same response pattern was obtained from the two listener groups who rated the speech samples regardless of the differences in their L1 profile (L1 English versus L1 Spanish/Catalan). The NLs and the NNLs were equally successful at distinguishing the English native speakers from the L2 learners, and they both seemed to perceive the same slight decrease in accentedness in the learners' samples after SA. Moreover, the FA ratings provided by the two listener groups presented a highly significant and strong correlation. This indicates a high degree of parallelism in the two listener groups' response to the speech samples, suggesting a stronger influence of SP, or speech characteristics, over LF,

or differences across listeners. Both inter-rater and intra-rater reliability were also remarkably high for the two groups.

These findings support the claim that the phenomenon of foreign accent can be accurately and reliably scaled (Højen, 2003; Piske et al., 2001; Southwood & Flege, 1999), and that these rating tasks can be performed not only by native listeners, but also by non-native listeners with a high degree of L2 proficiency, as suggested by previous research (Flege, 1988; Derwing & Munro, 2013). In fact, Flege (1988) points out that with increased L2 experience listeners may develop 'phonetic category prototypes' which they would use to assess the goodness of the phones they hear in terms of their degree of deviation from those prototypes.

Despite the clear parallelism between the foreign accent judgements provided by both the native and non-native listener groups, the NLs were found to accord significantly better FA ratings than the NNLs. In other words, the NNLs turned out to be more severe judges than the NLs. This is in line with findings in other studies addressing native and non-native assessment of L2 speech, for example, for fluency (Derwing & Munro, 2013; Rossiter, 2009; Valls-Ferrer, 2011).

Valls-Ferrer (2011) found that her non-native listeners presented a general tendency to assign worse fluency judgements than the native listeners. In Derwing and Munro (2013), a group of native listeners and a group of non-native listeners perceived fluency gains in a group of L2 English learners, but whereas the improvement was significant according to the native listeners' ratings, it did not reach significance according to the non-native listeners' ratings. Similarly, Rossiter (2009) found that a group of novice native listeners (university students without specific linguistic training) accorded better fluency ratings than a group of proficient L2 native listeners (English majors), although ratings from a group of expert

native listeners (teachers of English trained in phonology and phonetics) were not significantly different from the ratings accorded by the other two groups. Rossiter connected these differences between native and non-native listeners' ratings with similar differences in the assessment of L2 grammatical errors, for which native listeners seem also to provide more lenient judgements than non-native listeners (Derwing, Rossiter, & Ehrensberger-Dow, 2002, cited in Rossiter, 2009).

Since the NNLs in the current study shared the same linguistic profile of the L2 learners, and were besides trained en English phonetics and phonology, they might have been more familiar with the type of mispronunciations produced by the L2 learners they were assessing, which they could recognise in their own speech, as well as more sensitive to the deviations from the TL norms they had been taught. All this could have enhanced their language awareness when performing the rating task, resulting in the more severe ratings obtained from this listener group.

To sum up, the analysis of the FA ratings seemed to indicate a positive trend of development towards a decrease in accentedness, although it was not significant. The response pattern obtained from the NNLs mirrored that obtained from the NLs, as the ratings from the two listener groups were highly correlated, pointing towards a greater influence of speech characteristics over differences between listeners in the FA ratings elicited from the two groups. The lack of significant improvement in perceived degree of foreign accent parallels the similar lack of significant improvement observed in the analyses of the acoustic measures (section 4.1.1), although it seems to contrast with the significant decrease in mispronunciations reported for the error rate scores (section 4.1.2). Previous research has also found that changes in specific aspects of pronunciation are not always reflected in a straightforward manner by FA ratings. For example, in Riney and Flege (1998) gains in global FA ratings

did not coincide with improvement in segmental production regarding liquid identifiability and accuracy. The authors noted that it "appears not to be the case that improvement in global accent necessarily proceeds in parallel with improvement in any particular smaller components of pronunciation" (p. 237). Taking this into account, the following section is intended to shed some light into the relationship between the acoustic-phonetic measures and the listeners' perceived FA ratings.

4.3. Relationship between acoustic-phonetic and FA measures

This section addresses RQ3: *Are acoustic-phonetic measures related to perceived FA ratings and to what extent are they good predictors of variance in FA ratings?*

4.3.1. Results

In order to answer this research question, the objective acoustic-phonetic measures (acoustic values for VOT and vowel quality and duration, and PrEr scores) and the subjective FA ratings obtained from the two listener groups (NLs and NNLs) were submitted to a series of correlational and multiple regression analyses. To perform these analyses, an additional new acoustic composite measure was created with the aim of having a tentative measure that could serve as an acoustic index (AIndex) gauging the learners' overall phonological competence in the production of the acoustic features of VOT and vowel quality. This composite acoustic index was thus created from the measures of VOT, ED for /i:-ɪ/ and ED for /æ-ʌ/. The measures of DD were not included in the overall acoustic index, since duration is a secondary feature in the production of English

contrasts such as /i:-ɪ/, the main distinctive feature being vowel quality (see section 3.4.1.1 above).

The measures of VOT and ED were first normalised following Mora & Valls-Ferrer (2012). By means of this procedure, the NSs' scores were converted so that the mean for each of the three measures equalled 1.0 (NSs' *M*=1.0). The NNSs' scores for each measure were subsequently recalculated based on the NSs' new mean score of 1.0 (*NormalisedScore = Score(*1)/NSs'MeanScore*), in such a way that all the new mean scores ranged between 0 and 1, 1 indicating native performance and 0 indicating performance which is far from native. Then, the values of the three standardised measures were averaged for each participant at both testing times [*AIndex* = (*norm*VOT+*norm*ED/i:-ɪ/+*norm*ED/æ-ʌ/)/3].

	RatNL_PreT	RatNL_PosT	RatNNL_PreT	RatNNL_PosT
PrEr_PreT	.799**		.814**	
VOT_PreT	-.566**		-.599**	
ED /i:-ɪ/ PreT	-.712**		-.635**	
ED /æ-ʌ/ PreT	-.234		-.250	
DD /i:-ɪ/ PreT	-.137		-.018	
OAIndex PreT	-.665**		-.644**	
PrEr_PosT		.739**		.730**
VOT_PosT		-.622**		-.533**
ED /i:-ɪ/ PosT		-.816**		-.759**
ED /æ-ʌ/ PosT		-.218		-.220
DD /i:-ɪ/ PosT		-.199		-.216
AIndex PosT		-.780**		-.718**

Table 5. Correlation coefficients between acoustic-phonetic measures (acoustic values and PrEr scores) and FA ratings provided by the NLs and NNLs. Shaded boxes indicate significant correlations (p<.001).

As shown in Table 5, strong *Pearson-r* correlations were found between the objective measures of PrEr, VOT, ED for /i:-ɪ/ and AIndex and the

subjective FA ratings assigned by the two groups of listeners. These correlations were all highly significant (p <.001), and arose both for the measures obtained before and after SA.

The correlation coefficients between PrEr scores and FA ratings were all positive, which shows that the production of fewer pronunciation errors resulted in the perception by the listeners of a lower degree of accentedness, whereas the larger the number of pronunciation errors, the higher the degree of accentedness perceived. In contrast, the correlations between the acoustic values VOT, ED for /i:-ɪ/ and the composite AIndex with the FA ratings were all negative, indicating that increased values in the production of VOT and vowel quality distinctions were associated with a decrease in the perception of accentedness by both listener groups.

These correlational analyses thus allowed us to identify four variables out of the five acoustic-phonetic measures that were apparently strongly related to the perception of accentedness in the participants' speech as evaluated through the FA ratings provided by our two listener groups. These measures were PrEr scores (number of pronunciation errors in terms of production of phonological insertions, deletions and substitutions, as well as errors in lexical stress, VOT (an acoustic measure of participants' production of VOT duration), ED for /i:-ɪ/ (an acoustic measure of participants' production of vowel quality distinctions), and AIndex (a composite acoustic measure comprising measures for VOT and vowel quality distinctions).

Next, a series of simple multiple regression analyses were conducted with the aim of assessing the specific contribution of the objective acoustic-phonetic measures to the prediction of variance in the subjective measures of perceived accentedness. The FA ratings provided by the two listener groups at both testing times were entered as the outcome variables. Two

out of the four previously identified measures were added to the models as predictors: PrEr and AIndex. PrEr is a phonetic measure of pronunciation accuracy which was found to present the strongest correlation with the FA ratings and was likely, therefore, to predict the greatest amount of variance in the outcome FA scores. AIndex, as a composite acoustic measure averaged across both VOT and ED values, was deemed as an appropriate overall measure of participants' production of VOT and vowel quality distinctions. Table 6 summarises the outcome for the regression analyses, which yielded the same pattern of results for all the generated models.

As observed in this table, both PrEr and AIndex contributed significantly to the regression models, accounting for significant amounts of variance in the listeners' FA ratings. Of the two predictors, PrEr was in general terms the one which explained the greatest amount of variance in the outcome FA measures, from a minimum of 13% in the NLs Post-test ratings, and a maximum of 31,5% in the NNLs Pre-test ratings. The additional variance predicted by AIndex, was generally lower, but likewise significant, a minimum of 0.8% in the NLs Pre-test ratings, and a maximum of 19.4 in the NLs Post-test ratings. On the whole, all the models successfully accounted for considerably large amounts of overall variance in the FA ratings accorded by the listeners (from a minimum of 67.1% in the NNLs ratings to a maximum of 74.1% in the NLs, both corresponding with the Post-test measures).

RatNL_Pre-test	*B*	*SE B*	*β*	Part. corr	R square
(Constant)	4,067	,633			.727**
PrEr_Pre-test	,343	,066	,622**	.533	
OAIndex Pre-test	-2,470	,854	-,345**	.296	

RatNL_Post-test	*B*	*SE B*	*β*	Part. corr	R square
(Constant)	5,183	,639			..741**
PrEr_Post-test	,251	,069	,439**	.363	
AIndex Post-test	-3,916	,888	-,533**	-.441	

RatNNL_Pre-test	*B*	*SE B*	*β*	Part. corr	R square
(Constant)	4,231	,714			.731**
PrEr_Pre-test	,412	,074	,656**	.562	
AIndex Pre-test	--2,497	,963	-,307*	-.264	

RatNNL_Post-test	*B*	*SE B*	*β*	Part. corr	R square
(Constant)	5,288	,817			.671**
PrEr_Post-test	,310	,088	,478**	.395	
AIndex Post-test	-3,741	1,135	-,449**	.371	

Table 6. Summary statistics of regression analyses for the FA ratings provided by the NLs (two upper tables), and by the NNLs (two lower tables). Note: *B*=unstandardized beta; *SE B*=standard error of beta; *β*=standardized beta. ***p*<.001.

These results indicate that two acoustic-phonetic measures such as AIndex scores, as operationalized in this study in order to assess participants' production of very specific, fine-grained acoustic features, and especially PrEr scores, which was operationalized to assess participants' pronunciation accuracy by targeting a wider range of segmental and stress-related phenomena, were good predictors of listeners' FA ratings, and they resulted in similar patterns signalling significant contributions to the regression models regardless of the listeners' L1 linguistic profiles.

4.3.2. Discussion

Taken together, these results show a clear relationship between listeners' FA ratings and stimulus properties, as measured by means of an error rate score (PrEr) computing different segmental mispronunciations and lexical stress errors in the speech samples rated for foreign accent. Results also pointed towards a relationship between listeners' FA ratings and a measure of participants' overall phonological competence (AIndex) in the production of VOT and vowel quality distinctions. This finding seems all the more robust as the same pattern of results was obtained across testing time and for the two listeners groups, regardless of their different L1 background (L1 English and L1 Spanish/Catalan).

Correlational analyses revealed significantly strong and positive correlations between FA ratings and PrEr scores, indicating that the production of more mispronunciations was associated with a higher degree of foreign accent, and the production of few mispronunciations was in turn associated with a lower degree of foreign accent. Although the significant improvement observed in PrEr scores contrasted with the lack of similar significant gains in FA ratings, these results suggest that the FA ratings did reflect to some extent the gains in pronunciation accuracy captured by the PrEr scores, as it seems that both the native and non-native listeners were able to 'correctly' rate the speech samples on the basis of their degree of deviation from native phonetic norms. These results are in line with previous research which has likewise found an effect of pronunciation errors occurring in learners' L2 speech on listeners' judgements of foreign accent (Brennan & Brennan, 1981; Cunningham-Anderson & Engstrand, 1989; Magen, 1988). In addition to the high correlations found between PrEr scores and FA ratings, a regression analysis further revealed that the measure of PrEr scores was a

good predictor of degree of FA, accounting for over 31% of the variance in the FA ratings for all the generated models (specifically between 13% and 31.5%).

These results are consistent with the view, pointed out in section 1.2.2.4, that stimulus properties play an important role in listeners' perception and rating of L2-accented speech. Although some variation can always be attributable to listener factors, it should be remembered that the PrEr scores used in the present study were mainly a measure for segmental errors, as well as errors in lexical stress. Therefore, part of the variation in FA ratings not explained by the models is very likely to have been caused by other stimulus properties at the prosodic level (such as rhythm, intonation, etc.), since prosodic aspects of speech are also known to have an effect on listeners' assessment of foreign accent (Munro, 1995; Munro & Derwing, 1999).

Correlational analyses yielded likewise significantly strong but negative correlation coefficients between FA ratings and AIndex scores. AIndex was calculated as a tentative measure of overall phonological competence in the production of English VOT and vowel quality distinctions between the contrasts /i:-ɪ/ and /æ-ʌ/ as measured through ED values. High AIndex scores thus indicated longer VOT duration and higher ED values showing more robust distinctions in vowel quality, and signalled therefore more native-like production. Low AIndex scores indicated shorter VOT duration and lower ED values which signalled less robust vowel quality distinctions, pointing to more L1-like production patterns. Accordingly, these strong negative correlations between FA ratings and AIndex indicate that less native production in terms of VOT and vowel quality distinctions was associated with an increased degree of foreign accent, whereas more target-like VOT duration and vowel quality patterns were associated with less perceived foreign accent. A regression analysis showed that AIndex

scores accounted also for an additional amount of variance in the FA ratings for all the generated models (from 0.8% to 19.4%), which was smaller than the amount of variance predicted by the PrEr scores, but significant as well.

On the whole, these results suggest that there might be also a relationship between participants' overall phonological competence in the production of VOT and vowel quality distinctions and listeners' perception of foreign accent. The difference between the larger amount of variance in FA predicted by the PrEr scores and the generally smaller amount of variance predicted by the AIndex scores would be a natural consequence of the fact that the AIndex is an overall measure of phonological competence addressing only VOT and vowel quality production, whereas the score of PrEr measured a wide range of phonological phenomena present in the very speech samples which were rated by the listeners. All things considered, it is worth pointing out the great degree of parallelism of the different models across time and listener group, which can be argued to render considerable robustness to the different measures utilised in the study.

5. Summary and conclusions

The aim of the present study has been to shed light on the actual impact of an SA learning context on L2 phonological changes in speech production, an area of research within the SA field which has received limited attention to date. In order to do so, different measures were used targeting different aspects of speech production: acoustic measurements for VOT and vowel quality and duration, error rate scores of phonemic and lexical stress mispronunciations, and listeners' assessment of FA. To our knowledge, this is the first study within the SA field which has adopted this multiple-measures design, involving the use of acoustic measures, phonetic analyses of learners' mispronunciations and FA ratings, and it constitutes therefore an important contribution to the SA literature as it provides a better and more global picture of the type of changes that SA may foster within the domain of pronunciation.

Regarding acoustic measures, no evidence was found in the present study of significant improvement in the production of VOT and vowel quality and duration. Measures of VOT reflected compromise values between L1 and L2 patterns, suggesting the existence of merged categories (Flege, 1995). Learners were found to produce significant duration differences between /i:-ɪ/ (although not within native patterns), but they produced the two contrasting vowels /i:-ɪ/ and /æ-ʌ/ very close within the vowel space, suggesting an inability to produce them as distinct vowels in terms of quality. This could be attributable to a case of L2 SC assimilation (Best & Tyler, 2007). These findings confirm those in Højen (2003), who reported similar lack of gains in segmental vowel production after an SA period. These results indicate that improvement in the production of very fine-grained phonetic features which are particularly constrained by L1

influence, such as VOT and vowel quality distinctions of the type between /i:-ɪ/ and /æ-ʌ/, might be difficult to accrue in a meaning-oriented context such as SA, especially if those features are unlikely to be enhanced during communication. When facing the socio-pragmatic and online processing demands of SA, learners may be unable to override the L1 filter or sieve so as to attend to those subtle L2 features, unless specific pronunciation instruction should explicitly direct their attention to them.

However, there was a positive impact of SA on pronunciation accuracy in terms of a significant decrease in pronunciation errors, as measured through an error rate score which targeted a wider range of phonological features (phonemic insertions, deletions, substitutions, and lexical stress misplacement). This error rate score was thus revealed as a more apt measure to capture gains in pronunciation as a result of SA. The reason is likely to be that the phonological features targeted were more salient auditorily than the fine-grained features of VOT and vowel quality. Alternatively, they could have been somehow enhanced during conversational interaction in the SA context and brought to the learners' attention, particularly in cases of communication breakdowns and misunderstandings, as these situations may hinder learners' main objective in an SA context of achieving fluent communication. Indeed, results in SA research consistently indicate that some of the greatest benefits that accrue as a result of SA can be found in those measures related to fluency (Freed, 1995b; Trenchs-Parera, 2009; Valls-Ferrer, 2011).

It was likewise found that improvement in pronunciation accuracy was the result of the gains experienced by the learners with lower onset pronunciation accuracy level, whereas no change was observed for the higher onset level learners. This contributes to previous findings on threshold level effects for development in different language skills, which

is one major issue in SA research. We have provided empirical evidence that even within an already upper-intermediate range of proficiency level, learners with lower onset level benefit more from SA than higher level learners in the domain of pronunciation accuracy, an outcome which is in accordance with general findings regarding onset level effects. DeKeyser (forthcoming) accounts for this general trend of findings in the SA literature within the skill acquisition framework, and notes that progress continues at a rather slow rate once a considerably high level of practice has been reached. Learners with higher proficiency level would thus need longer SA periods in order for noticeable improvement to accrue.

Regarding listeners' assessment of FA, findings yielded no significant effect of SA on FA ratings, although a positive trend of improvement towards less accented speech was observed in the ratings provided by two different groups of listeners. The same pattern of response was obtained from the two groups, which would indicate that the FA ratings reflected the properties of the speech stimuli, rather than differences between listener groups. The FA ratings from the two groups of listeners were also found to highly correlate with the acoustic measures and with pronunciation errors, and to predict more than 50% of the variance in error rate scores. This suggests that the significant improvement observed in pronunciation errors was eventually reflected in the listeners' ratings, and further signals the robustness and reliability of the error rate scores.

In sum, the main findings and contributions of the present study to the body of SA research can be summarised as follows:

- We have assessed phonological development in speech production by means of a multiple-measures design which included acoustic measures, phonetic analyses of error rate scores and FA ratings, and

which, to our knowledge, has not been previously utilised in the study of SA effects on pronunciation.

- Findings did not yield significant changes in acoustic measures nor in FA ratings, suggesting that three months may not be enough for changes to take place that can be captured by those measures. However, significant improvement was observed in error rate scores, indicating that this measure may be better to capture the type of changes that might accrue as a result of a short SA period of three months.
- Our findings confirm the existence of a threshold level for pronunciation, in the sense that learners with lower pronunciation accuracy level were the ones who obtained significant gains. Higher-level learners, in turn, might need a longer LoS to obtain observable gains.
- The FA ratings obtained from two groups of listeners with different L1 backgrounds were found to follow exactly the same response pattern. They indicated a tendency towards less accented speech after SA, and presented strong correlations with the acoustic measures and especially so with the error rate scores, pointing towards the robustness of the different measures utilised.

Further research

We would like to point out some issues which would deserve to be addressed in further research, taking into account the findings in the present study, and some limitations and methodological issues derived from it.

First, the lack of significant gains regarding VOT and vowel production and listeners' FA ratings could have been related to LoS issues, as a 3-

month SA programme may not be long enough for substantial improvement to accrue in those aspects of pronunciation. Højen (2003) found improvement in FA ratings for those learners with longer LoS (up to 11 months), but no gains for learners with shorter LoS (3-4 months). However, he reported no significant improvement in vowel production as a function of LoS. Avello and Lara (forthcoming) failed to find a significant effect of LoS on L2 VOT and vowel production when comparing two SA programmes of three and six months. These results should, nonetheless, be taken with caution due to the small sample size of the 6-month group. In this sense, more studies are needed assessing LoS effects on SA outcomes, particularly taking into account that general findings of research on L2 speech production indicate that most progress in pronunciation seems to take place throughout the first year of exposure within an L2 immersion setting, and that the scarce SA research analysing LoS effects on other linguistic skills has generally found that 'the longer the better' for L2 improvement.

Secondly, whereas most SLA studies collect only L2 data from the learner population analysed, in the case of the analysis of oral production it would be particularly useful, when feasible, to collect L1 data from the same learner group as well in order to assess the nature of the interaction between the L1 and L2 systems as it is hypothesised, e.g., within the framework of models such as Flege's SLM. The use of both L1 and L2 data would be necessary to confirm the existence of merged categories, and the extent to which they reflect the properties of the assimilated L1 and L2 sounds, as these merged categories are assumed to equally underlie both L1 and L2 production. Data collection and analysis of learners' L1 oral production have been likewise recommended for the assessment of other dimensions of L2 speech, such as fluency (Segalowitz, 2010).

Finally, since findings in the present study failed to yield significant effects of the SA context on VOT and vowel production, it would also be interesting to assess the effects of SA as compared with those of the FI period which preceded it in order to explore whether the FI context had a more positive effect on the learners' L2 phonological competence, and whether different patterns of development arise across the two learning contexts. Results from such analyses on speech production could thus be related to results from analyses on perceptual data addressing the perception of the same English vowel contrasts which are available from a similar pool of participants, and which have revealed significant improvement in the perception of these vowels as a result of FI, but no effect of SA (Mora, forthcoming). A parallel improvement during FI in vowel production data would be in line with previous results which have likewise failed to found a superiority of SA over FI, e.g., for grammar (Collentine, 2004; DeKeyser, 1991), pointing to possible limitations of SA for the development of discreet units of language for which acquisition might be more successfully fostered by FI.

References

Aliaga-García, C. 2013.Effects of identification and articulatory training methods on L2 vowel production. *7th International Symposium on the Acquisition of Second Language Speech (New Sounds),* Concordia University.

Allen, H. W., & Herron, C. 2003. A mixed-methodology investigation of the linguistic and affective outcomes of summer study abroad. *Foreign Language Annals, 36*(3), 370-385.

Altarriba, J., & Basnight-Brown, D. M. 2009. An information-processing approach to second language acquisition. In W. C. Ritchie, & T. K. Bhatia (Eds.), *The New Handbook of Second Language Acquisition* (pp. 116-137). Bingley, UK: Emerald.

Anderson-Hsieh, J., Johnson, R., & Kohler, K. 1992. The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. *Language Learning, 42*(4), 529-555.

Archibald, J. 2005. Second language phonology as redeployment of L1 phonological knowledge. *The Canadian Journal of Linguistics, 50,* 285-314.

Archibald, J. 2009. Phonological feature re-assembly and the importance of phonetic cues. *Second Language Research, 25*(2), 231-233.

Asher, J. J., & García, R. 1969. The optimal age to learn a foreign language. *The Modern Language Journal, 53*(05), 334-341.

Aslin, R. N., Pisoni, D. B., Hennessy, B. L., & Perey, A. J. 1981. Discrimination of voice onset time by human infants: New findings and implications for the effect of early experience. *Child Development, 52,* 1135-1145.

Aslin, R. N., Saffran, J. R., & Newport, E. L. 1998. Computation of conditional probability statistics by 8-month-old infants. *Psychological Science, 9*(4), 321-324.

Avello, P. 2010a.The effect of study abroad on Catalan/Spanish bilinguals' production of English /i: - ɪ/ and /æ - ʌ/ contrasting pairs. *Paper presented at the 6th International Conference on Language Acquisition - CIAL,* Universitat de Barcelona, Spain.

Avello, P. 2010b.Study abroad and foreign accent: The production of /i: - ɪ/ and /æ - ʌ/ by Catalan/Spanish bilinguals. *Paper presented at the 20th International Conference of the European Second Language*

Association - Eurosla, University of Modena and Reggio Emilia, Italy.

Avello, P. 2011. Measuring perceived pronunciation gains in study abroad: Methodological issues. *Paper presented at the 29th AESLA International Conference,* Universidad de Salamanca, Spain.

Avello, P., & Lara, A. R. Forthcoming. Phonological development in L2 speech production during study abroad programmes differing in length of stay. In C. Pérez-Vidal (Ed.), *Language Acquisition in Study Abroad and Formal Instruction Contexts*. Amsterdam: John Benjamins.

Avello, P., Mora, J. C., & Pérez-Vidal, C. 2012. Perception of FA by non-native listeners in a study abroad context. *Research in Language, 10*(01), 63-78.

Barquin, E. 2012. *Writing Development in a Study Abroad Context.* PhD Thesis. Universitat Pompeu Fabra. Retrieved from http://www.tdx.cat/handle/10803/96196

Barron, A. 2006. Learning to say 'you' in German: The acquisition of sociolinguistic competence in a study abroad context. In A. DuFon, & E. Churchill (Eds.), *Language Learners in Study Abroad Context*s (pp. 59-90). Clevedon: Multilingual Matters.

Beattie, J. 2008. Evaluating the acquisition of English listening comprehension skills in the framework of the SALA project. In *Monografías De Aula Abierta 3*2 (pp. 301-310). Oviedo: Uniovi.

Beattie, J. Forthcoming. The 'ins and outs' of a study abroad programme: The SALA exchange programme. In C. Pérez-Vidal (Ed.), *Language Acquisition in Study Abroad and Formal Instruction Context*s. Amsterdam: John Benjamins.

Beattie, J., Valls-Ferrer, M., & Pérez-Vidal, C. Forthcoming. Listening performance and onset level in formal instruction and study abroad. In C. Pérez-Vidal (Ed.), *Language Acquisition in Study Abroad and Formal Instruction Context*s. Amsterdam: John Benjamins.

Best, C. T. 1990. Adult perception of non-native contrasts differing in assimilation to native phonological categories. *Journal of the Acoustical Society of America, 8*8(S1), S177-S178.

Best, C. T. 1995. A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Researc*h (pp. 171-204). Timonium, MD: York Press.

Best, C. T., McRoberts, G. W., & Sithole, N. M. 1988. Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology, 1*4(3), 345-360.

Best, C. T., & Strange, W. 1992. Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics, 2*0, 305-330.

Best, C. T., & Tyler, M. D. 2007. Nonnative and second-language speech perception. In O. S. Bohn, & M. J. Munro (Eds.), *Language Experience in Second Language Speech Learning. in Honour of James Emil Fleg*e (pp. 13-34). Amsterdam: John Benjamins Publishing Company.

Birdsong, D. 2007. Nativelike pronunciation among late learners of French as a second language. In O. S. Bohn, & M. J. Munro (Eds.), *Language Experience in Second Language Speech Learning: In Honor of James Emil Fleg*e (pp. 99-116). Amsterdam: John Benjamins Publishing Company.

Bleyhl, W. 2009. The hidden paradox of foreign language instruction or: Which are the real foreign language learning processes? In T. Piske, & M. Young-Scholten (Eds.), *Input Matters in SL*A (pp. 137-155). Buffalo, NY: Multilingual Matters.

Boersma, P., & Weenink, D. 2008. *Praat: Doing Phonetics by Computer, Version 5.0*. http://www.fon.hum.uva.nl/praat/.

Bohn, O. S. 1995. Cross-language perception in adults: First language transfer doesn't tell it all. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Researc*h (pp. 379-410). Timonium, MD: York Press.

Bohn, O. S., & Flege, J. E. 1990. Interlingual identification and the role of foreign language experience in L2 vowel perception. *Applied Psycholinguistics*, 11(03), 303-328.

Bohn, O. S., & Flege, J. E. 1992. The production of new and similar vowels by adult german learners. *Studies in Second Language Acquisition*, 14, 131-158.

Bongaerts, T., Planken, B., & Schils, E. 1995. Can late starters attain a native accent in a foreign language? A test of the critical period hypothesis. In D. Singleton, & Z. Lengyel (Eds.), *The Age Factor in Second Language Acquisition: A Critical Look at the Critical Period Hypothesis* (pp. 30-50). Clevedon: Mutlilingual Matters.

Bongaerts, T., van Summeren, C., Planken, B., & Schils, E. 1997. Age and ultimate attainment in the pronunciation of a foreign language. *Studies in Second Language Acquisition,* 19(04), 447.

Brecht, R., Davidson, D., & Ginsberg, R. 1995. Predictors of foreign language gain during study abroad. In B. Freed (Ed.), *Second Language Acquisition in a Study Abroad Contex*t (pp. 37-66). Amsterdam: John Benjamins Publishing Company.

Brennan, E. M., & Brennan, J. S. 1981. Measurements of accent and attitude toward Mexican-American speech. *Journal of Psycholinguistic Research,* 10(05), 487-501.

Brown, C. 2000. The interrelation between speech perception and phonological acquisition from infant to adult. In J. Archibald (Ed.), *Second Language Acquisition and Linguistic Theor*y (pp. 4-63). Oxford: Blackwell.

Bundgaard-Nielsen, R. L., Best, C. T., & Tyler, M. D. 2011. Vocabulary size is associated with second-language vowel perception performance in adult learners. *Studies in Second Language Acquisition,* 33(03), 433-461.

Cebrián, J. 2006. Experience and the use of non-native duration in L2 vowel categorization. *Journal of Phonetics,* 34, 372-387.

Cebrián, J. 2007. Old sounds in new contrasts: L2 production of the English tense-tax vowel distinction. In J. Trouvain, & W. J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Science*s (pp. 1637-1640). Dudweiler: Pirrot.

Cebrián, J., & Carlet, A. 2012.Investigating the contribution of visual and auditory cues in the perception of native and non-native consonants. *Paper presented at the 22th International Conference of the European Second Language Association - Eurosla,* Adam Mickiewicz University, Poland.

Cebrián, J., Mora, J. C., & Aliaga-García, C. 2011. Assessing crosslinguistic similarity by means of rated discrimination and perceptual assimilation tasks. In M. Wrembel, M. Kul & K. Dziubalska-Kolaczyk (Eds.), *Achievements and Perspectives in the Acquisition of Second Language Speech: New Sounds 2010, Volume* I (pp. 41-52). Frankfurt am Main: Peter Lang.

Coleman, J. A. 1997. Residence abroad within language study. *Language Teaching, 3*0(1), 1-20.

Coleman, J. A. 1998. Language learning and study abroad: The European perspective. *Frontiers: The Interdisciplinary Journal of Study Abroad,* 4, 167-203.

Collentine, J. 2004. The effects of learning contexts on morphosyntactic and lexical development. *Studies in Second Language Acquisition,* 26(2), 227-248.

Collentine, J. 2009. Study abroad research: Findings, implications and future directions. In M. Long, & C. J. Doughty (Eds.), *The Handbook of Language Teachin*g (pp. 218-233). Malden, MA: Blackwell.

Collentine, J., & Freed, B. 2004. Learning context and its effects on second language acquisition: Introduction. *Studies in Second Language Acquisition,* 26(2), 153-171.

Cook, V. 1999. Going beyond the native speaker in language teaching. *TESOL Quarterly, 33*(2), 185-209. doi:10.2307/3587717

Cunningham-Andersson, U., & Engstrand, O. 1989. Perceived strength and identity of foreign accent in Swedish. *Phonetica,* 46(4), 138-154.

Cutting, J., & Rosner, B. 1974. Categories and boundaries in speech and music. *Perception & Psychophysics, 16*(3), 564-570.

de Boysson-Bardies, B., Halle, P., Sagart, L., & Durand, C. 1989. A crosslinguistic investigation of vowel formants in babbling. *Journal of Child Language, 16*(01), 1-17.

de Boysson-Bardies, B., Sagart, L., & Durand, C. 1984. Discernible differences in the babbling of infants according to target language. *Journal of Child Language, 11*(01), 1-15.

DeCasper, A., & Fifer, W. 1980. Of human bonding: Newborns prefer their mothers' voices. *Science, 208*(4448), 1174-1176.

DeKeyser, R. 1991. Foreign language development during a semester abroad. In B. Freed (Ed.), *Foreign Language Acquisition Research and the Classroom* (pp. 104-119)

DeKeyser, R. 2000. The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition, 22*(04), 499-533.

DeKeyser, R. 2007a. Introduction: Situating the concept of practice. In R. DeKeyser (Ed.), *Practice in a Second Language: Perspectives from Applied Linguistics and Cognitive Psychology* (pp. 1-18). New York: Cambridge University Press.

DeKeyser, R. 2007b. Skill acquisition theory. In B. VanPatten, & J. Williams (Eds.), *Theories in Second Language Acquisition: An Introduction* (pp. 97-114). Mahwah, NJ: Lawrence Erlbaum Associates.

DeKeyser, R. 2007c. Study abroad as foreign language practice. In R. DeKeyser (Ed.), *Practice in a Second Language: Perspectives from Applied Linguistics and Cognitive Psychology* (pp. 208-226). New York: Cambridge University Press.

DeKeyser, R. Forthcoming. Methodological considerations about research on language development during study abroad. In C. Pérez-Vidal (Ed.), *Language Acquisition in Study Abroad and Formal Instruction Contexts*. Amsterdam: John Benjamins.

Derwing, T. M., & Munro, M. J. 1997. Accent, intelligibility, and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition, 20*(1), 1-16.

Derwing, T. M., & Munro, M. J. 2013. The development of L2 oral language skills in two L1 groups: A 7-year study. *Language Learning, 63*(2), 163-185.

Derwing, T. M., Rossiter, M. J., & Ehrensberger-Dow, M. 2002. 'They speaked and wrote real good': Judgements of non-native and native grammar. *Language Awareness, 11*, 84-99.

Díaz-Campos, M. 2004. Context of learning in the acquisition of Spanish second language phonology. *Studies in Second Language Acquisition, 26*, 249-273.

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, & T. L. Face (Eds.), *Selected Proceedings of the 7th Conference on the Acquisition of Spanish and Portuguese as First and Second Languages* (pp. 26-39). Sommerville, MA: Cascadilla Proceedings Project.

Dörnyei, Z., & Skehan, P. 2003. Individual differences in second language learning. In C. J. Doughty, & M. Long (Eds.), *The Handbook of Second Language Acquisition* (pp. 589-630). Malden Mass.: Blackwell.

Doughty, C. J. 2003. Instructed SLA: Constraints, compensation, and enhancement. In C. J. Doughty, & M. Long (Eds.), *The Handbook of Second Language Acquisition* (pp. 256-310). Malden Mass.: Blackwell.

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

Eckman, F. R. 2008. Typological markedness and second language phonology. In J. G. Hansen Edwards, & M. L. Zampini (Eds.), *Phonology and Second Language Acquisition* (pp. 95-116). Amsterdam: John Benjamins Publishing Company.

Eimas, P. D. 1974. Auditory and linguistic processing of cues for place of articulation by infants. *Perception & Psychophysics, 16*(3), 513-521.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. 1971. Speech perception in infants. *Science, 171*(3968), 303-306.

Eimas, P. D. 1975. Auditory and phonetic coding of the cues for speech: Discrimination of the [r-l] distinction by young infants. *Perception & Psychophysics, 18*(5), 341-347.

Ellis, N. C. 2002. Frequency effects in language processing. A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition, 24*(02), 143-188.

Ellis, R. 1993. The structural syllabus and second language acquisition. *TESOL Quarterly, 27*(1), 91-113.

Flege, J. E. 1984. The detection of French accent by American listeners. *The Journal of the Acoustical Society of America, 76*(3), 692-707.

Flege, J. E. 1987. The production of 'new' and 'similar' phones in a foreign language: Evidence for the effect of equivalent classification. *Journal of Phonetics, 15*, 47-65.

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. 1993. Production and perception of a novel, second-language phonetic contrast. *Journal of the Acoustical Society of America, 93*(3), 1589-1608.

Flege, J. E. 1995. Second language speech learning theory, findings, and problems. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research* (pp. 233-277). Timonium, MD: York Press.

Flege, J. E. 2002. Interactions between the native and second-language phonetic systems. In P. Burmeister, T. Piske & A. Rohde (Eds.), *An Integrated View of Language Development: Papers in Honor of Henning Wode* (pp. 217-243). Trier: Wissenschaftlicher Verlag.

Flege, J. E. 2009. Give input a chance! In T. Piske, & M. Young-Scholten (Eds.), *Input Matters in SLA* (pp. 175-189). Buffalo, NY: Multilingual Matters.

Flege, J. E., Bohn, O. S., & Jang, S. 1997. Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics, 25*(4), 437-470.

Flege, J. E., & Eefting, W. 1987. Production and perception of English stops by native Spanish speakers. *Journal of Phonetics, 15*, 67-83.

Flege, J. E., & 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.

Flege, J. E., Frieda, E. M., & Nozawa, T. 1997. Amount of native-language (L1) use affects the pronunciation of an L2. *Journal of Phonetics, 25*(2), 169-186.

Flege, J. E., Frieda, E. M., Walley, A. C., & Randazza, L. A. 1998. Lexical factors and segmental accuracy in second language speech production. *Studies in Second Language Acquisition, 20*(2), 155-187.

Flege, J. E., & Hammond, R. M. 1982. Mimicry of non-distinctive phonetic differences between language varieties. *Studies in Second Language Acquisition, 5*(01), 1-17.

Flege, J. E., and Hillenbrand, J. M. 1984: Limits on Pronunciation Accuracy in Adult Foreign Language Speech Production. *Journal of the Acoustical Society of America, 76*(3), 708-721.

Flege, J. E., & Liu, S. 2001. The effect of experience on adults' acquisition of a second language. *Studies in Second Language Acquisition, 23*(4), 527-552.

Flege, J. E., MacKay, I. R. A., & Meador, D. 1999. Native Italian speakers' perception and production of English vowels. *Journal of the Acoustical Society of America, 106*(5), 2973-2987.

Flege, J. E., Munro, M. J., & Mackay, I. R. A. 1995. Effects of age of second-language learning on the production of English consonants. *Speech Communication, 16*(1), 1-26.

Flege, J. E., Munro, M. J., & MacKay, I. R. A. 1995. Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America, 97*(05), 3125-3134.

Flege, J. E., Schirru, C., & MacKay, I. R. A. 2003. Interaction between the native and second language phonetic subsystems. *Speech Communication, 40*(4), 467-491.

Flege, J. E., Yeni-Komshian, G. H., & Liu, S. 1999. Age constraints on second-language acquisition. *Journal of Memory and Language, 41*(1), 78-104.

Foster, P., & Tavakoli, P. 2009. Native speakers and task performance: Comparing effects on complexity, fluency, and lexical diversity. *Language Learning, 59*(4), 866-896.

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

Freed, B. 1995b. What makes us think that students who study abroad become fluent? In B. Freed (Ed.), *Second Language Acquisition in a Study Abroad Contex*t (pp. 123-148). Amsterdam: John Benjamins Publishing Company.

Freed, B., Dewey, D. P., Segalowitz, N., & Halter, R. 2004. The language contact profile. *Studies in Second Language Acquisition, 2*6(2), 349-356.

Fullana, N. 2005. *Age-Related Effects on the Acquisition of a Foreign Language Phonology in a Formal Setting.* PhD thesis. Universitat de Barcelona. Retrieved from http://www.tdx.cat/handle/10803/1663

Fullana, N., & MacKay, I. R. A. 2003. Production of English sounds by EFL learners: The case of /i:/ and /ɪ/. In D. Recasens (Ed.), *Proceedings of the 15th International Congress of Phonetic Science*s (pp. 1525-1528). Australia: Causal Productions.

Fullana, N., Miralpeix, I., & MacKay, I. R. A. 2012.Examining the contribution of L2 vocabulary size to segmental perception in EFL learning. *Paper presented at the 22th International Conference of the European Second Language Association - Eurosla,* Adam Mickiewicz University, Poland.

Gass, S., and Mackey, A. 2006: Input, Interaction and Output. *AILA Review, 19*, 3-17.

Gass, S., & Mackey, A. 2007. Input, interaction and output in second language acquisition. In B. VanPatten, & J. Williams (Eds.), *Theories in Second Language Acquisition: An Introductio*n (pp. 175-200). Mahwah, NJ: Lawrence Erlbaum Associates.

Gass, S., & Varonis, E. 1984. The effect of familiarity on the comprehensibility of nonnative speech. *Language Learning, 3*4, 65-89.

Goldstein, L., & Fowler, C. A. 2003. Articulatory phonology: A phonology for public language use. In N. O. Schiller, & A. S. Meyer (Eds.), *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities* (pp. 159-207). Berlin: Mouton de Gruyter.

Gor, K., & Long, M. 2009. Input and second language processing. In W. C. Ritchie, & T. K. Bhatia (Eds.), *The New Handbook of Second Language Acquisition* (pp. 445-472). Bingley, UK: Emerald.

Grosjean, F. 1989. Neurolinguistics, beware! The bilingual is not two monolinguals in one person. *Brain and Language, 36*, 3-15.

Guion, S. G., Flege, J. E., & Loftin, J. D. 2000. The effect of L1 use on pronunciation in Quichua-Spanish bilinguals. *Journal of Phonetics, 28*, 27-42.

Halle, P., Best, C. T., & Levitt, A. G. 1999. Phonetic vs. phonological influences on french listeners' perception of american English approximants. *Journal of Phonetics, 27*(3), 281-306.

Hazan, V., & Barrett, S. 1999. The development of phonemic categorization in children aged 6-12. *Journal of Phonetics, 28*, 377-396.

Hillenbrand, J. M., Clark, M. J., & Houde, R. A. 2000. Some effects of duration on vowel recognition. *Journal of the Acoustical Society of America, 108*(6), 3013-3022.

Hogan, J. T., & Rozsypal, A. J. 1980. Evaluation of vowel duration as a cue for the voicing distinction in the following word-final consonant. *Journal of the Acoustical Society of America, 67*(5), 1764-1771.

Højen, A. D. 2003. *Second-Language Speech Perception and Production in Adult Learners before and After Short-Term Immersion.* Unpublished PhD thesis. University of Aarhus.

Howard, M. 2005. Second language acquisition in a study abroad context: A comparative investigation of the effects of study abroad and foreign language instruction on the L2 learner's grammatical development. In A. Housen, & M. Pierrard (Eds.), *Investigations in Instructed Second Language Acquisition* (pp. 495-530). Berlin: Mouton de Gruyter.

Howard, M. 2006. The expression of number and person through verb morphology in advanced French interlanguage. *International Review of Applied Linguistics in Language Teaching, 44*(1), 1-22.

Hulstijn, J. 2003. Incidental and intentional learning. In C. J. Doughty, & M. Long (Eds.), *The Handbook of Second Language Acquisition* (pp. 349-381). Malden Mass.: Blackwell.

Ife, A., Vives Boix, G., & Meara, P. 2000. The impact of study abroad on the vocabulary development of different proficiency groups. *Spanish Applied Linguistics, 4*(1), 55-84.

IPA. 1999. *Handbook of the International Phonetic Association.* Cambridge: Cambridge University Press.

Isaacs, T., & Trofimovich, P. 2010. Falling on sensitive ears? the influence of musical ability on extreme raters' judgments of L2 pronunciation. *TESOL Quarterly, 44*(2), 375-386.

Isaacs, T., & Trofimovich, P. 2011. Phonological memory, attention control, and musical ability: Effects of individual differences on rater judgments of second language speech. *Applied Psycholinguistics, 32*(1), 113-140.

Iverson, P., & Kuhl, P. K. 1995. Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *Journal of the Acoustical Society of America, 97*(1), 553-562.

Juan-Garau, M., & 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.

Jusczyk, P., Rossner, B. S., Cutting, J. E., Foard, C. F., & Smith, L. B. 1977. Categorical perception of nonspeech sounds by 2-month-old infants. *Perception & Psychophysics, 21*(1), 50-54.

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

Klapper, J., & Rees, J. 2003. Reviewing the case for explicit grammar instruction in the university foreign language learning context. *Language Teaching Research, 7*(3), 285-314.

Krashen, S. D. 1985. *The Input Hypothesis: Issues and Implications*. New York: Longman.

Kuhl, P. K. 1991. Perception, cognition, and the ontogenetic and phylogenetic emergence of human speech. In S. E. Brauth, W. S. Hall & R. J. Dooling (Eds.), *Plasticity of Development* (pp. 73-106). Cambridge, MA: MIT Press.

Kuhl, P. K. 1994. Learning and representation in speech and language. *Current Opinion in Neurobiology, 4*, 812-822.

Kuhl, P. K., & Iverson, P. 1995. Linguistic experience and the "perceptual magnet effect". In W. Strange (Ed.), (pp. 121-154). Timonium, MD: York Press.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. 1992. Linguistic experience alters phonetic perception in infants by 6 months of age. *Science, 255*, 606-608.

Ladefoged, P. 2003. *Phonetic Data Analysis: An Introduction to Fieldwork and Instrumental Techniques*. Malden, MA: Blackwell Publishing.

Ladefoged, P. 2005. *Vowels and Consonants*. Malden, MA: Blackwell Publishing.

Lado, R. 1957. *Linguistics Across Cultures*. Ann Arbor: University of Michigan Press.

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* (pp. 97-122). Amsterdam: John Benjamins Publishing Company.

Lafford, B. 2006. The effects of study abroad vs. classroom contexts on Spanish SLA: Old assumptions, new insights and future research directions. In C. A. Klee, & T. L. Face (Eds.), *Selected Proceedings of the 7th Conference on the Acquisition of Spanish and Portuguese as First and Second Languages* (pp. 1-25). Somerville, MA: Cascadilla Proceedings Project.

Lafford, B., & Collentine, J. 2006. The effects of of study abroad and classroom contexts on the acquisition of Spanish as a second language: From research to application. In R. Salaberry, & B. Lafford (Eds.), *The Art of Teaching Spanish: Second Language Acquisition from Research to Practice* (pp. 103-126). Washington DC: Georgetown University Press.

Lapkin, S., Hart, D., & Swain, M. 1995. A Canadian interprovincial exchange: Evaluating the linguistic impact of a three-month stay in Quebec. In B. Freed (Ed.), *Second Language Acquisition in a Study Abroad Context* (pp. 67-94). Amsterdam: John Benjamins Publishing Company.

Larsen-Freeman, D., & Long, M. 1991. *An Introduction to Second Language Acquisition Research*. London: Longman.

Lee, S. A., Davis, B., & MacNeilage, P. 2010. Universal production patterns and ambient language influences in babbling: A cross-linguistic study of Korean- and English-learning infants. *Journal of Child Language, 37*(02), 293-318.

Lenneberg, E. H. 1967. *Biological Foundations of Language*. New York: John Wiley & Sons.

Lisker, L., & Abramson, A. S. 1970.The voicing distinction: Some experiments in comparative phonetics. *6th International Congress of Phonetic Sciences*, Prague.

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., & Muñoz, C. 2009. A short stay abroad: Does it make a difference? *System, 37*(3), 353-365.

Logan, J. S., Lively, S. E., & Pisoni, D. B. 1991. Training Japanese listeners to identify /r/ and /l/: A first report. *Journal of the Acoustical Society of America, 89*, 874-886.

Long, M. 1983. Native speaker/non-native speaker conversation and the negotiation of comprehensible input. *Applied Linguistics, 4*(2), 126-141.

Long, M. 1990. Maturational constraints on language development. *Studies in Second Language Acquisition, 12*(03), 251-285.

Long, M. 1996. The role of linguistic environment in second language acquisition. In W. C. Ritchie, & T. K. Bhatia (Eds.), *Handbook of Second Language Acquisition* (pp. 413-468). San Diego, CA: Academic Press.

Lord, G. 2010. The combined effects of immersion and instruction on second language pronunciation. *Foreign Language Annals, 43*(3), 488-503.

MacKay, I. R. A., Flege, J. E., & Imai, S. 2006. Evaluating the effects of chronological age and sentence duration on degree of perceived foreign accent. *Applied Psycholinguistics, 27*, 157-183.

MacKay, I. R. A., Flege, J. E., Piske, T., & Schirru, C. 2001. Category restructuring during second-language speech acquisition. *Journal of the Acoustical Society of America, 110*(01), 516-528.

MacKay, I. R. A., Meador, D., & Flege, J. E. 2001. The identification of English consonants by native speakers of italian. *Phonetica, 58*, 103-125.

Magen, H. S. 1998. The perception of foreign-accented speech. *Journal of Phonetics, 26*(4), 381-400.

Major, R. C. 2001. *Foreign Accent: The Ontogeny and Phylogeny of Second Language Phonology*. Mahwah, NJ: Lawrence Erlbaun Associates.

Marriott, H. 1995. The acquisition of politeness patterns by exchange students in japan. In B. Freed (Ed.), *Second Language Acquisition in a Study Abroad Context* (pp. 197-224). Amsterdam: John Benjamins Publishing Company.

McAllister, R., Flege, J. E., & Piske, T. 2002. The influence of L1 on the acquisition of swedish quantity by native speakers of Spanish, English and estonian. *Journal of Phonetics, 30*(2), 229-258.

McGurk, H., & MacDonald, J. 1976. Hearing lips and seeing voices. *Nature, 264*, 746-748.

McLaughlin, B., & Heredia, R. 1996. Information-processing approaches to research on second language acquisition. In W. C. Ritchie, & T. K. Bhatia (Eds.), *Handbook of Second Language Acquisition* (pp. 213-228). San Diego, CA: Academic Press.

Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. 1988. A precursor of language acquisition in young infants. *Cognition, 29*, 143-178.

Merino, E., & Avello, P. Forthcoming. Contrasting intercultural awareness at home and abroad. In C. Pérez-Vidal (Ed.), *Language Acquisition in Study Abroad and Formal Instruction Contexts* (). Amsterdam: John Benjamins.

Milton, J., & Meara, P. 1995. How periods abroad affect vocabulary growth in a foreign language. *ITL Review of Applied Linguistics, 107*, 17-34.

Mitchell, R., McManus, K., & Tracy-Ventura, N. 2013.Impact of social networking and social integration on development of L2 accuracy and fluency during residence abroad: A longitudinal study. *Paper presented at the 23th International Conference of the European Second Language Association - Eurosla,* ACLC, University of Amsterdam.

Mitchell, R., Tracy-Ventura, N., McManus, K., Richard, L., Romero de Mills, P., & Dewaele, J. 2013.The influence of social networks, personality and placement type on language learning during residence abroad: Preliminary findings of the LANGSNAP project. *Paper presented at the Residence Abroad, Social Networks and Second Language Learning Congress,* University of Southampton, UK.

Mora, J. C. 2006.Learning context effects on the production of English stop consonants by bilingual Catalan/Spanish learners of English. *24th AESLA Proceedings,* Madrid (Spain).

Mora, J. C. 2008. Learning context effects on the acquisition of a second language phonology. In C. Pérez-Vidal, M. Juan-Garau & A. Bel (Eds.), *A Portrait of the Young in the New Multilingual Spain* (pp. 241-263). Clevendon: Multilingual Matters.

Mora, J. C. Forthcoming. The role of onset level on L2 perceptual phonological development during formal instruction and study abroad. In C. Pérez-Vidal (Ed.), *Language Acquisition in Study*

Abroad and Formal Instruction Contexts. Amsterdam: John Benjamins.

Mora, J. C., & Fullana, N. 2007. Production and perception of English /i:/-/ɪ/ and /æ/-/ʌ/ in a formal setting: Investigating the effects of experience and starting age. In J. Trouvain, & W. J. Barry (Eds.), *Proceeding of the 16th International Congress of Phonetic Sciences* (pp. 1613-1616). Dudweiler: Pirrot.

Mora, J. C., & Nadeu, M. 2012. L2 effects on the perception and production of a native vowel contrast in early bilinguals. *The International Journal of Bilingualism, 16*(4), 484-500.

Mora, J. C., & Valls-Ferrer, M. 2012. Oral fluency, accuracy and complexity in formal instruction and study abroad learning contexts. *TESOL Quarterly, 16*(4), 484-500.

Morales-Front, A. 2006: Acquisition of Syllable Structure in Spanish. In F. Martínez-Gil, and S. Colina (Eds.), *Optimality-Theoretic Studies in Spanish Phonology*. Philadelphia: John Benjamins Publishing Company, pp.497-524.

Morse, P. A. 1972. The discrimination of speech and nonspeech stimuli in early infancy. *Journal of Experimental Child Psychology, 14*(3), 477-492.

Moyer, A. 2004. *Age, Accent and Experience in Second Language Adquisition :An Integrated Approach to Critical Period Inquir*y. Clevedon: Multilingual Matters.

Moyer, A. 2009. Input as a critical means to an end: Quantity and quality of experience in L2 phonological attainment. In T. Piske, & M. Young-Scholten (Eds.), *Input Matters in SL*A (pp. 159-174). Buffalo, NY: Multilingual Matters.

Munro, M. J. 1995. Nonsegmental factors in foreign accent: Ratings of filtered speech. *Studies in Second Language Acquisition, 17*, 17-34.

Munro, M. J. 1998. The effects of noise on the intelligibility of foreign-accented speech. *Studies in Second Language Acquisition, 2*0(2), 139-154.

Munro, M. J. 2008. Foreign accent and speech intelligibility. In J. G. Hansen Edwards, & M. L. Zampini (Eds.), *Phonology and Second Language Acquisitio*n (pp. 193-218). Amsterdam: John Benjamins Publishing Company.

Munro, M. J., & Derwing, T. M. 1995a. Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning, 4*5(1), 73-97.

Munro, M. J., & Derwing, T. M. 1995b. Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech, 3*8(3), 289-306.

Munro, M. J., & Derwing, T. M. 1998. The effects of speaking rate on listener evaluations of native and foreign-accented speech. *Language Learning, 4*8(2), 159-182.

Munro, M. J., Derwing, T. M., & Morton, S. L. 2006. The mutual intelligibility of L2 speech. *Studies in Second Language Acquisition, 2*8(1), 111-131.

Munro, M. J., Flege, J. E., & Mackay, I. R. A. 1996. The effects of age of second language learning on the production of English vowels. *Applied Psycholinguistics, 1*7(03), 313-334.

Munro, M. J., & Derwing, T. M. 1999. Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning, 4*9, 285-310.

Newell, A., & Rosenbloom, P. S. 1981. Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive Skills and their Acquisitio*n (pp. 1-55). Hillsdale: Lawrence Erlbaum.

Nittrouer, S. 2002. Learning to perceive speech: How fricative perception changes, and how it stays the same. *Journal of the Acoustical Society of America, 11*2(2), 711-719.

Ohala, D. K. 2008. Phonological acquisition in a first language. In J. G. Hansen Edwards, & M. L. Zampini (Eds.), *Phonology and Second Language Acquisitio*n (pp. 19-40). Amsterdam: John Benjamins Publishing Company.

Oller, D. K., & Eilers, R. E. 1982. Similarity of babbling in Spanish-and English-learning babies. *Journal of Child Language,* 9(03), 565-577.

Oyama, S. 1976. A sensitive period for the acquisition of a nonnative phonological system. *Journal of Psycholinguistic Research,* 5(3), 261-283.

Patkowski, M. S. 1990. Age and accent in a second language: A reply to James Emil Flege. *Applied Linguistics, 1*1(1), 73-89.

Pérez-Vidal, C. 2011. Language acquisition in three different contexts of learning: Formal instruction, study abroad and semi-immersion (CLIL). In Y. Ruiz de Zarobe, J. M. Sierra & F. Gallardo del Puerto (Eds.), *Content and Foreign Language Integrated Learning: Contributions to Multilingualism in European Context*s (pp. 25-35). Berlin: Peter Lang.

Pérez-Vidal, C. Forthcoming: Study Abroad and Language Acquisition: The SALA Project. In C. Pérez-Vidal (Ed.), *Language Acquisition in Study Abroad and Formal Instruction Contexts*. Amsterdam: John Benjamins.

Pérez-Vidal, C., & Barquin, E. Forthcoming. Comparing progress in academic writing after formal instruction and study abroad. In C. Pérez-Vidal (Ed.), *Language Acquisition in Study Abroad and Formal Instruction Context*s. Amsterdam: John Benjamins.

Pérez-Vidal, C., & Juan-Garau, M. 2009. The effect of study abroad on written performance. In L. Roberts, D. Véronique, A. Nilsson & M. Tellier (Eds.), *EUROSLA Yearbook Volume 9 (2009*) (pp. 269-295). Amsterdam: John Benjamins Publishing Company.

Pérez-Vidal, C., & 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, 4*9(2), 157-185.

Pérez-Vidal, C., Juan-Garau, M., & Mora, J. C. 2011. The effects of formal instruction and study abroad contexts on foreign language development: The SALA project. In C. Sanz, & R. P. Leow (Eds.), *Implicit and Explicit Conditions, Processes and Knowledge in SLA and Bilingualis*m (pp. 115-138). Washington D. C.: Georgetown University Press.

Peterson, G. E., & Lehiste, I. 1960. Duration of syllable nuclei in English. *The Journal of the Acoustical Society of America, 3*2(6), 693-703.

Piske, T. 2007. Implications of James E. Flange's research for the foreign language classroom. In O. S. Bohn, & M. J. Munro (Eds.), *Language Experience in Second Language Speech Learning: In Honor of James Emil Fleg*e (pp. 301-314). Amsterdam: John Benjamins Publishing Company.

Piske, T., Flege, J. E., MacKay, I. R. A., & Meador, D. 2002. The production of English vowels by fluent early and late Italian-English bilinguals. *Phonetica, 5*9, 49-71.

Piske, T., MacKay, I. R. A., & Flege, J. E. 2001. Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics, 2*9(2), 191-215.

Polka, L. 1992. Characterizing the influence of native experience on adult speech perception. *Perception & Psychophysics, 5*2, 37-52.

Polka, L., & Werker, J. F. 1994. Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance, 2*0(2), 421-435.

Purcell, E. T., & Suter, R. W. 1980. Predictors of pronunciation accuracy: A reexamination. *Language Learning, 3*0(2), 271-287.

Quilis, A. (1993). *Tratado de fonología y fonética española*s. Madrid: Gredos.

Regan, V. 1995. The acquisition of sociolinguistic native speech norms: Effects of a year abroad on second language learners of French. In B. Freed (Ed.), *Second Language Acquisition in a Study Abroad Contex*t (pp. 245-268). Amsterdam: John Benjamins Publishing Company.

Riney, T., & Flege, J. E. 1998. Changes over time in global foreign accent and liquid identifiability and accuracy. *Studies in Second Language Acquisition, 2*0, 213-243.

Roach, P. 2004. British English: Received pronunciation. *Journal of the International Phonetic Association, 3*4(02), 239-245.

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

Rossiter, M. J. 2009. Perception of L2 fluency by native and non-native speakers of English. *The Canadian Modern Language Review, 6*5(3), 395-412.

Saffran, J. R., Newport, E. L., & Aslin, R. N. 1996. Word segmentation: The role of distributional cues. *Journal of Memory and Language, 3*5(4), 606-621.

Sanz, C. Forthcoming. Contributions of study abroad research to our understanding of SLA processes and outcomes: The SALA project, an appraisal. In C. Pérez-Vidal (Ed.), *Language Acquisition in Study Abroad and Formal Instruction Context*s. Amsterdam: John Benjamins.

Sanz, C., and Grey, S. 2012: The Effect of Immersion Experience on L2 Lexical and Morphosyntactic Development. *Paper presented at the American Association for Applied Linguistics- AAAL 2012,* Boston,MA.

Sanz, C., Morales-Front, A., Nagle, C., and Moorman, C. 2013: Learning without Attention: Improvements in Pronunciation in a 6 Week Study Abroad Program. *Paper presented at the Residence Abroad, Social Networks and Second Language Learning Congress,* University of Southampton, UK.

Sasaki, M. 2004. A multiple-data analysis of the 3.5-year development of EFL student writers. *Language Learning, 5*4(3), 525-582.

Sasaki, M. 2009. Changes in English as a foreign language students' writing over 3.5 years: A sociocognitive account. In R. M. Manchón (Ed.), *Writing in Foreign Language Contexts: Learning, Teaching, and Researc*h (pp. 49-76). Clevedon: Multilingual Matters.

Schmidt, R. W. 1990. The role of consciousness in second language learning. *Applied Linguistics, 1*1(2), 129-158.

Schmidt, R. W. 2001. In P. Robinson (Ed.), *Cognition and Second Language Instruction* (pp. 3-32). New York, NY: Cambridge University Press.

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

Scovel, T. 1988. *A Time to Speak. A Psycholinguistic Inquiry into the Critical Period for Human Speech*. Cambridge, MA: Newbury House Publishers.

Segalowitz, N. 2010. *Cognitive Bases of Second Language Fluency*. New York: Routledge.

Segalowitz, N., & 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.

Segalowitz, N., & Hulstijn, J. 2005. Automaticity in bilingualism and second language learning. In J. F. Kroll, & A. M. B. de Groot (Eds.), *Handbook of Bilingualism :Psycholinguistic Approaches* (pp. 127-190). New York: Oxford University Press.

Selinker, L. 1972. Interlanguage. *International Review of Applied Linguistics in Language Teaching, 10*(3), 209-231.

Serrano, R., Tragant, E., & Llanes, À. 2012. A longitudinal analysis of the effects of one year abroad. *Canadian Modern Language Review, 68*(2), 138-163.

Simões, A. R. M. 1996. Phonetics in second language acquisition: An acoustic study of fluency in adult learners of Spanish. *Hispania, 79*(1), 87-95.

Southwood, H. M., and Flege, J. E. 1999: Scaling Foreign Accent: Direct Magnitude Estimation Versus Interval Scaling. *Clinical Linguistics and Phonetics, 13*(5), 335-349.

Spencer, A. 1996. *Phonology: Theory and Description*. Oxford: Blackwell.

Stern, H. H. 1983. *Fundamental Concepts of Language Teaching*. Oxford: Oxford University Press.

Strange, W. 2007. Cross-language phonetic similarity of vowels. theoretical and methodological issues. In O. S. Bohn, & M. J. Munro (Eds.), *Language Experience in Second Language Speech Learning. in Honour of James Emil Flege* (pp. 35-55). Amsterdam: John Benjamins Publishing Company.

Strange, W., Akahane-Yamada, R., Kubo, R., Trent, S. A., & Nishi, K. 2001. Effects of consonantal context on perceptual assimilation of american English vowels by japanese listeners. *Journal of the Acoustical Society of America, 109*, 1691-1704.

Swain, M. 1985. Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass, & C. Madden (Eds.), *Input in Second Language Acquisition* (pp. 235-256). Rowley, MA: Newbury House.

Swain, M. 1993. The output hypothesis: Just speaking and writing aren't enough. *Canadian Modern Language Review, 50*, 158-164.

Swain, M., & Lapkin, S. 1998. Interaction and second language learning: Two adolescent French immersion students working together. *The Modern Language Journal, 82*(3), 320-337.

Syrdal, A. K., & Gopal, H. S. 1986. A perceptual model of vowel recognition based on the auditory representation of American English vowels. *Journal of the Acoustical Society of America, 79*(4), 1086-1100.

Tees, R. C., & Werker, J. F. 1984. Perceptual flexibility: Maintenance of recovery of the ability to discriminate non-native speech sounds. *Canadian Journal of Psychology, 38*, 579-590.

Towell, R. 2002. Relative degrees of fluency: A comparative case study of advanced learners of French. *International Review of Applied Linguistics, 40*(2), 117-150.

Trehub, S. E. 1973. Infants' sensitivity to vowel and tonal contrasts. *Developmental Psychology, 9*(1), 91-96.

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., & 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.

Trubetskoy, N. S. 1969. *Principles of Phonology*. Berkeley, CA: University of California Press.

Valls-Ferrer, M. 2011. *The Development of Oral Fluency and Rhythm during a Study Abroad Period.* PhD thesis. Universitat Pompeu Fabra. Retrieved from http://www.tdx.cat/handle/10803/52064

Van Els, T., & De Bot, K. 1987. The role of intonation in foreign accent. *The Modern Language Journal, 7*1(2), 147-155.

VanPatten, B. 2007. Input processing in adult second language acquisition. In B. VanPatten, & J. Williams (Eds.), *Theories in Second Language Acquisition: An Introductio*n (pp. 115-135). Mahwah, NJ: Lawrence Erlbaum Associates.

VanPatten, B. 2009. Processing matters in input enhancement. In T. Piske, & M. Young-Scholten (Eds.), *Input Matters in SL*A (pp. 47-61). Buffalo, NY: Multilingual Matters.

VanPatten, B., & Cadierno, T. 1993. Input processing and second language acquisition: A role for instruction. *The Modern Language Journal, 7*7(1), 45-57.

Varonis, E., & Gass, S. 1982. The comprehensibility of nonnative speech. *Studies in Second Language Acquisition,* 4, 114-136.

Verspoor, M., Lowie, W., & De Bot, K. 2009. Input and second language development from a dynamic perspective. In T. Piske, & M. Young-Scholten (Eds.), *Input Matters in SL*A (pp. 62-80). Buffalo, NY: Multilingual Matters.

Weil, S. A. 2003. *The Impact of Perceptual Dissimilarity on the Perception of Foreign Accented Speech.* Unpublished PhD thesis. Ohio State University.

Werker, J. F., & Tees, R. C. 1984a. Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development,* 7(1), 49-63.

Werker, J. F., & Tees, R. C. 1984b. Phonemic and phonetic factors in adult cross-language speech perception. *Journal of the Acoustical Society of America, 7*5(6), 1866-1878.

Williams, J. N. 2009. Implicit learning in second language acquisition. In W. C. Ritchie, & T. K. Bhatia (Eds.), *The New Handbook of Second Language Acquisitio*n (pp. 319-356). Bingley, UK: Emerald.

Wode, H. 1981. Language-acquisitional universals: A unified view of language acquisition. *Annals of the New York Academy of Sciences, 37*9(1), 218-234.

Wrembel, M. 2010.Third language phonological acquisition: An investigation into metalinguistic awareness and foreign language accent in L3. *Paper presented at the 20th International Conference of the European Second Language Association - Eurosla*, University of Modena and Reggio Emilia, Italy.

Wrembel, M. 2011. Cross-linguistic influence in third language acquisition of voice onset time. *Proceedings of the 17th International Congress of Phonetic Sciences. 17-21 August 2011. Hong Kong*, 2157-2160.

Wrembel, M. 2013: Cross-Linguistic Influence in the Acquisition of VOT in a Third Language. *Paper presented at the 7th International Symposium on the Acquisition of Second Language Speech (New Sounds),* Concordia University.

Wright, R. 2004. A review of perceptual cues and cue robustness. In B. Hayes, R. Kirchner & D. Steriade (Eds.), *Phonetically Based Phonology* (pp. 34-57). Cambrige,NY: CUP.

Yamada, R. A. 1995. Age and acquisition of second language speech sounds: Perception of American English /ɹ/ and /l/ by native speakers of Japanese. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Researc*h (pp. 305-320). Timonium, MD: York Press.

Ylinen, S., Uther, M., Latvala, A., Vepsäläinen, S., Iverson, P., Akahane-Yamada, R., & Näätänen, R. 2010. Training the brain to weight speech cues differently: A study of Finnish second-language users of English. *Journal of Neuroscience, 2*2(6), 1319-1332.

Zampini, M. L. 2008. L2 speech production research: Findings, issues, and advances. In J. G. Hansen Edwards, & M. L. Zampini (Eds.), *Phonology and Second Language Acquisitio*n (pp. 219-249). Amsterdam: John Benjamins Publishing Company.

Appendices

Appendix 1. SALA reading aloud task

Appendix 2. Sentence for FA ratings and phonetic analyses

Appendix 1. SALA reading aloud task

SALA / COL / ATT / SU code: _____ Date:_____
SURNAMES and name: _____ T1

READING ALOUD TASK– 2'

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 2. Sentence for FA ratings and phonetic analyses

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	ˈt̪ɹæv(ə)lə	tʊk	ˈɒf	(h)ɪz̪	kl̪əʊk