CHAPTER 2

LITERATURE REVIEW

Chapter 2 sets up the framework within which the current dissertation was conducted by defining a number of key issues related to second language acquisition (SLA) first. A review of the factors most often examined in SLA is then presented, with a special emphasis on the age factor. This is followed by a section on the acquisition of second language (L2) phonology, which includes a more in-depth review of three models of speech perception and their respective supporting studies. A representative sample of research on the perception and production of target language sounds by Romance language speakers of English is described next. The following section offers a summary of the main findings of foreign accent research. This chapter concludes with an outline of a number of methodological issues to be taken into consideration when conducting L2 phonological acquisition research.

2.1. General overview

The aim of this section is to provide an account of several key issues concerning second language acquisition (SLA), a research area that developed in academia in the 1960s with a growing interest ever since.

Taking Ellis' (1994) definition of SLA as the starting point – i.e. "the study of how learners learn an additional language after they have acquired their mother tongue" (p. 5) – the following needs consideration.

First, the additional language that learners learn after they have acquired their mother tongue – or first language (L1) – is normally referred to as second language (L2). L2 and, by extension, SLA are terms that comprise both L2 and foreign language (FL) acquisition. The difference between L2 and FL lies in the geographical context in which a language is spoken. For instance, Johnson and Johnson (1999) make the following distinction between English as an L2 (ESL) and English as an FL (EFL):

> An ESL situation is one where English is widely used in commerce, administration and education. It is a foreign language (EFL) in a country where English plays no such role. When English is taught to non-native speakers in an English-speaking country, ESL usually refers to people who are long-stay or permanent residents, whereas EFL is taught to those who return after a period of time to their own country. (pp. 133-134)

Alternatively, the term target language (TL) can be used to refer to the language that is the focus or object of study on the learner's side, be it an L2 or an FL.

Another concept related to the additional language being learned is that of interlanguage (IL) (Selinker, 1972). IL is the language that learners possess during SLA that is neither their L1 nor the TL. To put it in Crystal's (2003) words, IL is

> The linguistic system created by someone in the course of learning a foreign language, different from either the speaker's first language or the target language being acquired. It reflects the learner's evolving system of rules, and results from a variety of processes, including the influence of the first language ('transfer'), contrastive interference from the target language, and the overgeneralization of newly encountered rules. (p. 239)

Overall, L2 learning (or acquisition) takes place in immersion – naturalistic – settings, while FL learning (or acquisition) occurs in formal – classroom – settings. A major difference between the two types of learning contexts has to do with the amount of input in the TL that learners are exposed to, which is by far greater in a naturalistic environment. Moreover, as Johnson and Johnson (1999) note, the type of input that learners receive in an immersion setting has not previously been selected and comes mainly from native speakers (NSs) of the TL. In contrast, in a tutored environment learners have limited and selected exposure to input in the TL, which, in turn, is usually delivered by non-native speakers (NNSs).

Last, embedded in Ellis' definition of SLA, there are two processes that often appear hand-in-hand in SLA research, namely acquisition and learning. Although the two terms have been used interchangeably in the literature (e.g. Ellis, 1994; Krashen, Long, & Scarcella, 1979/1982), originally the distinction between acquisition and learning is ascribed to Krashen's work (Singleton, 1989, p. 7), whereby acquisition consists of "picking up a second language unconsciously through exposure", while learning is "the conscious study of an L2" (Ellis, 1994, p. 6). The latter distinction has not been exempt from criticism, as summarised by Johnson and Johnson (1999). However, it should be noted that all the key issues defined in this section illustrate scholars' majority view in the field of SLA.

Besides, it is within the framework outlined above that the present dissertation was conducted. More precisely, this dissertation *Age-related Effects on the Acquisition of a Foreign Language Phonology in a Formal Setting* examines the perception and production of sounds by Spanish and Catalan learners of English as an FL, differing in starting age of FL learning and amount of instruction in the FL. In the present investigation, then, the TL under study has the role of FL, involving, as it does, subjects having limited and selected exposure to English. Accordingly, FL will be used to refer to the language – English – being learned by the subjects in the current study. However, in those sections reviewing previous research the generic term L2 will be used to refer to "any language being learned other than the first language" (Singleton, 1989, p. 7, footnote 2). Moreover, in the present dissertation subjects had only been exposed to the TL in a formal learning context.

In accordance with most work cited above (e.g. Ellis, 1994; Singleton, 1989), there will be no explicit distinction as to whether a subconscious or conscious process is taking place. In other words, *acquisition* and *learning* will be used on a similar synonymous basis.

Furthermore, the aspect of language being investigated is phonology in broad terms. This is in agreement with current trends in SLA research, while examination of morphological and syntactic aspects of the TL was the primary focus of study until recently (for a review, see Ellis, 1994, among others).

Finally, among the various factors that have been studied on the attainment (or lack of attainment) of native-like proficiency in the TL, emphasis will be placed on starting age of SLA, and to a lesser extent, on amount and type of exposure and subjects' L1 and gender.

In what follows a summary of the main variables examined in SLA research will be presented with special reference to the acquisition of phonological skills, which is of interest to this dissertation.

2.2. Age and SLA

In his introduction to *Language Acquisition: The Age Factor* Singleton (1989) comments on the importance that age[1] has received in SLA research. He further states that this interest is likely to have sprung originally from the popular observation that children are better than adults at learning an L2; thus, leading to numerous studies of what the ideal or optimal age to begin learning an L2 is.

In the literature one of the first allusions to the optimal age to learn an L2 can be found in Penfield's work during the 1950s, where he mentioned an approximate ideal age range from nursery school to before age 10 or 14: "Let the first years, from nursery school and kindergarten on to grades for children of eight or ten, be conducted by foreign-born teachers… (p. 202)" (Penfield, 1953; as cited in Dechert, 1995, p. 79).

> If, before the age of 10 or 14, the child associates with those who speak a second and even a third language, he can learn by a similar technique [to the mother's method of teaching his first language] two or three languages with no evident increase in his effort. (p. 207) (Penfield, 1953; as cited in Dechert, 1995, p. 79)

A decade later – 1967 to be exact – E. H. Lenneberg published *Biological Foundations of Language* that included a (better defined) reference to an optimal period for language acquisition: "language development thus runs a definite course on a definite schedule; a critical period extends from about age 2 to age 12, the beginning and the end of resonance" (Lenneberg, 1970, p. 4). Although Lenneberg concentrated for the most part on first language acquisition by aphasic patients and children born to deaf parents, he turned to his incipient studies of foreign accent[2] in the L2 in order to illustrate the fact that progress in language development normally ended after the ages of 12 or 13 – in other words, after puberty. Thus, he pointed out a relationship between foreign accent and starting age of L2 learning, whereby a child learning to speak the TL at the age of 3 or 4 was thought to speak it with no foreign accent. By contrast, children starting to learn the TL at puberty – about 12 years of age – failed to speak the L2 accent-free. According to

[1] When talking about *age* or *the age factor* in SLA research, it is meant age at which L2 learning/acquisition started. Other terms used synonymously are age of onset/onset age (of L2 learning or acquisition) and starting age (of L2 learning or acquisition). In naturalistic research, the term AOA (age of arrival [in the TL country]) is often used in the same sense. As a result, these terms will be used interchangeably in the present dissertation (see also footnote 50 in the Method chapter).

[2] The notion of *foreign accent* will be dealt with in more detail in section 2.3.3. Suffice it to say, for the time being, that "foreign-accented speech … can be defined as nonpathological speech produced by second language (L2) learners that differs in partially systematic ways from the speech characteristic of native speakers of a given dialect." (Munro, 1998, p. 139)

Lenneberg (1967, 1970), the 4- to 12-year-old period was characterised as not showing a noticeable decline in the children's ability to speak the TL accent-free.

Johnson and Newport (1989) suggested that Lenneberg's notion of a critical period (CP) for language acquisition might be interpreted in the light of either the exercise hypothesis or the maturational state hypothesis, which they outlined as follows:

> *Version One: The exercise hypothesis*. Early in life, humans have a superior capacity for acquiring languages. If the capacity is not exercised during this time, it will disappear or decline with maturation. If the capacity is exercised, however, further language learning abilities will remain intact throughout life.
>
> *Version Two: The maturational state hypothesis*. Early in life, humans have a superior capacity for acquiring languages. This capacity disappears or declines with maturation. (p. 64)

They also indicated that Lenneberg's formulation of a CP was concerned with first language acquisition, whereas the two amendments Johnson and Newport made to the CPH involved not only L1 acquisition but also L2 acquisition. In the case of L1 acquisition, both hypotheses above agreed in predicting that children are superior to adults when it comes to learning their mother tongue. As far as L2 acquisition was concerned, each version hypothesised a different outcome depending on whether learners started to acquire the L2 as children or adults. On the one hand, the exercise hypothesis predicted that learners would be able to fully acquire an L2 no matter their starting age of L2 learning, provided that they had acquired their L1 during childhood (and therefore "exercised" this capacity for learning languages). What is more, adults were likely to be even better L2 learners than children, for they already had greater cognitive skills in the L1. On the other hand, the maturational state hypothesis suggested that children would be better L2 learners than adults, since maturation brought about a diminishment in the ability to learn languages.

Thus, in order to find out if there was an age effect on SLA – or a CP for SLA – as characterised by either the exercise hypothesis or the maturational state hypothesis, Johnson and Newport (1989) examined the performance of 46 Chinese or Korean speakers differing in age of arrival (AOA) in the US (from 3 to 39 years old) on a grammatical and morphological task. As in the Patkowski (1980) study below (cited in Patkowski, 1990), the distinction between earlier and later arrivals was set at the age of 15. In both cases, subjects should have had a minimum of five years of exposure to English in the target country, of which they should have lived three continuous years in the US right before testing.

The results were consistent with the predictions of the maturational state hypothesis in that age and ultimate performance in the L2 grammar were strongly correlated – i.e. the earlier a subject started to learn the L2, the more native-like their performance was – and in the nature of the decline in performance as a function of maturation. Therefore, within the pre-pubescent group, 3- to 7-year-olds performed like the English NS control group. After age 7 up to puberty a gradual decline in performance was observed, resulting in progressively lower (or less native-like) scores for the 8- to 10-year-olds and the 11- to 15-year-olds. The post-pubescent group as a whole (17- to 39-year-olds) performed at the lowest range, though L2 learning was shown to be possible in adulthood. Besides, there were no age subgroup differences in decline within the later arrival group, as the authors had hypothesised "for presumably there are not many important maturational differences between, for example, the brain of a 17-year-old and the brain of a 27-year-old" (Johnson & Newport, 1989, p. 79). Unexpectedly, though, noticeable individual differences surfaced as to adults' ultimate attainment in the TL.

Apart from maturational constraints, other factors considered such as motivation, amount and quality of English exposure, and L1 were not found to account significantly for earlier and later arrival group differences in performance. Nor were they a plausible explanation for the variability encountered in adults' performance.

In conclusion, the Johnson and Newport (1989) study confirmed the existence of a CP for L2 acquisition that manifests itself as the maturational state hypothesis predicts. In the authors' words, "human beings appear to have a special capacity for acquiring language in childhood, regardless of whether the language is their first or second." (p. 95)

Lenneberg's notion of a CP for language acquisition constituted a turning point in SLA research. In fact, the Critical Period Hypothesis (CPH) led to a still ongoing debate between those investigators who support its existence (e.g. Eubank & Gregg, 1999; Hurford & Kirby, 1999; Patkowski, 1979/1982, 1990, 1994; Weber-Fox & Neville, 1999) and those who reject it (e.g. Bialystok, 1997; Bialystok & Hakuta, 1999; Bongaerts, 1999; Bongaerts, van Summeren, Planken, & Schils, 1997; Flege, 1981, 1987a, 1999a, 2005; Major, 1987b; Snow & Hoefnagel-Höhle, 1977/1982, 1978/1982).

In the case of L2 pronunciation, Flege's (1987a) article 'A critical period for learning to pronounce foreign languages?' and Patkowski's (1990) rebuttal 'Age and accent in a second language: a reply to James Emil Flege' illustrate these two opposing views on the non-existence and existence of a CP for language learning, respectively.

According to Flege (1987a), the development of the notion of a CP to initially account for animal – not human – behaviour; findings of adult learners perceiving and/or producing L2 sounds on a similar basis to (or even better than) children (e.g. Flege, 1981; Snow & Hoefnagel-Höhle, 1978/1982), together with the observation of a linear decline – and not an abrupt drop-off – in learners' pronunciation of the L2 after the passing of a CP (e.g. Oyama, 1978/1982); and the insufficient neurological evidence at the time regarding the effects, if any, of age at which lateralisation (or hemispheric specialisation) was completed on language learning, are all facts that disconfirm the existence of a CP for language acquisition.

Besides, Flege (1987a) lists a number of factors that are often interrelated or confounded with age, and therefore might be the cause of child-adult differences in the pronunciation of the TL, rather than neurophysiological maturation. First, there are developmental factors in relation to "differences in size and functioning" (p. 167). For instance, Flege (1987a) notes that anatomical differences between adults and children may affect the articulation of segments in favour of the latter. Motivation and affective factors as well as social factors may also play a role in the differences encountered between younger and older learners' ultimate attainment, again to children's advantage. Another cause listed is the quantity and quality of L2 input, which varies depending on whether it is addressed to children or adults. Last, incomplete learning might be the reason for child-adult differences in the production of L2 sounds, for "differences in the rate of learning might be misinterpreted as representing a difference in the extent of learning" (Flege, 1987a, p. 171). Flege claims that successful command of the production of new L2 sounds requires time both on the side of child learners and adult acquirers. Thus, differences in rate of learning – e.g. children learning L2 sounds at faster rates than adults apparently – should be understood as a consequence of motivation, affective and social variables, and not as support for the existence of a CP for language learning.

Based on all of the above, Flege does not only conclude that the evidence available is not in favour of the existence of a CP for language acquisition, but he also puts forward two alternative, though related, hypotheses to the CPH. Thus, he hypothesises that differences in child-adult production of L2 sounds are due to children's preference of an auditory mode to a phonetic mode for speech processing (hypothesis 1), and/or due to children's still developing L1 phonetic categories – unlike adults who already have firmly established L1 phonetic categories – which allow them to develop new phonetic categories for new L2 sounds provided that there is sufficient exposure to

the L2 (hypothesis 2[3]). As a result of all this, and not neurophysiological maturation, children's production of L2 sounds has been deemed better than that of adults.

On the contrary, Patkowski (1990) argues that there exists enough evidence for a CP for language acquisition. In his view, the studies Flege reviews to support his dismissal of the CPH do not examine learners who have reached their ultimate attainment in the TL, whereas "ultimate proficiency is the real test of the CPH (and not rates of acquisition, whether initial or otherwise)" (p. 80). Thus, Patkowski suggests that the findings reported in Flege (1987a) of adult learners' advantage over younger learners in the pronunciation of the L2 are more consistent with Krashen, Long, & Scarcella's (1982) generalisation that confers an advantage to older learners in the performance in the L2 in the initial or first stages of SLA. He then goes on to comment on research that agrees with his line of thought, among others, Asher and García (1969/1982), Seliger, Krashen, and Ladefoged (1975/1982), and Tahta, Wood, and Loewenthal (1981a) concerning the phonological domain; and the above mentioned Johnson and Newport (1989) study related to the morpho-syntactic area. As for L2 phonology, the results indicated that learners exposed to the L2 before age 6 (Asher & García, 1969/1982; Tahta et al., 1981a) or at age 9 or under (Seliger et al., 1975/1982) either produced the L2 with no trace of FA (Tahta et al., 1981a) or at near-native rates (Asher & García, 1969/1982; Seliger et al., 1975/1982), whereas those starting to learn the L2 after puberty – around age 13 – pronounced the L2 with a consistent degree of FA. These findings, together with that of a decline in the ability to pronounce the L2 accent-free in the 7- to 12-year-old group as a function of starting age of L2 learning, agree with the assumptions of the CPH and the notion of maturational limitations.

Although Patkowski admits that investigations such as Mack (1984; as cited in Patkowski, 1990) (where she found that both English monolinguals and French-English early bilinguals performed similarly on the production and perception of English consonant segments) may cast doubt upon the validity of the CPH, he points out that studies supporting that hypothesis outnumber those few studies that disagree with the assumptions of the CPH. In addition, he presents findings of younger learners' imitation of FL sounds and intonation at better rates in their initial stages of SLA than those of adults (e.g. Loewenthal & Bull, 1984; Tahta, Wood, & Loewenthal, 1981b), which are in opposition to the findings reported above by Flege (1987a) to reject the CPH.

[3] Flege fully developed this hypothesis in his Speech Learning Model (see 2.3.1. Theoretical Frameworks).

Further arguments against Flege's viewpoint come from Patkowski's re-examination of his own 1980 study (as cited in Long, 1990; Patkowski, 1990). As in his previous results for the syntactic domain, he finds that there is a manifest discontinuity in the accent ratings that learners received according to whether they started to learn the TL before or after the passing of a sensitive period (see Oyama's definition of this term below); the latter finding constituting evidence for the CPH. All things considered, Patkowski states that the evidence presented in his article is at odds with Flege's arguments against the validity of the CPH; rather, it "is demonstrably consistent with the hypothesis" (p. 87).

Besides the CPH debate here exemplified in the area of L2 phonology, refinements of the CPH were not slow to appear; being sensitive period (Oyama, 1976/1982, 1978/1982) the formulation most widely accepted. Oyama's (1978/1982) preference of the term "sensitive" to "critical" lies in the fact that the former "reflects more accurately the gradual nature of such phenomena, as well as their responsiveness to variation in experience" (p. 40).

Oyama's hypothesis about the existence of a sensitive period, rather than a critical period, for the acquisition of an L2 phonology was tested by means of examining 60 Italian (male) immigrants' production and perception of English speech samples in her 1976 and 1978 studies, respectively. The 1976 study neither confirmed nor rejected the existence of a sensitive period, though that investigation, together with the 1978 study, did show a relationship between the subjects' ability to speak and comprehend the TL (English) at a near-native or native-like level and their age upon arrival in the host country (the US). Therefore, the subject group arriving between the ages of 6 and 10 performed on a similar basis to English NS controls, whereas the 11- to 15-year-old and 16+-year-old groups' scores differed from those of English NSs (yielding significant differences in the case of the older group). Moreover, results indicated that after age 11 subjects' accent and comprehension scores decreased – i.e. were less native-like – linearly, instead of suddenly dropping as predicted by the CPH[4]. This finding then led Oyama (1976/1982) to conclude that

> This sensitive period is obviously not an all-or-nothing phenomenon; adults can and do learn to speak new languages, and often very well. Whether the efficiency of the acquisition process simply decreases as a person grows older or whether the

[4] See Bialystok and Hakuta (1999) as well as Birdsong and Molis (2001) later in this section as to how they claim the finding of a linear decline in L2 performance by post-puberty learners should be interpreted in relation to the existence of a sensitive or critical period.

> process actually ceases to function, so that one must use other abilities whose efficiency is not so closely tied to maturation, learning to "speak like a native" seems to be quite difficult for all but the very young. (pp. 33-34)

Evidence in favour of the existence of a sensitive period[5] for SLA was also given by Patkowski (1979/1982), who further summarised the distinction between critical period and sensitive period as follows:

> The term 'critical period' refers to the notion that the age limitation is absolute. In theory, first language acquisition is not possible past the critical point. The term 'sensitive period', on the other hand, refers to the fact that age limitation is not absolute. It is indeed possible to acquire a foreign language at an adult age, but it is not possible to do so to the extent of being able to 'pass for native'. (p. 52)

In that case, Patkowski (1979/1982) looked at 67 immigrants' syntactic production in the L2 differing in age of arrival in the US – before and after age 15 – and in L1, being Spanish, Polish, and Chinese the most common mother tongues. Results supported the idea that a sensitive period existed for the acquisition of syntax in an L2 and established that native-like attainment of L2 syntax would only be possible if learning had started before age 15.

Further support for the existence of a sensitive period in the form of maturational constraints was provided by Long's (1990) review of a series of studies on L2 acquisition in addition to L1 acquisition[6]. More precisely, Long claimed that the available evidence indicated that there existed several sensitive periods – and not just one – for the acquisition of the various linguistic abilities in the TL.

> There are sensitive periods governing the ultimate level of first or second language attainment possible in different linguistic domains, not just phonology, with cumulative declines in learning capacity, not a catastrophic one-time loss, and beginning as early as age 6 in many individuals, not at puberty, as is often claimed. (Long, 1990, p. 255)

---

[5] Note that whenever Patkowski makes use of the term "critical" (e.g. in his 1990 reply to Flege, 1987a, alluded to above and in his 1994 review article on the CPH and IL phonology), he specifies it is used on a synonymous basis with what is understood by "sensitive" period, rather than contradicting himself as to his views on critical period vs. sensitive period:

> Since it is the purpose of this paper to argue against the rejection of the notion of an age-based constraint on the acquisition of full native fluency in a second tongue, the term 'critical period' is taken here only to refer to this particular notion and is used as an approximate synonym with 'sensitive period'. (Patkowski, 1990, pp. 73-74)

[6] L1 acquisition studies reviewed include the well-known case of Genie, a girl who had been deprived of social interaction from about 20 months of age until her discovery at the age of 13;7, and clinical cases involving hearing children of deaf adults and vice versa.

Thus, based on findings of studies such as Asher and García (1969), Fathman (1975), Oyama (1976), Scovel (1981), Seliger et al. (1975), and Tahta et al. (1981a) (all cited in Long, 1990), he hypothesised that learners should begin acquiring the L2 phonology before the age of 6 if they were to attain a native-like level in the L2 phonology. Moreover, if learning began after age 6 and before age 12, few (if any) learners would be able to attain native-like proficiency in the L2 phonology. Finally, after age 12 no learner was expected to perform in the L2 phonology within the NS range.

It should be mentioned that, like Patkowski (1990), Long points out that the real test for the existence of sensitive period(s) comes from investigations on learners who have reached their ultimate attainment stage in SLA. Birdsong (1999) observes that learners' ultimate attainment does not necessarily mean native-like attainment in the TL. In fact, in most L2 acquisition research *ultimate attainment* has been characterised in the following way: "Ultimate attainment is to be understood as synonymous with the end state or asymptotic L2A [SLA], however close or far from native-like that state may be" (Birdsong, 1999, p. 10, footnote 3). The distinction between ultimate attainment and rate of acquisition studies is of importance not only to the notion of a sensitive period, but also to account for earlier divergent findings regarding older learners' and/or adults' advantage over younger learners in the acquisition of the L2 phonology (e.g. Ekstrand, 1976, 1978; Ervin-Tripp, 1974; Olson & Samuels, 1973; Snow & Hoefnagel-Höhle, 1977; as cited in Long, 1990) vs. the expected younger learners' advantage in SLA (Asher & García, 1969; Oyama, 1976, 1978; as cited in Long, 1990; see also Patkowski's reply to Flege above). Besides, by drawing this distinction, Krashen et al. (1982) made the following three generalisations about child-adult differences in SLA (whereby generalisations 1 and 2 refer to rate, and generalisation 3 to ultimate attainment):

1. Adults proceed through early stages of syntactic and morphological development faster than children (where time and exposure are held constant).
2. Older children acquire faster than younger children (again, in early stages of syntactic and morphological development where time and exposure are held constant).
3. Acquirers who begin natural exposure to second languages during childhood generally achieve higher second language proficiency than those beginning as adults. (p. 159)

The three generalisations above also extend to the acquisition of L2 phonology, though Long (1990) notes that the rate advantage for adults or older children over younger children in L2 phonological acquisition is shorter in time than that of syntax and

morphology. Thus, studies reporting on later starters' advantage over earlier beginners (e.g. Snow & Hoefnagel-Höhle, 1977, 1978; Olson & Samuels, 1973; Schmid, 1986; as cited in Long, 1990), which, in turn, would question the claim that age 6 was critical for L2 phonology, are dismissed by Long on the grounds that what they actually showed was a short-lived rate advantage on the part of older learners. Furthermore, Long observes that none of the three generalisations makes a claim about the degree of native-likeness of learners' ultimate attainment – i.e. the fact that child beginners outperform adult starters in the long run does not necessarily imply that children will attain native-like proficiency in the TL.

Counter-evidence to a sensitive period for phonology ending at age 6 derived from a series of ultimate attainment studies conducted by Neufeld in the late 1970s (1978, 1979; as cited in Long, 1990), where English adult learners of French were judged as native French speakers after a long period of naturalistic exposure to French as an L2. However, as Long (1990) and Patkowski (1990) observed, Neufeld's studies contained a number of methodological caveats – for example, subjects recorded their oral production many times until they were satisfied with their performance – that did not falsify the notion of a sensitive period for phonology as outlined by Long (1990) above.

The issue of falsification for the CPH is discussed in more detail by Birdsong (1999). He points out that when Long (1990) first raised this idea of falsifying the hypothesis, nobody expected falsification to happen for real. However, later research has challenged the validity of the CPH. Thus, in the case of L2 phonological acquisition, Bongaerts et al. (1997) and Palmen, Bongaerts, and Schils (1997; as cited in Bongaerts, 1999) report on two groups of successful Dutch late learners (11 subjects in the Bongaerts et al. study, and 9 in the Palmen et al. study) who began their extensive exposure to the TL (English and French in Bongaerts et al., 1997, and Palmen et al., 1997, respectively) from age 12 onwards. Thirteen NSs of British English and ten NSs of French rated nearly half of the subjects' L2 pronunciation in their corresponding group as native-like. Therefore, the finding of late L2 learners' successful mastery of the TL pronunciation is in contradiction to the premises of the CPH. Based on these results, together with further findings of successful late L2 learners in other linguistic domains (e.g. syntax in Birdsong, 1992; Cranshaw, 1997; as cited in Birdsong, 1999; and White & Genesee, 1996), Birdsong (1999) claims that those successful learners cannot be considered mere outliers, for they also comprise a "significant" percentage of the sample under study, namely between 5% – 20% (cf. Scovel, 2000, pp. 216-217). He finally

proposes that, if the CPH is to be falsified, the number of learners acquiring the L2 under favourable conditions and attaining native-like proficiency should be determined (see also Marinova-Todd, Marshall, & Snow's claim (2000, 18ff.) about the study of the successful adult L2 learner population). As of yet, figures have not been agreed upon.

Finally, in support of a sensitive period for L2 phonological acquisition, Long (1990) stated that the finding in the Payne (1980), Ramsey and Wright (1974), and Scovel (1981) studies, among others, that after starting age 6 learners' chances to attain native-like phonological skills in the L2 diminished – irrespective of being linear or not as starting age increased – lent support to maturation effects on SLA.

As in the delineation of a sensitive period for phonology, Long reviewed a number of studies – namely, Coppieters (1987), Johnson and Newport (1989), and Patkowski (1980) – that led him to establish a sensitive period for the acquisition of L2 syntax and morphology at age 15. Maturational constraints were found to play a part in learners' acquisition, as well, for younger starters achieved higher levels of syntactic and morphological proficiency.

Therefore, the review of available research undertaken by Long (1990) supported not only the existence of a sensitive period for SLA, but of several sensitive periods for different language areas, as well as the existence of maturational constraints on SLA. As he noted,

> There is considerable evidence of maturational constraints and a sensitive period for first language development. Data on various feral children, Genie, deaf children of hearing adults producing "home sign", and deaf individuals learning ASL all show that language learning is typically somewhat irregular and incomplete if begun late (around age 6-8) … The very same late first language starters often exhibit an accelerated *rate* of development compared to younger learners. … The picture for SLA is similar. While some learners who begin late reach very high standards, the SL [L2] data, too, are consistent with the existence of one or more sensitive periods for SLA, with approximately the same lower and upper age bounds as those for FL [L1] development. The ability to attain native-like phonological abilities in an SL begins to decline by age 6 in many individuals and to be beyond anyone beginning later than age 12, no matter how motivated they might be or how much opportunity they might have. Native-like morphology and syntax only seem to be possible for those beginning before age 15. As with first language development, adults and older children learn early SL morphology and syntax faster than younger children. The rate advantage is generally only temporary, however, and, "faster" again does not mean "better", as shown by the fact that only young starters can attain native-like proficiency. The first and second language data are consistent, in other words, with the so-called "maturational state hypothesis". (pp. 279-280)

At this point it is worth commenting on the finding of a linear decline in subjects' performance after the end of a critical or sensitive period (e.g. Oyama, 1978/1982). As just stated above, Long (1990) appears to consider any type of diminishment – gradual or abrupt – in performance as a function of onset age of L2 learning as solid evidence for the existence of maturational constraints on L2 acquisition. However, recently it has been argued that the finding of linearity after the offset of a CP runs counter to the neurological basis of the CPH. In particular, Bialystok and Miller (1998; as cited in Bialystok & Hakuta, 1999) and Birdsong and Molis (2001) (see also Singleton, 2003, p. 9) in their partial and complete replication studies of Johnson and Newport (1989), respectively, have observed post-maturational age-related effects. On the basis of this result, Bialystok and Hakuta (1999) have concluded that

> If there is a critical period, then the relation between age of learning and proficiency will be nonlinear because of a sharp break at the critical period; if there is no critical period, the relation will be linear. Regarding other factors, if there is a critical period, then age will be the exclusive or primary factor accounting for proficiency; if there is no critical period, then other factors will be significant. (p. 173)

By the same token, Birdsong and Molis (2001) have come to the conclusion that their findings of maturational effects on the morpho-syntactic judgement task in post-pubescent Spanish learners of English – i.e. subjects with an AOA after age 17 obtained progressively lower scores on the test as their AOA increased – and results of some late arrivals' performance within or close to the NS range, are not in accordance with the account of maturational constraints in SLA, and would also falsify the CPH.

> Our strict replication of J&N89 [Johnson & Newport, 1989] yielded evidence of postmaturational age effects and native language effects. Modest evidence of nativelike performance was also observed.
>
> Researchers have argued these results would be evidence for falsification of the Critical Period Hypothesis as it relates to L2 acquisition. The postmaturational age-related effects we found are fairly robust and are grounds for refutation according to Pulvermüller and Schumann (1994), who maintain that 'if the decrease in grammatical proficiency with greater age in postpuberty starters could be confirmed, the present [neurobiologically-based] proposal would have to be modified' (p.273). (Birdsong & Molis, 2001, p. 248)

Both Bialystok and Hakuta (1999) and Birdsong and Molis (2001) further indicate that the different terms put forth in SLA literature all advance that language learning is feasible and similar among individuals during a critical period, and that once the CP has come to an end, language learning differs in form and success. However, it should be

noted that usage of "critical period" does not entail researchers opting for the phenomenon as originally conceived by Lenneberg (1967). In fact, "critical period" is often used in a broader sense, embracing the notion of sensitive period, as well. Or, as Birdsong and Molis (2001) put it, "[c]ustomarily, these accounts are lumped under a single category name for the sake of convenience." (p. 236, footnote 1)

Thus, although Long's (1990) and Oyama's (1976/1982, 1978/1982) characterisation of a sensitive period for language acquisition may reflect differences in child and adult L2 learning more accurately, the term "critical period" – and the CPH – has prevailed in the literature over that of "sensitive period", regardless of scholars' acceptance or rejection of the existence of such notion. As a matter of fact, refinements and/or reformulations of the CPH never cease to appear. Another two instances are worth commenting on, namely Kellerman (1995) and DeKeyser (2000), who have put forward two amendments to the CPH as understood by Johnson and Newport (1989) – i.e. the maturational state hypothesis.

In the first place, Kellerman's (1995) refinement assigns the subjects' L1 a relevant role in SLA.

> There is an interaction between L1 and L2 features and age of acquisition, such that learners attempting to acquire certain (but not all) features in the L2 which no L1 equivalents must have acquired those features by the age of x[12] ["Fill in your own favored upper bound for the CPH" (p. 230, footnote 12)] or they will never acquire them. Features of the L2 with clear L1 analogies, on the other hand, can in principle be mastered whatever the age of onset of learning. (p. 222)

DeKeyser (2000), for his part, advocates the role of implicit learning mechanisms in SLA in order to account for the age effects found in Johnson and Newport (1989) and his own replication study of Johnson and Newport (1989) with Hungarian NS immigrants to the US. In addition, the DeKeyser (2000) study examined the Fundamental Difference Hypothesis (Bley-Vroman, 1988; as cited in DeKeyser, 2000), which states that

> Whereas children are known to learn language almost completely through (implicit) domain-specific mechanisms, adults have largely lost the ability to learn a language without reflecting on its structure and have to use alternative mechanism, drawing especially on their problem-solving capacities, to learn a second language. The hypothesis implies that only adults with a high level of verbal analytical ability will reach near-native competence in their second language, but that this ability will not be a significant predictor of success for childhood second language acquisition. … According to this hypothesis, adults can no longer rely on the innate mechanisms for implicit language acquisition and must, therefore, rely on alternative, problem-solving mechanisms. (DeKeyser, 2000, pp. 499-500)

The results of his replication study led DeKeyser to conclude that native-like acquisition of an L2 is constrained by a CP in its more strict sense, provided that the CPH is limited to implicit learning mechanisms.

> Most importantly, however, this study suggests that there really is a critical, and not just a sensitive or optimal, period for language acquisition, provided that the Critical Period Hypothesis is understood narrowly enough, that is, applying only to implicit learning of abstract structures. As long as L2 competence is assessed without regard for the learning mechanisms that produced it, it may appear that there is merely an optimal age for language learning, in the sense that there is a sizable negative correlation between age of acquisition and ultimate attainment. If the Critical Period Hypothesis is constrained, however, to implicit learning mechanisms, then it appears that there is more than just a sizable correlation: Early age confers an absolute, not a statistical advantage – that is, there may very well be no exceptions to the age effect. Somewhere between the ages of 6–7 and 16–17, everybody loses the mental equipment required for the implicit induction of the abstract patterns underlying a human language, and the critical period really deserves its name. (DeKeyser, 2000, p. 518)

However, instead of discussing about the existence of the CPH for language acquisition, many scholars (e.g. Oyama, 1978/1982; Patkowski, 1994) have argued that efforts should be directed to explaining what the basis for a CP – or sensitive period – is; or, in other words, what causes child-adult differences in SLA.

> It seems, to this writer at least, that the time has come for the controversy which has surrounded this issue [the claim that a sensitive period exists for the acquisition of phonology in a second language] to move to the area of providing an explanation for the observed phenomenon, rather than to continue questioning the existence of the phenomenon itself. (Patkowski, 1994, p. 216)

Earlier attempts at explaining child-adult differences in SLA were originally classified into neurological, cognitive, input, and affective accounts (Krashen, 1973/1982; Long, 1990). More recently, Singleton and Ryan (2004) have further elaborated on the taxonomy of factors that might account for the differences observed between child and adult L2 acquisition. Thus, in addition to the above mentioned neurological account (e.g. Penfield & Roberts, 1959; Pulvermüller & Schumann, 1994; Scovel, 1988; Walsh & Diller, 1981; cf. Dechert, 1995)[7], affective explanation (e.g. Gardner & Lambert, 1972; Guiora, 1992; Krashen, 1985), input explanation (e.g. Hatch, 1978, 1983; Scarcella & Higa, 1982), and cognitive account (e.g. Ausubel, 1964;

[7] In this paragraph all supporting studies for the different accounts of age differences are as cited in Singleton and Ryan (2004).

Cummins, 1979; Piaget & Inhelder, 1969; Rosansky, 1975; cf. Schumann, 1978), age differences might be explained in terms of a diminishment of sensory acuity (e.g. Hatch, 1983; Joiner, 1981; cf. Gleiss et al., 1979), competition between languages (e.g. Bever, 1981; Flege, 1981; Jia & Aaronson, 1999), nativism or innatism (e.g. Chomsky, 1972, 1978; cf. DeKeyser, 2003), and the de-coupling hypothesis (e.g. Bever, 1981).

It should be noted that no account per se is exempt from opposing viewpoints, or, alternatively, a given explanation about age differences might be lacking in concluding evidence. Take, for instance, the view that considers input – specifically, quality of input – as the key factor leading to age differences in L2 acquisition. In particular, comprehensible input (Krashen, 1978, 1980a, 1980b; cited in Scarcella & Higa, 1982) appears to account for the finding of children's superior ultimate attainment in the TL to that of older language learners (e.g. Hatch, 1977; as cited in Long, 1990; and in Singleton & Ryan, 2004).

According to Krashen (1980b; as cited in Scarcella & Higa, 1982), comprehensible input is the key type of input that fosters language acquisition. It is characterised as being simplified input, given in a sufficient amount at a level that is only a bit beyond the learner's IL stage, and delivered in a supportive environment.

Based on this type of input which has normally been addressed to child L2 learners, younger starters have been considered "better" language acquirers in the long run. However, Scarcella and Higa (1982; also summarised in Krashen, 1982; Long, 1990; and Singleton, 1989) found that older learners – adolescents in their study – managed to obtain more comprehensible input than child L2 learners, resulting in older learners' faster acquisition rates in the first stages of L2 learning. To be precise, Scarcella and Higa examined 10-minute recorded interactions of 14 younger and older Spanish learners of English (seven learners aged 8.5–9.5 and another seven aged 15.5–16.5, respectively) with seven English NSs. In spite of the fact that English NSs provided children with more comprehensible input, adolescents did obtain a larger amount of comprehensible input than children (or better input for L2 learning) by means of negotiation work. And, consequently, they became faster learners in the initial stages of SLA.

Thus, as for type of input, it seems that comprehensible input offers a two-fold explanation about age-related differences in SLA. On the one hand, the provision of comprehensible input to children results in their higher proficiency in the TL. On the other hand, comprehensible input leads to older starters' faster rate of L2 acquisition as a

function of their superior "negotiating" skills, which allow them to obtain more and better input (Krashen, 1982).

Nevertheless, the evidence presented in favour of the importance of input – to be exact, type of input – in explaining child-adult differences has been considered scarce (e.g. Scarcella & Higa, 1982; Singleton, 1989; Singleton & Ryan, 2004) and inconclusive. For example, Long (1990) points out that the account of age differences based on the quality of input is problematic in that "there is a relative lack of effect for quite major input differences in normal and abnormal FL [L1] acquisition." (p. 276)

When contemplating input as the key factor to explaining age differences in SLA, one might hypothesise that it is not only type but also amount of input that should be taken into account. On top of that, a distinction needs to be drawn between amount of input delivered in L2 immersion settings, which has often been equated with length of residence (LOR) or number of years spent in the TL country (e.g. Krashen et al., 1979/1982; Long, 1990; Piske, MacKay, & Flege, 2001), and amount of input provided in FL learning settings, which usually takes the form of formal instruction.

With reference to amount of input or L2 experience delivered in L2 immersion environments, research reporting on an LOR effect (for the most part) on the degree of FA includes the already mentioned Asher and García (1969/1982) study, in addition to Ekstrand (1976/1982) and Seliger et al. (1975/1982), and a number of studies carried out by Flege et al. – among them, Flege, Bohn, and Jang (1997); Flege and Fletcher (1992); Flege and Liu (2001); and Flege, Munro, and MacKay (1995a).

In those investigations LOR ranged from an averaged minimum of 5 months to an averaged maximum of 20.5 years, and the US was the TL country in the majority of the studies. For example, Asher and García (1969/1982) looked at 71 Cuban immigrants to the US differing in AOA (1–6, 7–12, and 13–19 years old) and LOR (1–4 and 5–8 years). Subjects who had lived in the TL country for more than five years had more chances to obtain a rating of near-native pronunciation in the L2. This (significant) effect of LOR was correlated with AOA, as well. Therefore, early arrivals (before age 6) with a longer LOR (5+ years) had the highest probability to be rated as near-native.

Ekstrand (1976/1982) also found that LOR (with a median of 15.5 months) was strongly related to the scores obtained on a free oral production task, and, to a lesser extent, to listening comprehension and free written production tasks that 2,189 immigrant students performed in Swedish. In spite of this, AOA (which ranged from 8 and 16.4 years of age) was a better predictor of Swedish SLA than LOR.

Likewise, in some of the investigations conducted by Flege and his colleagues, the factor of LOR has yielded significant results on the production and/or perception of TL speech. For instance, Flege and Fletcher (1992) examined the degree of FA on the production of English sentences by three groups of Spanish NSs. The groups differed in age of learning (AOL) or age of arrival (AOA): one group consisting of early L2 learners with an AOL before 5–6, and two groups of late L2 learners with AOA from 20 years old onwards. The two late L2 learner groups were further divided into experienced and inexperienced learners according to the average number of years spent in the US: 14.3 years and 7 months for the experienced and inexperienced learners, respectively. The results showed that as LOR in the US increased, the pronunciation in English of the late L2 learner group improved. In other words, experienced late L2 learners obtained higher accent scores (more native-like) than inexperienced late L2 learners, though the former did not attain a native-like pronunciation in English. In fact, the main factor in determining degree of FA in the TL was AOL, and not LOR, as the early L2 learner group received native-like accent ratings. Another finding worth mentioning that involved LOR was the significant correlation found between degree of accent, on the one hand, and AOA, LOR, and years of formal instruction in English, on the other hand. That is to say, the earlier subjects had arrived and the longer they had lived in the US, together with the more formal education in English they had received, the more native-like their pronunciation in English was rated.

In the Flege, Munro, and MacKay (1995a) study, length of residence in Canada (between 15 and 44 years) significantly accounted for a small amount of variance in the accent ratings obtained by 240 Italian NSs distributed into ten age groups (AOL ranged between 2 and 23 years of age). However, as in Flege and Fletcher (1992), the main factor that accounted for variance in accent ratings was AOL (about 78%), instead of LOR.

Flege, Bohn, and Jang (1997) also found significant LOR effects on the perception and production of English vowels by four different language groups (German, Spanish, Mandarin, and Korean NSs) of 20 learners each, all of whom had an AOA in the US from 14 years old onwards. Although all the subjects in the group had received the same amount of formal instruction (between 6.7 and 7.2 years on average), groups differed in LOR: 7.3 years for ten learners (experienced learners) and 7 months for the remaining ten learners (inexperienced learners) in each of the four language groups. The difference in LOR was significant in that adult learners perceived and produced L2

vowels more accurately as their experience in the L2 (or LOR) increased. Furthermore, Flege, Bohn, et al. (1997) observed that even the most experienced groups in the study failed to perceive and produce English vowel segments at native-like rates. The latter finding led the authors to hypothesise that the amount of L2 input or experience in the L2 in the TL country should be higher than 7 years if native-like L2 phonological skills were to be attained.

Finally, Flege and Liu (2001) examined the performance on a final consonant production task, grammaticality judgement test, and listening comprehension task of 60 Chinese NSs matched for "late" age of arrival in the US (between 16 and 40 years old), but differing in LOR[8]. Short LOR was characterised as the period comprised between 5 months and 3.8 years, and long LOR consisted of the period between 3.9 years and 15.5 years. The average number of years of instruction in the subjects' home country was nine (range between 1 and 18 years). In addition, subjects were classified according to their further instruction in the US: those who had pursued a university degree (students) and those who did not (non-students). LOR effects were found in the student population, but they were absent in the non-student population. More precisely, students with a longer LOR obtained significantly higher scores (more native-like) than students with an LOR of 3.8 years and below. Flege and Liu further hypothesised that the LOR effects observed in the student population of their investigation might have been due to the kind of English-language input or formal education that the subjects had received in the US. As in Flege, Bohn, et al.'s (1997) research reported above, a longer LOR advantage did not result in the subjects' native-like performance in the TL[9].

Moreover, Seliger et al. (1975/1982) found an LOR effect, but, contrary to Asher and García (1969/1982) and Ekstrand (1976/1982), subjects with a shorter LOR were considered to have a better pronunciation in the TL than that of subjects who had lived for a longer period of time in the host country. That is, the 10–15-year-old starters in the Seliger et al. (1975) study with a shorter LOR – namely, 15.03 years on average – self-rated their production in the TL (English or Hebrew) as native-like; whereas the same age group with a longer LOR – i.e. 20.5 years on average – judged their production in the TL as far from being native-like.

On the other side, a number of studies conducted in L2 naturalistic settings have found no effect of LOR on the differences encountered in SLA between child and adult

[8] A minimum LOR of 5 months was required to take part in the experiment.
[9] In fact, Flege and Liu (2001) did not specify subjects' starting age of L2 learning in their home country.

learners. These studies include, among others, Oyama (1976/1982, 1978/1982), Piper and Cansin (1988), and Tahta et al. (1981a) concerning L2 phonological acquisition; and Fathman (1975/1982) and Patkowski (1979/1982) as far as L2 syntactic acquisition is concerned.

In her 1976 and 1978 studies Oyama did not find any effect of LOR (5–11 years vs. 12–18 years) on the comprehension scores and mean accent scores obtained in paragraph reading and story telling tasks obtained by any of the three age groups she examined (6–10, 11–15, and 16–20 year olds).

Piper and Cansin (1988) looked at the production of 29 advanced ESL students (comprising twelve different L1s) who differed in AOA in Canada (0–5, 6–11, 12–17, 18–22, 23–27, 28+ years) and in LOR (from 2 months to 37 years). The main finding had to do with a significant AOA effect on the accent scores on a read-aloud short story and an oral retelling of a personal experience. So, the earlier subjects arrived in the TL country, the better their accent in English was rated (in special, differences were larger between the group with an AOA below 6 years old and the group with an AOA between 12 and 17 years old). Surprising though it may seem, the wide range difference in LOR was not significant in the slightest for the accent ratings.

Tahta et al. (1981a) examined the accent ratings on a read-aloud paragraph of English prose obtained by 109 subjects (NSs of nine different L1s in total) who differed in AOL (from 6 years old to 15+ years old) and with a minimum LOR in the UK of two years. LOR had no significant effect on the accent ratings subjects received. Actually, the authors pointed out that those subjects with an LOR shorter than two years had a noticeable degree of FA in English, no matter how early they had started to learn the L2. Consequently, they were discarded for the purpose of analysis. As reported above, AOL was the main predictor of FA, which accounted for nearly half the variance (43%) in accent scores and which yielded the following findings, too. First, subjects with an AOL before 6 years old were rated within the NS range. Those with an AOL between 7 and 11 years old might produce the L2 with a certain degree of FA, while subjects with an AOL from 12 onwards had a marked accent in English.

At the syntactic level, Fathman (1975/1982) looked at the performance of 200 L2 learners aged between 6 and 15 on an oral morphological and syntactic test. They were grouped on the basis of AOA in the US – 6–10 and 11–15 years old – and LOR – 1, 2, 3 years. Some subjects also differed in the language programmes offered in the schools they attended: one-hour day ESL classes vs. no ESL instruction. Although Fathman did

not include an actual result about the lack of effect of LOR by itself on the subjects' scores, she reported on a significant interaction between age, language programmes, and LOR. The only significant finding was concerned with AOA, which in that case favoured older starters.

Last, Patkowksi (1979/1982; see also above) distributed the 67 subjects of the study into two main onset age groups (before and after age 15) and gave specific details about each subject group's LOR, hours of formal instruction and informal instruction. The 33 pre-puberty learners in the study had an LOR of 20.4 years, 345 hours of formal instruction, and 84,452 hours of informal instruction. By contrast, the 34 post-puberty subjects had an LOR of 18.4 years, a total of 1,201 hours of formal instruction, and 58,479 hours of informal instruction. Neither LOR nor number of hours of formal and informal instruction had a significant effect on the younger learners' higher scores obtained for the morphological and syntactic task.

Many of the studies above have shown that amount of input may also be measured as number of hours of formal instruction in the TL. As Flege and Liu (2001) point out, "differences in formal education must also be considered when investigating the role of LOR on L2 acquisition" (p. 532). Their study further suggests an effect (though not clearly stated) of formal instruction on experienced students' better performance.

The effects of quantity of L2 input in the form of formal instruction on child vs. adult L2 acquisition are more evident in FL learning contexts. Take, for example, (in order of publication) the studies by Olson and Samuels (1973/1982), Elliott (1995a), Bongaerts, Planken, and Schils (1995), Bongaerts et al. (1997), and Moyer (1999) (cf. Oller & Nagato, 1974, among others).

Olson and Samuels (1973/1982) provided phonetic training in German over a period of three weeks to three different age groups of 20 English NSs each (elementary, high school, and college) with no previous formal FL instruction. Results showed that older subjects – junior high and college students – obtained significantly higher accent ratings than elementary students. All factors being equal, such as amount of time spent on training and quality of the model, the authors concluded that older learners would benefit more significantly from formal instruction in the learning of an L2.

By the same token, Elliott (1995a) tested the effects of a pronunciation instruction treatment on 43 adult English NSs learning Spanish in the US (experimental group). The method consisted of devoting 10–15 minutes of 21 class periods over a semester to

formal instruction and practice of pronunciation. The experimental group significantly differed in their pronunciation scores between the pre-test and post-test after treatment administration. In other words, their pronunciation of Spanish segments improved as a result of the phonetic treatment. On the other hand, the 23 English NSs comprising the control group did not obtain significantly different accent scores between the pre-test and post-test (if so, their pronunciation of TL sounds was rated a little less native-like on the post-test, but the difference in scores was nonsignificant).

The two investigations mentioned above would fall within the categorisation of short-term studies. Then, the finding of beneficial amount of formal instruction for older learners' pronunciation of L2 segments (especially in Olson & Samuels, 1973/1982) might be considered as a short-lived advantage in the first stages of L2 acquisition (Long, 1990). Moreover, neither quantity of formal instruction resulted in learners' native-like pronunciation of the L2. In recent years, however, several studies on successful ultimate attainment of L2 phonology by late learners have appeared, in which formal instruction accounts, to a certain degree, for the subjects' mastery of the TL phonology.

One such study is Bongaerts et al. (1997), which examined the accent scores of two groups of late Dutch learners of English – one labelled as highly proficient learners (11 subjects) and the other as mixed-proficiency level learners (20 subjects) – obtained on a sentence production task as rated by a total of 13 English NS judges (6 experienced and 7 inexperienced) on a 5-point scale of FA. Both learner groups started to be significantly exposed to British English from age 12 onwards: in high school they received a 2-hour weekly instruction mostly from Dutch NSs. But it was not until they entered university at about age 18 that they began to have a fair amount of English NS input, as well as completing a one-year stay at a British university. The authors also stated that during the subjects' first year at college, they were provided with intensive training in British Received Pronunciation, though the total quantity is not specified, as Piske et al. (2001) state. The results showed that the highly successful learner group obtained accent scores similar to those of the English NS control group (M = 4.61 and 4.84, respectively), while the mixed-ability group obtained a much lower mean accent score (2.59). Moreover, a closer look at each of the subjects' performance in the successful learner group revealed that five subjects were taken as native speakers of English, since they obtained accent ratings like those of English NS foils. Therefore, the results in Bongaerts et al. (1997) would disconfirm the hypothesis that it is not possible to acquire a native-like accent in the TL if L2 learning starts at a late age. The authors

further suggest that the amount of formal instruction – in this case, "intensive training both in the perception and in the production of the speech sounds of British English" (Bongaerts et al., 1997, p. 463) – might have led to the five late learners' successful attainment of native-like pronunciation in English.

Another study conducted in an FL learning context in which formal instruction (amount of input) has been thought to influence late learners' better pronunciation in the TL is Moyer (1999). The author investigated the performance of 24 English late learners of German in the US on a series of phonological tasks. All the subjects in the experiment were highly motivated learners, who had been first exposed to the TL in a formal context at age 11 and for an average of 9.5 years (range = 3–15 years), and who had spent a mean LOR in Germany of 2.7 years (range = from no time at all to 11 years). The learners' mean accent ratings given by four German NS judges were found to be more foreign-accented than those of the four German controls (M = 4.98 and 2.63, respectively, on a 6-point scale), except for one NNS subject. Moreover, age at which first exposure took place and age of first instruction were significantly correlated, as well as age of first exposure and mean accent rating, and phonological performance on tasks and type of phonological feedback the subjects had received (i.e. segmental only vs. segmental and suprasegmental). As regards the latter correlation, a beneficial effect of input was observed in that subjects who had received both segmental and suprasegmental feedback (instead of segmental feedback alone) obtained a mean accent rating that was closer to that of NS controls. Thus, although most late learners did not perform at a native-like level on any of the tasks in the L2 phonology, the "phonological" input subjects had received played a role in achieving native-like pronunciation in some instances. As Moyer (1999) indicates,

> Based on the results of this analysis, phonological instruction (and SLA research) should expand to incorporate suprasegmental training and its effects. As has been demonstrated here, intonational and stress errors frequently mark the speaker as nonnative, perhaps more often than segmental errors, due to their significance for discourse fluency. If any immediate implication is appropriate, it is the need for further exploration and experimentation regarding both the process of second language phonological acquisition, as well as the ultimate effects of overt training. (p. 100)

Despite the finding of a positive effect of formal – phonological – instruction on late L2 learners' attainment of native-like pronunciation, the input received cannot be considered the only contributor to successful L2 phonological acquisition. Therefore,

other factors such as late learners' high motivation and greater neurocognitive flexibility due to an exceptional brain organisation for language (Bongaerts et al., 1997; Marinova-Todd et al., 2000) are likely to play a role, as well. Moreover, subjects in the Bongaerts et al. (1997) and Moyer (1999) studies had actually been immersed in the TL country for an average range of 1 to 2.7 years. Although Moyer (1999) did not find a significant correlation between accent ratings and years spent in the TL country, one might doubt whether the period of time spent in an L2 immersion context has indeed no consequences for late L2 acquirers, based on findings of LOR effects in the studies mentioned above (e.g. Asher & García, 1969/1982; Ekstrand, 1976/1982; Flege, Bohn, et al., 1997), and also in the light of Singleton's (1995) estimates of quantity of input in an L2 naturalistic setting (based on Snow & Hoefnagel-Höhle, 1978) and in an FL formal learning context.

> … a period of, say, five years of exposure to second language in a second language environment would in most circumstances involve a very great deal more exposure of the language than five years of formal second-language instruction, where the target language was being treated simply as one school subject among many. Thus, if the amount-of-exposure variable is held constant, the concepts of 'initial advantage' and 'eventual attainment' in a formal instructional setting need to be associated with much longer real-time periods than in a largely informal exposure situation. To be more precise, the present author has estimated (Singleton, 1989: 236) that more than 18 years would need to be spent in a formal instructional setting in order to obtain the same amount of second language input as seems (according to the Snow and Hoefnagel-Höhle studies) to be required for older learners' 'initial advantage' to begin to disappear. No one would want to postulate a literal equation between a given quantity of input over 12 months and the same amount of exposure over 18 years, but the point is that when comparing different categories of language learners one does have always to keep in mind the varying relationship between real time and exposure time. (p. 3)

To sum up, divergent results have been found as to the effects of amount of input on L2 naturalistic learning, as Piske et al. (2001) indicate. What is more, when different age groups' performance is compared, AOL/AOA has been considered the main factor responsible for age differences, independently of LOR (Long, 1990). LOR effects are often significant when comparing same age groups differing in LOR in the host country. But even in those instances where length of residence extends over a long period of time, more experienced subjects fail to attain native-like skills in the TL. In formal FL environments, amount of input does not always account for child-adult differences in language acquisition, either. In this context, it seems that quality of input (e.g. specific

phonetic training) is of more importance especially to native-like attainment of L2 phonological skills (Piske et al., 2001).

Taken all together, the studies on the amount and type of input have not provided enough convincing evidence upon which child-adult differences might be solely based.

To conclude, none of the accounts of age differences reviewed in Singleton and Ryan (2004) has offered convincing or conclusive evidence that can fully characterise the differences observed between child and adult L2 acquisition. Furthermore, all the accounts seem to be interrelated, rather than single, independent explanations (Krashen, 1973/1982). As Singleton (2001) concludes,

> Bialystok (personal communication) takes the view that talking about an age factor is misconceived; we should rather be thinking in terms of a range of age factors. This coincides with my own conclusion of more than a decade ago which I have not discarded: that 'various age-related phenomena … probably result from the interaction of a multiplicity of causes' (Singleton, 1989, p.266). Such a perspective can encompass the notion that decreasing cerebral plasticity and/or other changes in the brain may play a role, but the notion that L2 age effects are exclusively neurologically based, that they are associated with absolute, well-defined chronological limits, and that they are particular to language looks less and less plausible. (p. 85)

2.3. Studies in L2 phonological acquisition

This section is concerned with SLA research conducted in the area of L2 phonology, and, more specifically, with perception and production of TL sounds (or segments). As summarised by Flege (2003), two general hypotheses have shaped most L2 phonological acquisition research: the CPH and an alternative hypothesis (or hypotheses) to the CPH.

In the previous section it was mentioned that the CPH traditionally attributed older child L2 learners' and adult L2 learners' failure to attain native-like proficiency in the TL (or lower command of the TL in comparison to younger child L2 learners) to a loss of neural plasticity after the passing of a CP. Moreover, advocates of the CPH considered L2 phonology the linguistic domain to be first influenced by the end of the critical (or sensitive) period, resulting in nonnative-like pronunciation by L2 learners starting to acquire the TL as early as age 6; and even some scholars (e.g. Scovel, 1969, 1988) regarded phonology the only language area subject to a CP.

By contrast, the alternative hypothesis (or hypotheses – see Theoretical Frameworks below) dominating L2 phonological acquisition research from the 1980s onwards emphasises the importance of the L1 phonological system in the learning of the L2 phonology. Thus, differences between child and adult acquisition of the TL phonology are thought to arise from the L2 learner's L1 and their stage of L1 phonological acquisition when L2 learning commences, rather than neurophysiological maturation (see also Flege, 1987a, above). The alternative hypothesis has been further developed into several models of L2 speech learning, namely, the Speech Learning Model (SLM) (Flege, 1991b, 1992, 1995a, 1999a, 2003), the Perceptual Assimilation Model (PAM) (Best, 1995; Best & Strange, 1992), and the Native Language Magnet (NLM) model (Kuhl, 1993; Kuhl & Iverson, 1995).

Another account that argues for the influence of L1 on L2 phonological acquisition is Major's Ontogeny Philogeny Model (OPM) (2001, 2002). As the author states, the OPM is a revision of his former Ontogeny Model (OM) (e.g. Major, 1987b, 1992), and covers the individual's language in addition to the development of whole languages. The basic claim of the OPM is that IL consists of parts of both L1 and L2, and parts of universals (U) which are not included in L1 and L2 (U refers to language universals comprising Universal Grammar (UG), markedness, stylistic variation, and developmental processes). The model then aims at determining the "weighting" of each component of IL throughout the process of L2 acquisition. Therefore, the OPM hypothesises that in the first stages of L2 acquisition the component of L1 is mostly present, whereas it is not until later acquisitional stages that the component of U is evident. Besides, the factors of similarity and markedness are key to the OPM, in as much as they relate to rate of acquisition, rather than ultimate attainment or degree of difficulty. Thus, according to an earlier hypothesis, i.e. the Similarity Differential Rate Hypothesis (SDRH) (Major, 1996; as cited in Major, 2001), Major (2001) predicts that "similar phenomena are acquired more slowly than dissimilar phenomena" (p. 105) due to the prevailing nature of L1 transfer. Likewise, markedness is considered to slow rate, hence "an unmarked similar phenomenon is acquired at a faster rate than a marked similar phenomenon" (p. 105).

In spite of Major's objection to establishing the OPM within a given linguistic framework, he suggests several theories from which the model's premises could be tested or interpreted, all of which share some sort of abstract linguistic nature (e.g. generative phonology, optimality theory, and parameter setting). This is in contrast to the SLM, the

PAM, and the NLM, which, as will be seen later, are segmentally-based. Since the present dissertation is concerned with perception and production of TL segments (i.e. more concrete units of speech), the OPM will not be reviewed in detail. Instead, where appropriate, reference will be made to Major's premises concerning similarity and foreign accent in subsequent sections – for instance, Sections 2.3.1.1/2.3.1.2, and 2.3.3, respectively – which might provide a new insight into any of the three models' inconclusive or unexpected findings[10].

Therefore, the basic premises of each of the three models – SLM, PAM, and NLM – or theoretical frameworks, together with a comparison of the relevant differences and similarities between the models, are presented in the following section (2.3.1).

2.3.1. Theoretical Frameworks

2.3.1.1. The Speech Learning Model (SLM)

The SLM (Flege, 1991b, 1992, 1995a) has as its starting point the basic assumption that phonetic learning ability or "the capacities underlying successful L1 speech acquisition remain intact across the life span" (Flege, 2003, p. 8), which is in contrast to the premises of the CPH. As Flege indicates, the model aims to account for age differences encountered in L2 speech learning, and, more precisely, in the production and perception of L2 vowel and consonant sounds. As he puts it,

> The aim of our research is to understand how speech learning changes over the life span and to explain why 'earlier is better' as far as learning to pronounce a second language (L2) is concerned. An assumption we make is that the phonetic systems used in the production and perception of vowels and consonants *remain adaptive over the life span*, and that phonetic systems reorganize in response to sounds encountered in an L2 through the addition of new phonetic categories, or through the modification of old ones. (Flege, 1995a, p. 233)

[10] Similarly, the developmental model of speech perception proposed by Wode (1992, 1993, 1994, 1995) within the universal theory of language acquisition (UTA) might help explain some of the unresolved issues in the SLM and the PAM (see below). Briefly, Wode's model of speech perception also places an emphasis on the learners' state of development of their L1 perceptual system on starting to learn an L2. He further claims that the categorical and continuous modes comprising the auditory system are necessary for speech perception, and thus for the development of L1 phonetic categories. Wode (1995) goes on to suggest that that these two modes "remain unchanged throughout life, but the original categorical sensitivities [innate perceptual discontinuities] become increasingly difficult to access as a function of L1 acquisition" (p. 323). Finally, when it comes to acquiring the L2, Wode refers to equivalence and assimilation theory.

The last updated version of the SLM goes back to 1995, though as he points out "this is a working model and is subject to further revision as new data are gathered." (Flege, 1995a, p. 238). Earlier versions of this theoretical framework can be found in a number of articles and studies by Flege (1981, 1987a, 1987b, 1991b, and 1992), many of whose points have been incorporated into the several hypotheses and postulates making up the current version of the SLM[11] (see below).

Therefore, one of the first hypotheses Flege put forth is concerned with the observation that older L2 learners pronounce the TL at more foreign-accented rates than younger L2 learners. According to Flege (1981), the perception of an FA might arise from a combination of segmental, suprasegmental, and subsegmental deviations from the TL phonetic norms that NNSs produce in their pronunciation of the L2. These deviations or mispronunciations include, among others, sound substitution and sound distortion at the segmental level, average speaking fundamental frequency at the suprasegmental level, and speech timing differences between languages at the subsegmental level. Although NSs of the TL do not seem to consider the different deviations as equally important in their assessment of L2 speech as foreign-accented (e.g. Flege & Hammond, 1981; cited in Flege, 1981), Flege notes that it is not clear which specific features contribute the most to NSs' perception of FA (but see more recent studies discussed in 2.3.3).

The author further states that both children and adults have the potential to learn to pronounce an L2 at a native-like level, though their pronunciation of the L2 will probably be accented if they base the production of L2 sounds on corresponding (or matching) L1 sounds. In addition, Flege suggests that both the L1 and TL provide acoustic models or input that will, in turn, shape the learners' acquisition of the L2 phonology. That is, learners "modify native-language patterns of phonetic implementation, and … superordinate acoustic models based on pairs of corresponding sounds or phonetic dimensions in two languages serve as input for phonetic learning in second language acquisition." (Flege, 1981, p. 452)

---

[11] It should be mentioned that in the articles comprising the first formulations of the SLM (e.g. Flege, 1992) the model is never referred to as Speech Learning Model (SLM) (Flege, 1995a). Rather, references to the model are in the form of "hypotheses" and "proposals" (e.g. "The general hypothesis that has motivated our program of speech research over the past decade" [p. 591]). However, there exists an earlier allusion to the "Speech Learning Model" as such in Flege (1991b, p. 283): "One theoretical model and methodological problem facing the speech learning model sketched here…".

For a similar observation about the use of the term "SLM" as such, see also Major (2001, p. 38, footnote 5).

All of the above developed into the "phonological translation hypothesis", which emphasises the role of the L1 (in preference to neurophysiological maturation) and a "tendency by mature speakers to interpret sounds occurring in a foreign language[12] in terms of sounds found in their native language" (Flege, 1981, p. 448) in order to explain the pronunciation differences found between child and adult L2 learners.

Finally, Flege (1981) makes two predictions whereby the phonological translation hypothesis might be rejected. One prediction states that bilingualism at the phonetic level will not be possible due to learners' "accommodation" of their L2 sound pronunciation between the acoustic models of the L1 and L2. That is, the L2 sounds as produced by learners will present intermediate phonetic values between those typical of their L1 and those of the L2. This prediction is in agreement with previous studies such as Caramazza, Yeni-Komshian, Zurif, and Carbone (1973; as cited in Flege, 1981) and Caramazza and Yeni-Komshian (1974; as cited in Flege, 1981). In those investigations voice onset time (VOT) values for voiceless stops in English /p, t, k/ as produced by French Canadian speakers were measured and then contrasted with the values English and French monolinguals obtained for English and French /p, t, k/, respectively. The results showed that French Canadian produced English /p, t, k/ with intermediate VOT values between those characteristic of English and French.

The second prediction suggests that the learning of an L2 will have an effect on the pronunciation of the L1. The Williams study (1980; as cited in Flege, 1981) constitutes evidence in favour of this prediction, where an L2 (English) effect was observed in Spanish learners of English on the production of Spanish stops. In that case, learners produced Spanish stops with more aspiration (that is, similar to English stops) than those produced by Spanish monolinguals.

Despite the fact that the available research reviewed is in agreement with the phonological translation hypothesis, Flege (1981) comments that further studies are still needed to accept the above-mentioned hypothesis.

---

[12] Note that in earlier articles (e.g. Flege, 1981, 1987a), Flege uses the term "FL" in the generic way the term "L2" has more often been used (see Section 2.1 above), although the TL under study appears to have the status of L2 as defined by Johnson and Johnson (1999) above (i.e. ESL, in the case of English as the TL). In later articles (e.g. 1991b, 1992), the author consistently talks about L2 to refer to this general sense comprising both L2 and FL, instead of FL. So, in this section and the following ones every effort has been made to use the terms "L2" and "FL" as outlined in Section 2.1. However, when including original quotations, each author's preference for "L2" or "FL" (at the general level) has been kept.

Similarly, the studies he discusses in a later article questioning the existence of a CP for the pronunciation of L2s (Flege, 1987a; see also 2.2 above) are insufficient to accept the two alternative hypotheses to the CPH he outlines for L2 phonological acquisition. Like the phonological translation hypothesis, these two hypotheses stress the importance of the learners' L1 in SLA. To be exact, the two formulations examine the effect(s) of whether L1 phonetic categories[13] have been firmly established when L2 learning starts on the acquisition of the L2 phonology.

> One hypothesis that could be tested is that children pronounce an L2 better than adults because they tend to process speech in an 'auditory' rather than a 'phonetic' mode more often, or to a greater extent, than adults, and that this enables them to develop more accurate perceptual 'targets' for L2 sounds. … Children may be more likely than adults to favor an auditory mode (or at least a pre-categorical auditory 'stage') in processing speech sounds because their central representations for sound categories are still evolving. … Since the accurate production of an L2 sound surely depends on an accurate assessment of its acoustic characteristics, this should make it easier for children than adults to pronounce L2 sounds accurately.
>
> A related hypothesis is that children may pronounce L2 sounds better than adults because they are more likely to develop new phonetic categories as a result of exposure to sounds in L2 which are acoustically non-identical to sounds found in L1. This hypothesis rests on the assumption that children's L1 categories are less firmly established or thoroughly elaborated than those of adults. … If the hypothesis just outlined is correct, it would mean the child L2 learners who come to recognize the existence of new categories in L2 would be more fully able to exploit the basic human ability to translate the sensory correlates of speech sounds into articulatory motor routines. Differences between adults and children, if observed, would therefore not be seen as arising from a difference in basic ability, so much as from differences in the extent to which that basic ability is tapped during the course of naturalistic L2 acquisition. (Flege, 1987a, pp. 172-174)

Flege (1987b) further specified the two hypotheses above by incorporating the notion of similar vs. new L2 sounds as well as the mechanism of equivalence classification.

As for the distinction between similar and new L2 sounds, Flege (1987b) states that L2 sounds may be acoustically classified into identical (e.g. /n/ in English and Spanish), similar (e.g. /t/ in English and Spanish), and new (e.g. English /ʌ/ in Spanish) with reference to the learners' L1. Earlier attempts at characterising the distinction between new and similar L2 sounds appeared in the work of Brière (1966; cited in Flege, 1992), Delattre (1964), and Wode (1978) (the latter two cited in Flege, 1991b); while

[13] A definition of phonetic category can be found in Flege (1995a, p. 239): "Language-specific aspects of speech sounds are specified in long-term memory representations called *phonetic categories*."

Flege (1987b, 1991b, 1992, 1997) offers more recent definitions, whereby a similar L2 sound is "an L2 phone which is realized in an acoustically different manner than an easily identifiable counterpart in L1", and a new L2 sound is "an L2 phone which does not have a counterpart in L1, and may therefore not be judged as being the realization of an L1 category." (Flege, 1987b, pp. 58-59). To take a further example of similar and new L2 sounds – i.e. French /t/ and /y/, respectively, for English NSs – Flege (1987b) indicates that "/t/ is found in both French and English, but it is implemented as a short-lag stop with dental place of articulation in French, and as a long-lag stop with alveolar place of articulation in English"; whereas "realizations of French /y/ are 'new phones' for native speakers of English" (pp. 48-49).

Besides, Flege (1991a, 1991b, 1992) discussed several methods to classify L2 sounds into new or similar sounds based on the following criteria: phonetic transcription, analysis of the acoustic vowel space occupied by L1 and L2 vowel segments, and phonetic distinctness tests. In the case of phonetic transcription, Flege (1992) suggested that if an L1 sound and L2 sound differ from each other on an auditory basis but are transcribed with the same IPA (International Phonetic Association) symbol, the L2 sound in question will be considered "similar". By contrast, if a different IPA symbol represents each the L1 sound and the L2 sound, the latter will be regarded as a "new" L2 sound. This method, though, has the problem that expert phoneticians might disagree on transcribing sounds due to different conventions and experience (e.g. English /i/-/ɪ/ distinction might be represented as /iː/-/i/, in which case Flege (1991b) notes that Spanish NSs will consider /ɪ/ a new L2 sound, but if /ɪ/ is transcribed as /i/, they will consider it a similar L2 sound). Moreover, an added drawback of this method is that even if experts agree on the transcriptions of sounds, there is doubt as to whether inexperienced L2 learners will relate L2 sounds to L1 sounds in the same way as predicted by expert phoneticians (Flege, 1991b).

According to Bohn and Flege (1989; cited in Flege, 1991b), the acoustic analysis approach suggests that if the realisation of an L2 vowel occupies a region of the acoustic vowel space that is not occupied by the realisations of any L1 vowel, the L2 vowel will be considered new[14].

[14] Useful though this approach might be, Flege (1997) pointed out that the portion or distance between L1 sounds and L2 sounds in an acoustic space still needs to be determined so that an L2 sound can be considered new.

Finally, another method to distinguish between new and similar L2 sounds takes into consideration phonetic distinctness tests together with perceived auditory differences between L1 and L2 sounds. In Flege's (1992) words,

> an L2 sound could be classified as similar if (1) it were shown to be phonetically distinct from L1 sounds using one of several 'phonetic distinctness' tests, and (2) it could be shown to differ auditorily from the nearest L1 sound(s). (p. 574)

Phonetic distinctness tests may consist of rhyming judgements – "a new vowel might be defined as one that rhymes less with any vowel in the L1 than does any pair of vowels drawn from two adjacent L1 vowel categories" (Flege, 1992, p. 574) – dissimilarity ratings[15] – "a new sound might be defined as an L2 sound that is judged to be more dissimilar when paired with realizations of the closest L1 category than pairs of sounds drawn from that L1 category and its closest neighbor" (Flege, 1992, p. 574) – and orthographic classification of sounds.

An example of the latter is Flege (1991a) who had three groups of Spanish NSs hear and classify four English vowels (namely /i, ɪ, ɛ, æ/) in one-syllable words in terms of letters representing Spanish vowels (that is, <i>, <e>, <a>, <o>, and <u>) and a "none" response option (to be used if a subject considered that the vowel heard did not exist in the Spanish vowel inventory). One subject group consisted of 20 Spanish monolingual speakers with no experience in English or very little experience (i.e. some speakers had a maximum LOR in the US of two weeks), while the two remaining groups comprised 20 Spanish adult learners of English (mean AOL was 25 years) differing in experience in English (average LOR in the US was 6.4 years and 0.8 years for the 10 experienced and 10 inexperienced subjects, respectively). Of interest to the classification or identification of new L2 sounds are the results obtained for English /ɪ/ and /æ/. Based on previous studies (e.g. Scholes, 1967; cited in Flege, 1991a) and the comparison of the acoustic vowel space occupied between those two vowels and the space occupied by any of the Spanish vowels, Flege hypothesised that Spanish NSs would not identify English /ɪ/ and /æ/ with any of the five letters standing for Spanish vowels; hence, these two English vowels would be considered new L2 sounds. However, the results showed that although Spanish learners of English used the label "none" for /ɪ/ and /æ/ more frequently

[15] For an example of a study examining dissimilarity ratings, see Flege, Munro, and Fox (1994) where English monolingual speakers and Spanish learners of English differing in exposure to English rated English-English vowel, Spanish-Spanish vowel, and English-Spanish vowel pairings for similarity on a 9-point scale (*1* meaning *very dissimilar* and *9 very similar*).

than Spanish monolinguals, the differences in frequency rates were nonsignificant. Nor were the somewhat higher frequency rates of "none" responses that experienced Spanish learners of English obtained in comparison to those of relatively less experienced Spanish learners of English. Furthermore, within the same study, a follow-up experiment was conducted where instructions were "manipulated" in order to make learners use the "none" response more often[16]. The findings did not confirm the hypothesis either that /ɪ/ and /æ/ are considered new sounds for Spanish NSs. In addition, despite an increase of "none" responses after a change in instructions, the percentage of frequency rates for "none" responses paralleled that of highly experienced learners in contrast to less experienced Spanish learners of English – i.e. experienced Spanish learners had a higher frequency rate of "none" responses for /ɪ/ and /æ/ than inexperienced Spanish NSs, but not at a significant level, as well as showing a high degree of inter-subject variability in the frequencies of "none" responses. Moreover, the latter experiment might also question the extent to which such a conscious "manipulation" kind of task yields reliable and valid results about the use of an orthographic approach to classify new and similar L2 sounds.

In conclusion, however useful the distinction between new and similar L2 sounds might be in predicting child-adult differences in the perception and production of L2 vowels and consonants (see Flege, 1991b, below), there are still no definite methods as to how to classify L2 sounds into new and similar (e.g. Flege, 1992).

Another relevant aspect to the SLM is the cognitive mechanism of equivalence classification[17], by which humans "perceive constant categories in the face of inherent sensory variability found in the many physical exemplars which may instantiate a category" (Flege, 1987b, p. 49). As Flege (1987b) indicated, equivalence classification is an important mechanism in L1 acquisition "because it permits children to identify phones produced by different talkers or in different phonetic contexts, as belonging to the same category" (pp. 49-50). However, in the case of older L2 learners equivalence classification may prevent them from establishing phonetic categories for L2 sounds as

[16] In the first experiments of the study subjects were informed that "the purpose of the experiment was to identify differences between English and Spanish vowels", whereas in the experiment with the manipulated instructions subjects were told that "many English vowels are unlike any vowel in Spanish" and thus they were encouraged to "use 'none' if they heard one of these 'different' English vowels, but to use one of the five letters if they thought they had heard a Spanish vowel." (Flege, 1991a, p. 720)

[17] In his developmental model of speech perception, Wode (1995) also assigns equivalence classification a relevant role in the acquisition of additional languages.

they already have a firmly established inventory of L1 phonetic categories; hence resulting in their foreign-accented pronunciation of the L2.

Taking all of the above into account, Flege (1987b) does not only outline two related hypotheses about equivalence classification and approximation to L2 phonetic norms (see below), but he also tests them, unlike previous articles mentioned above, where the formulation of hypotheses was mostly based on findings of available studies[18]. Thus, the first hypothesis deals with the effects of equivalence classification on the acquisition of L2 sounds, specifically on the production of French plosive /t/ and vowels /u/ and /y/ by English learners of French:

> [E]quivalence classification prevents adult L2 learners from establishing a phonetic category for similar but now new L2 phones. … As mentioned earlier, an assumption made here is that native English speakers will eventually recognize that /y/ is a separate category. Another important assumption is that L2 learners will be unable to produce authentically L2 phones that differ acoustically from phones in L1 unless they establish a phonetic category for the L2 phones. (Flege, 1987b, p. 50)

The second hypothesis, named the "merger hypothesis", derives from the assumption that all learners maintain their phonetic learning ability through which they acquired their L1 phonological system, no matter whether they are early or late starters. It also considers the new vs. similar L2 sound distinction and the role of phonetic input in perceiving and producing the two sound types successfully. To be exact, the merger hypothesis establishes that

> L2 learners will approximate but not achieve the phonetic norms of L2 for similar L2 phones as they gain experience in the L2. This hypothesis rests on two assumptions: namely, that L2 learners are able to detect auditorily the acoustic differences distinguishing similar L1 and L2 phones; and the phonetic representations which guide segmental articulation continue to be modifiable throughout life as the result of phonetic input. (Flege, 1987b, p. 50)

[18] In addition to suggesting a general hypothesis – "namely, that as the result of the development of the L1 phonetic system, the effect of a mechanism called equivalence classification prevents adults from producing L2 phones authentically by rendering them unable to make effective use of sensory input in speech learning" (p. 49) – Flege (1987b) mentions two other main hypotheses that researchers have put forth in order to account for L2 learners' inability to produce L2 sounds accurately. These are the phonological filtering hypothesis (Trubetzkoy, 1939, 1969; cited in Flege, 1987b) by which the L1 seems to act as a "sieve" through which "perceptually acoustic differences that are not phonemically relevant" are filtered out (pp. 49-50); and the hypothesis that points to motoric reasons for the difficulty older L2 learners experience on trying to produce L2 segments accurately (e.g. Kalikow & Swets, 1972; cited in Flege, 1987b).

What is more, an expected consequence of the merger hypothesis, in conjunction with equivalence classification, is that there will be a bi-directional effect between the L1 and the L2. That is to say, there will be an L1 effect on L2 sound production (or what traditionally has been called interference), as well as an L2 effect on L1 sound production.

In order to test the merger hypothesis and the hypothesis about equivalence classification, Flege (1987b) examined English NSs' production of French /t/ in sentence-initial position by measuring VOT (similar sound to English /t/), and the second formant (F_2) values of French vowels /u/ and /y/ (the former would be considered a similar L2 sound for English NSs, while the latter would be regarded as a new L2 sound) in the words *tous* and *tu*, all of which were uttered in both a phrase and a sentence speaking condition. Six subject groups participated in the study: two groups of monolingual speakers[19] (one of English NSs and another one of French NSs), three groups of English learners of French mainly differing in experience in the TL (average LOR in France was 9 months, 1.3 years, and 11.7 years from the least to the most experienced L2 learners), and one group of French learners of English matched for the same experience in the TL and subject characteristics as the most experienced group of English learners of French (i.e. a mean LOR of 12.2 years in the TL country and married to NSs of the L2). All subjects performed similarly on the two speaking conditions, but the various learner groups varied in their production of new and similar L2 sounds. As predicted, it was observed that highly experienced L2 speakers (of both French and English) approximated the norms for the TL sound /t/, but failed to produce the L2 sound at a native-like level. Furthermore, a merging effect between the typical VOT values of English and French /t/ as produced by the corresponding monolingual NSs was evident in the most experienced L2 learners' /t/s.

[19] The inclusion of monolingual NSs' performance on the TL under study is standard practice in L2 phonological acquisition studies. On the one hand, Flege (1991b) notes that mean values monolinguals obtain for the specific phonetic dimensions under investigation are a point of reference or "benchmarks" used to measure the extent to which L2 learners approximate the phonetic norms of the L2, in addition to "L2 learners' progress in modifying previously established patterns of L1 production as they gain experience in the L2" (p. 251). On the other hand, and as will be seen later, NSs of the TL (preferably monolinguals) are included in FA research in order to give validity and reliability to NS listeners' judgements or ratings. That is, by incorporating NSs of the L2 researchers make sure that listeners or judges are indeed able to do the rating task correctly by identifying the NSs accordingly, as well as preventing biased accent judgements, had all the samples been produced by NNSs only.

As far as approximation to the phonetic norms for TL sounds was concerned, the highly experienced English speakers of French produced an average VOT value of 43 ms for French /t/, whereas the French monolinguals' VOT was 33 ms. In the case of English, the English monolingual speakers' mean VOT value was 77 ms, and that of the highly experienced French learners of English was 49 ms. Regarding the bi-directional effect between L1 and L2, the highly experienced English speakers of French produced English /t/ with a mean VOT value of 56 ms – i.e. an intermediate value between the VOT values of /t/ obtained by English and French monolinguals: 77 ms and 33 ms, respectively. Likewise, the highly experienced French learners of English produced French /t/ with intermediate VOT values – 49 ms – between those typical of monolingual speakers.

Further support for the hypothesis (or hypotheses) above came from the highly experienced subjects' production of French vowels /u/ and /y/. While the most experienced learners of French produced the new L2 sound /y/ accurately, their production of /u/ (similar L2 sound) was not native-like. Instead, they approximated the F_2 frequency values for French /u/: 1508 Hz for English NSs (i.e. F_2 values for /u/ were more English-like) vs. 1196 Hz for French monolinguals.

Finally, none of these findings of phonetic approximation to similar L2 sound norms and intermediate phonetic values in the production of native language sounds resulting from a mutual effect between the L1 and the TL was observed for English learners of French with less amount of experience in the L2. In addition to confirming the hypotheses of the 1987b study, the latter finding suggested that an increase in experience in the TL might help learners discern the phonetic differences between L1 sounds and new L2 sounds, and thus lead to the establishment of additional phonetic categories for L2 sounds.

The variable of experience in the L2 as well as starting age of L2 learning – apart from the already mentioned distinction between new and similar L2 sounds – were examined in more detail in Flege (1991b). In the first place, the proposal that as L2 learners gain experience in the L2, they become more able to discern the phonetic differences between L1 and L2 sounds is further supported by two studies: Flege and Hammond (1982) and Flege (1988b) (all cited in Flege, 1991b). Flege and Hammond (1982) had English NSs produce English sentences with what the speakers believed was a typical Spanish accent in English. It was found that those speakers with more knowledge

of Spanish-accented speech in English produced a larger number of the expected deviations from English phonetic norms that NSs normally produce in English (in that case, sound substitutions and shortening of VOT values for English voiceless stops). Besides showing that NSs of the TL do have a clear notion of how their L1 should sound, the results were consistent with the hypothesis that the more L2 experience subjects have, the better able they are to identify differences between the L1 and L2 segments. This finding in favour of increasing experience in the L2 was corroborated by Flege (1988b; cited in Flege, 1991b) who looked at NSs' and NNSs' perception of speech samples as produced by English NSs and two groups of Chinese NSs differing in experience in English (mean LOR was 1 and 5 years) and in AOL (adult vs. 7.6-year-old starters). The results showed that English NSs distinguished the NS- from the NNS-produced sentences at significantly higher rates than both Chinese talker groups. Similarly, the most experienced Chinese speakers of English also identified NSs' and NNSs' productions at significantly higher rates than the Chinese group with less experience in English. Moreover, this investigation showed that even if L2 learners themselves speak the L2 with a certain degree of FA, as was the case of the two Chinese groups of the study, they are able to recognise instances of FA in the L2.

The later finding led Flege to hypothesise that prior to authentic production of L2 sounds, learners will first need to perceive L2 segments accurately (though Flege, 1995a, mentions some exceptions such as Sheldon & Strange, 1982; see also Gass, 1984). What is more, learners must establish additional phonetic categories for L2 sounds that are the same or resemble those of NSs of the L2 in order to produce L2 sounds at a native-like level.

In sum, experience in the TL plays an important part in the successful formation of phonetic categories for L2 sounds. However, as noted by Flege, experience may have an opposite effect if learners received what he calls accented L2 input, instead of NS input. According to Flege's (1991b) "accented L2 input hypothesis", L2 learners will fail to perceive and produce L2 sounds accurately if they are not provided with adequate L2 phonetic input, regardless of their starting age of L2 learning. For instance, Puerto Rican Spanish subjects who started to learn English at the age of 5 to 6 years, in addition to Puerto Rican adult L2 starters, produced significantly shorter VOT values for English /p, t, k/ than English monolinguals (Flege, 1988a; cited in Flege, 1991b). While the adult L2 starters' performance was in accordance with previous findings (e.g. Flege, 1987b), the

results for child L2 learners did not agree with the hypothesis that if L2 learning starts by the age of 5–6 years, they will be able to produce both similar and new L2 sounds accurately (Flege & Eefting, 1987a; cited in Flege, 1991b). Further evidence for the interpretation of the Flege (1988a) results in light of the accented L2 input hypothesis was provided by Flege (1990c; cited in Flege, 1991b). In that study, the author also looked at a group of Spanish child L2 learners of English, but who differed from the Puerto Rican children in that the former were exposed to English native-like input at the age of 5–6 years in Texas. A group of adult Spanish speakers learning English in the same place was also included in the experiment. Contrary to Flege (1988a), children in Flege (1990c) produced the same VOT values for English /p, t, k/ as English monolinguals. Additional studies in agreement with the accented L2 input hypothesis Flege (1991b) comments on are Flege and Eefting (1987a), where child L2 learners, who had received NS input in the TL, performed at a native-like level; and Flege and Eefting (1988), where the same Puerto Rican subjects who had received accented L2 input failed to reproduce the VOT values on the English continuum /da/ to /ta/ on a native-like basis.

It was indicated above that age of learning has an effect on the establishment of (or failure to establish) additional phonetic categories for L2 sounds. According to Flege (1990b; cited in Flege, 1991b) age 5–7 draws a distinction between early and late L2 learners. This age delimitation derives from the observation concerning the age at which FAs appear to emerge in the L2. Based on all this and the distinction between new and similar L2 sounds, Flege hypothesised that if L2 learning begins before or at age 5–7, L2 learners will be able to establish phonetic categories for both new and similar L2 sounds. However, if L2 learning begins after the age of 5–7 years, L2 learners will not be able to establish additional phonetic categories for similar L2 sounds. In that case, L2 learners will "reuse" phonetic categories they have already established for L1 sounds.

> The hypothesis offered here is that humans' speech learning ability changes with age because phonetic systems remains [*sic*] sufficiently flexible to permit the establishment of additional phonetic categories for sounds similar but not identical to those already in the phonetic repertoire only up to about the age of 5 to 7 years. When children learning an L2 are exposed to a class of auditorily similar phones unlike any in the L1 (what we have been calling a "similar sound"), they create a new category for the unfamiliar L2 sound. … Unlike young children, adults who are exposed to a class of auditorily similar phones unlike any in the L1 tend to **reuse** existing categories. (Flege, 1991b, p. 279)

To conclude, all the points discussed in Flege (1991b) in conjunction with the several hypotheses, predictions, and observations the author discusses in earlier articles and studies above are delineated in a proposal for a model of L2 phonological learning. The model then can be summarised as follows. First, it states that phonetic learning ability is present in all human beings and remains intact through the life span. However, as the L1 phonological system develops and the L1 phonetic categories become firmly established, the addition of new phonetic categories is more difficult, especially in the case of similar L2 sounds and if L2 learning starts after the age of 5–7 years. In that case, L2 learners will start substituting L2 sounds for L1 sounds. In spite of this, and as a sign that phonetic learning ability does not cease as a function of age, the model hypothesises that, regardless of age of onset of L2 learning, phonetic categories for new TL sounds might be established provided there is sufficient input in the L2. Finally, it is thought that authentic production of L2 sounds will depend on their prior accurate perception, which, in turn, involves that phonetic categories for L2 sounds must have been established previously. Nevertheless, the establishment of L2 phonetic categories will not necessarily entail accurate L2 sound production, as FA research studies have shown. Below are all the premises in full that comprise Flege's first outline of an L2 phonological learning model.

Outline of a model of L2 phonological learning (From Flege, 1991b, pp. 281-283)

1. Every human being is born with "phonetic learning ability", i.e., the ability to learn to identify the phonetic categories of an input language(s) and to produce speech with acoustic properties closely conforming to the phonetic characteristics of those categories.
2. L1 phonology "develops" in the following way:
    (a) Phonetic categories are established before phonemic categories. The number of phonemic categories will be smaller than the phonetic categories used to implement them when phonemes are produced with clearly identifiable allophones that are not phonetically conditioned; The number of phonemic categories is determined only after a sufficiently large lexicon has been established.
    (b) The perceptual representations for phonetic categories are elaborated until they conform closely to those of mature speakers in the surrounding community.
    (c) The basic motor plans specified in each phonetic category, and the sensorimotor realization rules used to translate the phonetic categories into articulatory gestures, are aligned with perceptual representations so that phonemes are produced in a language-appropriate manner.
    (d) To varying degrees, all of the above are dependent on the quantity and quality of the input received by the learner.
3. Phonetic learning ability remains intact through the lifespan. Speech learning in L2 differs from L1 speech learning, however, because:
    (a) The phonetic system is gradually optimized for the encoding and decoding of the sounds in L1 (as outlined in 2);
    (b) The phonetic system becomes resistant to the addition of new phonetic categories, which is partly a result of the optimization process.

4. Whether L2 learners identify phones of L2 as "similar" or "new" has important consequences:
    (a) If similar, learners will substitute sounds from their L1 repertoire;
    (b) If new, learners will eventually produce them independently of sounds in the L1 repertoire.
5. An L1 "accent" in the L2 may result from:
    (a) Immature attempts at a new sound, which may lead to "developmental processes" resembling those of children learning L1;
    (b) Equivalence classification of similar sounds, which may lead to transfer errors in production;
    (c) Incorrect lexical representations (e.g., /lak/ for *rock*);
    (d) Correct central representations but immature realization rules;
    (e) Some combination of (a)-(d).
6. Age of learning will determine how similar but not new sounds in an L2 are treated:
    (a) Similar sounds are identified increasingly as being inside the phonetic repertoire and less often as being outside the repertoire as age of learning (AOL) increases.
    (b) For individuals who begin learning an L2 before the age of about 5-7 years, additional phonetic categories are established for similar L2 sounds. The corresponding L1 and L2 sounds will be implemented using different phonetic categories and phonetic realization rules. The production of similar L1 and L2 sounds will be authentic.
    (c) For individuals who begin learning an L2 after about the age of 5-7 years, additional phonetic categories will **not** be established. The corresponding L1 and L2 sounds will mutually influence one another because they are implemented using the same phonetic category. Differences in production may result from the application of different realization rules, but the L2 sounds will not be produced authentically.
    (d) Given sufficient L2 input, L2 learners remain able, even as adults, to establish additional phonetic categories for new L2 sounds. Many of those who do so will go on to produce new L2 sounds authentically. Attitudes and motivation, as well as psychosocial factors may play a role in defining phonetic input in these instances, and so may have an impact on how well new L2 sounds are produced.

Additional studies in support of the speech learning model outlined above are offered in Flege (1992). Thus, investigations such as Flege (1992a; cited in Flege, 1992), Flege, Munro, and Fox (1992; cited in Flege, 1992), and Major (1987a) provide evidence for L2 learners' varying performance in the L2 phonology, according to whether they identify the L2 sounds as being new or similar (point 4 in the model). In the case of Flege (1992a) and Major (1987a), Brazilian Portuguese NSs were found to master the production of English /æ/ at native-like levels, for this vowel does not exist in the L1 inventory (and therefore it is a new L2 sound for that NS group). However, the same subjects failed to produce English /ɛ/ within the NS range, probably due to the fact that Portuguese has a similar vowel sound. As for Flege, Munro, and Fox (1992), they found an effect of equivalence classification on Spanish NSs' learning of English. Those subjects rated English /i/ and Spanish /i/ as being very similar, preventing them from establishing a phonetic category for the similar L2 vowel, which, in turn, led to a non-native production of the specific sound.

Besides the studies of L2 vowel and stop consonant production and perception discussed previously (e.g. for L2 vowels, Flege, 1988a; Flege & Eefting, 1987b, 1988; and for L2 stop consonants, Flege, 1987b, 1991b; Flege & Hammond, 1982; Flege & Hillenbrand, 1984), the findings of the Bohn and Flege (1990c, 1992a; cited in Flege, 1992) and Flege and Bohn (1992; cited in Flege, 1992) investigations agree with the hypothesis that AOL of 5–7 years delineates the distinction between early and late L2 learners, and what this distinction implies for L2 sound learning together with the exposure to input in the TL (points 6(b), 6(c), 6(d) in the model).

Bohn and Flege (1990, 1992a) reported on an advantageous effect of experience on the production of similar L2 sounds. In that study, inexperienced German NSs produced English /æ/ at significantly lower rates than English NSs. On the other hand, experienced German NSs, though presenting some noticeable differences from English /æ/, did not differ significantly from the English NSs' production of /æ/. Moreover, the observed differences between German subjects' /æ/ production and that of English NSs was taken as evidence that late L2 learners will only approximate the phonetic norms for similar L2 sounds.

In another study, Flege and Bohn (1992) examined the differences in the production of English /i, ɪ, ɛ, æ/ in the context [bVt] between a group of 10 Spanish early learners of English, on the one hand, and two groups of each 10 Spanish late learners differing in English experience (average LOR was 0.4 years and 9.0 years for the inexperienced and experienced subjects, respectively), on the other hand. The results showed that the three learner groups produced temporal contrasts between English vowels – i.e. /i/ longer than /ɪ/, and /æ/ longer than /ɛ/ – on a similar basis to that of the two groups of English monolingual participants included in the study. Spanish early learners also produced English vowels with little spectral overlap like English NSs. On the contrary, Spanish late learners produced a larger spectral contrast between English /ɛ/-/æ/ than English monolinguals. Furthermore, three English NS listeners were asked to identify the vowels as produced by the five groups under investigation. Unexpectedly, the results obtained for inexperienced and experienced late learners were not consistent with the hypothesis that late L2 learners will manage to master new L2 sounds as exposure to the L2 increases. Therefore, for the hypothesised new English vowels /ɪ/ and /æ/, the correct identification rates were 61% and 73% for experienced Spanish late learners of

English, and 51% and 70% for inexperienced Spanish late learners. That is, both groups of learners produced English /ɪ/ and /æ/ at very similar rates, regardless of their previous linguistic experience in the TL. Despite this, the study was not taken as evidence to fully reject that hypothesis (6(d)), for it was observed that some experienced L2 learners did master the production of the new English vowels. In other words, a great deal of individual subject variability was found in late L2 learners' performance. In fact, the Flege and Bohn (1992) study pointed to the problem of how to account for the wide range of individual differences found in late L2 learners' production of L2 vowels (a problem already stated in Flege, 1991b).

An extension of this study to three different L1 late learner groups – German, Korean, and Mandarin – further corroborated the prediction that late L2 learners will only approximate the accurate production of similar L2 sounds. At the same time, the number of experienced subjects who correctly produced English /i, ɪ, ɛ, æ/ in more than 90% of instances was overall larger than that of inexperienced late L2 learners (Flege, Bohn, & Jang, 1997).

Another relevant aspect of the 1991b version of the model that Flege (1992) goes into more detail is the hypothesis that the ability to learn to identify phonetic categories and produce speech accordingly remains intact across the life span (point 3 in the model above). He indicates that between the ages of 5 and 7 years children's L1 phonetic system experiences an important change that, in turn, influences L2 sound learning considerably. That change in the L1 phonetic system is concerned with the firm establishment of L1 phonetic categories. As Flege hypothesises, during the ages of 5 and 7 central representations of phonetic categories become better defined in terms of their acoustic properties and their weighting. Besides, that change is also hypothesised to bring about a better and wider delineation of the boundaries for each phonetic category, for there is an increase in the number of phones that are identified as being realisations of each phonetic category in the L1 sound inventory.

> We propose that as children's awareness of segments increases during the period from five to seven years, the categories comprising their phonetic system undergo two broad types of change: (1) the core acoustic properties of prototypical exemplars of each phonetic category, and the weighting of those properties, will become better defined; (2) the *range* of phones that are identifiable as a realization of each category will increase. Note that the first hypothesis pertains to category centers, the second to the boundaries between categories. (Flege, 1992, p. 593)

These two hypotheses/changes are further illustrated in Flege (1992, p. 594, footnote 8). The charts show a hypothetical situation of an acoustic vowel space with four phonetic categories of L1 vowels at time 1 (before changes take place) and at time 2 (after changes have occurred). When both illustrations are compared, it can be seen that at time 2 the centres of phonetic categories are better defined. So are their boundaries (i.e. thicker lines), in addition to being wider (meaning that more variants have been identified as realisations of each specific category). Moreover, a consequence of the developmental shift in speech processing is that the available acoustic vowel space that is free or not occupied by any L1 vowel is reduced significantly. This consequence is of great importance to L2 learning, for it poses the condition that the majority of realisations of an L2 vowel must be found in this "uncommitted" space if learners are to add/create a new phonetic category for the L2 vowel. Otherwise, L2 learners will be unable to establish an additional phonetic category, as they will identify the L2 vowel based on an L1 phonetic category. This, in turn, will likely lead to L2 learners' foreign-accented pronunciation in the TL.

Last, two comments from Flege (1992) are worth mentioning. First, concerning FA, Flege notes that other aspects than segments might influence L2 pronunciation, such as suprasegmental and lexical factors. Despite this acknowledgement, his learning model focuses mainly on segments, namely L2 sound perception and production. Second, and in relation to Flege's model, he justifies the use of the terms "speech" and "learning" that comprise the name of the model – speech learning model – as follows.

> We use the term L2 *learning* rather than *acquisition* because of the view that phonetic systems, even those of adults, undergo constant change in the face of new phonetic input. Thus, speech is never fully *acquired*. The term *speech learning* as used here refers to all aspects of learning that affect the production and perception of the sounds making up words. It is used in preference to *phonological learning* because much or our research has focused on phonetic-level processes. (Flege, 1992, p. 565, footnote 1)

As indicated earlier, all of the above hypotheses and predictions have been included in the most recent version of the Speech Learning Model (SLM) (Flege, 1995a); hence the resulting premises and hypotheses that make up the current version of the SLM are presented below.

Postulates and hypotheses forming a speech learning model (SLM) of second language sound acquisition (From Flege, 1995a, p. 239)

Postulates

P1	The mechanisms and processes used in learning the L1 sound system, including category formation, remain intact over the life span, and can be applied to L2 learning.
P2	Language-specific aspects of speech sounds are specified in long-term memory representations called *phonetic categories*.
P3	Phonetic categories established in childhood for L1 sounds evolve over the life span to reflect the properties of all L1 or L2 phones identified as a realization of each category.
P4	Bilinguals strive to maintain contrast between L1 and L2 phonetic categories, which exist in a common phonological space.

Hypotheses

H1	Sounds in the L1 and L2 are related perceptually to one another at a position-sensitive allophonic level, rather than at a more abstract phonemic level.
H2	A new phonetic category can be established for an L2 sound that differs phonetically from the closest L1 sound if bilinguals discern at least some of the phonetic differences between the L1 and L2 sounds.
H3	The greater the perceived phonetic dissimilarity between an L2 sound and the closest L1 sound, the more likely it is that phonetic differences between the sounds will be discerned.
H4	The likelihood of phonetic differences between L1 and L2 sounds, and between L2 sounds that are noncontrastive in the L1, being discerned decreases as AOL increases.
H5	Category formation for an L2 sound may be blocked by the mechanism of equivalence classification. When this happens, a single phonetic category will be used to process perceptually linked L1 and L2 sounds (diaphones). Eventually, the diaphones will resemble one another in production.
H6	The phonetic category established for L2 sounds by a bilingual may differ from a monolingual's if: 1) the bilingual's category is "deflected" away from an L1 category to maintain phonetic contrast between categories in a common L1-L2 phonological space; or 2) the bilingual's representation is based on different features, or feature weights, than a monolingual's.
H7	The production of a sound eventually corresponds to the properties represented in its phonetic category representation.

In comparison to previous formulations, and, in particular, to the first outline of the model above (Flege, 1991b), Flege points out that the 1995 version introduces a new hypothesis, namely hypothesis 6[20]. Thus, as bilinguals have been shown (to try) to keep a perceptual distinction between L1 and L2 sounds found in a common phonological space (postulate 4), the categories they establish for both L1 and L2 sounds might deviate even further from monolinguals' categories, so that the contrast between L1 and L2 sounds

[20] In a more recent study (Flege, 2002), hypothesis 6 is renamed as "dissimilation hypothesis", which states that "differences between bilinguals and L1 monolinguals may exist even when a category has been established for an L2 sound. The production of such a sound may dissimilate from the perceptually closest sound(s) in the L1 inventory." (p. 132)

persists. Consequently, Flege (1995a) states that the bilinguals' establishment of a phonetic category for an L2 sound will not necessarily entail its accurate production in the same way as monolingual NSs'. Moreover, Flege (1995a) points to the fact that "the model no longer predicts 'mastery' of certain L2 sounds, and the model is now congruent with Grosjean's view of bilingualism[21]" (p. 243).

In addition to this new change in the predictions of the SLM, the following traits that characterise the present SLM (with reference to other models of L2 phonological acquisition) are important to note. First, Flege (1995a) indicates that the hypotheses and premises of the SLM are concerned with learners who have reached their ultimate attainment in L2 phonological acquisition. Moreover, the subjects examined have learned the L2 in naturalistic settings. Another characteristic of the SLM is that the TL under study normally refers to English. And, finally, the model maintains that learners need to perceive L2 sounds accurately in order to "guide the sensorimotor learning of L2 sounds" (Flege, 1995a, p. 238) and to produce segments in the L2 authentically. Despite this prior condition to successful L2 sound production, this assumption does not necessarily mean that all L2 sound production errors will have an auditory basis.

Next, Flege (1995a) applies the SLM hypotheses, which refer to sounds in general, to making specific predictions about the perception and production of L2 vowels and consonants.

As for L2 vowel perception and production, it is hypothesised that any L2 learner, including adult learners, may discern the phonetic differences between L1 and L2 vowels, no matter the size of the L1 sound inventory. If differences are discerned, additional phonetic categories for L2 vowels will be established. In turn, the representations of those categories will constitute the basis upon which the L2 vowels are produced. Moreover, there is a higher probability for a new phonetic category for an L2 vowel to be established as a function of a greater perceived distance between an L1 vowel and an L2 vowel. In addition to a greater perceived distance between L1 and L2 vowels, new phonetic categories for L2 vowels will be more likely to be established the earlier the L2 learner starts acquiring the TL. For example, Munro, Flege, and MacKay (1995; cited in Flege, 1995a) examined native Italian speakers' production of English vowels /i, ɪ, ɛ, æ,

[21] Grosjean (1989) states that there is a continued interaction and co-existence of the two languages of a bilingual speaker, resulting in "a unique and specific linguistic configuration" (p. 6). As for the use of the term "bilingual" in Flege et al.'s research, it refers to "individuals who use two languages on a regular basis" (Flege, personal communication, August 2004). Moreover, Flege (2005) has observed that the bilinguals examined within the SLM are sequential bilinguals, in contrast to simultaneous bilinguals.

ɒ, ʌ, ɚ, o, ʊ, u/. The subjects differed in AOL – ranging from 3 to 21 years of age – but all had reached their ultimate attainment in English (as they had lived in the host country for an average period of 32 years). Results showed that the production of English /ʌ/ by L2 speakers with an AOL after 10 years old was often misidentified by English NS listeners. Similarly, the majority of learners with an AOL greater than 10 years produced /ɚ/[22] with a consistent degree of FA. Moreover, in the case of learners with a starting age of L2 learning greater than 10 years old, the accent ratings obtained for English vowels both with and without an Italian counterpart (e.g. /u/ and /æ/, respectively) were increasingly lower (i.e. less native-like). Flege (1995a) takes these results to demonstrate the effect of varying AOLs on the perception and production of L2 vowels at native-like levels. According to the author, no other factor such as the subjects' insufficient or non-existent experience in the L2 is likely to account for these findings, for their average LOR in the host country was 32 years. However, experience effects as predicted by the SLM have been observed on the adult L2 learners' production of new L2 vowels (e.g. Flege, 1987a, mentioned above).

Concerning category formation for L2 vowels, it then follows that if a new category is established for an L2 vowel, the L2 vowel will be produced accurately. However, if the new phonetic category is not established, the SLM hypothesises that the L2 vowel will be produced with the properties of the closest L1 vowel, and so will the L1 vowel in relation to the L2 vowel. Evidence in favour of what Flege calls the "merging" of phonetic properties between L1 and L2 vowel sounds resulting from L2 learners' inability to form additional phonetic categories for TL sounds comes from, among others, Major (1987a; cited in Flege, 1995a), who found that Portuguese NSs produced English /ɛ/ (with a similar sound in Portuguese) increasingly less accurately as their production of English /æ/ improved; and Bohn and Flege (1992; cited in Flege, 1995a), who also reported on inexperienced German speakers' English /ɛ/ production on a poorer basis than that of experienced German learners, though their L1 has a similar vowel sound.

Similar hypotheses to those for L2 vowel perception and production can be made about the perception and production of L2 consonants. Thus, it is predicted that L2 learners are able to detect phonetic differences between L1 and L2 consonant segments, which may lead to the establishment of new phonetic categories for L2 consonants. As in

[22] /ɚ/ corresponds to /ɜ/ in Standard British English.

the establishment of phonetic categories for L2 vowels, category formation for L2 consonants can be triggered on the basis of a greater perceived phonetic distance between L1 and L2 consonants. For instance, based on the finding that Japanese adult NSs perceive the distance between Japanese /r/ and English /ɹ/ to be greater than that of Japanese /r/ with English /l/ (Sekiyama & Tohkura, 1993; Takagi, 1993; cited in Flege, 1995a), together with the model's predictions, Flege, Takagi, and Mann (1995a, 1995b; cited in Flege, 1995a) hypothesised and confirmed in their studies that Japanese adult learners of English would be more likely to establish a new category for English /ɹ/, rather than English /l/, and consequently they would produce English /ɹ/ more accurately.

In the case that the phonetic distance between an L1 consonant and L2 consonant sound is perceived to be smaller – that is, in those instances of "continued perceptual linkage of L1 and L2 sounds – i.e. by equivalence classification" (Flege, 1995a, p. 258) – it has been noted that L2 learners produce the TL sounds with intermediate values between the typical values of L1 sounds and those of L2 sounds. For example, Flege (1987b) and Flege and Eefting (1987b) (cited in Flege, 1995a; see above also) observed that French and Spanish learners of English, respectively, produced English word-initial /p, t, k/ in a nonnative-like manner. More precisely, the learners' production of VOT values for English word-initial /p, t, k/ approximated the "standard" English VOT values for the voiceless stops in word-initial position. Another finding consistent with hypothesis 5 above is that experienced L2 learners produced both their L1 stops (Spanish or French /p, t, k/) and the English L2 stops with VOT values very similar to each other, as a result of equivalence classification and subsequent restructuring of the L1 phonetic categories to "accommodate" the perceptual linked sounds in the L1 and L2.

Like L2 vowel perception and production, the earlier a learner starts acquiring the L2, the more likely they will be to establish additional phonetic categories for L2 consonants. Therefore, Flege (1991b; cited in Flege, 1995a) interpreted the result that early Spanish NS learners of English produced the same VOT values for English /p, t, k/ as English NSs as evidence that they had established a category for those sounds. On the other hand, late Spanish NS learners produced English word-initial /p, t, k/ with the expected compromise VOT values between those typical of Spanish VOT and English VOT. Similarly, late Italian-speaking learners of English with an AOL of 21 years

produced English word-initial /p, t, k/ with nearly intermediate VOT values between English and Italian (Flege, Munro, & MacKay, 1995b; also cited in Flege, 1995a). Besides, favourable effects of earlier starting age of L2 learning on accurate production of English VOT values were found in Italian subjects' production of English /p/ with an AOL of 16 years and below. In the case of English /t/, subjects with an AOL of 10 years and below produced VOT values similar to those of native English /t/. Flege et al. (1995b) further stated that onset age of L2 learning of 10 years was also the upper-limit age for native Italian subjects to produce English fricatives in word-initial position with the same correct frequency rates as English NSs'. Additional starting age effects have been observed in Japanese native speakers' perception of a synthetic continuum of English /ɹ/-/l/. So, Yamada and Tohkura (1992b; cited in Flege, 1995a) reported on the finding of native-like perception of English /ɹ/-/l/ by Japanese NSs starting to learn English before age 5. On the other side, subjects with an AOL between 5 and 10 years did not always perceive English /ɹ/-/l/ accurately (about two thirds were "successful"), and very few subjects (about 25%) with an AOL after 10 years managed to perceive English /ɹ/-/l/ correctly.

Apparent counter-evidence to the hypotheses of the SLM – to be exact, hypotheses 1 and 7 – comes from Flege, Munro, and Skelton (1992; also cited in Flege, 1995a). According to the first hypothesis in the model, Flege (1995a) suggests that L2 learners who lack consonant sounds in certain allophonic positions in their L1 sound inventory, as is the case of the English voiced plosives in word-final position for Spanish NSs, will be able to establish new phonetic categories for those sounds. Consequently, they will perceive and produce the L2 sounds in question accurately. However, Flege, Munro, and Skelton (1992) found that all the English word-final stops produced by Spanish and Mandarin NSs were not identified correctly as often as those of English NSs, regardless of the subjects' experience in the TL (9 yrs vs. 2 yrs for Spanish NSs, and 6 yrs vs. 1 yr for Mandarin NSs). Another study disconfirming hypothesis 1 is Flege, Munro, and MacKay (1995b; also cited in Flege, 1995a) where nearly half the number of Italian subjects with an AOL of 15 years and greater devoiced English word-final /b, d, g/, despite all subjects' long-term exposure to English in the host country (average LOR was 32 years).

With the exception of the two latter studies, Flege's review of research is consistent with the hypotheses of the SLM. It should also be mentioned that throughout the formulation and illustration of the hypotheses, there appears a recurrent notion, namely category formation or the establishment of phonetic categories. As outlined in the postulates of the SLM, a phonetic category is a long-term memory representation comprising language-specific aspects of speech sounds (see also footnote 13 in this section). Flege (1992) makes a proposal as to how L1 phonetic categories are formed or become firmly established between the ages of 5 and 7 years (see above). The author also notes that

> the notion of "phonetic category" implies the perceptual ability to: 1) identify a wide range of different phones as being "the same", despite auditorily detectable differences between them along dimensions that are not phonetically relevant; and 2) ability to distinguish the multiple exemplars of a category from realizations of other categories, even in the face of noncriterial commonalities (Kluender, Diehl, and Killeen 1987). (Flege, 1995a, p. 244)

Furthermore, Flege (1995a; see also Flege 1999a, 1999b) comments on various tasks that might be used to test for category formation by L2 learners, such as the two-alternative forced-choice (2AFC) identification test, oddity discrimination task, and the speaking rate paradigm.

Both the 2AFC identification test and the oddity discrimination task have been administered to assess whether L2 learners have created new phonetic categories for L2 vowels. However, Flege (1995a) illustrates that the former is not an appropriate test for category formation, based on the results from the Flege and Bohn study (1989; as cited in Flege, 1995a). In that investigation, the authors found that although Spanish NSs were able to divide each of the /i-ɪ/ and /ɛ-æ/ continua into two response categories, they failed to identify the members of the /i-ɪ/ continuum in a native-like fashion. That is, while the native English-speaking subjects identified instances of /i/ and /ɪ/ based on spectral quality, the Spanish NSs made use of vowel duration – "a readily available property" (p. 244) – in order to identify both members of the continuum. In the case of the /ɛ-æ/ continuum, Spanish learners of English recognised both members on the continuum like English NSs did. However, Flege (1991c; as cited in Flege, 1995a) presented data (not evident at first hand in Flege & Bohn, 1989) showing that native Spanish speakers' successful recognition of /ɛ-æ/ was based on two different Spanish vowel categories – /e/

and /a/, respectively – which was unlike English NSs' recognition of /ɛ-æ/. Thus, the 2AFC identification test seems to overlook certain aspects or processes that might lead to the wrong assumption that L2 learners' category formation for L2 vowels has been successful. Then, a more suitable test of category formation is an oddity discrimination task, in particular "an ABX (three stimuli: A,B always different) "categorial" discrimination task" (Flege, 1995a, p. 245). As an example, Flege (1995a) comments on a slightly modified categorial discrimination task to that originally devised by Gottfried (1983; as cited in Flege, 1995) that Flege, Munro, and Fox (1994; also cited in Flege, 1995a) used to test for category formation in two groups of Spanish adult NSs differing in experience in English (4.1 yrs and 0.9 yrs of L2 input operationalised as LOR x % English use), in addition to an English NS group. First, the categorial discrimination task consisted of 45 different types of triads – 3 types of Spanish-Spanish vowel pairings, 21 types of English-English vowel pairings, and 21 English-Spanish vowel pairings – each with three stimuli spoken by different talkers. Second, there was a longer inter-stimulus interval (1.2 seconds) between the three stimuli of each triad to avoid subjects' responses based on auditory short-term memory. And, third, "catch" triads (i.e. three different talkers' realisations of a single phonetic category) were included, so that subjects would base their responses on phonetically relevant differences.

The results showed that there was no effect of experience on Spanish subjects' identification of the 45 vowel pair types. However, a difference in the perception of adjacent versus nonadjacent vowel pairs was observed between English and Spanish NSs. Thus, in the case of nonadjacent vowel pairs (vowels that are relatively distant in an F_1-F_2 acoustic space), both English and Spanish NSs rated the vowel pairs as dissimilar on the same basis, whereas Spanish subjects rated vowels in adjacent pairs (vowels that are less distant in an F_1-F_2 acoustic space) to be less dissimilar than English NSs. The latter finding led the authors to conclude that Spanish NSs identified vowels in adjacent pairs (e.g. Spanish /a/ – English /æ/) based on one category only.

Finally, the speaking rate paradigm (Miller & Volaitis, 1989; Volaitis & Miller, 1992; as cited in Flege, 1995a, and Flege & Schmidt, 1995) is another technique to test for category formation of L2 sounds, particularly voiceless stops. Flege and Schmidt (1995; also cited in Flege, 1995a) state that "the paradigm is based on the observation that in speech production the duration of VOT intervals, especially those of voiceless stops, is shortened as speaking rate increases" (p. 93). Thus, to test for category formation

of English /p/ by 40 Spanish adult learners of English varying in proficiency in English – proficient vs. nonproficient – Flege and Schmidt (1995) used the original speaking rate paradigm from Miller and Volaitis (1989) and explained how it works as follows.

> Speaking rate variations are simulated by varying overall syllable duration in two VOT continua. The VOT of syllable-initial bilabial stops ranges from short-lag values typical for English /b/ to values that exceed those typical for English /p/. English monolinguals have rated the goodness of bilabial stops as realizations of the English /p/ category. Their ratings have been observed to increase systematically as VOT increases, then to decrease as VOT values extended beyond typical of English values. (p. 93)

Furthermore, Flege and Schmidt (1995) hypothesised that in order to prove that L2 learners have formed a category for English /p/ three types of evidence/results are needed. First, like English NSs, Spanish subjects should increase their goodness ratings systematically as VOT increased from 10 to 50 ms. The second condition is that Spanish subjects should assign the highest goodness ratings to stimuli containing the VOT interval values typical of English NSs; as well as decreasing their goodness ratings in a systematic manner as a result of an increase in VOT beyond the values typical for English /p/. Last, rate-dependent processing should be evident in Spanish NSs' performance.

Only the first condition was met by both proficient and nonproficient Spanish speakers. As to the second requirement, proficient Spanish learners fulfilled the condition just like English NSs. However, the proficient speaker group did not meet the third condition (nor did the nonproficient group). Therefore, Flege and Schmidt (1995) drew a tentative conclusion, which stated that failure to meet the third condition, even if the two other requirements for category formation had been met, prevented the proficient Spanish subjects from establishing a perceptual category for English /p/. Moreover, those speakers' single category for /p/ was thought to share the acoustic characteristics of both English and Spanish /p/. As just mentioned, the findings of the study were not conclusive in that a great deal of individual differences was noted in late learners' performance, and thus some speakers might indeed have established a new category for English /p/.

The studies of category formation tests reviewed above focus on NSs of Spanish only. However, tests of category formation have been examined in other L1 speakers, in addition to being administered for other purposes (e.g. Flege, 1995a; Flege, MacKay, &

Meador, 1999). For instance, Flege (in preparation; cited in Flege, 1995a) used the ABX categorial discrimination task to investigate whether different L1 speakers (Arabic, Dutch, German, Korean, Portuguese, and Spanish) would all discriminate English vowels similarly to English NSs, as well as establishing whether subjects' discriminative failures would vary according to each group's L1 vowel inventory size. None of the L1 groups discriminated English vowels at English native-like rates, though German and Dutch NSs obtained higher correct rates than the remaining L1 groups. Differences among the various L1 groups were also found in the number and type of discriminative failures. For example, Spanish speakers had difficulty discriminating five vowel pairs – /ɑ/-/ʌ/, /ʊ/-/u/, /ɛ/-/ɪ/, /æ/-/ɑ/, and /ɛ/-/æ/ – whereas Germans' discriminated only one pair with difficulty – /ɛ/-/æ/ – and Dutch speakers two vowel pairs – /ʊ/-/u/ and /ɑ/-/ʌ/. This finding agreed with the hypothesis that if two vowels are identified in terms of one single category, L2 learners will have difficulties in discriminating the pair of English vowels in question.

Another likely explanation for the findings of the study might be that non-native speakers of English did not have sufficient experience yet (mean LOR was 7 years) to detect the phonetic differences between the English vowel pairings, whereby category formation might have been triggered (for effects of experience understood as NS input see also Flege, 1991b, above, and Flege, 2003). As Flege (1995) suggests,

> The full range of L2 sounds may at first be identified in terms of a positionally defined allophone of the L1, but as L2 learners gain experience in the L2, they may gradually discern the phonetic difference between certain L2 sounds and the closest L1 sound(s). When this happens, a phonetic category representation may be established for the new L2 sound that is independent of representations established previously for L1 sounds. (p. 263)

As noted earlier, the hypotheses of the SLM have been confirmed for the most part by empirical studies (as reviewed by Flege, 1991b, 1992, 1995a). In spite of this, Flege (1991b, 1992, 1995a, 1999a, 1999b, 2003) comments on a number of "unresolved issues" within the theoretical framework of the SLM.

First of all, the SLM establishes that perceived cross-language phonetic distance and AOL are of great importance to trigger the process of category formation. Thus, the greater the distance between an L2 and the closest L1 sound is perceived, the more likely the formation of an additional phonetic category. In addition, the earlier L2 learning starts, the smaller the perceived phonetic distance is needed to prompt category formation. However, as far as cross-language phonetic distance is concerned, Flege

(1995a) notes that the attempts to measure cross-language phonetic distance either based on the visual and acoustic characteristics associated with L1 and L2 sounds, or on differences in perceived gestures, have failed to provide sufficient satisfactory evidence. Therefore, means or methods to assess cross-language phonetic distance objectively are still lacking (Flege, 1995a, 2003).

By the same token, a reliable classification technique to characterise L2 sounds as new or similar is yet to be found (for the several proposals made to date and their disadvantages, see Flege, 1991b, 1992, and above; cf. Flege, 1997)[23].

As for the role of AOL in category formation, Flege (1995a) mentions that the findings of starting age-related effects on L2 speech learning derive from L2 production studies mainly. Consequently, the effects of AOL on L2 sound perception need further examination by carrying out the same types of tasks administered in the investigation of AOL effects on L2 sound production.

Additional research should also be conducted to determine whether the perception of L2 sounds changes as a function of AOL; and, if so, how it changes. Moreover, Flege (1995a) indicates that if L2 sound perception does change, the effects (if any) of such changes on L2 sound production should be typified. All these research questions are motivated by the assumption of the SLM that the degree of accurate production of L2 sounds depends on the extent to which L2 sounds are perceived in a native-like fashion. As a result, the model also predicts that many L2 production errors are based upon whether L2 learners perceive the TL segments at a native-like or nonnative-like level.

Bearing this in mind, recent articles by Flege (1999a, 1999b, 2003) aim to illustrate the relation between perception and production in L2 phonological acquisition. After a review of studies examining the relation between perception and production of L2 vowels (e.g. Flege, Bohn, et al., 1997; Flege, et al., 1999; also cited in Flege, 1999a, 1999b) and L2 consonants (e.g. Flege, 1993; Flege & Schmidt, 1995; Schmidt & Flege, 1995; all cited in Flege, 1999b), Flege (1999b) concludes that significant correlations

---

[23] Wode (1995) further comments on a "paradoxical shortcoming" of the SLM and the PAM. In addition to problems outlined above as to how to asses cross-language phonetic distance objectively and reliably and techniques on how to classify L2 sounds as new, similar, or identical, he notes that

> they [SLM and PAM] make no reference to what may enable learners to solve a given learning task, that is, how does the learner within Flege's model decide whether a given L2 phone is identical, similar, or new with respect to his or her prior L1 categories? Or how, according to Best's view, is the learner to determine whether to go for single category assimilation, category goodness, or whether to treat the target as nonassimilable? (p. 332)

He then proposes that the interaction between categorical and continuous modes might hold the key to this shortcoming.

between perception and production of L2 segments do exist, though the correlations observed are "modest" – that is, the average correlation coefficient is r = .50. Furthermore, he lists a number of reasons that might account for the finding of modest correlation coefficients. One reason has to do with the fact that some subjects who manage to perceive an L2 sound accurately might fail to modify the production of the L2 according to the new phonetic category for that specific sound. A second cause points to the existence of age constraints on learning new forms of articulation, based on the hypothesis of the SLM that the production of some L2 sounds in a nonnative-like way does not have a perceptual basis. Another reason for the relatively modest correlation between perception and production of TL sounds lies in methodological factors such as speech clarity (in particular, perception tests lack in their characterisation of rate and clarity, unlike production tests) and the dimensions measured (while often L2 phonological research measures one single dimension such as VOT, it is thought that other dimensions are likely to change in perception and production in parallel). The last cause Flege (1999b) advances is concerned with category formation. According to the author, category formation is "the most meaningful perceptual variable" (p. 1276). However, most investigations in L2 speech acquisition do not contemplate the study of category formation and its relation to accurate perception and production of L2 sounds (cf. Flege, 1999a).

Additional research is also needed in order to determine the effects of training on L2 perception and production, in addition to assessing the quantity and quality of L2 phonetic input and their subsequent effects on L2 phonological learning (Flege, 1991b). As noted by Flege (1999a), there exists preliminary evidence that training on the perception of specific L2 sound contrasts results in a more accurate perception of the L2 sound contrast, which, in turn, might lead to an improvement in the production of the particular sounds (e.g. Bradlow, Pisoni, Yamada, & Tohkura, 1996; Yamada, Tohkura, Bradlow & Pisoni, 1996; cited in Flege, 1999a).

Last, the SLM still needs to provide an account of the individual subject differences encountered in the various studies conducted within this theoretical framework (Flege, 1995a). That is, as of yet there has been no definite explanation as to why some adult L2 learners produce similar L2 sounds in a native-like fashion, whereas several adult L2 learners as well cannot produce new L2 similar sounds accurately. Several possibilities (pending further investigation) might be suggested based on the selection criteria for participants in L2 research (Flege, 2003). They include identification

of factors that are often confounded with AOL, L2 input and language patterns, and language dominance (for a recent review of studies examining these variables in L2 sound perception and production, see Flege, 2003).

2.3.1.2. The Perceptual Assimilation Model (PAM)

The Perceptual Assimilation Model (PAM) (Best, 1995, 1999; Best & Strange, 1992; Best, Traill, Carter, Harrison, & Faber, 2003) originated from the observation that adults have difficulty perceiving certain sound contrasts in the TL, whereas very young children succeed in the correct discernment of those same non-native sound contrasts. Additionally, the finding of both adults' and children's accurate perception of some other non-native contrasts to which neither age group had previously been exposed – e.g. Zulu click contrasts as perceived by English-speaking adults and infants (Best, McRoberts, & Sithole, 1988) – contributed to the formulation of the PAM. That is to say, the Perceptual Assimilation Model was developed in an attempt to account for child-adult differences noted in the perception (or various degrees of perceptual difficulty) of non-native contrasts.

Like the SLM, the PAM assumes that the mechanisms present in the acquisition of L1 speech do not cease to function and/or to be accessible throughout the lifespan, or in Best's (1995) words "perceptual learning continues into adulthood" (p. 198). In the case of the PAM, these mechanisms consist of integrated perceptual systems and their exploratory activities (Best, 1995). Thus, the characterisation of perceptual mechanisms within the PAM agrees with that of the direct realist approach to speech perception. In fact, Best (1995) points out that the Perceptual Assimilation Model is based on the ecological theory of perception developed by Gibson and Gibson (1955, 1972; as cited in Best, 1995) and its direct realist hypotheses about perceptual learning.

Briefly, and as summarised by Best (1995), the ecological account claims that humans perceive and obtain information from the world about objects and events directly – distal articulatory gestures being the basic units of perceptual primitives that comprise this information. As regards speech learning, infants are thought to originally perceive non-linguistic information (again in the form of gestures). However, early on infants start (and need) to discover critically distinctive features in the L1 – i.e. "invariants of

language-specific articulatory gestures and constellations of intergestural phasing at all levels from segments to syllables, words, and so forth" (Best, 1995, pp. 184-185). Consequently, the notion of perceptual learning is central to direct realism, as it involves (infant) perceivers' detection and discovery of optimal gestural invariants comprising the L1 structures in an effective economical manner, along with the attunement of their perceptual system to native speech.

By extension, Best (1995) states that the PAM assumes that, as a result of perceivers' attunement to language-specific (or native) speech, listeners are able to detect higher-order invariants that specify the gestural constellations making up the native phonological inventory. In addition to the notions of "gestures" and "gestural constellations" mentioned above, the author comments on the concepts of "universal phonetic domain" and "native phonological space" considered within the direct realist approach to speech perception, all of which are relevant to outlining the theoretical foundations of the PAM. Furthermore, the delineation of such notions derives from Browman and Goldstein's (1986, 1989; as cited in Best, 1995) model of gestural phonology.

Specifically, the gestural model proposes that "phonological patterning in languages obeys the constraints provided by the physical structure of the vocal tract and the movement that its biomechanical components afford" (Best, 1995, p. 187). Moreover, the model considers articulatory gestures as the "primitives, or 'atoms', of phonological structure" (p. 187). From this perspective, a simple articulatory gesture is defined as "the formation (and release) of some degree of relative constriction at some location along the vocal tract[24]" (p. 187). In turn, phonological structures "are stable constellations, or 'molecules', assembled with those atoms" (p.187). Therefore, phonological structures are mostly made up of several multiple gestures, rather than simple gestures.

All the gestures that human languages can possibly produce and employ constitute the universal phonetic domain, while the native phonological space is formed of only those gestures (both simple gestures and constellations) that serve phonological purposes in a given language. Then, the limits of the native phonological space are defined not only by the static spatial dimension of constriction location, but also by the spatiotemporal dimension of simple constriction gestures. Despite the many gestures and constellations that coincide within native phonological spaces, Best (1995) indicates that

[24] In the case of consonants, "constriction" and "location" might stand for *manner* and *place of articulation*, respectively.

languages are thought to differ from each other phonologically; hence in the delineation of their corresponding native phonological space. Such delineation further results from the simple gestures that languages select, in conjunction with the combination of those simple gestures and the specific phasing languages adopt between the gestures that are used to form constellations (or segments).

What is more, as speakers become attuned to their L1, they are better able to efficiently detect the critically distinctive features of gestural constellations comprising the native phonological space. A corollary of this perceptual attunement to the L1 is the fact that speakers/perceivers "discard" simple gestural features (i.e. those which do not have a phonological function in the L1) that they noticed at first. Best goes on to observe that this attunement to native language(s) and subsequent experience with the L1 will determine the perception of non-native gestural constellations (or non-native segments) – "those [segments] whose gestural elements or interlingual phasing do not match precisely any native constellations" (Best, 1995, p. 193).

As far as the PAM is concerned, its basic premise is that L1 experience determines the perception of TL sounds. More precisely, the model hypothesises that listeners tend to perceive non-native segments based on their similarities and dissimilarities to native gestural constellations that are closer in native phonological space. Both the coincidences and discrepancies between native and non-native phones listeners perceive result in three different patterns of perceptual assimilation of non-native constellations to native constellations. In the first place, a non-native segment may be assimilated to one particular native phone, and therefore be heard as either a good instance, or an acceptable though not ideal, or a deviant instance of the native phone. Secondly, non-native segments might be assimilated to a speech-like sound, but uncategorisable within the listener's native phonological space. Lastly, non-native segments might not be assimilated to any sound at all, and, as a consequence, they will be heard as nonspeech sounds. As described by Best (1995), the available patterns of perceptual assimilation of non-native segments to L1 sounds are presented below.

1. *Assimilated to a native category*: clearly assimilated to a particular native segmental category, or perhaps to a cluster or string, in which case it may be heard either as:
    a. a good exemplar of that category
    b. an acceptable but not ideal exemplar of the category
    c. a notably deviant exemplar of the category
2. *Assimilated as uncategorizable speech sound*: assimilated within native phonological space as speechlike gestural constellation, but not as a clear

exemplar of any particular native category (i.e., it falls within native phonological space but in between specific native categories)

3. *Not assimilated to speech (nonspeech sound)*: not assimilated into native phonological space at all; heard, instead, as some sort of nonspeech sound.

(Best, 1995, pp. 194-195)

It was mentioned earlier that the differences observed in the perception of non-native contrasts between infant and adult listeners led to the delineation of the PAM. Thus, Strange (1995a) emphasises that the perceptual model specifically addresses sound contrasts rather than individual segments, unlike the SLM. The three patterns of perceptual assimilation described above are therefore applicable to each sound of the non-native contrast under study. In consequence, the model allows for various combinations of assimilation patterns derived from the perceptual assimilation of each non-native segment included in the contrast. According to Best (1995), the following assimilation patterns are possible: (a) two-category assimilation (TC type), where each TL segment is assimilated to a different L1 category; (b) category-goodness difference assimilation (CG type), where both segments are assimilated to the same L1 sound, but one non-native segment is heard as a better instance of the L1 category than the remaining non-native segment in the pair; (c) single-category assimilation (SC type), where both segments are assimilated to the same L1 category and the two segments are heard just as both good or deviant instances of the L1 segment; (d) both uncategorisable assimilation (UU type), where both segments are heard as possible sounds within listeners' native phonological space, but cannot be assimilated into (or identified as) any L1 phoneme; (e) uncategorised vs. categorised assimilation (UC type), where one non-native segment is assimilated to one L1 category and the remaining sound in the contrast is not assimilated into any native category, though it is recognised as part of the native phonological space; and (f) non-assimilable pattern (NA type), where both non-native segments cannot be assimilated to any L1 sound and are not recognised as any possible sound that can be produced within listeners' native phonological space.

Furthermore, the PAM makes predictions about the discrimination levels or degrees of perceptual difficulty associated with each pattern of perceptual assimilation of non-native contrasts. It should be noted that the discrimination levels predicted refer to initial perceptual difficulties (Strange, 1995a), for the model examines "naive" listeners – i.e. listeners with no prior exposure to the TL or very little experience in the L2/FL (Best, 1995). This differs from the subject population investigated within the SLM framework –

namely, highly experienced learners in the L2 or learners who have reached their ultimate attainment in the L2 (Flege, 1995a) – and its subsequent hypotheses regarding L2 speech perception. As Strange (1995a) puts it,

> Two working models have been offered that attempt to predict (and explain) the relative perceptual difficulty of non-native phonetic categories. They complement each other in that Best's Perceptual Assimilation Model (PAM) [36][25] focuses on initial perceptual difficulties, while Flege's Speech Learning Model (SLM) [37] proposes an account of perceptual reorganization in both L2 and L1 as a function of L2 experience. (pp. 79-80)

Therefore, the perceptual assimilation model predicts that discrimination will be excellent for the TC type, and very good for UC type assimilations. NA type patterns will yield good to very good discrimination results. Furthermore, discrimination ranging from poor or modest to good is predicted for CG and UU type assimilations, respectively. Finally, SC type assimilations will result in poor discrimination levels. Best's (1995) outline of the perceptual assimilation patterns that are most likely to occur on perceiving non-native contrasts, together with their discrimination levels, are summarised as follows.

> *Two-Category Assimilation (TC Type)* Each non-native segment is assimilated to a different native category, and discrimination is expected to be excellent.
>
> *Category-Goodness Difference (CG Type)* Both non-native sounds are assimilated to the same category, but they differ in discrepancy from native "ideal" (e.g., one is acceptable, the other deviant). Discrimination is expected to be moderate to very good, depending on the magnitude of difference in category goodness for each of the non-native sounds.
>
> *Single-Category Assimilation (SC Type)* Both non-native sounds are assimilated to the same native category, but are equally discrepant from the native "ideal"; that is, both are equally acceptable or both equally deviant. Discrimination is expected to be poor (although it may be somewhat above chance level).
>
> *Both Uncategorizable (UU Type)* Both non-native sounds fall within phonetic space but outside of any particular native category, and can vary in their discriminability as uncategorizable speech sounds. Discrimination is expected to range from poor to very good, depending upon their proximity to each other and to native categories within native phonological space.
>
> *Uncategorized versus Categorized (UC Type)* One non-native sound assimilated to a native category, the other falls in phonetic space, outside native categories. Discrimination is expected to be very good.
>
> *Nonassimilable (NA Type)* Both non-native categories fall outside of speech domain being heard as nonspeech sounds; discrimination is expected to be good to very good.
>
> (Best, 1995, p. 195)

[25] [36] and [37] refer to Best (1995) and Flege (1992), respectively.

Just as the notion of category formation was frequently alluded to within the SLM framework, in the formulation of the PAM there also appears a recurrent notion, namely, perceptual assimilation. Best (1995) first specifies that perceptual assimilation of non-native segments to native segments is determined by phonetic (gestural) similarity. Moreover, Best et al. (1988) note that the concept of perceptual assimilation is to be understood as "phonetic analogy" and "phonic interference", instead of "phonemic assimilation".

> The perceptual process described here may be similar to the notion of 'phonetic analogy' mentioned by Eilers *et al*. (1982) and also to the concept of 'phonic interference' that has been used to describe the spoken errors made by learners of a second language (e.g., Weinreich, 1953). Note that the process proposed here is not the same phenomenon as 'phonemic assimilation' in speech production whereby, for example, <pocke*t*book> is pronounced as though it were <pocke*p*book>. (p. 347, footnote 1)

Finally, perceptual assimilation has been measured and assessed through identification or labelling tests, classification tasks, and categorisation (among them goodness ratings and AXB oddity discrimination) tests (Best, 1995).

PAM's predictions have found support in a number of studies undertaken by Best and her colleagues. Initially, the investigations focused on non-native consonant contrasts (e.g. Best, 1990; cited in Best, 1995; Best & Strange, 1992; Best, McRoberts, & Sithole, 1988; Polka, 1992). However, Best (1995) illustrates that in an attempt to further test and extend the findings of the Perceptual Assimilation Model concerning non-native consonant contrast discrimination, recent research has examined perception of non-native vowel contrasts (e.g. Best, Halle, Bohn, & Faber, 2003). Additionally, the authors have looked at infants' perception of non-native contrasts in order to depict the nature of developmental changes in TL sound perception.

Based on a review of developmental cross-language studies, Best (1995, 1999) and Strange (1995a, 1995b) have concluded that a marked decline in infants' native-like perception of non-native contrasts is evident between 6 and 12 months of age. To be exact, difficulties in perceptual discrimination of non-native consonants and vowels arise between 8 and 10 months of age and by 6 months of age, respectively. The findings then suggest that "the developmental reorganization of speech perception implicit in these adult-infant differences has begun before the emergence of the child's first words" (Best, 1999, p. 1261), and, above all, they provide the basis for the important (unavoidable) effect of the L1 on L2 speech perception (as postulated by the PAM).

As for adults' perception of non-native consonant contrasts, one of the first studies to test PAM's predictions (Best et al., 1988) centred on the NA type assimilation pattern by examining nine English NSs' perception of non-nasalised click contrasts from Zulu. The high correct discrimination scores (mean range = 80.6% – 99.1%) Ss obtained on an AXB discrimination task[26] consisting of 18 different non-nasalised click contrasts were consistent with the hypothesis that listeners would perceive click contrasts as non-assimilable speech sounds, leading to very good discrimination levels.

Another perceptual assimilation pattern – the TC type – was supported by native English listeners' near 100% correct discrimination of the Ethiopian ejective contrast /pʼ/-/tʼ/, which had been hypothesised to assimilate into two different L1 categories (English /p/-/t/, respectively) (Best, 1990; as cited in Best, 1995).

Best and Strange (1992) further investigated the TC type assimilation pattern, in addition to the CG and SC type patterns, and their corresponding predicted degrees of perceptual difficulty, in two Japanese and English NS groups' (of nine subjects each) discernment of three English approximant contrasts – i.e. /w/-/j/, /w/-/r/, and /r/-/l/ – on an AXB discrimination task and a two-forced choice identification test. Since both /w/-/j/ and /w/-/r/ are phonological contrasts in Japanese, the former contrast was hypothesised to assimilate as a TC type (into Japanese /w/ and /j/), whereas the latter was predicted to assimilate into a single Japanese category but with a category-goodness difference in fit – or CG type (English /w/ and /r/ being considered as a good and poor exemplar of Japanese /w/, respectively). Based on these perceptual patterns, very good discrimination of English /w/-/j/ was expected and finally achieved (77% correct discrimination), which was, in turn, better than the good level of discrimination obtained for English /w/-/r/ (65% correct discrimination). By contrast, the lack of phonemic relevance of English /r/-/l/ in Japanese led to the supposition that both sounds in the contrasting pair would be assimilated into one Japanese L1 category, whereby both sounds would be equally perceived as poor exemplars of Japanese /r/. In that case, Japanese Ss discriminated

[26] In this case, each trial consisted of three stimuli, where A and B represented two different categories (presented as the first and third stimulus, respectively) and X might be the same category as either A or B.

English /r/-/l/ at lower correct rates (64%), corroborating the higher degree of perceptual difficulty as a result of an SC type assimilation.

As stated previously, PAM research largely examines subjects who had minimal or no prior exposure to the TL. Moreover, listeners have often been assessed in their home country and testing procedures have been carried out in their L1. However, Best and Strange (1992) suggest that L2 category formation (and hence a more accurate perception of non-native speech) might be aided by the listener's increased awareness of discrepancies between native and non-native sounds as a function of gaining experience in the L2.

Therefore, the Best and Strange (1992) study further aimed at determining the effect of L2 experience on Japanese NSs' perception of English approximant contrasts. In this case, L2 experience was understood as intensive English conversation instruction (8h–10h/week) and an LOR in the US between 18 and 48 months. Although a larger amount of L2 experience did not fully confirm the authors' hypothesis about category formation (i.e. neither the 4 experienced nor the 5 inexperienced Japanese Ss performed within the NS range), an increase in experience did result in significantly higher correct discrimination scores for experienced Japanese on contrasts assimilated as CG and SC type patterns. In those instances as well, the experienced Japanese Ss performed on a more similar basis to that of English NSs. In contrast, L2 experience effects were not observed for TC patterns, irrespective of NNSs' amount of exposure to the TL.

Recently, Best, Traill, Carter, Harrison, and Faber's (2003) examination of two !Xóõ contrasts – that is, [ʘx]-[|x] and [!x]-[ǂx] – as perceived on an AXB discrimination task and categorisation task by 16 English NSs and 13 Ss of each Isizulu and Sesotho[27] has further corroborated PAM's predictions about SC, TC, CG, and NA assimilation patterns, as well as providing evidence for UC and UU type patterns. Specifically, English listeners assimilated both click contrasts into NA speech sounds, hence discriminating them at high correct rates (80% for [ʘx]-[|x] and 89% for [!x]-[ǂx]). Moreover, results obtained by this language group confirmed the hypothesis that they would outperform the two African language groups on the discrimination scores for [!x]-[ǂx], based on the prediction that Isizulu and Sesotho NSs would display an SC assimilation pattern (as opposed to English Ss' NA pattern). Finally, the UC and UU

[27] Among the various features of these two African languages described, Best, Traill, et al. (2003) indicate that both Isizulu and Sesotho contain click consonants in their phonetic inventory.

assimilation patterns were revealed in Isizulu and Sesotho Ss' perception of [ʘx]-[ǀx], leading to a somewhat better discrimination level on the part of Isizulu listeners (81% vs. 75%), which was consistent with the hypothesis of the PAM.

Additional support for the PAM comes from Polka's (1992) assessment of the effects of phonemic status and phonetic familiarity on English and Farsi listeners' perception of non-native Salish place contrasts through a same-different (AX) discrimination task and an identification (ID) test. In spite of corroborating assimilation patterns of Salish /k'/-/q'/ (Experiments 1 and 2) and Farsi /ɢ/-/g/ (Experiment 2), several unpredicted results arose as to the expected influence of the factors of phonemic status and phonetic familiarity. Regarding phonemic status, Farsi NSs did not perceive the Salish contrast at better rates than English NSs, and to a certain extent, they seemed to have more difficulty in successful discrimination. Besides, noticeable individual differences were observed in both groups of listeners. With reference to phonetic familiarity effects, English listeners did not systematically behave as expected in their discrimination of Salish /k'/-/q'/ and "more familiar" Farsi /ɢ/-/g/. Moreover, other factors appeared to interact with phonetic familiarity. Such was the case of the order of testing procedures which resulted in significant differences in the perception of Salish and Farsi contrasts only in the group that was first administered with Farsi and then Salish contrasts (i.e. their performance on the Salish contrast was worse).

Even though research on non-native vowel contrast perception is so far more limited, findings are in line with PAM's predictions. Particularly, the results show a "strong association between individual listeners' actual perceptual assimilation patterns and their level of discrimination" (Best, 1995, p. 197). Take, for instance, Best, Faber, and Levitt (in preparation; as cited in Best, 1995) who examined English NSs' perception of two Norwegian vowel contrasts, as well as one Thai and three French vowel contrasts. The vowel contrasts in question were assimilated as ranging from TC – CG to UC – SC types, which revealed various degrees of accurate discrimination accordingly. In other words, better discrimination was reported for TC assimilation patterns than for CG assimilations, which, in turn, were better discerned than SC assimilations.

Lately, Best, Halle, Bohn, and Faber (2003) have looked at three language groups' perception of non-native vowel contrasts in an effort to extend (and specify) perceptual assimilation patterns to naive listeners other than English NSs. For that matter, the performance on an AXB discrimination task and a categorisation task comprising

Norwegian /i/-/y/, /y/-/u/, /y/-/ʉ/, and /ʉ/-/u/ by 16 English and Danish NSs and 24 French speakers was examined. As expected, Danish Ss correctly discriminated each of the four vowel contrasts in virtually all instances. On the other hand, English and French listeners obtained higher discrimination scores for three vowel contrasts, while they displayed a poorer rate of discrimination accuracy on Norwegian /i/-/y/. The latter constituted evidence for a prior assumption that a CG assimilation pattern would surface for this vowel contrast: in both cases, Norwegian /i/ and /y/ were hypothesised to assimilate to each English and French /i/ as imperfect fits.

It should also be mentioned that one prediction of the PAM was not fully supported in that Danish NSs excelled in their discrimination of Norwegian /y/-/ʉ/, when poorer discrimination levels had been ventured according to the CG assimilation pattern. Rather than falsify the PAM's hypothesis, Best, Halle, et al. (2003) take this unexpected finding as indicative of their continued search for objective means that can predict perceptual assimilation in a distinct manner.

In sum, research on both non-native consonant and vowel contrasts has for the most part confirmed the predictions of the Perceptual Assimilation Model (cf. Best, Halle, et al., 2003; Polka, 1992).

It is worth noting that many of the findings reported within the PAM framework have also been interpreted in light of other speech perception models (and vice versa) in an effort to highlight the validity and generalisability of the given model over the hypotheses made by the remaining models. Of special interest is the comparison between the PAM and the SLM, as both models have been considered to complement each other (Strange, 1995a; see above, as well).

At a more theoretical level, both models assume that L1 experience influences L2 speech perception (especially perception of TL segments). In addition, the notion of perceived phonetic similarity is taken as the starting point upon which the various hypotheses and postulates are based (Strange, 1995a; see also, Major, 2001). However, the SLM and PAM differ in their definition of phonetic similarity from an acoustic and articulatory/gestural point of view, respectively; yet again the two models agree with the need for devising an objective means that characterises phonetic similarity from either perspective (e.g. Best, 1995; Flege, 1995a).

> According to PAM, learners perceive the gestures used to form sounds in the L2. For the SLM, the objects of cross-language perception are vowel and consonant

> segments as perceived via a set of phonetically relevant features. Both models require empirical evidence bearing on the perceived relation, if any, of phonic elements in the L2/FL and those in the L1 before they can generate specific predictions. (Flege, Bohn, & Jang, 1997, p. 443, footnote 4)

Moreover, the PAM and the SLM confer L2 experience – whether in the form of formal training or immersion/naturalistic exposure to the TL – an important role in learners' perception of L2 sounds in contrast to L1 sounds and in the attainment of native-like perception of TL sounds (or, in its absence, in more accurate L2 sound perception). Even so, experience effects have to date been examined to a lesser extent in the PAM.

> The results of feedback training experiments have suggested that language-specific perceptual patterns are modifiable to some extent. This suggests that the perceived relation of L1 and L2 sounds may change during naturalistic L2 learning. (Flege, 1995a, p. 237)

> [I]ncreased L2 experience may foster improved recognition of the *discrepancies* between L1 and L2 phones. This could lead to a decline in degree of assimilation of L2 phones to L1 categories, and perhaps ultimately to the emergence of a separate L2 phoneme category due to improved recognition of phonetic properties within the L2 phonological system. (Best & Strange, 1992, p. 307)

Category formation of L2 sounds is also addressed by the two models, though the circumstances under which category formation might be triggered vary from one model to the other. Thus, according to the SLM, the more dissimilar an L2 sound is perceived to be from its closest L1 sound, the greater the likelihood for a new phonetic category to be formed. In Flege's (2003) words, "the SLM predicts that the greater is the perceived phonetic dissimilarity of an L2 speech sound from the closest L1 sound, the more likely it is that a new category will be created for the L2 sound" (p. 10). By contrast, the PAM hypothesises that for L2 category formation to take place an L2 sound should be perceived as "discrepant" of an L1 phonetic category rather than "dissimilar" or "distant".

> [A]dult L2 learners should be expected to form new phonetic categories most readily for L2 phones perceived as discrepant exemplars of a native category, i.e., for the non-prototypical member of a contrast that is assimilated as a category goodness difference from a native phoneme. If no discrepancies are perceived between the L2 sound and L1 phone – that is, for the L2 phone that is perceived as a good exemplar of the native phoneme – it should be quite difficult for the L2 learner to form a new category. Conversely, if the L2 phone is so *dis*similar from L1 phonemes that it cannot readily be related to any L1 category, we may expect the L2 learner to have some difficulty forming a new phonetic category, because a

> clear *contrast* between a specific familiar phoneme and an unfamiliar phone may be particularly informative to the learner. (Best & Strange, 1992, p. 327)

L2 sound category formation is relevant to the SLM and PAM, as well, for it shapes the models' hypotheses about L2 speech learning. So, in the SLM, if new categories are formed for L2 sounds, learners are expected to perceive and produce L2 sounds accurately (or at higher correct levels) than learners who have not established an additional phonetic category (Flege, 1995a). On the other hand, category formation does not appear to be crucial to successful discrimination of L2 sounds from the perspective of the PAM. As Flege (2003) indicates,

> The Perceptual Assimilation Model … proposes that the accuracy with which L2 speech sounds are discriminated will depend on how, or if, they are perceptually "assimilated" by L1 speech sounds. … However, it appears that most L2 speech sounds are perceptually assimilated by an L1 category, at least initially. That being the case, L2 speech sounds will be discriminated more accurately if they are assimilated by two distinct L1 speech sounds than if they are assimilated by a single L1 speech sound category. (p. 5)

The models are further distinguished by their predictions about production of non-native sounds. While the SLM and the PAM assume that perception precedes production (e.g. Best, 1999; Flege, 2003) and both "take the view that many L2 production errors have a perceptual origin" (Flege, Bohn, et al., 1997, p. 443, footnote 4), the PAM does not make any specific claim as to the production of non-native segments (or, in this case, non-native contrasts) based on the perceptual assimilation patterns. Moreover, it was noted earlier that the SLM's hypotheses focus on the perception and production of individual L2 segments, whereas those of the PAM are concerned with the perception of non-native sound contrasts.

In relation to actual studies conducted to give support to each model's predictions, two different aspects are worth mentioning. Firstly, (as mentioned above) the subject population consists of naive infant and adult listeners in the PAM in contrast to highly experienced L2 learners and speakers who have reached their ultimate attainment in the L2 – and with a wider range of AOL – in the SLM. Secondly, it is often the case that English is the subjects' L1 in PAM studies, while English is the TL/L2 in most SLM research.

To finish, both frameworks concur with a number of unresolved questions such as how to characterise phonetic similarity objectively, how to typify training effects on TL sound perception and production both in the short and long term, and how to account for

the emergence of individual differences that do not comply with the hypotheses of either of the two models. All of the above contributes to highlighting the authors' consideration of the SLM and the PAM as "working models", which might be modified as new findings come to the fore (Best, 1995; Flege, 1995a).

2.3.1.3. The Native Language Magnet Model (NLM)

The development of the Native Language Magnet (NLM) model or theory (Kuhl, 1993; Kuhl & Iverson, 1995) was motivated by the finding of infants' reorganisation of speech perception from original language-general phonetic perception to language-specific phonetic perception by 10–12 months of age (e.g. Werker & Tees, 1984; as cited in Kuhl, 1993). Therefore, the primary aim of NLM[28] is to account for the manner in which infants reorganise speech perception over their first year of life by means of the "perceptual magnet effect" (PME) mechanism (Kuhl & Iverson, 1995). In addition, Kuhl (1993) illustrates how NLM might be applied to L2 acquisition, and thus offer an account of adult L2 learners' degrees of perceptual difficulty in their discernment of TL sounds.

As for its main objective, NLM claims that infants are born with innate abilities related to general auditory perception mechanisms. These innate abilities are composed of the capability of dividing "the sound stream into gross categories separated by natural boundaries" (Kuhl & Iverson, 1995, p. 139), and allow infants to perceive all the world's language sounds successfully. Moreover, the model suggests that, as a function of the surrounding language(s), by 6 months of age infants have started forming representations of L1 sounds and show magnet effects, which derive from their analysis of linguistic input.

> [NLM] holds that infants are born with language-relevant abilities that are innate and attributable to general mechanisms of auditory perception. … To this innate component, the model adds a component that derives from language experience. The model holds that by six months of age infants develop stored representations of speech information, based on their perception of ambient language input. These representations of native-language sounds constitute the beginnings of language-

[28] When referring to the Native Language Magnet Model in abbreviated form, a consistent notation appears to be lacking – e.g. "development of *the NLM theory*" (Kuhl & Iverson, 1995, p. 122), "*NLM* holds that" and "*NLM theory* holds that" (Kuhl & Iverson, 1995, p. 141). Particularly, in the "NLM" notation both "magnet" and "model" seem to be implicitly included in "M". Thus, in the present dissertation NLM will stand for "Native Language Magnet Model", unless otherwise specified (i.e. NLM model).

> specific speech perception and play a critical role in infants' perception of native- and foreign-language sounds. (Kuhl, 1993, p. 128)

Linguistic experience then is considered to play a crucial part in speech perception, for it tunes infants' perception to those features that are only relevant to the L1. As Kuhl and Iverson (1995) state, "learning a primary language results in alterations of the underlying perceptual mechanisms that affect the processing of language from that time forward" (p. 122).

The changes (or alterations) in perceptual mechanisms resulting from exposure to a given language should be interpreted within the prototype theory (Rosch, 1978; as cited in Kuhl & Iverson, 1995). In the case of NLM, Kuhl and Iverson (1995) define a prototype as "a good instance of a [phonetic] category" (p. 123). Prototypes have been found to be pertinent to speech perception in that they function "as perceptual magnets for other sounds in the category" (p. 123). That is, a given prototype (P) of a phonetic category acts as a magnet attracting sounds in a surrounding acoustic space to itself. Besides, this magnet function involves a reduction or shrinking of the perceptual distance between the prototype and more remote sounds (i.e. sound variations which belong to the same L1 category). Conversely, poorer instances of the same category or nonprototypes (NP) are not subject to the same magnet effects of the prototype in the sense that they are not pulled towards the magnet. Instead, the perceptual space or distance between the prototype and the nonprototypes is enlarged.

The existence of prototypes of L1 phonetic categories was demonstrated in a series of adult studies conducted by Kuhl and her colleagues, and, especially, in investigations focusing on the English vowel /i/ (Grieser & Kuhl, 1989; Kuhl, Williams, & Meltzoff, 1991a; as cited in Kuhl & Iverson, 1995). Based on the finding that adult English NSs succeeded in identifying the prototypes of English /i/ on a 7-point rating scale, a further study was undertaken to determine the hypothesised magnet effect displayed by prototypes. In that case, both adults' and 6-month-old infants' perception of P /i/ and NP /i/, together with their corresponding 32 acoustically-modified variants, confirmed the hypothesis that P /i/ would be perceived as more similar to its variants than NP /i/.

What is more, the presence of magnet effects in 6-month-old infants were subsequently investigated and found to be dependent on language experience. To be

exact, Kuhl, Williams, Lacerda, Stevens, and Lindblom (1992; as cited in Kuhl, 1993) showed that English and Swedish infants' perception of English /i/ and Swedish /y/ revealed a stronger magnet effect only for their respective L1 prototype; further corroborating the premise that perception is altered as a result of language experience.

Additional evidence for the influence of linguistic experience on L1 speech perception comes from a number of studies on Motherese[29], where infants have been reported to rate vowels in Motherese talk as better exemplars than the same vowels included in speech addressed to adults (see also Liu, Tsao, & Kuhl, 2003).

As mentioned above, the magnet effects displayed by prototypes of phonetic categories entail a distortion or "warping" of the perceptual space in the areas of the prototype. More precisely, it is held that the perceptual distance between the prototype and the best instances shrinks, while it is the reverse when the worst instances are considered. This implication has found empirical support in several studies Kuhl and her colleagues have conducted on English adult speakers' perception of L1 vowel sounds (/i/ and /e/) by means of employing an objective measure of sensitivity (d′) and multidimensional scaling (MDS) techniques (for a review of those studies, see Kuhl & Iverson, 1995). More recently, research on English and Mandarin adult NSs' perception of English and Mandarin fricative and affricate consonants (Tsao, Liu, & Kuhl, 2003) has lent further support to the perceptual magnet effects that NLM predicts.

Despite the fact that the NLM account still presents some unresolved issues such as whether magnet effects exist from birth or develop with exposure to a particular language, and whether speech representations are, in fact, created (and, if so, what form they take); Kuhl (1993) argues that the PME mechanism outlined above does indeed provide an explanation for infants' reorganisation of speech perception by 10–12 months of age found in the Werker and Tees (1984) study. Thus, once language-specific magnets (or prototypes or categories) have reorganised the innate natural boundaries for sounds, infants will no longer be able to discriminate FL sounds with the same degree of accuracy as they had previously discerned them. In Kuhl's (1993) words, since "magnets perceptually pull other instances towards the prototype, [t]his will cause certain perceptual distinctions to be minimised (those near the magnets themselves), while others are maximised (those far from a native language magnet)" (p. 130).

[29] Kuhl and Iverson (1995) define Motherese as "the special speech style used by caregivers when addressing infants" (p. 130).

Regarding L2 acquisition, the premises of the NLM about perceptual magnet effects and their subsequent distortion of the perceptual space, as well linguistic experience effects on speech perception, might account for not only infants' difficulty in perceiving certain FL sounds, but also that of adults. More importantly, NLM assumes that adults (and older adolescents) retain their phonetic abilities that foster L1 speech acquisition, which is also in line with the basic claim of the SLM and the PAM (Bongaerts, 2003; Flege, 2003; see above as well). Recently, Wang and Kuhl (2003) have further demonstrated this assumption.

The NLM model then hypothesises that FL sound contrasts will be difficult to discriminate if they are very similar or close to a prototype of an L1 phonetic category. Like the SLM and the PAM, phonetic similarity – here operationalised as "the proximity principle" – will determine the degree of difficulty in the discernment of two TL sounds. It should be noted, as well, that although the proximity principle is not defined from a gestural point of view, it is similar to the PAM's interpretation of phonetic similarity and its derived perceptual patterns. In fact, Kuhl (1993) observes that "NLM thus provides a model that is compatible with Best's (in press) levels of relative difficulty in discriminating various foreign-language contrasts" (p. 131).

Furthermore, the NLM account suggests that extensive training might lead to an improvement in adults' performance when asked to perceive FL sound contrasts (Kuhl & Iverson, 1995), which is consistent with the importance the SLM and the PAM assign to phonetic training in the attainment of accurate L2 sound perception.

Additionally, NLM considers that both perception and production are influenced by early linguistic experience.

> NLM argues that exposure to language early in life produces a change in perceived distances in the acoustic space underlying phonetic distinctions, and this subsequently alters both the perception of spoken language and its production. (Kuhl & Iverson, 1995, p. 122)

However, just as the PAM illustrated, and unlike the SLM, studies within the NLM framework mainly centre on sound perception. Similarly, the subjects examined in NLM research are naive listeners.

Finally, and as opposed to both the PAM and the SLM, the NLM is primarily concerned with L1 speech perception during the child's first year of life. Nonetheless, recent articles on L2 phonological learning (e.g. Bongaerts, 2003; Flege, 2003) and current research conducted from either the PAM or the SLM perspective (e.g. Best,

Halle, et al., 2003; Best, Traill, et al., 2003; Guion, Flege, Akahane-Yamada, & Pruitt, 2000) contrast the findings obtained in light of the hypotheses of the three models[30] in an effort to fully depict L2 speech learning.

2.3.1.4. Main similarities and differences between the SLM, the PAM, and the NLM

The two tables below summarise, on the one hand, a number of similarities between the SLM, the PAM, and the NLM; and, on the other hand, several differences between the three models discussed in Sections 2.3.1.1 through 2.3.1.3.

Table I. Some similarities between the SLM, the PAM, and the NLM

	SLM	**PAM**	**NLM**
Similarities	(1) Phonetic learning abilities underlying L1 speech acquisition are accessible during L2 acquisition. (2) Phonetic similarity constitutes the basis from which the models' hypotheses derive. (3) Category formation of L2 sounds shapes predictions about L2 sound perception and production. (4) L1 experience influences L2 speech perception. (5) L2 experience might lead to better L2 sound perception. (6) Perception is generally thought to precede production of TL segments. (7) Incipient findings suggest beneficial (phonetic) training effects on more accurate L2 sound perception. (8) Further research is still in need in order to (a) typify phonetic similarity through objective means, and (b) account for individual subject differences that do not conform to the models' hypotheses.		

[30] Other suggested models or accounts have been included for comparison purposes, particularly in the PAM studies, though to a lesser extent. See, for example, Strange's (1995a) review of current theories of L2 speech perception that includes Pisoni's (1995) Exemplar-Based Model; and Best, Halle, et al. (2003) and Best, Traill, et al. (2003) studies which make reference to Burnham's (1986) Fragile-Robust Hypothesis.

Table II. Some differences between the SLM, the PAM, and the NLM

Differences	**SLM**	**PAM**	**NLM**
(1) Definition of phonetic similarity	Auditory-acoustic viewpoint	Articulatory/gestural viewpoint	Auditory-acoustic viewpoint
(2) Scenarios triggering L2 sound category formation	New additional category more likely to be formed, the greater the perceived phonetic dissimilarity of an L2 sound from its closest L1 sound.	New L2 phonetic category predicted to be formed if L2 phones are perceived as discrepant (not dissimilar) exemplars of an L1 category.	Not addressed explicitly concerning L2 acquisition, rather L1 acquisition. However, NLM claims that a new sound closer to an L1 magnet will be more assimilated by L1 category, and thus the new sound will become undistinguishable from the L1 sound.
(3) Effects of category formation on L2 speech perception	L2 sound category formation is a requirement for successful perception of TL sounds.	In case an L2 sound category is not formed, accurate perception of TL sounds might still be possible based on perceptual assimilation of L2 sounds to L1 sounds (as CG type, for example).	As above
(4) Hypotheses about production of TL sounds	L2 segments will be produced accurately only if learners perceive L2 sounds in a native-like way. However, not all production errors are perceptually-based.	Not addressed in the PAM, except for a brief reference to production of TL sounds (Best, 1999).	Not addressed in the NLM, except for a short reference to production of TL sounds.
(5) Focus of study	Perception and production of TL segments	Perception of TL sound contrasts	Perception of L1 prototypes and TL segments
(6) "Status" of English	TL under study	Subjects' L1 mostly	Subjects' L1 mostly
(7) Subject population	Highly experienced L2 learners and learners who have reached ultimate attainment in L2	Naive listeners	Naive listeners
(8) Age of first exposure to the TL	From very early childhood (through adolescence) to adulthood	Infants and adults	Infants and adults

To summarise, Section 2.3.1 has reviewed three alternative accounts – the SLM, the PAM, and the NLM – to the neurophysiological maturation explanation on child-adult differences in L2 speech learning, of which the SLM was discussed in by far the greatest detail. The lengthy consideration of Flege's Speech Learning Model might be justified on the grounds that the SLM is concerned with both perception and production of English TL sounds and that a wealthy body of research has proven most of the SLM's hypotheses. In addition, Flege's model has often been adopted as the theoretical framework within which studies in rather formal learning contexts have been conducted, in spite of the SLM's focus on subjects who learned the L2 largely in naturalistic settings. In fact, no other specific account that addresses learners' TL phonological acquisition in a strictly formal setting has to date been put forth. For all of the above reasons, most of the findings of the present dissertation, in conjunction with the formulation of research questions, will be interpreted in light of the SLM.

2.3.2. Perception and production of TL segments: Focus on NSs of Romance languages

The present section aims to summarise the main findings reported on the perception and production of English segments by non-native speakers. A special emphasis will be given to literature on native speakers of Romance languages who have learned English, as those results are presumed to be most relevant to the populations studied in the present research. This section supplements previous discussion of the many studies that are relevant to the theoretical frameworks outlined in 2.3.1 – in particular, the SLM. Consequently, those studies will be mentioned briefly in this section.

As far as perception and production of English consonants are concerned, two phonetic features in which English and Romance languages differ have been extensively researched, namely, VOT in voiceless (and voiced) stops in word-initial position and the voicing contrast (between stops, primarily) in word-final position.

VOT studies have shown that early L2 learners, with AOLs ranging from 2 to 7 years, are generally capable of partitioning synthetic /ba/-/pa/ and /da/-/ta/ continua differing in VOT values, and as delivered in identification and discrimination tasks, at identical or similar boundaries to those of English monolinguals (e.g. Spanish-English bilinguals in Bohn & Flege, 1993; Flege & Eefting, 1987; Flege & Schmidt, 1996; and

Flege, Schmidt, & Wharton, 1996). Moreover, the same learners have produced /p, t, k/ within native English range. For example, Schmidt and Flege (1996) stated that the VOT values for /p/ and /t/ in initial-sentence position at a normal speaking rate were 47 ms and 57 ms, respectively, for English monolinguals; and 48 ms and 61 ms for early bilinguals.

In contrast, late L2 learners, especially those with an AOL greater than 13 years, often fail to perceive and produce VOT in English /p, t, k, b, d, g/ at native-like levels, even in those instances where Ss had benefited from long-term exposure to English. For instance, Flege, Munro, and MacKay (1995b) found that, after a mean LOR of 32 years in Canada, native Italian subjects with AOLs of 11 to 21 years and 17 to 21 years still produced /t/ and /p/, respectively, with significantly shorter VOT values than English NSs'. Moreover, the Flege et al. (1995b) study provided evidence on adult learners' production of /p, t, k/ and perception/identification of category boundaries at VOT values that were intermediate to those of the subjects' L1 and English monolinguals, which, in turn, was consistent with early findings of adult French and Spanish NSs' production of VOT values for /p, t, k/ halfway between French and Spanish, respectively, and English (for French, see, e.g., Caramazza et al., 1973; cited in Flege & Schmidt, 1995; for Spanish, see Flege & Eefting, 1987; Williams, 1977, 1979, among others).

Finally, noticeable individual differences have been observed in adult Romance language learners of English, resulting in some late learners' successful production of VOT values for English /p, t, k/. In addition to Italian Ss in Flege et al. (1995b), several adult Spanish and Catalan NSs appeared to perform at English native-like levels, as reported by Schmidt and Flege (1996) and Rallo (1999), respectively. To be exact, Schmidt and Flege (1996) reported that three Spanish late learners of English (with an AOA between 16 and 28 years) produced English /p, t/ with native-like VOT values, while the remaining seven Ss in the late learner group produced shorter VOT values for /p, t/ than did English monolinguals. Rallo's (1999) study extended the finding of adult individual differences to a formal learning environment. Thus, she reported on two adult Catalan NSs' near-native production of English /t/ VOT under varying speaking rate conditions. Conversely, another three Catalan Ss produced English /t/ with VOT values typical of their L1. Besides, the finding of one adult learner's inconsistent usage of

Catalan and English phonetic norms in English /t/ production contributed to emphasising the existence of noticeable individual differences in late L2 starters[31].

Lastly, as stated in 2.3.1.1, VOT in English voiceless stops, produced by L2 learners at slow, normal, and fast speaking rates, has been further examined in finer-grained studies testing for category formation of L2 sounds (for main results, see Flege & Schmidt, 1995; and Schmidt & Flege, 1996, above; see also Magloire & Green, 1999).

To conclude this brief review of VOT investigations, a more recent study by MacKay, Flege, Piske, and Schirru (2001) might be taken as a summary of the main findings of VOT research with NSs of Romance languages. That is, the performance by 72 native Italian Ss with an LOR of at least 28 years, but differing in AOL (7–8 years vs. 18–20 years) and L1 use (low use: 6%–10%; high use: 40%–50%), suggests that the learning of English /p, t, k, b, d, g/ as L2 sounds is plausible, irrespective of learners' establishment of new phonetic categories. However, failure to form additional phonetic categories will likely result in learners' approximation to English phonetic norms only. Moreover, NNSs' chances to perceive and produce VOT in a similar manner to English monolinguals will increase if learners start acquiring the L2 at an early age, provided that they have been extensively exposed to NS input, and, where applicable (i.e. immersion settings), they hardly ever use their L1.

With reference to the distinctive feature of voicing, the difficulty shown by Romance language NSs in perceiving and producing English consonant voicing in word-final position has often been attributed either to the lack of word-final consonants in their L1(s) (e.g. Spanish and Italian) or to the non-occurrence of voiced consonants – especially stops – in absolute word-final position (e.g. Catalan)[32]. In line with these assumptions, Flege and Davidian (1984) found that adult Spanish learners of English with an average AOA in the US of 19.9 years and LOR of 7 years tended to delete word-

[31] The same pattern of individual differences in adults' L2 performance on VOT stop perception and production arises when the TL under examination is a Romance language and the learners are English NSs. For example, Major (1992, 1997) looked at five American English speakers who were immigrants to Brazil in adulthood. Ss' production of both Portuguese and English VOT in /p, t, k/ varied from each other. So, two Ss appeared to maintain native L1 pronunciation proficiency while they did not achieve native L2 pronunciation proficiency. On the other hand, one subject achieved native-like proficiency in the L2 phonology, though the subject lost native L1 pronunciation proficiency. Lastly, two other subjects did not perform at native-like levels in either their L1 English or the L2 Portuguese phonology.

What is more, just as NSs of Romance languages learning English, Flege and Hillenbrand (1984), among others, illustrated that English NSs' perception and production of Romance language /p, t, k/ displayed intermediate VOT values between English and French.

[32] See Yavaş (1994) for a summary of some studies on English consonant devoicing by speakers of different languages, apart from Romance language speakers.

final stops when asked to produce /p-b/, /t-d/, and /k-g/ contrasts in CVC syllables both in isolation and in the carrier phrase "*Now I say* ____". However, in the event that stops were produced, Spanish subjects either devoiced or fricativised voiced stops (English /b, d, g/) in 43% or 19% of instances, respectively. In fact, when taken together, results revealed Spanish NSs' preference of devoicing of English final stops to deletion and fricativisation, which was interpreted as an indication of L2 learning.

On a similar basis, a larger amount of experience in English appears to enhance adult Spanish L2 learners' production of English /t/-/d/ contrast in word-final position. To be exact, ten Spanish adults with 9 years of experience in English approximated the English-like production of /t/-/d/ opposition in CVC words at the end of a carrier phrase. Although they did not attain native-like production levels for English /t/-/d/, they performed more accurately than ten Spanish late L2 learners with a significantly lower amount of experience in English (i.e. 0.4 years) (Flege, Munro, & Skelton, 1992).

A further study by Flege, Munro, and MacKay (1995b) examining highly experienced Italian L2 learners – that is, learners with an average LOR in Canada of 32 years – showed that those native Italian Ss with an AOL of 3 to 13 years produced both English final /t/ and /d/ tokens accurately, while those Ss with an AOL of 15 to 21 years failed to produce the /t/-/d/ contrast within the native English range. Additionally, late L2 learners' performance on English /p-b/ and /k-g/ contrasts differed significantly from that of English NSs (learners' AOL of 19 to 21 years, and 17 to 21 years, respectively). In spite of this, and as was the case of VOT studies, Flege et al. (1995b) comment on the fact that several individual late L2 learners produced the voicing contrast between /p/ and /b/, /t/ and /d/, and /k/ and /g/ in word-final position at native-like rates.

Similar results to those stated above about Romance language NSs exposed to English in an L2 naturalistic setting have been obtained in formal learning contexts. One such study is Cebrian (2000), who examined the production of voicing contrasts in English by 12 Catalan NSs (undergraduate students of English). The English test words ended in either /p-b/, /t-d/, /k-g/, /f-v/, /θ-ð/, /s-z/, /ʃ-ʒ/, or /tʃ-dʒ/, and in four different environments: in absolute final position, before a voiceless consonant, before a voiced consonant, and in prevocalic position. As expected, Catalan Ss produced unvoiced consonants in absolute final position with 100% accuracy (based on the fact that Catalan

possesses unvoiced consonants in word-final position), while devoicing all English voiced consonants /b, d, g, v, ð, z, ʒ, dʒ/ between 92% and 100% of the time (based on their L1 neutralisation rule). On the other hand, Cebrian (2000) noted that when the segments under study were produced before a voiced consonant or a vowel, Ss exhibited a variety of response patterns. Despite this, it was found that for the most part Catalan NSs continued devoicing /b, d, g/ in those two phonetic contexts; hence further supporting learners' application of their L1 neutralisation rule to English.

Apart from the features of VOT and voicing contrast, investigations on Romance language NSs have also looked at perception and production of English consonants that do not exist in the Ss' L1 inventories in initial, medial, and final positions. For example, English /ð/ and /θ/ production in word-initial position by Italian NSs with an AOL of 11 to 21 years has been found to differ from English NSs, regardless of their long-term naturalistic exposure (i.e. 32 years on average) (Flege et al., 1995b). With a more limited amount of experience in English, 20 Catalan learners of English in a formal setting failed to produce /ð/ in initial position at native-like levels, though their production of medial /ð/ was significantly better (based on L1 production rules probably) (Cortés, 2002, 2003). It is interesting to note that the common substitute for /ð/ in both Italian and Catalan L1 groups was English /d/.

Finally, in the comparison of perception and production of English consonants in initial position vs. word-final position (e.g. Flege et al., 1995b), both early and late L2 learners are more successful at word-final consonants than word-initial consonant segments, even performing very similarly to English NSs when consonants are presented under non-ideal (or noise) conditions (e.g. MacKay, Meador, & Flege, 2001).

When it comes to vowel perception and production, English front vowels – especially /i, ɪ, ɛ, æ/ – have been by far the segments mostly examined (e.g. for Catalan NSs: Cebrian, 2002a, 2002b, 2002c, 2003; for native Spanish Ss: Escudero, 2002; Flege, 1991a, 1992; Flege, Bohn, & Jang, 1997; and García Lecumberri & Cenoz, 1998; for Italian learners of English: Flege, 2002; Flege, MacKay, & Meador, 1999; and Munro, Flege, & MacKay, 1996). In addition, the fact that Romance languages lack the distinctive feature of tense/lax vowel contrasts has further contributed to the study of English front vowels, with special reference to nonnatives' discernment and implementation of English /i-ɪ/ contrast. Moreover, a common trait of English L2 vowel

perception and production research is its focus on late L2 learners' performance – i.e. learners with a minimum AOL of 12 years – both in L2 immersion and formal learning settings – except for the Italian studies conducted by Flege and his colleagues who have investigated early and late L2 starters (for Spanish/Basque younger and older bilinguals learning English within the L1 community, see also Gallardo, García Lecumberri, & Cenoz, 2002; and García Lecumberri & Gallardo, 2003).

Acoustic measurements of the formant values of English /i, ɪ, ɛ, æ/, Spanish /i, e, a/, and Catalan /i, e, ɛ/ suggest that English /i/ is very similar to Spanish and Catalan /i/ (Flege, 1991a; Cebrian, 2002a); English /ɪ/ falls in between the acoustic space of Spanish /e/ and /i/, and Catalan /e, ɛ, i/; English /ɛ/ might be identified primarily with Spanish /e/ or Catalan /ɛ/; whereas English /æ/ does not appear to have a direct or similar equivalent in Spanish and Catalan[33].

Consistent with the identification of English /i/ with Spanish and Catalan /i/ on the basis of acoustic measurements, adult Spanish and Catalan L2 learners have been reported to equate English /i/ with Spanish and Catalan /i/ at high correct frequencies ranging from 84%[34] to 99% (Flege, 1991a; Cebrian, 2002a, 2002b) as presented in one-syllable English and nonsense words comprising classification and perceptual assimilation tasks. Similarly, adult Catalan NSs mainly identified English /ɛ/ with Catalan /ɛ/ (78% of the time) and, to a lesser extent, with Catalan /e/ (22% of the time). However, Spanish Ss did not consistently identify English /ɛ/ with Spanish /e/ (range: 44%-48%), as Spanish /a/ was often equated with English /ɛ/ (range: 22%-39%). Likewise, both Catalan and Spanish participants exhibited various identification responses for English /ɪ/, in contrast to Spanish monolinguals, namely, Catalan /e/, /ɛ/ and /i/ (58%, 21%, and 18% of the time), and Spanish /e/, /i/, and "none[35]" (in 39%, 36%, and

[33] See also Mott (1991) for an acoustic comparison of English, Spanish, and Catalan vowels.

[34] Flege (1991a) also had 20 Spanish monolinguals label English vowels. In comparison to the Spanish learners of English included in the study, the Spanish monolinguals identified English vowels with Spanish vowels at higher rates. Thus, English /i/ was identified with Spanish /i/ in 94% of instances (vs. 84% for Spanish learners of English), English /ɪ/ with Spanish /i/ in 68% of instances, English /ɛ/ with Spanish /e/ 81% of the time, and English /æ/ with Spanish /a/ in 71% of tokens.

[35] As mentioned in 2.3.1.1, Ss in the Flege (1991a) study were provided with a "none" response option, in addition to orthographic letters <i>, <e>, <a>, <o>, <u>, to indicate that the vowel they had just heard was not equivalent to any Spanish vowel. The "none" option was not available in the perceptual assimilation

21% of instances). Last, according to the predictions derived from acoustic measurements on vowel formants, English /æ/ was surprisingly associated with Spanish /a/ in 82% of instances[36], and only regarded as a new sound in 18% of instances.

It is worth noting that in the studies above all Catalan Ss had an average LOR of 24.5 years in Canada, while Spanish Ss differed in LOR in the US, and thus in experience in English (6.8 years vs. 0.8 years). In spite of those differences, relatively experienced Spanish Ss obtained the same identification scores as relatively inexperienced Spanish Ss, and therefore failed to show a better discernment or identification of certain English vowels – i.e. /ɪ/ and /æ/ – as new L2 sounds along with an increase in English L2 experience.

Parallel nonsignificant effects of L2 experience were replicated in Flege, Munro, and Fox (1994) where 15 experienced adult Spanish learners (mean LOR of 7 years) did not perceive English vowels as being more dissimilar to Spanish vowels than 15 inexperienced adult Spanish subjects (mean LOR of 1.8 years).

Moreover, somewhat divergent findings from those above concerned with the identification of English vowels in terms of the learners' L1 vowels have been observed when adult Spanish and Catalan Ss were asked to produce and perceive English /i, ɪ, ɛ, æ/. For instance, Flege, Bohn, and Jang (1997) pointed out that Spanish late learners of English differing in experience (7.3 years vs. 0.7 years of LOR in the US) were judged to have produced English /i/ in one-syllable words only in 57% and 69% of instances (for experienced and inexperienced Ss, respectively), which is in opposition to Flege's (1991a) findings. Besides, the three native English listeners in the study identified learners' productions of English /ɛ/ as intended 99% (for experienced Ss) and 91% of the time (for inexperienced Ss), unlike Flege (1991a) and Cebrian (2002a, 2002b). Relatively high intelligibility scores on English /æ/ were also obtained for experienced and inexperienced Spanish speakers of English (73% and 70% correct productions), taking into account that in Flege (1991a) Spanish NSs had failed to perceive /æ/ as a completely different sound from any Spanish vowel. Therefore, it was only Ss' /ɪ/ productions that

task delivered in the Cebrian (2002a, 2002b) studies; hence the different resulting identifications by Spanish vs. Catalan NSs.

[36] Cebrian (2002a, 2002b, 2002c) did not address English /æ/ perception and production; instead, Canadian English /e^{I}/ was examined.

resembled the above interlingual identification scores, being rated as accurate in 61% and 51% of instances (for experienced and inexperienced Ss).

Furthermore, learners' failure to produce English /i/ and /ɪ/ accurately in the Flege, Bohn, and Jang (1997) study pinpoints Romance language NSs' well-attested difficulty perceiving and producing English tense/lax vowel contrasts such as /i/-/ɪ/ at native-like levels. Among the many investigations focusing on tense/lax vowel contrasts in English, all agree in ascribing late learners' failure to perceive English /i/-/ɪ/ successfully to their reliance on temporal cues vs. English monolinguals' reliance on both spectral (or quality) and temporal (or quantity) differences (e.g. Cebrian, 2002a, 2002b; Escudero, 2002; Fox, Flege, & Munro, 1995). By the same token, subjects are thought to make use of the same temporal cues on producing English /i/-/ɪ/ (e.g. Cebrian, 1999; as cited in Cebrian, 2002b), which, in turn, results in learners' common mispronunciation of English /ɪ/ as /i/.

In sum, late L2 learners in both naturalistic and formal learning settings rarely achieve native-like perception and production of English /i, ɪ, ɛ, æ/ (but see several Italian individuals in late low- and high-use groups who perceived /i/-/ɪ/ at native-like rates in Flege & MacKay, 2004). In addition, an increase in English experience does not always lead to a significantly better perception (and, as a consequence, production) of those English vowel segments, and particularly English /i/-/ɪ/ (cf. Flege, Bohn, & Jang, 1997). However, recent studies (e.g. Cebrian, 2003; García Lecumberri, 1999) suggest that a higher amount of experience in the form of explicit phonetic instruction appears to enhance perception of English tense/lax vowel contrasts by Catalan and Spanish NSs. As regards English /ɛ/, both Catalan and Spanish learners of English perceive and produce this vowel within the NS range. Finally, English /æ/ is often better perceived than produced, reaching native-like accuracy levels when presented in contrast to /ɛ/ (e.g. in /ɛ/-/æ/ vowel continuum as in Flege, Bohn, & Jang, 1997, vs. significantly lower discrimination scores when delivered in /æ/-/ʌ/ and /æ/-/ɑ/ pairs in Flege, Munro, & Fox, 1994).

A different set of findings emerges if early L2 learners are examined. As in the review of English consonant perception and production above, Flege et al.'s studies on

highly experienced Italian NSs of English demonstrate that mastery of English /i, ɪ, ɛ, æ/ is attainable if L2 learning starts at an early age. Take, for instance, Munro, Flege, and MacKay (1996) where subjects with a mean AOL of 3.1 to 11.6 years produced not only English /i, ɪ, ɛ, æ/ but also the remaining Canadian English vowel sounds accurately. Moreover, even Italian NSs with an AOL of 21.5 years managed to produce English segments with no equivalent counterpart in Italian – i.e. /ɪ, æ, ʊ, ɝ/[37] – between 70% and 83% of the time. Findings of early L2 learners' (maximum AOL: 7-8 years) native-like perception and/or production of English /i, ɪ, ɛ, æ/ are further corroborated in Flege (2002); Flege and MacKay (2004); Flege, MacKay, and Meador (1999); and Flege, Schirru, and MacKay (2003), among others. What is more, early bilinguals behave just like NE monolinguals in their discrimination of English tense/lax vowel contrasts, and even in those vowel oppositions that had proven to be problematic for Spanish NSs (e.g. /æ/-/ʌ/). Conversely, late L2 learners with an AOL greater than 14 years resemble adult Spanish and Catalan learners of English in that they fail to perceive tense/lax vowel contrasts of English /i/-/ɪ/ and /u/-/ʊ/ in a native-like fashion. Furthermore, Flege (2002), Flege, MacKay, and Meador (1999), as well as Flege, Schirru, and MacKay (2003), indicate that high L1 use (40%-50%) leads to Italian late L2 learners' significantly poorer performance on English vowel discrimination and production in comparison to those late starters who hardly ever resort to their L1 (6%-10%). The influence of L1 use is also noticeable in early L2 bilinguals as Flege and MacKay (2004) have recently reported. Therefore, while early-low (i.e. low L1 use) bilinguals discriminate English /i/-/ɪ/ like NSs of English in both an oddity task and discrimination task involving short phrases extemporaneously produced, early-high (i.e. high L1 use) bilinguals significantly differ from native English speakers in /i/-/ɪ/ discrimination.

Last, research on early L2 starters' perception and production of English vowels in a formal language setting is so far more limited (e.g. Gallardo, García Lecumberri, & Cenoz, 2002; García Lecumberri & Gallardo, 2003). In addition, the available evidence is not consistent with findings of Italian early L2 learners of English. For example, García Lecumberri and Gallardo (2003) have found that after 6–7 years of exposure to English as an FL, 11-year-old starters discriminated all English vowel sounds at higher correct

[37] As Munro, Flege, and MacKay (1996) point out, the only exception was /ʌ/ with a correct identification rate of 25%.

rates than 8- and 4-year-old beginners. However, as the authors state, these results should be taken with caution, for the older group did not perform at native-like levels. They further hypothesise that the late onset age advantage observed might have resulted from the lesser quantity of input delivered in FL formal settings than in L2 immersion environments. That is, García Lecumberri and Gallardo (2003) suggest[38] that the average 3-hour/week instruction in English that Spanish/Basque bilinguals have received over a period of 6–7 years is insufficient for an early start advantage to surface, contrary to what happens to early L2 starters immersed in the L2 community after one year of naturalistic exposure (e.g. Snow & Hoefnagel-Höhle, 1978/1982).

A final note on Romance language speakers of English addresses the relation between perception and production of TL sounds. Thus, irrespective of the language learning context, several studies reviewed above point out that the degree to which nonnatives perceive both English vowels and consonants accurately determines the way TL sounds will be produced (e.g. Cebrian, 2002a; Flege, Bohn, & Jang, 1997; Flege, MacKay, & Meador, 1999; Flege & Schmidt, 1995). This is consistent with Flege's (1999a, 1999b) observation about the existing relation between L2 sound production and perception, though he illustrates that only modest correlations between these two processes have been found so far (see also 2.3.1.1 above). Moreover, the finding of Romance language speakers' better perception of English sounds than their production in some studies (e.g. Cebrian, 2002b; Cortés, 2002) constitutes further evidence that "perception precedes production[39]" in L2 learning, which agrees with the SLM's predictions, as well (e.g. Flege, 1999b, 2003), and parallels L1 acquisition research findings (e.g. Whalen, 1999; Wode, 1995, 1999).

[38] The authors explicitly state that they base their suggestion on Singleton's (1995) estimates of amount of input delivered in an L2 immersion setting vs. FL instructed-classroom setting (see also 2.2 above).

[39] It should be mentioned that the view that accurate sound perception is attained before accurate sound production in L2 acquisition is not unanimous. For example, Llisterri's (1995) review of research on L2 sound perception and production relationships (see also Strange, 1995a) shows that in some instances production of L2 sounds has proven to be more native-like than their perception (e.g. Mack, 1989; Sheldon & Strange, 1982; all cited in Llisterri, 1995; see also Gass, 1984). Llisterri (1995) further notes that even when studies point to the precedence of perception over production "direct inferences about pronunciation accuracy can not probably be made from perceptual abilities in a straightforward manner" (p. 94); hence suggesting that perception and production are related in complex ways and influenced by other factors such as onset age of L2 learning, L2 experience, phonetic context, and training.

2.3.3. FA research

In addition to perception and production of L2 segments, a great deal of L2 phonological acquisition research has investigated foreign accent (FA) in order to determine whether L2 learners attain native-like proficiency in the TL phonology.

As mentioned earlier (see Footnote 2, Section 2.2), Munro (1998) defines FA (or foreign-accented speech, for that matter) as "nonpathological speech produced by … L2 learners that differs in partially systematic ways from the speech characteristic of native speakers of a given dialect" (p. 139). Besides, Major's (2001) depiction of FA includes the role of NS listeners in differentiating NS from NNS speech. According to Major, global FA then refers to the overall judgement NS listeners make about whether a speaker sounds native or not (and to what degree) in relation to the listeners' L1.

It is commonly held that FA results from deviations in TL segmentals, suprasegmentals, syllable structure, and voice quality (e.g. Flege, 1981, and above; Major, 2001; Munro, 1995, 1998). Moreover, Major (2001) indicates that failure to master just one of these levels will result in a foreign-accented pronunciation of the TL.

By the same token, NS listeners, whether "naive" or expert, are thought to be able to perceive FA in learners' productions in the TL ranging from long speech samples (e.g. three-minute excerpts in García Lecumberri & Gallardo, 2003) to "minimal" speech samples (e.g. release burst of English /t/ in Flege, 1984). What is more, non-native speakers appear to discern accented productions from accent-free productions in the TL (e.g. Flege, 1988b; cited in Flege, 1991b; Scheuer, 2002). Nonetheless, in the latter case, non-native listeners need have had considerable exposure to the TL to identify accented speech successfully (as illustrated by the five native Polish teachers of English in Scheuer, 2002).

With reference to NS listeners, the following observation deserves further attention. As was already noted in 2.2 above, Flege (1981) indicates that on rating NNSs' productions for degree of FA, NS judges do not deem segmental, suprasegmental, syllable structure, and voice quality deviations from NS phonetic norms as equally salient (see also Flege, 1984). In spite of this, few studies to date have been conducted to ascertain which types of differences contribute the most to NSs' perception of FA.

One example is Anderson-Hsieh, Johnson, and Koehler (1992) who looked at the relationship between accent ratings on 60 Ss' English speech productions and deviations

in segmentals, suprasegmentals (or prosody, in this case), and syllable structure found in the subjects' speech samples. Specifically, three experienced ESL teachers rated the productions of 60 subjects differing in L2 proficiency (high and low) and L1 (among them, Arabic, Chinese, German, Greek, Hindi, Korean, and Spanish) on a 7-point scale[40]. Their accent scores were then correlated to prosody scores (according to a 4-point scale from *0*, *least native-like*, to *3*, *most-native like*, for stress, rhythm, phrasing, intonation, and overall prosody), as well as to the deviations or errors the subjects made at segmental (i.e. sound substitutions and modifications in vowels and consonants) and syllable structure levels (i.e. vowel and consonant epenthesis, vowel/syllable deletion, consonant deletion, and metathesis). All correlational analyses yielded significant results, the relationship between divergences from TL prosody and accent scores showing the strongest correlation coefficients. Therefore, the significant relevance of prosody in the evaluation of FA corroborated previous findings of prosody as the main determinant for accent detection, in contrast to segmental and syllable structure errors (e.g. James, 1976; Johansson, 1978; cited in Anderson-Hsieh et al., 1992). In turn, deviance in segmentals and syllable structure was found to be less prominent in NSs' perception of FA, since the varying degrees of significance appeared to be subject to the speakers' L1.

In that respect, Magen (1998) centres exclusively on one L1 speaker group – Spanish – to assess the relative weighting that English NSs confer to segmental and suprasegmental deviations in two subjects' production of Spanish-accented English sentences. In her study, 10 native English listeners were asked to rate on a 7-point scale (*1 = closer to native English*, *7 = less close to native English*) 96 original sentences uttered by the speaker with the heaviest accent in English, as presented simultaneously in aural and written form, together with the corresponding acoustically edited productions where all deviations in segmentals (i.e. presence/absence of word-final /s/ or /z/, /tʃ/-/ʃ/ and intervocalic voicing /s/-/z/ distinction, stop voicing, vowel reduction/non-reduction in unstressed syllables, and tense/lax vowel contrast) and suprasegmentals (i.e. presence/absence of epenthetic /ə/ in word-initial and non-initial position[41], lexical and phrasal stress) had been modified to conform to English NS norms. As predicted, judges rated the edited productions as significantly less foreign-accented than the original

[40] Unlike most accent rating scales, the points on the scale differed by 0.5. Therefore, in that study the scale points ranged from *0* (*heavily accented and unintelligible speech*), *0.5*, *1*, *1.5* (*accented but intelligible speech*), *2*, *2.5*, to *3* (*near-native speech*).
[41] Note that Magen considers syllable structure phenomena as suprasegmentals.

productions, except for those acoustically edited samples containing stop voicing, vowel reduction, and /s/-/z/ alternation in intervocalic position. Besides, similar findings were obtained in the assessment of the remaining Spanish NS's production of English sentences (with a lesser degree of FA) by a different pool of listeners. Therefore, Magen concluded that the use of acoustic editing proved to be successful in determining that English NS listeners assigned more weight to suprasegmental deviations in the detection of FA, while other divergences from NS speech often regarded as indicative of FA, such as voicing differences (e.g. Flege & Hammond, 1984), failed to be significantly salient in FA perception. Yet again Magen noted that the findings of the study might only apply to Spanish L1 speakers. Furthermore, orthographic interference seemed to have played a part in the rating task; hence the lack of significance in /s/-/z/ alternation in intervocalic position and vowel reduction might be the result of orthography.

To sum up, the study of the relative weighting of segmentals, suprasegmentals, syllable structure, and voice quality in accent detection suggests that suprasegmentals[42] exert a stronger influence on NSs' FA perception, though research is still scarce (and thus inconclusive as to the exact contribution of the various phenomena to accent perception), as Piske, MacKay, and Flege (2001) have noted (and already indicated in Flege, 1981). Additionally, in spite of the salience of prosody, Piske et al. (2001; also Piske, Flege, & MacKay, 2002) point to the existence of numerous FA studies on segmental perception and production, in opposition to the reduced number of investigations addressing the role of prosody (whether alone or in comparison to segmentals)[43] in the assessment of global foreign accent.

As to the factors affecting degree of global FA, Major's (1987) earlier review of FA studies identified age of L2 learning, L1 interference, developmental factors, and style as potential variables leading to FA. Recently, Piske et al. (2001, 2002) have offered an updated, thorough review of FA research. Since that is a very complete state of the art already, the content about factors likely to predict degree of FA in the following paragraphs is basically a digest of that review article[44]. In some instances, though, supplementary comments have been incorporated.

---

[42] For the salience of prosody in FA perception, see also Missaglia (1999) and Wayland (1997), as reported by Piske et al. (2001), in addition to Munro (1995).

[43] Scheuer's (2002) study of the weighting of various segmental parameters in FA detection is a further example of the emphasis on segmentals placed by FA research.

[44] Note that many studies mentioned in this digest were already discussed in previous sections of the present dissertation (mainly 2.2 and 2.3.1.1).

Among all the factors that might affect degree of FA – i.e. age of onset of L2 learning, formal instruction, gender, L1/L2 use, language learning aptitude, LOR, and motivation – Piske et al. (2001) have concluded that age of onset of L2 learning is the main determinant for FA (e.g. Asher & García, 1969/1982; Flege & Fletcher, 1992; Flege, Munro, & MacKay, 1995a; Moyer, 1999; Piper & Cansin, 1988; Tahta et al., 1981a). Thus, accent has been found to emerge between the ages of 5 and 8 years in L2 naturalistic settings (e.g. Flege & Fletcher, 1992; see, in particular, Flege et al.'s large-scale study, 1995, examining 240 Italian speakers of English who represented a wide AOL range – from 3 to 22 years).

To a lesser extent, a longer LOR in the TL country has resulted in native-like or less accented pronunciation (e.g. Asher & García, 1969/1982; Flege et al., 1995a; Flege, Bohn, & Jang, 1997; Flege & Fletcher, 1992; Flege & Liu, 2001; Seliger et al., 1976/1982). So has L1 use, "a variable recently introduced into the literature more recently" (Piske et al., 2001, p. 192). That is, the higher the percentage learners' L1 use on a daily basis within the L2 community, the more foreign-accented the TL pronunciation is on the part not only of late L2 learners but also of early L2 learners (e.g. Flege, Frieda, & Nozawa, 1997; Piske & MacKay, 1999; see also Piske et al., 2002). The authors further illustrate that, in the event of significant gender differences in FA scores, females' L2 speech production has often been rated as less foreign-accented than that of males (e.g. Tahta et al., 1981a, 1981b). Also, formal instruction seems to play a role in the attainment of less accented pronunciation, but only if it consists of specific phonetic training (e.g. Bongaerts et al., 1997; Moyer, 1999). However, Piske et al. (2001) note that "L2 pronunciation receives little attention in most foreign language classrooms. This might explain why instructional variables seem to have had so little effect in the studies just cited"[45].

Even in those studies where the factors of LOR, formal instruction, gender, and L1 use (as well as motivation and language learning aptitude) lead to a significantly lower degree of FA in the L2, Piske et al. (2001) argue that onset age of L2 learning is the prevailing factor in the prediction of FA. What is more, many other studies have yielded divergent results from the findings of factors influencing FA mentioned above.

Thus, the more limited FA research so far conducted in formal learning contexts poses a challenge to the finding that late starting age will lead to foreign-accented L2

[45] For an extensive review of how pronunciation teaching was implemented – or even disregarded – in ESL classrooms until the beginning of the 1990s, see Morley (1991).

pronunciation (e.g. Bongaerts, 1999; Bongaerts et al., 1997), as well as disagreeing with results conferring an early starting age advantage to attaining accent-free pronunciation in the TL – or, at least, younger child starters' less foreign-accented FL pronunciation than older child beginners' (e.g. García Lecumberri & Gallardo, 2003).

Similarly, nonsignificant LOR effects have been observed in studies such as Elliott (1995a, 1995b), Moyer (1999), Oyama (1978/1982), Piper & Cansin (1988), Tahta et al. (1981a). Gender differences are not relevant, either, to attaining native-like pronunciation of the TL (or less accented pronunciation) in Elliott (1995a) and Snow & Hoefnagel-Höhle (1977/1982), among others.

In light of the divergent results above, Piske et al. (2001) further hypothesise that the inconclusive findings of factors influencing FA might be partly due to the different methodologies of earlier accent studies (see also McAllister, 1995).

One difference then has to do with the inclusion or non-inclusion of a control group of NSs. According to the authors, the inclusion of a group of NS participants is recommended in order to determine whether NSs perform on a particular task as expected (which, in turn, may affect the interpretation of NNS results), on the one hand; and to establish whether listeners in a given study are capable of identifying native and nonnative speech successfully, on the other hand. It should also be added that listeners' accent ratings are subject to range effects. That is to say, L2 learners, particularly late starters, receive increasingly higher-accented scores as the sample of NSs of the L2 under investigation enlarges (Flege & Fletcher, 1992). Accordingly, Flege and Fletcher (1992) have shown that when Spanish late L2 learners' sentences in English were presented without the English NS group, listeners judged their productions as less foreign-accented than when presented together with English foils' productions.

Regarding listeners (or judges or raters), it was mentioned above that both non-expert and expert (e.g. trained phoneticians and ESL teachers, among others) can accurately assess degree of FA. However, Piske et al. (2001) note that when both types of listeners are asked to rate the same speech samples, divergent results have emerged. Thus, while some studies reported on non-expert listeners' ratings on nonnative speech as more accented than those of expert judges, other investigations demonstrated that both listener types rate speech productions on a similar basis (Thompson, 1991, vs. Bongaerts et al., 1997; as cited in Piske et al., 2001). Besides, in some instances a further distinction drawn between listeners who are familiar with foreign-accented speech and those who are not have also yielded inconclusive results (e.g. Flege & Fletcher, 1992; Flege, Frieda, &

Nozawa, 1997). Finally, Piske et al. (2001) observe that the "ideal" number of listeners required to assess FA reliably has not yet been specified.

Another varying methodological aspect in accent studies is connected to rating techniques. On assessing degree of FA, judges have mostly been asked to use either rating or sliding scales, the major differences among studies being located in the points comprising rating scales (ranging from 3 to 9 points). In general terms, one end-point on the rating scale (and sliding scale) stands for "*no FA*", while the other refers to "*heavy or very strong FA*"[46] (but see García Lecumberri & Gallardo, 2003, where no end-point represents "*no FA*", instead "*slight accent*"). Unlike the unresolved number of listeners needed to assess degree of FA, Piske et al. (2001) highlight that a recent study by Southwood and Flege (1999) has provided evidence that foreign accent is best assessed by means of rating or equal-appearing interval (EAI) scales, based on the finding of FA as a metathetic continuum[47]. Moreover, the authors recommend using 9- or 11-point scales if degree of FA is to be depicted accurately.

To conclude the overview of varying methodological issues in FA studies, Piske et al. indicate that numerous tasks have been used to obtain subjects' speech samples for subsequent FA rating, including reading aloud of word lists, sentences, and short passages; direct or delayed repetitions of L2 speech after stimulus presentation via a NS model voice; and production of extemporaneous speech in the L2 (for disadvantages of these elicitation tasks, which are likely to affect subjects' performance, see 2.3.4).

All in all, the claim that starting age of L2 learning is the main determinant for FA is mostly supported by a great deal of FA research, as reviewed by Piske et al. (2001). By contrast, the influence of factors such as amount of L1 use and L2 experience on the degree of perceived FA has yielded divergent results. Moreover, the lack of conclusive findings has been hypothesised to arise in part from methodological differences among the studies, which involve the presence of a NS control group, listener type, rating techniques and elicitation tasks. Finally, the authors have noted that the findings of most FA research apply to learners' perception and production of L2 segments, while suggesting that other less studied phenomena, such as suprasegmentals, should be considered in further research. Along those lines, several studies conducted by Munro

[46] It appears that listeners' ratings are not affected by whether the accent scale goes from *no FA* (e.g. point 1 on the scale) to *heavy accent* (e.g. point 9 on the scale), or from *heavy accent* (e.g. point 1 on the scale) to *no FA* (e.g. point 9 on the scale); or, at least, this issue has never been addressed in FA research.

[47] Piske et al. (2001) define *metathetic continuum* as "a continuum like pitch that can be divided into equal intervals ranging from high to low" in contrast to *prothetic continuum* or "a continuum like loudness that is not amenable to linear partitioning" (p. 195) (see also Southwood & Flege, 1999, pp. 336-337).

and Derwing have recently provided incipient evidence that on assessing FA the dimensions of intelligibility and comprehensibility should be further distinguished from that of accentedness (e.g. Derwing & Munro, 1997; Munro & Derwing, 1995a, 1995b; 2001), in addition to preliminary findings on the perception of foreign-accented speech in noisy conditions (Munro, 1998; see also Mayo, Florentine, & Buus, 1997).

2.3.4. Methodological issues in L2 phonological acquisition research: main tasks and training effects

As stated in 2.3.3, the use of different methodologies in FA studies to a certain extent appears to be responsible for the inconclusive or divergent findings of factors influencing degree of FA. Therefore, the objective of this last section of the Literature Review chapter is to comment on several methodological considerations in L2 phonological acquisition research. Specifically, it will centre on the main effects of tasks and training on learners' performance in the TL[48].

Generally, the two main types of tasks administered to assess younger and older L2 learners' perception of TL sounds are discrimination and identification tasks. As Polka, Jusczyk, and Rvachew (1995) indicate, discrimination tasks test for subjects' "ability to differentiate two physically different stimuli, regardless of their category membership" (p. 75), whereas identification (or labelling) tasks aim at examining listeners' categorisation of stimuli. As follows, several characteristics of these two task types and effects (shortcomings) on learners' performance are presented.

First, discrimination tasks are further classified into same-different (AX) or 2IAX tasks, ABX/AXB discrimination tasks, oddity discrimination tasks, 4IAX discrimination tasks, and categorial or name identity tasks (Beddor & Gottfried, 1995; Logan & Pruitt, 1995). In AX tasks, two stimuli which might be the same or different are presented at a time. While AX tasks do not involve high cognitive demands on the part of the listeners, Beddor and Gottfried (1995) argue that one "negative" effect of 2IAX tasks is the subjects' tendency to consider difficult pairs to discern as "the same" at significant rates. Thus, in order to avoid listeners' bias towards "the same" responses, AXB/ABX tasks

[48] For other methodological variables that should be taken into account in the implementation of L2 speech perception and production research, such as stimulus materials (e.g. natural, edited, synthetic stimuli), task instructions and subject characteristics, see, among others, Beddor and Gottfried (1995) and Logan and Pruitt (1995), in addition to Sato (1985).

might be more appropriate. In this case, subjects are required to decide whether one (X) of the three stimuli delivered on a given trial is identical to either one of two acoustically different stimuli (A or B). However, in contrast to AX tasks, ABX/AXB tasks are more demanding at the cognitive level. Similar cognitive demands are expected from subjects' performance on oddity discrimination tasks, where they are asked to discern which one of the three stimuli presented at a time is acoustically different from the remaining two[49]. As an alternative to overcoming both higher cognitive demands and listeners' biased responses towards "the same", Beddor and Gottfried (1995) suggest employing 4IAX tasks. In those tasks, two pairs of stimuli are delivered at a time, one pair containing two acoustically identical stimuli, and the other containing two differing acoustic stimuli. Another type of discrimination task often administered in L2 speech perception research is the categorial AX task or name identity task, whereby listeners identify the phonetic category of the natural stimuli delivered. That is, in a name identity task "same" pairs contain different tokens from the same phonetic category, while "different" pairs are composed of stimuli belonging to different phonetic categories (Logan & Pruitt, 1995).

With regard to identification tasks, listeners are asked to identify each single stimulus presented on a trial by means of choosing only one of the response alternative provided. On top of that, participants are frequently asked to make category goodness or confidence judgements (Beddor & Gottfried, 1995). One limitation of labelling tasks that might influence learners' performance lies in the difficulty in providing listeners with the appropriate response alternatives for each stimulus in the task.

Although discrimination and identification tasks have been administered to L2 learners of different ages, Polka et al. (1995) specify that both tasks are generally adapted when testing preschool children aged between 3 and 5 years. Moreover, in the case of infants under one year of age, four different tasks have been commonly employed to test for infants' perception of native (primarily) and non-native sounds: the High Amplitude Sucking (HAS) Procedure, the Conditioned Headturn (HT) Procedure; the Visual Habituation Procedure (VHP), and the Head-turn Preference Procedure (HPP) (for a detailed account of these tasks, see Polka et al., 1995).

In the study of TL sound production, the main tasks administered consist of either repeating/imitating or reading aloud of speech samples that contain the L2 sounds

[49] Often, oddity discrimination tasks contain "catch" or "no-change" trials (e.g. Flege, MacKay, & Meador, 1999; and Flege & MacKay, 2004, respectively) consisting of three physically different instances of a single phonetic category.

under investigation. Therefore, L2 learners have been asked to produce a wide variety of speech samples ranging from words in isolation, in word-sequence lists, and in carrier phrases (where target words are to be inserted either in initial, medial, or final sentence position) to sentences between 7 and 11 syllables long on average. To a lesser extent, subjects have been asked to read paragraphs or short texts aloud. Additionally, picture stories or questions might be delivered in order to have subjects produce extemporaneous L2 speech (e.g. in the form of narratives and retelling of a personal anecdote).

As in perceptual tasks, the various production tasks are not exempt from shortcomings. Thus, one of the limitations of reading aloud tasks is the influence that orthography might exert on subjects' L2 speech production, "which can act either as a guide or a hindrance to pronunciation" (Tench, 1996, p. 250). Similarly, production tasks that involve direct imitation from a NS model (delivered either via technical support or the experimenter) might indicate short-term phonological memory effects rather than learning effects. As Tench (1996) notes, "[e]ducational psychologists … inform us that any accurate acoustic image of a spoken stimulus is retained by learners for five seconds at the most. Hence, if an imitation is attempted within those five seconds, it will be as its most accurate" (p. 250). A way then to overcome the disadvantages of reading aloud and direct imitation tasks is by means of a delayed repetition task, as Flege et al. have shown (e.g. Flege et al., 1995a; Piske et al., 2001). This technique consists of repeating a series of sentences (or sequences of words), which are presented both in written and aural formats by a NS model voice in a question-answer context. Then, the question is delivered a second time, immediately after which participants have to utter the answer they have previously heard.

All the elicitation techniques in the preceding paragraph are rather formal in nature, which, in turn, might have an effect on learners' performance. Among others, Tench (1996) observes that "[f]ormal procedures put people on their best behaviour" and goes on to state that "[m]ost speech is not formal, and so a formal style … is not a true indicator of the total range of competence in phonology" (p. 250). Thus, the elicitation of extemporaneous L2 speech samples might characterise learners' phonological competence stage better. However, extemporaneous speech has limitations, too. For instance, Piske et al. (2001) comment on the fact that when asked to produce free extemporaneous speech, certain L2 sounds or sound sequences that pose a difficulty for learners might be avoided; thereby learners' assessment of production of specific L2 target sounds is deemed impossible. Moreover, learners' instances of morphosyntactic

and lexical deviations from NS norms, in addition to hesitation and pause phenomena, might result in learners' poorer performance of TL speech if, for instance, their speech samples are to be rated for degree of FA.

In sum, L2 learners' TL sound perception and production have been examined by means of discrimination and identification tasks and through reading aloud and imitation tasks. At times, tasks have been adapted to suit younger learners' cognitive state of development. In fact, Beddor and Gottfried (1995) highlight the need for further development of tasks that can be administered to child and adult language learners simultaneously (i.e. with no prior changes in task characteristics based on the subjects' age). Along those lines, Polka et al. (1995) have provided preliminary evidence on the application of HP and HT Procedure used to test infants' speech perception to the assessment of adult L2 learners' perceptual abilities in the TL. Likewise, Rvachew and Jamieson (1995) have employed the same techniques to compare TL (= L1) sound production in children with speech disorders to TL (= L2) segmental production of adult L2 learners.

As far as the variable of training is concerned, it has been reported that training in the form of explicit phonetic instruction may result in learners' better perception and production of TL segments, especially in formal language learning settings (e.g. Bongaerts et al., 1997; García Lecumberri, 1999; Moyer, 1999). Even then, the available evidence to date is still scarce. Moreover, the few investigations comparing training procedures (e.g. segmental vs. suprasegmental treatment) have yielded divergent findings. For instance, Champagne-Muzar, Schneiderman, and Bourdages (1993) and Derwing, Munro, and Wiebe (1998) found that both segmental and suprasegmental training led to an improvement in learners' perception and production of TL sounds (together with intonation in Champagne-Muzar et al., 1993; and intelligibility and fluency in Derwing et al., 1998). Conversely, Macdonald, Yule, and Powers (1992) did not report on any significant change in learners' speech production as a function of any of the three training procedures implemented. Recently, Cebrian (2003) hypothesised that one of the reasons why Catalan learners of English in the L1 community performed similarly to native Catalan long-term residents of Canada might have been due to instruction in English phonetics that the former had received.

Based on all of the above, the long-stated need for further research on the role of phonetic training in adult learners' attainment of L2 phonological skills (e.g. Flege, 1995a) does still hold. Thus, if studies are to be conducted in order to determine the

effects of training on learners' phonological acquisition of the L2, several methodological issues should be borne in mind, as is the case of perceptual and production tasks discussed previously.

In this regard, Logan and Pruitt (1995) have offered a good summary of the main aspects characterising the factor of training. To begin with, it should be mentioned that the authors' overview focuses on perceptual training based on their assumption that perception precedes production and that both processes are related; thereby hypothesising that any changes in perception will transfer to production. Thus, they suggest that "the primary goal of perceptual training studies may be to facilitate the long-term development of a novel phonemic category that is potentially usable among a variety of phonetic contexts, talkers, and other sources of variability" (p. 353). For that purpose, training might be administered on a short-term (one session) or long-term basis (typically, 15 training sessions over a three-week period). Generally, perceptual training is delivered by means of identification tasks, in preference to discrimination tasks, for they are considered to promote better transfer to new stimuli (but see Flege, 1995b). Moreover, the authors illustrate that training sessions should provide subjects with information about their own performance or feedback, either after each trial – trial-by-trial feedback – or after a number of trials, a block, or training session – cumulative feedback. In order to assess the effectiveness of training, a pre-test and post-test are delivered before and after the actual administration of training, respectively. Additionally, when evaluating the role of training, Logan and Pruitt (1995) recommend the inclusion of a control group (subjects who have not received training) to discard any positive result in the experimental group's performance on the post-test (in comparison to pre-test results) that might be the consequence of practice effects, rather than training effects. Finally, any type of training administered should aim at the generalisation of beneficial effects to new stimuli, tasks, and talkers. In line with the latter requirement, Rochet (1995) and Bradlow et al. (1997) provide preliminary evidence on the fact that perceptual training does not only generalise to the perception of novel L2 sounds and phonetic contexts, but it also transfers to L2 sound production.