Bridging the gap between L2 speech perception research and phonological theory

Paola Escudero & Paul Boersma, July 9, 2003

Abstract. A series of experiments shows that Spanish learners of English acquire the *ship-sheep* contrast in a way specific to their target dialect (Scottish or Southern British), and that many learners exhibit a perceptual strategy found in neither Spanish nor English. To account for these facts as well as for the findings of earlier research on L2 speech perception, we provide an Optimality-Theoretic model of phonological categorization that comes with a formal learning algorithm for its acquisition. Within this model, the dialect-dependent and L2-specific facts provide evidence for the hypotheses of Full Transfer and Full Access.

Introduction

Part of the phonology of a language consists of sound distinctions that the speakers perceive and produce. The sound distinctions of a language are signalled by a number of auditory properties (duration, static and dynamic spectral features, periodicity, noise, intensity) that integrate to constitute phonological contrasts. For instance, the contrast between the English vowels /i/ and /ɪ/ is one of vowel height (or tenseness) as well as length, i.e., these vowels differ in spectral features (Peterson & Barney, 1952) as well as in duration (Peterson & Lehiste, 1960) and native speakers rely on both of these auditory cues when having to categorize these vowels (Bohn & Flege, 1990).

Perceptual cue weighting has cross-linguistic as well as developmental aspects. Cross-linguistically, the attention paid to the cues that signal a contrast varies between adult speakers of different languages (Bradlow, 1995; Fox, Flege & Munro, 1995; Gottfried & Beddor, 1988). For instance, Gottfried & Beddor (1988) show that unlike American English speakers, for whom vowel contrasts involve duration in production as well as in perception, Parisian French speakers produce only small durational differences and do not use durational information at all when categorizing vowels. Developmentally, babies have to learn what aspects of the phonetic signal serve as cues in their language and how much importance to attach to each cue (Scobbie 1998). Since all infants start out with identical perception systems, we expect that the cross-linguistic variation in adults must have been brought about by developmental changes in cue weighting, and indeed several studies have shown that the use of the cues that signal a certain phonological contrast can be different for adults, infants and children (Gerrits, 2001; Nittrouer, 1992, 1996; Nittrouer, Manning & Meyer, 1993?; Nittrouer & Miller, 1997). For instance, Gerrits (2001) shows that 4-year-old Dutch children attend to duration cues much more than adult listeners do, for vowels as well as consonants.

Both cross-linguistic and developmental variation are also attested in research on second language (L2) speech. It has been shown that learners may weigh the cues to phonological contrasts differently from native speakers of the target language, in production as well as in perception. For instance, Bohn (1995) and Flege, Bohn & Jang (1997) showed that Mandarin learners of English use temporal information more than spectral information when differentiating between American English /i/ and /ɪ/, and that Spanish speakers use the two dimensions equally, whereas American English speakers have a preference for the spectral cues. It has also been shown that learners change their cue weighting as their experience with

the target language increases. For instance, Morrison (2002) showed that Spanish learners of Canadian English change their cue weightings for the /i/–/ɪ/ contrast from a very fuzzy reliance on duration after one month of presence in Canada to a reasonably good reliance on spectrum or duration (but not on both) after six months. The question of interest to L2 researchers now is how to explain the perceptual behaviour of the learners.

It has long been acknowledged that structural properties of the first language can be transferred to the interlanguage system. In the realm of phonological perception, Polivanov (1931) gives the example of the 'European' word *drama*, which is perceived and therefore produced by Japanese learners of European languages as *dorama* or *dzurama*, in line with the Japanese ban on syllable-initial consonant clusters. Within the formal framework of generative grammar, second language research has led to the formulation of several hypotheses. The most radical, explicit, and simple proposal is that of Full Transfer (Schwartz & Sprouse 1996), according to which an L2 learner uses her entire first-language (L1) system (except the phonological makeup of the lexicon) as a starting point for her L2 development. Schwartz and Sprouse combined this proposal with the hypothesis of Full Access, according to which an L2 learner has access to all the principles of Universal Grammar and her entire language acquisition device, as she had when acquiring her L1.

It has turned out that most apparent counterexamples to the Full Transfer/Full Access hypothesis can be reanalysed as supporting cases if one looks into the data deeply enough. When we return to the case of Bohn (1995) and Flege et al. (1997), for instance, we note that whether their learners' behaviour can be explained by grammatical transfer from Mandarin or Spanish to the English interlanguage is a matter of debate. Bohn (1995) claims that transfer does not play a major role in the explanation of the Spanish learners' larger-than-native use of duration: since the English vowels /i/ and /ɪ/ fall within a single Spanish spectral category ("high") as well as in a single Spanish duration category (because there is no length contrast in Spanish), Bohn expected Spanish learners to have trouble both with the spectral contrast and with the duration contrast; instead, they turned out to have less trouble with the duration contrast than with the spectral contrast. But whether the learners' behaviour constitutes a case of grammatical transfer depends on the theoretical framework that one assumes. Within the L2 speech perception tradition that Bohn (1995) is part of, the Perceptual Assimilation Model (Best, 1995) describes the behaviour that listeners show when confronted with foreign-language sounds, with implications for the initial state of L2 learning, and the Speech Learning Model (Flege, 1995) advances generalizations and predictions about the L2 development of phonetic categories. Both models make claims about transfer (and, for Flege, about access to acquisition devices), but neither model invokes the formal devices of linguistic theory, so that the precise underlying systems (i.e. lexical representations and the mappings between these and the overt language data) and learning mechanisms remain unclear. But since phenomena like the weighting of auditory cues in the categorization of vowels are language-specific, we argue that speakers have systematic knowledge that underlies their perceptual behaviour, and that this knowledge, like other kinds of language-specific knowledge, is therefore a natural subject matter for linguistic theory. The present paper thus aims at filling the gap between L2 perception research and phonological theory by expressing language-specific perception phenomena by means of formal perception grammars. In §1.6 and §3.2, we will show that in a theoretical framework that handles perception with linguistic devices and is compatible with the findings of speech perception research, the behaviour of the learners in Bohn's (1995) experiments can be regarded as grammatical transfer from the L1 and subsequent access to L1-like acquisition devices.

The perception grammar that we will take to underlie L1 and L2 perceptual behaviour will be formalized within the constraint-based framework of Optimality Theory (Prince & Smolensky, 1993). The application of Optimality Theory (OT) to L2 phonological acquisition

is not new, but nearly all of the work has concentrated on production, i.e. the mapping of underlying forms to surface forms. One of the virtues of the constraint-based framework of OT in L2 production is that it can explain some interlanguage phenomena that cannot be explained by frameworks based on sequential rule applications. Broselow, Chen & Wang (1998), for instance, argue that the tendency of Mandarin learners of English to delete underlying coda obstruents is hard to explain with these rule-based theories, because neither Mandarin nor English has a rule of coda consonant deletion: according to Eckman (1981), English lacks such a rule because underlying obstruents surface faithfully in that language, while Mandarin lacks such a rule because this language has no underlying coda obstruents to begin with, so there is nothing to delete. Thus, the Mandarin-to-English interlanguage seems to exhibit a rule that belongs neither to the L1 nor to the target language, which implies that within rule-based theories the assumption of transfer would predict that Mandarin learners of English should be able to produce final coda obstruents. Within OT, by contrast, the fact that Mandarin does not allow coda obstruents in surface forms is expressed *in the grammar* as the high ranking of a constraint against coda obstruents, a situation that does not cause overt coda deletion in Mandarin but would cause coda deletion in a language with underlying coda obstruents like English. Thus, within an OT framework the Mandarin-to-English interlanguage simply exhibits an effect of a transfer from the L1, while for Eckman (1981) this was not a instance of transfer. By strengthening the case for Full Transfer and thereby simplifying explanations for L2 phenomena, Optimality Theory must therefore be regarded as a serious candidate for expressing generalizations about L2 phonology. Other OT work in L2 production includes Davidson (1997), Hancin-Bhatt & Bhatt (1997ab), Hancin-Bhatt (1997, 2000), and Hayes (2000). In all of these studies, the hypothesis of Full Transfer is taken to imply that an L2 learner brings to the learning task the constraint set of her first language as well as the ranking of these constraints. For instance, Broselow et al.'s explanation relies on the idea that the high ranking of a constraint against coda obstruents is *transferred* from the learners' L1 Mandarin to their Mandarin-to-English interlanguage. The hypothesis of Full Access is taken to imply that an L2 learner has access to a device that changes the rankings of the constraints on the basis of incoming data, such as Tesar & Smolensky's (1998, 2000) Robust Interpretive Parsing / Constraint Demotion algorithm or Author & SomebodyElse's (2001) Gradual Learning Algorithm (discussed with respect to L2 acquisition by Davidson 1997 and Hayes 2001, respectively). For example, Broselow et al. assume that hearing surface coda obstruents in English will ultimately lead Mandarin learners of English to lower the ranking of their constraint against coda obstruents.

The applicability of Optimality Theory is not restricted to the modelling of production. It is possible to apply constraint-ranking methods to the modelling of language-specific perceptual knowledge, i.e. the listener's knowledge of how to map continuous auditory features to discrete phonological surface structures such as segments and syllables. Polivanov (1931), for instance, explains the Japanese perception of *drama* as /dorama/ or /dzurama/ by the interaction of five constraints: a Japanese well-formedness constraint against consonant clusters rules out /drama/, and a well-formedness constraint against /d/ before the vowel /u/ rules out /durama/. This leaves the listener with two well-formed candidates to choose from: /dzurama/ and /dorama/. Each of these forms has a shortcoming: the form /dorama/ contains a full vowel /o/ that corresponds to no auditory cue in the input (the form /durama/, where only the reducible vowel /u/ is inserted, would do better in that respect), and the form /dzurama/ violates faithfulness to perceived plosiveness. Accounts like these can be straightforwardly formalized as Optimality-Theoretical perception grammars, as was done by Author (1998) for L1 and by Hayes (2001, 2002) for L2. Hayes performed discrimination and identification experiments on the Japanese consonant length contrast, with monolingual speakers of English, English-speaking learners of Japanese, and native speakers of Japanese,

and found that the natives show categorical perception, the non-natives show continuous perception, and the learners show something in between these two kinds of perception. Hayes formalizes the three kinds of perception with Optimality-Theoretic analyses, and concludes that the OT perception model accounts for the attested fact that the learners' perception changes during development.

The current paper investigates the acquisition of the English /i/–/ɪ/ contrast by L1 Spanish learners, who have problems with it in production as well as in perception (Flege, Bohn & Jang 1997). This paper differs from Hayes' in five respects, all of which pose stronger challenges to the OT perception theory. First, for native speakers of English the /i/–/ɪ/ contrast is one of duration as well as spectrum, so that the L2 perceptual learning task for the Spanish involves an integration of multiple auditory cues rather than a categorization of a single auditory continuum as in Hayes' case, and it is not clear from the start that the theory can handle such more complicated cases. Second, we will find a typical L2 learning stage that is attested neither in the learner's native language nor in the target language, and is difficult to label as 'intermediate'. Third, we consider the acquisition of two closely related target languages, Scottish Standard English and Southern British English, rather than a single target language, and the theory will have to account for the large differences that we will find between the L2 developments of the Spanish learners of these varieties. Fourth, in order to successfully test the model, all these four different developmental patterns will have to be replicated in a computer simulation that uses the initial state and the learning algorithm that come with the theory. Fifth, we think that the applicability of our theoretical model should extend beyond the experiments that we report on in this article, i.e., we have to take into account the generalizations advanced by other empirical researchers on L2 speech perception.

We believe that the stochastic version of OT (Author, 1998), together with the Gradual Learning Algorithm (GLA) that is associated with it (Author & SomebodyElse, 2001), constitutes a reasonable model of phonological development. For instance, the GLA can account for the order of constraint demotions attested by Broselow et al. (1998): the fact that many learners go through a stage in which they can produce obstruent codas but devoice them, means that the constraint against obstruent codas in general is demoted earlier or faster than the constraint against voiced obstruent codas; the GLA, which works by small learning steps, explains this because it demotes a constraint against voiced coda obstruents only if it encounters a final voiced obstruent, and it demotes the general constraint against coda obstruents every time it encounters *any* coda obstruent.

We will show that for a complicated case with multiple auditory continua and multiple varieties of the target language, the OT/GLA perception model is capable of accounting for the following observed behaviour:

1. Listeners optimize their perception in accord with the productions that they encounter, i.e., their perception becomes such that it copes well with the specific characteristics of the ambient language.
2. L1 listeners arrive at an optimal perception, i.e., they come to use the available auditory information in the best possible way.
3. Depending on the target dialect, L2 listeners may reach optimal perception or may manifest sub-optimal optimization strategies that are specific to L2 acquisition.

Each of these observations will be established with experiments involving real speakers and listeners (§1), modelled with OT perception grammars (§2, §3), and confirmed in computer simulations (§3) . In §4, we will discuss to what extent our findings yield evidence in favour of or against the Full Transfer / Full Access hypothesis.

1. L1 and L2 vowel categorization in two English dialects

It is an empirical question whether L2 speakers can learn to perceive non-native sounds in a way that resembles the behaviour of native speakers of the target language, and if they cannot, in what particular ways their perception may differ. To be able to compare L1 with L2 perception, we examine the categorization of the English /i/-/ɪ/ contrast by Spanish learners of Scottish and Southern English, and compare it to how the contrast is perceived by adult Scottish and Southern English speakers. We will see that while the perceptual behaviour of the native listeners is *optimal* in the sense that their cue weighting in perception matches the relative use of the cues in the production of their dialect, many Spanish listeners behave in a way that is not found in either of the native groups, nor in their L1 Spanish.

Several studies have shown that Spanish speakers have problems perceiving some English vowels. Fox, Flege & Munro (1995) show that Spanish listeners are not sensitive to durational information when categorizing vowels, and therefore may have problems with English vowels that differ in both duration and quality (for the same claim, see Bradlow, 1995). For the English /i/–/ɪ/ distinction, Flege (1991) shows that monolingual Spanish listeners associate both English /iː/ and English /ɪ/ with Spanish /i/ (in both cases they prefer the Spanish /i/ category over the "none" category in an orthographic labelling experiment). According to Flege, this may explain why early Spanish learners of English use /i/ indiscriminately in their mental representations for English /i/ and /ɪ/, which again would explain why they do not differentiate the two vowels in L2 *production*. However, Bohn (1995) shows that Spanish learners of English are able to distinguish between /i/ and /ɪ/ in an identification task (from which it seems to follow that they must have separate lexical representations for the vowels in *bit* and *beat*, since they are capable of using these words as labels); nevertheless, they make such a distinction by relying on auditory cues that are less important for English listeners. Specifically, the Spanish learners of English in Bohn's study pay as much attention to durational as to spectral differences when having to categorize English /i/ and /ɪ/ (spectral cue reliance: 50%, duration: 44%); the English listeners exhibited a strong preference for spectral cues (spectral cues: 88%, duration: 9%).[1] In contradiction with Bohn (1995), however, Flege, Bohn & Jang (1997) concluded from the same data set that the Spanish learners were no different from the English listeners: a statistical reanalysis of the data showed no significant difference between the Spanish learners of English (both inexperienced and experienced) and the native English listeners with respect to the attention paid to the spectral information that cues the /i/–/ɪ/ contrast (47% versus 88%). However, if a measured difference of 47% versus 88% is compatible with no underlying difference, the power of the statistical test must be so low that the measured difference will also be compatible with a large underlying difference (e.g. 27% versus 97%), so that Flege et al.'s results must be considered inconclusive.

In sum, the literature seems to be inconclusive in two respects. First, it does not clearly answer the question as to whether Spanish learners of English can or cannot learn to perceive the English /i/–/ɪ/ contrast in a way that resembles native speaker perception. Second, it presents contradictory conclusions with respect to the auditory cue weighting in Spanish learners of English. This section attempts to resolve these contradictory findings of previous studies. First, we show that the initial problems in the perception of Spanish listeners of English as well as the cue weighting patterns shown in L2 development are shaped by the learners' L2 target variety. More specifically, we show that Spanish learners of English manifest a different starting point and a different developmental cue reliance depending on

---

[1] In §1.3, we explain how such percentages are computed.

whether their target L2 is Southern British English or Scottish English. Second, we show that L2 perception becomes optimal or near-optimal for the target L2 environment.

1.1. Native production of the contrast in the two dialects

In a production experiment (described and analysed in more detail in Authors 2001), we find that there is a considerable difference in the production of the /i/-/ɪ/ contrast for Scottish versus Southern English speakers with respect to the relative use of the acoustic dimensions that signal the contrast. The cues that we investigate are spectral vowel height, as expressed by the first formant frequency (F1), and vowel duration. We recorded four minimal word pairs containing /i/ and /ɪ/: *ship*, *sheep*, *Snicker*, *sneaker*, *lid*, *lead*, *filling*, *feeling*. These words were chosen in order to obtain some realistic variation with respect to the voicing of the following consonant and the number of syllables. Each of the eight words, embedded in the carrier sentence *THIS is a __ as well*, was produced 50 times by a male Scottish English speaker (grown up in Edinburgh, 20 years old) and by a male Southern British English speaker (grown up in Reading, 20 years old), after some practice in order to get familiar with the ambiguous or near-nonsense words. There were also ten distracter words produced ten times each. Each speaker pronounced the 500 sentences in a balanced randomized order, in about 30 minutes. The recordings were made with a DAT recorder in a soundproof room, then digitized into a computer for the acoustic analyses.

    The vowels were segmented by the two authors with the help of the Praat program. The averages of the time markings of the two segmenters were used for an automatic analysis of duration and F1. The results are in Figure 1, expressed along logarithmic axes so that we can easily work with fractional differences, e.g. the distance in the figure between 50 and 60 ms is equal to the distance between 150 and 180 ms.[2] Likewise, standard deviations of F1 and duration will be expressed in terms of relative frequency differences (in units of octaves) and relative duration differences (in units of duration doublings) rather than in absolute differences in units of Hertz and milliseconds.
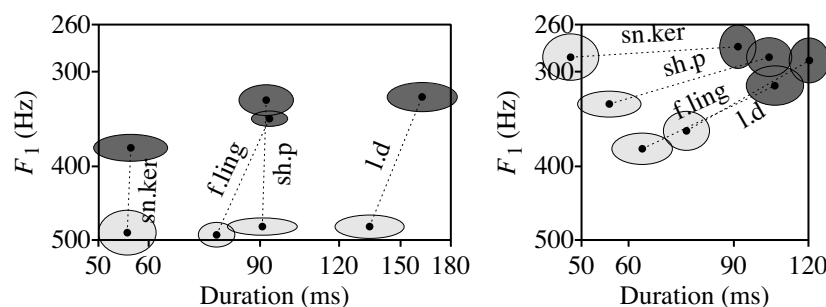

Fig. 1. Scottish (left) and Southern (right) production of /ɪ/ (light) and /i/ (dark). The axes are logarithmic. The ellipses show the standard deviations.

In order to single out the correlation between the vowel contrast and the two acoustic dimensions, we average (the logarithms of) the duration and F1 values for the two vowels in the two dialects across consonantal context and across number of syllables. The averaged data (converted back to Hz and ms) are shown in Table 1.

---

[2] The use of logarithmic scales must be considered a standard procedure, since both acoustic dimensions have only positive values (in Hertz and milliseconds). The purpose is to make it more likely that we can model the F1 and duration distributions as normal. In the figure, we indeed see that the sizes of the ellipses are comparable across words. In addition, a logarithmic scale is perceptually appropriate for the duration continuum (Ruhm, Mencke, Milburn, Cooper, & Rose, 1966).

Table 1. Duration and F1 for /ɪ/ and /i/ for the Scottish and Southern English speaker, averaged across the four contexts.

Scottish			Southern		
/ɪ/	duration	84.8 ms	/ɪ/	duration	59.7 ms
	F1	485 Hz		F1	337 Hz
/i/	duration	94.0 ms	/i/	duration	104.6 ms
	F1	343 Hz		F1	292 Hz

The attested large difference in production is to be ascribed mainly to a regional difference rather than to speaker idiosyncrasies. As for the spectral dimension, the large height difference between Scottish and Southern /ɪ/ is obvious, e.g., Labov (1994, p. 169) shows a Cockney speaker whose /ɪ/ is higher than her /i/. As for the duration dimension, the Scottish speaker follows more or less the Scottish vowel length rule (Hewlett, Matthews, & Scobbie, 1999; McClure, 1977; Scobbie, Turk, & Hewlett, 1999), according to which /i/ and /ɪ/ (in the Scots language, and hence in many varieties and styles of Scottish English) are equally long before plosives and sonorants. The large duration difference depending on the voicing of the following consonant (more than 50% for the word pairs in Fig. 1) is judged by Hewlett et al. as 'less Scottish' but not uncommon; similarly, we find a small but reliable length difference between /ɪ/ and /i/ before the voiced consonants /d/ and /l/; both of these properties may be due to a 'less Scottish' regional influence (Edinburgh) or style (educated).

1.2. Optimal perception of the contrast in the two dialects

We claim that *optimal* vowel categorization involves a dependency of perception on the specific production environment, i.e. that the optimal way of perceiving the /i/-/ɪ/ contrast depends on how the differences between /i/ and /ɪ/ are produced in the language variety at hand. The perception strategy must depend on the production environment because to optimize perception the listener has to minimize the probability of miscomprehension, and the best way to do this is by making decisions that lead to maximum-likelihood behaviour (Helmholtz 1910). For speech perception, this means that the best thing for the listener to do is to perceive any incoming auditory event as the phonological category that was most likely to have been intended by the speaker. This is what we call an *optimal* strategy for perception.

Figure 2 shows how likely it is for any auditory event (duration-F1 pair) to have been intended as the category /i/, if the distribution of the auditory events is given by the average duration and F1 values of Table 1 and by rather arbitrarily chosen standard deviations of 0.4 duration doublings and 0.2 octaves. The black curve in the figure is the *equal likelihood line*; auditory events on this line have a 50% probability of having been intended as /i/, and a 50% probability of having been intended as /ɪ/. Suppose, now, that a Scottish and a Southern English listener are confronted with the same auditory event, for example [74 ms, 349 Hz], shown as diamonds in Figure 2. If the two listeners exhibit optimal perception (i.e. if they follow a maximum-likelihood strategy), then it is clear that the Scottish listener will best perceive this auditory event as /i/, since Scottish *speakers* are more likely to intend this auditory event as /i/ than as /ɪ/. Likewise, the Southern listener will best perceive the same event as /ɪ/, since in a Southern environment this auditory event is more likely to have been intended as /ɪ/ than as /i/. More generally, both listeners will perceive everything above their own equal-likelihood line (in the figure) as /i/, everything below as /ɪ/.
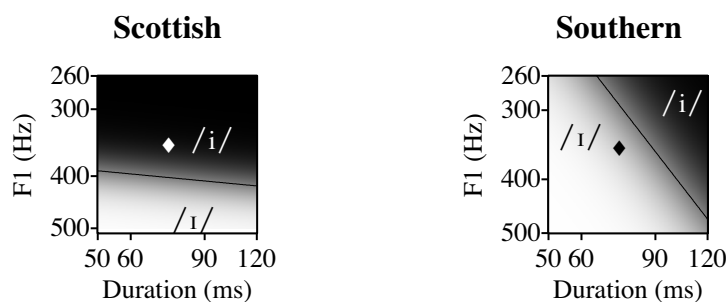
Fig. 2. The probability that any given duration-F1 pair was intended as /i/ in a Scottish or Southern English environment. Values range from 0% (white) to 100% (black); the 50% line is also shown. In the pictures, "/i/" and "/ɪ/" depict the average F1 and duration values in the two dialects, taken from Table 1. The diamond is the auditory event [duration = 74 ms, F1 = 349 Hz], discussed in the text.

The optimal perceiver will therefore have a category boundary in perception that coincides with the equal-likelihood line in her production environment. If the boundary is horizontal, this means that the listener relies on the spectral distinction only, and if the boundary is vertical, it means she relies only on the duration difference. A diagonal boundary, as in the case of the optimal Southerner, implies that the listener relies both on the spectral and on the duration differences between the two vowels. We see that the optimal relative reliance on spectrum and duration directly reflects the reliability of the cues in the production environment.[3]

1.3. The listening experiment

Since the current study, to the best of our knowledge, is one of the first in comparing the perceptual development of two different L2 varieties, our L2 subject population needed to exhibit variation both with respect to the variety of the target language and with respect to the learners' degree of experience with the target language. Therefore, we selected 30 Spanish learners of English who differed among each other in experience level and in target dialect (either Scottish or Southern British English). They were 15 women and 15 men, from various regions within Spain and various countries in South America, aged between 18 and 58. They were middle and upper class students (undergraduate and postgraduate) and employees, had started their L2 learning after the age of 12, and were visiting or living in Edinburgh when participating in the study. To compare their results to those of native listeners, we also selected 20 speakers of Scottish Standard English (10 women and 10 men who reported to have lived in Edinburgh for most of their lives, aged between 23 and 35) and 21 speakers of Southern British English (10 women and 11 men, aged between 19 and 55; all reported to have grown up in the South of England and were judged to have various Southern accents, although ten of them lived in Edinburgh at the time of the experiment). All of the Spanish and Scottish subjects and ten of the Southern English subjects were tested at the University of Edinburgh. The remaining Southern English subjects were tested at the University of Reading. The subjects were tested by the same experimenter, the first author of this paper.

---

[3] An anonymous reviewer argues that if the relationship between production and perception is as predictable as is assumed here, it would not be very parsimonious to posit separate grammars for perception and production, as we seem to do here. This is a good point. Our perception grammar model is a part of a wider model of phonology (Author 1998), in which faithfulness constraints in the speaker's production grammar evaluate the extent to which the listener's perception grammar will be able to reconstruct the hidden phonological structure. This ensures that in this model the two grammars do not replicate each other (perception is primary).

F1 (Hz)

344
366
388
410
433
456
480

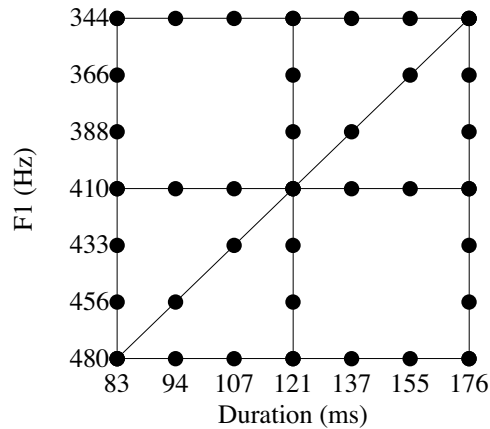83 94 107 121 137 155 176

Duration (ms)

Fig. 3. The 37 stimuli.

The stimuli for the experiment were isolated synthetic vowels. As in most other cue weighting studies (Bohn, 1995; Flege, Bohn & Jang, 1997; Gerrits, 2001; Nittrouer, Manning & Meyer, 1993), the reason for using synthetic stimuli was the possibility of varying acoustic dimensions separately, equally and consistently. Gerrits (2001) has suggested that there is a possible problem with synthetic stimuli, in that they may be perceived differently from naturally produced stimuli, in the sense that subjects categorize natural stimuli more easily; but a more recent study (Nittrouer 2001) has shown that neither children nor adults are adversely affected by the use of synthetic stimuli, since synthetic and naturally produced exemplars of the English /s/–/ʃ/ contrast were categorized equally well. The reason for using isolated vowels was that cue weighting studies that use CVC stimuli (Gottfried, 1984; Gottfried & Beddor, 1988; Strange, Akahane-Yamada, Kubo, Trent & Nishi, 2001) show contextual effects, i.e., vowels are perceived differently when presented in different consonantal environments. The use of isolated vowels avoids these effects and makes the listeners rely more on their abstract representation of the vowels at hand; the drawbacks of presenting isolated vowels are generally that cues deriving from the consonantal context are lost (but that may actually be an advantage in our case) and that an isolated /ɪ/ sounds unnatural because this vowel never occurs utterance-finally in English (but a series of pilot studies revealed that subjects had no large bias against identifying /ɪ/ in isolation).

The stimuli were synthesized on the basis of the auditory properties of natural exemplars of the vowels /i/ and /ɪ/, produced 10 times each by two Scottish English speakers. The average F1 of the naturally produced vowels was 484 Hertz for /ɪ/ and 343 Hertz for /i/. The average F2 (second formant frequency) was 1890 Hertz for /ɪ/ and 2328 Hertz for /i/. These values were taken as the basis for the top and bottom edges of the stimulus rectangle (Figure 3 shows the F1 values only). The six vertical steps, which led to seven spectrally different stimuli, were equal on the mel scale,[4] ranging from 480 to 344 Hz for F1 and from 1893 to 2320 Hz for F2 (F1 is the direct auditory correlate of vowel height, and F2 has to be covaried with F1 in order to make the stimuli sound like natural front vowels). Seven duration values were also considered: they ranged from 83 ms (the left edge of the stimulus rectangle) to 176 ms (the right edge) in six equal fractional steps of 1.1335.[5] In total, 37 vowel stimuli (the

[4] It has been shown that a mel scale resembles the way our ears perceive auditory data more closely than a Hertz scale (Stevens, Volkmann & Newman 1937).

[5] Other cross-linguistic studies have considered different number of steps when constructing stimuli continua. For instance, both Bohn (1995) and Flege, Bohn & Jang (1997) used 11 separate F1/F2 values but only three separate durations. As Bohn (1995) acknowledges, having fewer and larger duration steps might have enhanced the perceptual reliance on the duration cue. To avoid this type of perceptual bias, we decided to use the same number of steps for both acoustic dimensions.

points in Figure 3) were created with the Sensyn version of the Klatt parameter synthesizer. Each listener had to classify 370 stimuli, since each vowel was presented 10 times.

The experimental design was created with the Psyscope software running on a Macintosh computer. The subjects listened to all stimuli under comfortable hearing conditions. The experiment consisted of a forced identification task.[6] The subjects were asked to press either of two buttons, one containing a picture of a ship, the other a picture of a sheep, depending on the vowel that they thought they heard. We used pictures rather than written words in order to avoid orthographic effects for both L1 and L2 perception: at least two cross-liguistic studies (Bohn 1995 and Flege et al. 1997) suggest possible orthographic causes for their attested response reversals (subjects that exchanged the labels *bit* and *beat* in their responses). Every time a sound was presented, the subject would see a visual display on the computer screen (similar to what they had on the button box) that reminded them of what to press. There were both verbal and written instructions, which did not use the words *ship* and *sheep* explicitly. The subjects were told that in case of uncertainty about the answer they should make a guess, and that they could take as much time as they thought convenient to make a decision. Each subject was presented with a different randomization of 10 blocks of 37 trials. After each block the subjects were allowed to take a short break.

1.4. L1 perception of the contrast in the two dialects

The results for the two native groups are in Figures 4 and 5. Dark areas indicate a predominance of /i/ responses, light areas a predominance of /ɪ/ responses. To get a continuous representation, the values in the 12 not-measured cells were interpolated from the values in the neighbouring cells. A *duration reliance* was computed as the percentage of /i/ responses along the right edge of the stimulus rectangle (i.e. the number of stimuli along the right edge that were responded to with /i/, divided by 70) minus the percentage of /i/ responses along the left edge. A *spectral reliance* was computed as the percentage of /i/ responses along the top edge minus the percentage of /i/ responses along the bottom edge. These computations resemble the analysis used by previous cross-linguistic studies that involve the relative reliance on different acoustic cues (Bohn, 1995; Flege et al., 1997). Thus, each of the two reliance values is computed from only 14 of the 37 cells. A *reliance ratio* was then computed as the ratio of the duration reliance and the spectral reliance; this ratio can be shown to be a measure of the slope of the 50% boundary line (Authors 2001). In the figures, the subjects are divided into groups on the basis of their reliance ratios: if a subject's ratio is larger than 4 (i.e. her boundary is more or less vertical), the subject is judged to rely "exclusively on duration"; if her ratio is between 2 and 4, she is judged to rely "mainly on duration"; if her ratio is between 1 and 2, she relies on "duration & spectrum"; if her ratio is between 1/2 and 1, she relies on "spectrum & duration"; if her ratio is between 1/4 and 1/2, she relies "mainly on spectrum"; and if her reliance ratio is less than 1/4 (her boundary is more or less horizontal), the subject relies "exclusively on spectrum". For the "duration & spectrum" and "spectrum & duration" cases, the boundary is nearly diagonal.

---

[6] Categorization experiments often involve both an identification task and a discrimination task. While we did run the discrimination task for all our subjects (Author 2002), the current paper does not report on it, since Gerrits (2001) has shown that boundary effects in discrimination reflect prior discrete labelling, so that we must consider identification as primary, discrimination as secondary. Like us, both Bohn (1995) and Flege, Bohn & Wang (1997) used an identification task only.

exclusively duration:	(none)

mainly duration:	(none)

duration & spectrum:	

spectrum & duration:	(none)

mainly spectrum:	

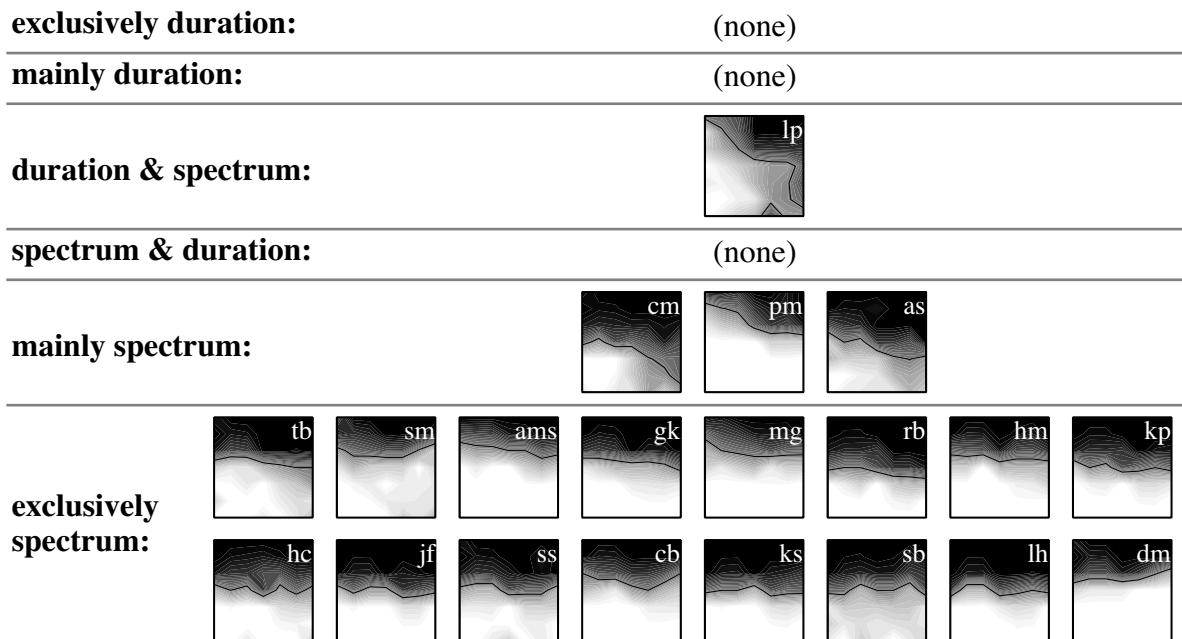|	
exclusively spectrum:	

Fig. 4. Identification results for each of the 20 **Scottish** listeners. In each square, duration runs from 83 ms (left) to 176 ms (right), and F1 runs from 480 Hz (bottom) to 344 Hz (top), as in Figure 3.

exclusively duration:	

mainly duration:	

duration & spectrum:	

spectrum & duration:	

mainly spectrum:	

exclusively spectrum:	

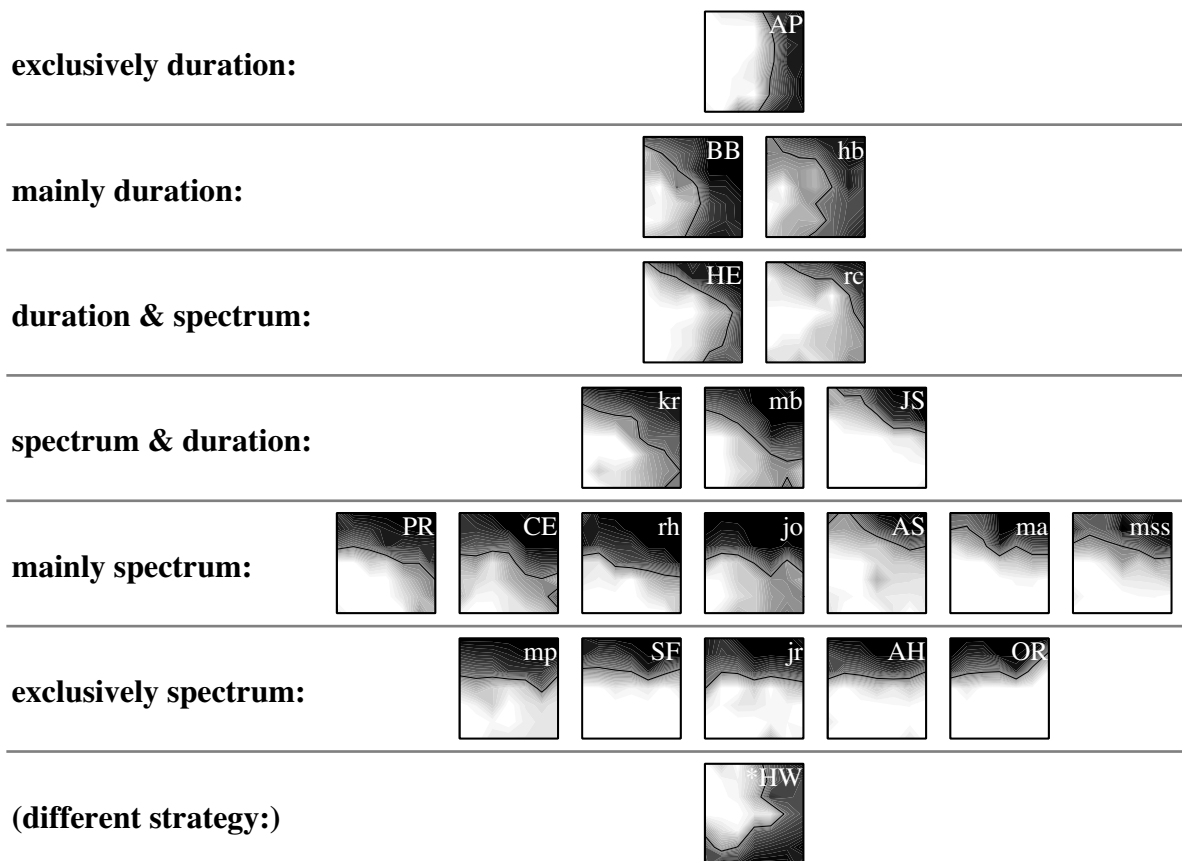|	
(different strategy:)	

Fig. 5. Identification results for the 21 **Southern English** listeners. In each square, duration runs from 83 ms (left) to 176 ms (right), and F1 runs from 480 Hz (bottom) to 344 Hz (top), as in Figure 3. The subjects whose initials are written in lower case were tested in Edinburgh, the others in Reading.

The Scottish listeners have a clear preference for the spectral cues: 16 out of 20 subjects have a reliance ratio of less than 1/4, meaning that they rely almost exclusively on the formants. The Southern English listeners show a different pattern. First, we excluded subject **HW** from further consideration, because she seemed to follow a perverse strategy (her spectral reliance was statistically reliably negative). Only 5 of the remaining 20 subjects relied almost exclusively on the formants, while 14 subjects used a combination of the two cues. One subject relied exclusively on duration.

The results of the Scottish and Southern English listeners confirm that the perception of /i/ and /ɪ/ in these two varieties show differences in the same direction as the production. Both types of listeners have an optimized vowel categorization because they integrate the cues to the /i/-/ɪ/ contrast in accord with their own production environment.
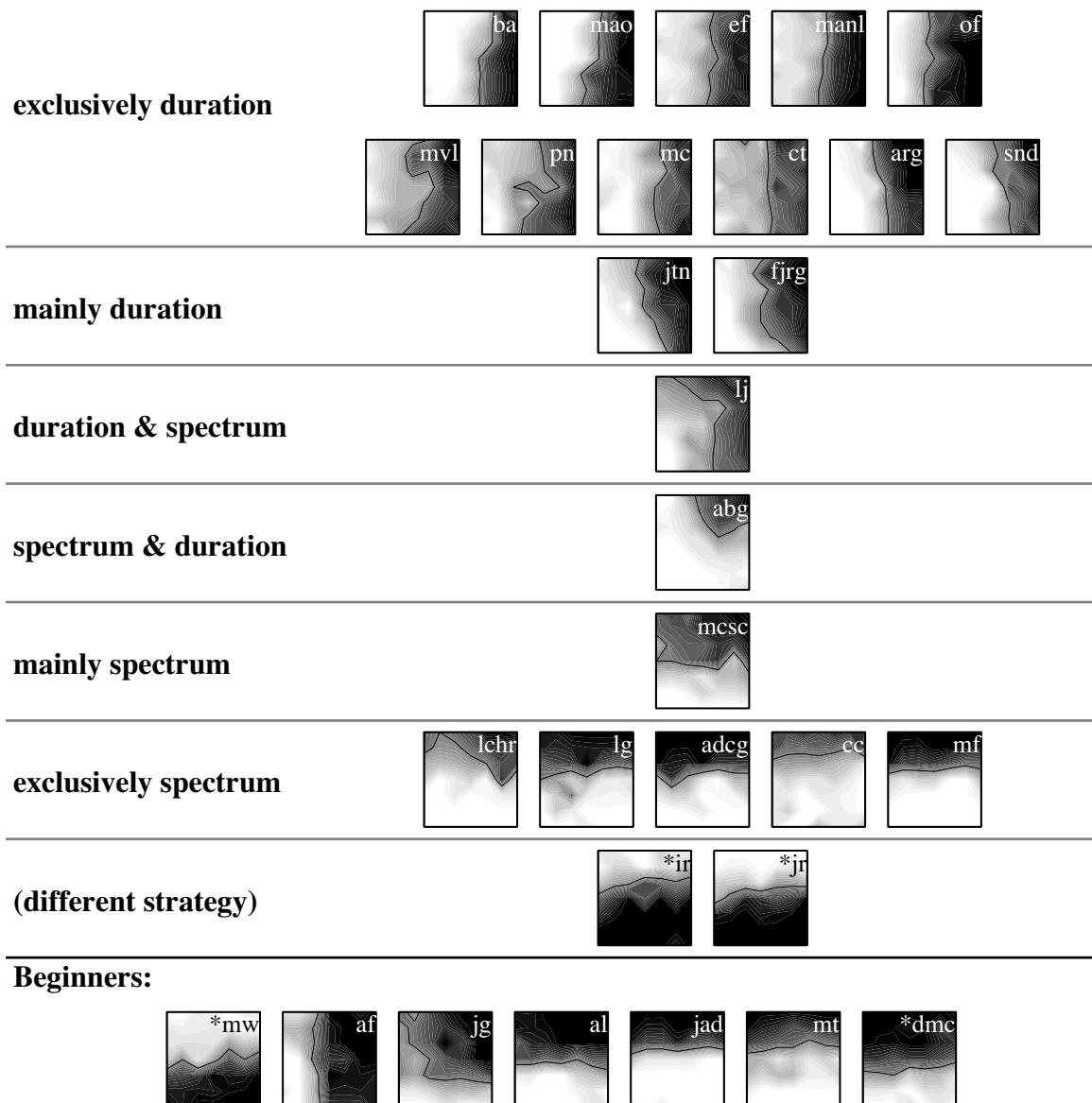
Fig. 6. Identification results of 30 Spanish listeners on the English /i/-/ɪ/ contrast. In each square, duration runs from 83 ms (left) to 176 ms (right), and F1 runs from 480 Hz (bottom) to 344 Hz (top), as in Figure 3.

1.5. L2 perception of the contrast in the two dialects

The same experiment was performed with the group of Spanish learners of English. The results are in Figure 6 and Table 2.

Table 2. Data on the 30 Spanish listeners. Org = origin (S = Spain, LA = Latin America), Edu = higher education in English (FCE = First Certificate in English, EP = student of English philology, CAE = Certificate in Advanced English, Dip = Diplomacy, - = no higher English education reported); Tim Sco/Eng/Ire/Zim = time spent in Scotland, England, Ireland, Zimbabwe (in months);[7] . * = perverse strategy.

Subject	Org	Edu	Tim Sco	Tim Eng	Tim Ire	Tim Zim	Duration reliance (%)	Spectral reliance (%)	Reliance ratio	Cues relied on
ba	S	-	4	0	0	0	94	-17	(-5.53)	
mao	S	Dip	0	1	0	0	96	-17	(-5.65)	
ef	S	EP	0	0	2	0	86	-10	(-8.6)	
manl	S	EP	46	0	0	0	90	-7	(-12.9)	
of	LA	-	1	108	0	0	99	-7	(-14.1)	
mvl	LA	-	48	36	0	0	86	-4	(-21.5)	**excl. duration**
pn	S	-	5	13	0	0	99	-4	(-24.8)	
mc	S	EP	0	0	2	0	79	-3	(-26.3)	
ct	S	FCE	0	0	6	0	67	9	7.444	
arg	S	EP	0	0	0	0	94	20	4.7	
snd	S	EP	0	0	0	0	89	20	4.45	
jtn	S	CAE	0	1	0	0	99	27	3.667	**mainly duration**
fjrg	S	EP	0	0	0	0	94	27	3.481	
lj	S	EP	0	6	0	0	70	36	1.944	**dur. & spectrum**
abg	S	-	0	2	0	0	40	47	0.851	**spectrum & dur.**
mcsc	S	EP	0	1	0	0	29	81	0.358	**mainly spectrum**
lchr	S	-	1	1	0	60	14	77	0.182	
lg	LA	-	114	0	0	0	0	96	0	
adcg	S	-	48	0	0	0	-1	100	(-0.01)	**excl. spectrum**
cc	LA	-	96	36	0	0	-1	66	(-0.02)	
mf	S	-	72	0	0	0	-6	100	(-0.06)	
*ir	LA	-	36	0	0	0	19	-90*		*
*jr	LA	-	6	312	0	0	13	-100*		*
Beginners:										
*mw	LA	-	1	0	0	0	16	-94*		*
af	S	-	0	0	0	0	87	11	7.909	excl. duration
jg	S	-	0	1	0	0	37	77	0.481	mainly spectrum
al	S	-	0	0	0	0	13	94	0.138	excl. spectrum
jad	S	-	0	0	0	0	-7	97	(-0.07)	excl. spectrum
mt	S	-	0	0	0	0	-10	100	(-0.10)	excl. spectrum
*dmc	S	-	0	0	0	0	-11	94	(-0.12)	excl. spectrum

---

[7] Time spent in the US was not included, because it was never more than a few weeks.

A language background questionnaire showed that the subjects had had 1 to 15 years of formal English instruction in their home countries and had spent 0 to 27.5 years in English-speaking countries. All reported using Spanish as well as using English with native speakers. We labelled seven subjects as 'beginners', namely those that had spent less than two months in English-speaking countries and reported no higher education in English. Most of the beginners did not seem to know the difference between the pronunciations of *ship* and *sheep* (though they must have learned the orthographies during formal instruction), and they probably created on the fly a strategy based on the only division that their L1 perception allows, namely that between Spanish /e/ and /i/. Thus, 5 of the 7 beginners exhibit a horizontal boundary; it is telling that subject **mw** perversely assigns the higher part of the stimulus continuum to the *ship* picture, the lower half to *sheep*. Subject **dmc** had the same reversal in an identification test on the diagonal of Figure 3 (which preceded the experiment reported on here), though the test on the whole square (Figure 6) had him guessing correct. Subjects **al**, **jad**, and **mt** made the correct guess throughout. Subjects **af** and **jg** showed duration reliance and cue integration, respectively; only these two may have had the correct lexical contrast.

The 23 more advanced subjects manifest a bimodal distribution: most use spectral information only or duration information only. The division falls primarily along the lines of their target dialect: Table 2 shows that students of English philology and others with a higher education in English tend to go in the duration group, probably because the variety usually taught in this setting in Spain is a southern English standard. Of the remaining listeners, those who had spent more time in Scotland (or in Zimbabwe, which has a low /ɪ/ in *ship* as well; Wells 1982b) than in other English-speaking countries tended to be spectral listeners, while those who had spent more time in England (or in Dublin, which has a high /ɪ/ as well; Wells 1982a) tended to be duration listeners. Only subjects **ir** and **jr** exhibited a completely diverging strategy (perhaps orthographic, thinking that <ee> must be a long mid vowel, and <i> a short high vowel, although the response categories were shown as pictures of a sheep and a ship).

Table 3 compares the cue reliances for the (non-beginner, non-perverse) Spanish listeners with those of the two groups of native English listeners (from §1.4). More than half of the Spanish listeners with a Southern English target rely mainly on duration, a pattern not found in either of the L1 groups. With Figures 1 and 2, we conclude that the differences in L2 categorization are in the same direction as the differences manifested in the two L2 production environments (which are the same as the L1 environments in §1.1), but that these differences in L2 categorization are far more extreme than the differences in L1 categorization we found in §1.4.

Table 3. Comparison of L1 and L2 cue integration on the English /i/-/ɪ/ contrast.

	Scottish L1	Southern L1	L2
Duration only	0	1	11
Duration mainly	0	2	2
Duration and F1	1	2	1
F1 and duration	0	3	1
F1 mainly	3	7	1
F1 only	16	5	5
Total	20	20	21

1.6. Interpretation of the results: L2 development specific to the target dialect

In Table 3, we saw that most L2 listeners (16 out of 21) use only one acoustic cue for categorizing /i/ and /ɪ/, unlike at least the Southern English natives, who typically integrate the two cues (14 out of 20). Therefore, it can be said that the L2 listeners typically use a different strategy from that of the L1 listeners when having to categorize vowels, at least those with a Southern English target.

We claim that the strategies of the L2 learners aim at minimizing the probability of perceptual confusion, as in §1.2. The first advantageous strategy is *not* to start where L1 learners start, namely without any vowel categories at all. It is much better for Spanish learners of English to *reuse* their five native vowel categories /i, e, a, o, u/ in their initial perception of English. Even if this is obviously far off from what is necessary to speak a target language with 13 vowels proficiently, the learners will be able to handle a five-way contrast, which is better than no contrast at all. This category reuse is an instance of transfer of language-specific structures (Polivanov, 1931). The Full Transfer hypothesis (Schwartz and Sprouse, 1996) holds that linguistic experience completely determines the initial state for the learners, i.e. that learners will transfer all of their L1 knowledge to their L2 grammar; this behaviour gives L2 learners a head start.

Despite their similar initial L2 grammars, all L2 listeners must tune their perceptual strategies to their specific new production environments, if they are to become optimal perceivers. If the beginning Spanish learner of English, with her five perceivable vowels, enters the Scottish production environment, she will map /i/ and /ɪ/ to her Spanish /i/ and /e/, respectively, since these are the two categories closest to the average[8] realizations of the vowels in the target language (from §1.4, the Scottish boundary is about 400 Hz; according to Bradlow, 1996, the Spanish boundary lies near 430 Hz). This is shown in Figure 7, which also shows that the Spanish learner of the Southern English variety will initially map both /i/ and /ɪ/ to her native /i/ category.

The perceptual behaviour of beginning L2 learners can be seen as a case of foreign-language speech perception. The Perceptual Assimilation Model (PAM) by Best (1995) distinguishes the two strategies depicted in Figure 7. For the L2 Scot, each of the two average vowel tokens falls inside the production/perception space of a separate L1 category; Best calls this *two-category assimilation*. For the L2 Southerner, by contrast, the two average vowel tokens fall inside the same L1 category; Best calls this *single-category assimilation*. Although the PAM was devised to account for cross-language speech perception and not for L2 development, it can be used to make predictions about the learnability of non-native contrasts (as suggested by Guion, Flege, Akahane-Yamada & Pruitt 2000 and Author 2001), and taken to predict the initial state of a beginning L2 learner.

---

[8] We would like to distinguish the concept of *average token value* from the concept of *prototype*. Prototypes are tokens that are least likely to be perceived as anything else than the category at hand. For that reason, prototypical values tend to be rather extreme. For instance, the F1 value of the /i/ prototype is much lower than the F1 value of the average /i/ token (Johnson, Flemming, & Wright, 1993; Frieda, Walley, Flege, & Sloane, 2000).
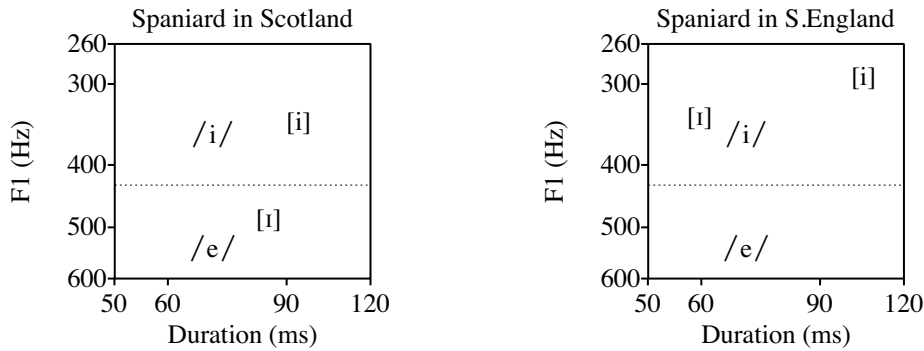
Fig. 7. Perceptual assimilation by Spanish listeners of different varieties of English. In square brackets: the average realizations of /i/ and /ɪ/ for the environment at hand (taken from the average values in §1.1). Between slashes: the average Spanish vowels (Bradlow, 1996: F1 values of 360 and 540 Hz). The dotted line is the Spanish category boundary (Bradlow, 1996: 430 Hz). To the left: two-category assimilation in Scotland. To the right: single-category assimilation in Southern England.

Although the two L2 listeners do the best thing they can do given their linguistic experience, their behaviour is not yet good enough for the new environments. Rather than staying with their five original L1 vowels, it would be advantageous for them to either move their category boundaries to match the target language boundaries, or to somehow create more categories. Spanish learners of Scottish English would have to move their category boundary, perhaps from 430 to 400 Hz, as in Figure 8 (left), which also shows that the centres of the two reused categories have shifted. The availability of this boundary shift is an instance of access to L1-like learning mechanisms. The Full Access hypothesis (Schwartz & Sprouse, 1996) holds that learners are able to make use of the same language acquisition strategies that are used in L1 acquisition in order to turn their initial state into a more target-like grammar on the basis of the input (strategies such as category formation, boundary shift, category reduction, and perhaps category split and merger).

Spanish learners of Southern English need a different strategy. They can choose between splitting the /i/ category into two new vowels /i/ and /ɪ/, or form a new feature contrast, perhaps a length distinction, i.e. a new /short/-/long/ contrast that does not occur in their native language. And indeed it seems that the latter is what we have seen L2 learners of Southern English actually do: the group of duration reliers does not distinguish the spectral properties of /i/ and /ɪ/, which suggests that their representations for these vowels are /i, long/ and /i, short/, respectively. Figure 8 (right) shows the results of applying this strategy.
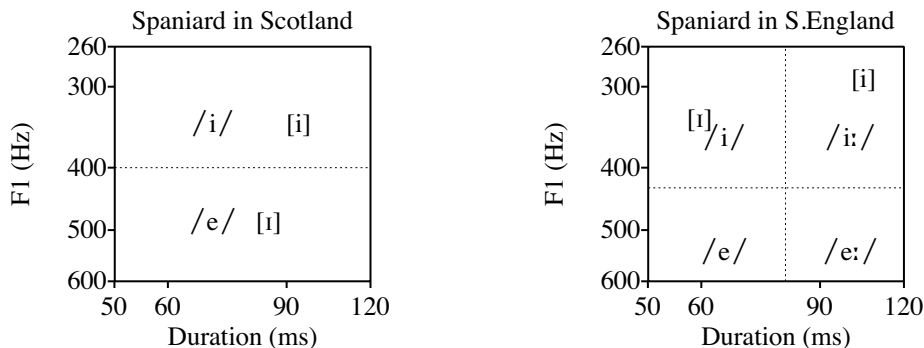

Fig. 8. Further L2 perceptual optimization in the two environments.

It remains to be explained why these listeners choose the new length distinction rather than split their /i/ category. Bohn (1995) claims that it has to do with a universal availability

of duration as a source of lexical distinctions, which is called upon if the L1 has insufficient spectral distinctions. Bohn tacitly assumes that Spanish speakers have a single duration category, and any explanation for their behaviour would have to address the question why learners split this single duration category rather than the single high front vowel category. For Bohn, the explanation is that there is something universally special about duration. But we claim that duration is *not* special: these learners start not with *one* duration category, but with *no* duration category at all, and they use L1 acquisition strategies to divide the duration continuum into two categories, probably by means of distributional learning (Authors & SomebodyElse, 2003). Thus, creating a length contrast is a case of *category formation*, which is a typical L1-like acquisition strategy, whereas creating a spectral contrast is a case of *category split*. Thus, the simple hypothesis of Full Access must already predict that category formation in general, and the formation of a duration contrast in this particular case, is available to L2 learners. By contrast, category split has, to our knowledge, not been proposed as a common L1 learning strategy, so the hypothesis of Full Access is compatible with a general unavailablity of category split to L2 learners. We predict that learners who implement a new length distinction will be able to cope with five short and five long vowels, thus at one stroke extending their mastery of English from five vowels to ten. Tentatively, we can say that the resulting assimilation pattern will be as follows:

/iː/ → /i, long/	/ɪ/ → /i, short/
/ei, ɜː/ → /e, long/	/ɛ/ → /e, short/
/ɑː/ → /a, long/	/ʌ, æ/ → /a, short/
/ou, ɔː/ → /o, long/	/ɔ/ → /o, short/
/uː/ → /u, long/	/ʊ/ → /u, short/

This situation of having a 5x2 vowel system is not found in the learners' native language, which has no length distinction at all, nor in the target language, in which nearly all vowels are distinguished by spectral differences as well. This L2-specific phenomenon seems to be a big challenge for a formal theory that relies on a gradual learning algorithm, since at first sight one would think that such a theory predicts a learning path that interpolates linearly between the native language and the target language. We will see, nevertheless, that if we separate the role of the grammar (the categorization system) from the role of the inventory of categories (the lexical symbols), our Optimality-Theoretic perception model with the Gradual Learning Algorithm does predict the perceptual behaviour of Spanish learners of English.

2. Modelling the knowledge behind vowel categorization

As mentioned in the introduction, our main goal is to link L2 speech perception research with linguistic theory. To that end, this section presents a formal account of the speaker's knowledge behind the production-perception relation observed in §1. The development of this knowledge in L1 and L2 is discussed in §3.

We claim that the knowledge behind the perception process is a formal *perception grammar* that determines an optimal output (e.g. a phonological category) on the basis of a given input (e.g. an auditory event). In our formalization of this grammar, the decision scheme works according to the constraint-based framework of Optimality Theory (OT, Prince & Smolensky, 1993), more specifically, its probabilistic version (stochastic OT; Author, 1998). While in the general case listeners map multiple auditory cues on multiple phonological contrasts simultaneously (e.g., vowel duration plays a role in the perception of the vowel contrast as well as in the perception of the voicing of the following consonant), we

only consider here the integration of *two* auditory cues into *one* phonological contrast, which is already one step more complicated than the accounts of one-dimensional categorization presented by Author (1998: ch. 8) and Hayes (2001).

2.1. Formalizing adult vowel perception in English and Spanish

Traditionally in the OT literature, grammars map underlying forms to surface forms. Such grammars can be called *production grammars*. Here we are interested in *perception grammars*, which, with different kinds of constraints, map auditory events to phonological structures. Author (1998:164) proposed continuous constraint families for mapping auditory continua to discrete phonological categories. If these categories are arbitrary symbols maintained for purposes of lexical storage, we must assume that any value of any auditory continuum could in principle be mapped to any phonological category. Under this assumption, the general formulation of a simple constraint is "a value *x* on the auditory continuum *y* should not be perceived as the phonological category *z*". For the F1 continuum in native English, we have constraints like "an F1 of 260 Hz should not be perceived as /ɪ/" and "an F1 of 500 Hz should not be perceived as /ɪ/", and analogously for all other F1 values, and a similar constraint set for the category /i/. For the duration continuum, we have constraints like "a duration of 50 ms should not be perceived as /ɪ/" and "a duration of 120 ms should not be perceived as /ɪ/" (and the same for all other duration values, and for /i/). So we use four families of negatively worded constraints[9] for modelling the categorization of the two English vowels on the basis of two auditory continua.

So how do Scottish and Southern English listeners implement an optimal categorization? The knowledge behind a Scottish listener's perception of the auditory event [74 ms, 349 Hz] can be represented as a ranking of constraints. One possible ranking that does the trick is shown in Tableau 1. The top-left cell shows the auditory event, which is the input to the perception grammar. The cells below it show the candidates for the output of the perception grammar; only the two relevant vowel categories are shown as candidates. Only four constraints are relevant here. The highest ranked of these could be "349 Hz is not /ɪ/", perhaps because of the large distance between 349 Hz and the mean F1 for /ɪ/ (§1.1). When the auditory event [74 ms, 349 Hz] arrives, the tableau will select the candidate /i/ as the winner (i.e. as the actually perceived category) because this candidate violates the least high-ranked constraints.

[74 ms, 349 Hz]	349 Hz not /ɪ/	74 ms not /i/	74 ms not /ɪ/	349 Hz not /i/
/ɪ/	*!		*	
☞ /i/		*		*

Tableau 1. The perception of the auditory event [74 ms, 349 Hz] by a Scottish listener.

The knowledge underlying the perception of the same auditory event for the Southerner can be shown with a ranking like the one in Tableau 2. If her two F1 constraints are ranked in the reverse order from those of the Scot, as in this example, she will choose to perceive /ɪ/.

---

[9] For the present paper, we could equally well have used positively worded constraints like "an F1 of 260 Hz should be perceived as /i/", but we happen to know that such constraints do not work for cases in which more than one auditory continuum is mapped to more than two phonological categories.

[74 ms, 349 Hz]	349 Hz not /i/	74 ms not /i/	74 ms not /ɪ/	349 Hz not /ɪ/
☞ /ɪ/			*	*
/i/	*!	*		

Tableau 2. The perception of the same auditory event [74 ms, 349 Hz] by a Southern English listener.

For native speakers of Spanish, the most important constraints are those for the spectral features. For the three front vowels, we have three continuous constraint families. Examples of constraints are:

"an F1 of 200 Hz should not be perceived as /i/"
"an F1 of 1000 Hz should not be perceived as /i/"
"an F1 of 200 Hz should not be perceived as /e/"
"an F1 of 200 Hz should not be perceived as /a/"

There are analogous constraint families for duration.

The constraints must be ranked in a way appropriate for the perception of Spanish, e.g., they should result in an F1 boundary of 430 Hz between /i/ and /e/ (Bradlow 1996), and in an F1 boundary of 630 Hz between /e/ and /a/ (from our own preliminary measurement). Figure 9 shows a possible continuous ranking of the constraints against perceiving /i/, /e/, and /a/, as functions of F1 (modelled in the same way as in Author, 1998, ch. 8).
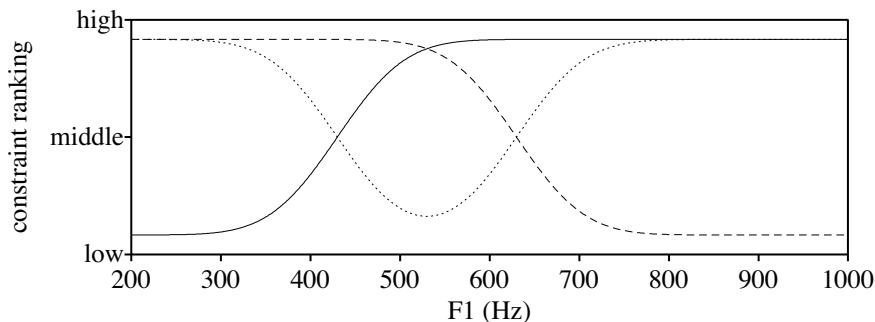

Fig. 9. The adult Spanish F1-to-front-vowel perception grammar. The solid curve is the continuous ranking of "don't perceive an F1 of *x* as /i/". The dotted curve is "don't perceive an F1 of *x* as /e/". The dashed curve is "don't perceive an F1 of *x* as /a/".

This figure contains enough information to read off the perception of Spanish front vowels on the basis of F1. For an input of [350 Hz], the figure shows that the highest ranked constraint is "don't perceive an F1 of 350 Hz as /a/". The constraint "don't perceive an F1 of 350 Hz as /e/" is ranked almost as high. The constraint "don't perceive an F1 of 350 Hz as /i/" is ranked lowest. As a result, the listener, when having to choose from the three candidates /i/, /e/, and /a/, will decide that /i/ violates the lowest ranked constraint and is therefore the best perception. The figure similarly shows that all F1 values below 430 Hz (one of the intersection points in the figure) are best perceived as /i/, F1 values between 430 and 630 Hz are best perceived as /e/, and F1 values above 630 Hz are best perceived as /a/.

Since Spanish vowels do not exhibit large duration differences, the duration constraints cannot be ranked too high. If they are ranked at "middle" in the ranking scale of Figure 9, they are hardly capable of contributing to the determination of the winning candidate.

2.2. The L2 initial state: perceptual assimilation

As we discussed in §1.6, the initial state of Spanish learners of English can be derived from their cross-language perceptual behaviour. We can derive from Figure 9 how a Spanish perception grammar would perceive the average tokens of Scottish and Southern English /i/ and /ɪ/, whose F1 values and durations were presented in Table 1. Tableaus 3 through 6 are based on the rankings that can be read off Figure 9, while assuming that all duration constraints are ranked at "middle". Thus, some F1 constraints will be ranked above the duration constraints, some will be ranked below.

[94.0 ms, 343 Hz]	343 Hz not /e/	94.0 ms not /e/	94.0 ms not /i/	343 Hz not /i/
/e/	*!	*		
☞ /i/			*	*

Tableau 3. The perception of the average Scottish /i/ token by a monolingual Spaniard (or by a beginning Spanish learner of English).

[84.8 ms, 485 Hz]	485 Hz not /i/	84.8 ms not /e/	84.8 ms not /i/	485 Hz not /e/
☞ /e/		*		*
/i/	*!		*	

Tableau 4. The perception of the average Scottish /ɪ/ token by a monolingual Spaniard.

[104.6 ms, 292 Hz]	292 Hz not /e/	104.6 ms not /e/	104.6 ms not /i/	292 Hz not /i/
/e/	*!	*		
☞ /i/			*	*

Tableau 5. The perception of the average Southern /i/ token by a monolingual Spaniard.

[59.7 ms, 337 Hz]	337 Hz not /e/	59.7 ms not /e/	59.7 ms not /i/	337 Hz not /i/
/e/	*!	*		
☞ /i/			*	*

Tableau 6. The perception of the average Southern /ɪ/ token by a monolingual Spaniard.

Through Tableaus 3–6, Figure 9 provides the explanation of the perceptual assimilation patterns in Figure 7.

Analogously to the usual OT interpretation of the Full Transfer hypothesis, we claim that the behaviour of beginning Spanish learners of Scottish or Southern English is explained by

their native Spanish set of vowel categories, their Spanish constraint set, and their ranking of the Spanish constraints. Thus, we claim that these learners perform according to Tableaus 3–6 as well. The Spanish in Scotland will perform rather well. They can reuse the Spanish /i/ as the arbitrary lexical symbol for storing words with English /i/, and they can reuse their native /e/ for storing words with English /ɪ/, i.e., the phonological part of their lexical representation of the word *ship* is |ʃep|. These learners thus exhibit two-category assimilation, the only problem being that their /i/-/ɪ/ boundary is at 430 Hz, not at 400 Hz as it is for the average native Scot of §1.4. By contrast, the Spanish in Southern England are in more trouble, since they will use the Spanish /i/ for storing English words with both /i/ and /ɪ/. Fortunately, Full Access will allow the learners to overcome entirely or partially these problems of boundary mismatch and single-category assimilation, as is modelled in §3.

2.3. Modelling gradual boundaries

Tableaus 1–6 suggest that a listener, when confronted with a certain auditory event, always maps it to the same category. Real listeners, however, exhibit gradual category boundaries (the grey areas in Figures 4–6). For this reason (and for learnability reasons, discussed in footnote 10), the actual version of OT that we use is *stochastic OT* (Author, 1998), in which *evaluation noise* is temporarily added to the ranking of each constraint at evaluation time. Thus, if two constraints are ranked at approximately the same height, the higher-ranked will win over the lower-ranked most of the time, but the lower-ranked will also win over the higher-ranked in a minority of evaluations. The effect of this in categorization is that the boundaries are gradual rather than rigid, just as they are in real listeners.

3. Modelling the L1 and L2 development of categorization

In §2.1, we showed that Optimality-Theoretic listeners can handle several examples of vowel categorization. But the number of constraints for the perception of continuous auditory dimensions is rather large, and the number of their possible rankings is extremely large. Our account would be unsatisfactory if we did not supply a theory of how listeners arrive at these constraints, and especially at optimal rankings of these constraints. Our answer is that listeners create categories (e.g. /ɪ/ and /p/) on the basis of distributional information (Authors & SomebodyElse, 2003), then use these categories to create phonological forms in their lexicon (e.g. |ʃɪp|) and mapping constraints in their perception grammar (e.g. "74 ms is not /ɪ/"), and finally optimize their constraint rankings by applying the Gradual Learning Algorithm (GLA: Author & SomebodyElse, 2001) to their perception grammar, driven by recognition in the lexicon.[10]

For L1 acquisition, we will start our modelling at the point where the infant has the lexicon in place. For example, a baby Scottish listener may at some point inadvertently entertain a grammar that would have been appropriate for a Southern British listener instead. As a consequence, she will perceive a token with a somewhat raised higher-mid vowel with cues [74 ms, 349 Hz] as /ɪ/, as shown with the pointing finger in Tableau 7.

---

[10] The Gradual Learning Algorithm (for stochastic OT) was preceded by Tesar & Smolensky's (1998, 2000) Error-Driven Constraint Demotion (for non-stochastic OT). That learning algorithm would not have worked for our findings, since it is not capable of handling variable mapping. The same token of [80 ms, 400 Hz] can represent /ɪ/ as well as /i/, and EDCD crashes if such tokens are presented to it in sequence.

[74 ms, 349 Hz]	349 Hz not /i/	74 ms not /i/	74 ms not /ɪ/	349 Hz not /ɪ/
☞ /ʃɪp/			←*	←*
√ /ʃip/	*!→	*→		

Tableau 7. Error-driven learning by the Gradual Learning Algorithm in an Optimality-Theoretic perception grammar.

However, her environment is Scottish, so this auditory event could well have been related to /ʃip/ rather than to /ʃɪp/. If so, the baby's recognition system will detect the error, perhaps by noting that the semantic context (a fluffy animal rather than a floating means of transportation) requires that she should have perceived /ʃip/, since that matches the phonological part of her lexical representation of the English word *sheep*. Now that the child knows that /ʃip/ would have been correct (as depicted with a check mark in Tableau 7), the child's GLA will change her perception grammar by raising the rankings of all the constraints violated in her incorrect winner and by lowering the rankings of all the constraints violated in the form that she now considers correct (as depicted by the arrows in Tableau 7). This increases the probability that she will perceive /ʃip/ on the next occasion of hearing an F1 of 349 Hz or a duration of 74 ms. The rankings are changed by only a small step (the *plasticity*) along the continuous ranking scale of stochastic OT (e.g. one thousandth of the high-low distance in Figure 9), but after a large number of perception errors involving auditory events containing either an F1 of 349 Hz or a duration of 74 ms, the rankings of the constraints will have become those of the adult Scottish listener in Tableau 1.

3.1. L1 simulations

We illustrate the development of L1 acquisition with the behaviour of a virtual Scottish listener and a virtual Southern English listener who grow up in virtual Scottish and Southern English environments, respectively. We will show how the virtual listeners' behaviour comes to be based on the relative reliability of the two cues in the virtual production environments.

In our simulations, we start at the stage in which the baby has just created different lexical representations for /ɪ/ and /i/. At that stage, distributional learning must have led to the creation of mapping constraints like "an F1 of 260 Hz is not /ɪ/", together with a reasonably good initial ranking of these constraints (Authors & SomebodyElse, 2003). Since the current article has no space to dwell on category emergence, we assume, rather artificially, a worst-case initial state in which both virtual babies start with all constraints ranked at the same height, so that they are equally likely to perceive any auditory event as /ɪ/ or as /i/ (which constraint is ranked highest at evaluation time is completely random).

We simulate the development of a Scottish and a Southern English listener by feeding them with input-output pairs drawn randomly from Gaussian distributions for the appropriate production environment. We assume that the distributions are centred about the mean F1 and duration values for the Scottish and Southern English speakers (Table 1). We choose fixed standard deviations of 0.2 octaves for F1 and 0.4 doublings for duration, for both vowels and both English varieties. These values are large enough to ensure that a wide range of duration-F1 pairs will occur in the Scottish and Southern environments (the distributions were shown in Figure 2). For computational reasons, both the duration continuum and the F1 continuum were divided up into 21 discrete values. The lowest duration value is 50 ms, the 21st is 120 ms, and the other 19 are evenly distributed between 50 and 120 ms along a logarithmic scale.

The same goes for the 21 F1 values, which are distributed evenly from 260 to 500 Hz along a logarithmic scale. The number of relevant constraints, therefore, was 84 (= 21 steps x 2 continua x 2 categories).

Each virtual listener received 1000 data per virtual month, and changed some constraint rankings every time she detected a mismatch between her perceived category and the correct category recognized by the lexicon. The standard deviation of the evaluation noise (§2.3) was held constant at 2.0, and the plasticity (the amount by which constraints are moved up or down the ranking scale after the detection of an error) decreased with age: it was 1.0 during the first 10 months, 0.1 during the next 90 months, and 0.01 during the remaining 900 months (therefore, learning is fast at the beginning, and slower but more accurate later on). Figure 10 shows the perceptual performance of the virtual Scot and the virtual Southerner in five stages. As in Figures 2, 4, 5 and 6, black areas stand for /i/ perceptions, white areas for /ɪ/ perceptions, and the black curve is the 50% category boundary line. These pictures were computed by running each of the 21x21=484 possible auditory events through the perception grammar a thousand times, keeping the evaluation noise at the same level as during learning. For instance, the token in the centre of the square (77 ms, 361 Hz) is perceived by a simulated Southern listener more often as /ɪ/ than as /i/, because the constraint against perceiving 77 ms as /ɪ/ and the constraint against perceiving 361 Hz as /ɪ/ are both ranked a bit below the constraint against perceiving 77 ms as /i/ (not too far below, since candidate /i/ does sometimes win, as we can see from the fact that the area in the centre of the square is light grey rather than white).
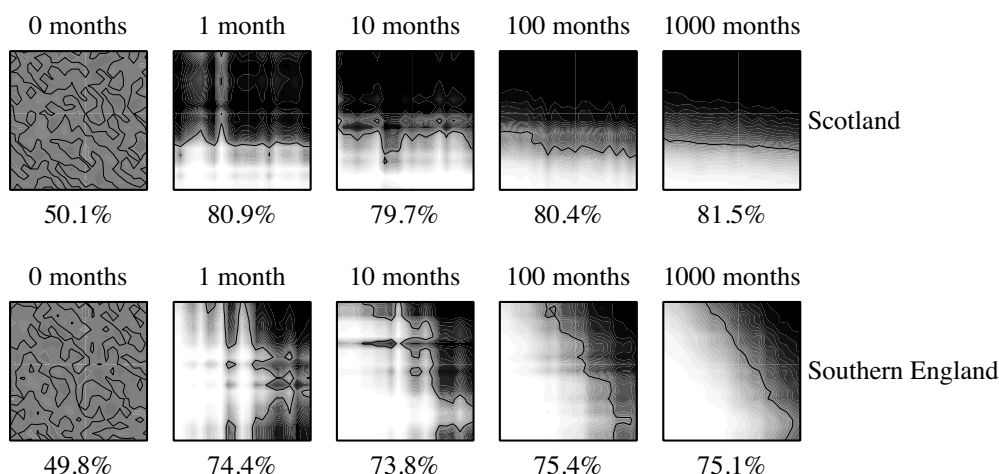

Fig. 10. Perceptual development of a virtual Scottish (top) and a virtual Southern (bottom) listener. Horizontal axes: duration from 50 ms (left) to 120 ms (right). Vertical axes: F1 from 500 Hz (bottom) to 260 Hz (top). Black = /i/, white = /ɪ/.

For each stage, a measure of perceptual proficiency is computed by drawing 100,000 tokens from the appropriate environment (i.e. 100,000 /i/ and /ɪ/ tokens with F1 and duration values distributed according to the Scottish or Southern environment) and running these tokens through the learner's perception grammar, leading to 100,000 perceptual outputs. By comparing these outputs to the intended vowels, we can calculate the fraction of correct responses. In the initial state, the learners score no better than 50% correct, since every token is as likely to be perceived as /ɪ/ as as /i/. Both virtual listeners gradually improve in distinguishing /ɪ/ from /i/. Even after 1 virtual month, tokens that lie around the centres of the categories are already identified quite well, which leads to a performance level that is close to that of the final stage. The locations and slopes of the boundaries in the final

stages in Figure 10 compare well with the optimal ones in Figure 2.[11] We conclude that we are able to model the knowledge behind an optimal strategy for vowel categorization, as well as the acquisition of this knowledge.

3.2. Modelling L2 development

The issues involved in L2 phonemic categorization have been mainly addressed by speech perception researchers, and hardly by formal linguistic theorists. The two dominant approaches in the study of L2 perception are the Speech Learning Model (SLM, Flege 1995) and the Perceptual Assimilation Model (PAM, Best 1995), which was already mentioned in §1.6. The two models predict that (and how) L1 linguistic experience determines the behaviour of non-native listeners (PAM) and L2 learners (SLM) confronted with L2 sound contrasts. Neither the SLM nor the PAM, however, is currently able to give an accurate and complete developmental account of L2 speech perception: Guion, Flege, Akahane-Yamada, & Pruitt (2000) conclude that the SLM will have to be extended non-trivially before being able to account for early stages in L2 development; and Best, McRoberts, & Goodell (2001) do succeed in accounting for the initial state for L2 learners, but do not address further development. We propose here a formal linguistic model of the underlying mechanism of L2 phonemic categorization, thereby accounting for PAM's generalizations about the initial state and SLM's generalizations about later developmental stages.

We will show that our model is capable of explaining and predicting the strategies and developmental stages manifested in learning to optimize L2 perception in the specific target environments. For neither group of Spanish learners of English is the initial behaviour entirely appropriate for the target language. According to the Full Access hypothesis, the learners will overcome these problems in a way similar to how L1 acquirers achieve this. In our case, they will use the Gradual Learning Algorithm and category formation.

For Spanish speakers in Scotland, a small mismatch will occur in perception. First-formant values around 420 Hz will have been intended as /ɪ/ in most of the Scottish utterances, but perceived as /i/ by Spanish beginners most of the time. In such a situation, the Gradual Learning Algorithm will change the perception grammar, as in Tableau 8.

[420 Hz] intended /e/	420 Hz not /e/	420 Hz not /i/
☞ /ʃip/		←*
√ /ʃep/	*!→	

Tableau 8. Acquisition of vowel categorization by a Spanish listener in Scotland.

The learner perceives /ʃip/, but notes that the semantic context requires that she should have perceived /ʃep/, since that is the phonological part of her lexical representation of the English word *ship*.[12] Tableau 8 shows that she will take action in such a way that she is more

---

[11] The more or less vertical boundary line in the lower right corner of the Southern figures is no coincidence; it is caused by the utter rarity of such high F1 values in the environment, and reminiscent of the behaviour of many of the Southern English listeners in Figure 5. Apparently, these Southerners do not necessarily judge something that spectrally sounds like /ɛ/ as being closer to /ɪ/ than to /i/. They therefore rely only on duration for such low sounds, judging [ɛː] as being more similar to [iː] than to [ɪ].

[12] It is of course not necessary to have actual minimal pairs in the lexicon. Perceiving an intended *shift* as the non-word *sheeft* also suffices for the recognition system to issue a protest, and the same learning step will take place.

likely to perceive /e/ at the next [420 Hz] token. Thus, boundary shifts are handled entirely by the learning algorithm.

For Spanish speakers in Southern England, the task seems to be more complicated, since they seem to have to 'notice' that their single /i/ category represents two different phonemes in the target language. Such 'noticing' *could* be based on the realization that the tokens of some lexical entries cluster around different auditory cue values than others. But we claim that they do not have to notice such complicated correlations: they only have to 'notice' that the distribution of English vowel durations has two peaks. As in §1.6, we propose that the L2 learners start with a virgin duration continuum (i.e. zero categories rather than one, with no duration-to-category mappings yet), like all infants do, and that two categories are created as soon as the learners establish that the distribution is bimodal (analogously to the explicit L1 OT modelling by Authors&SomebodyElse 2003). At that point, the learners will add a binary length contrast to their grammar, by introducing 'phonetic' mapping constraints like the following:

"a duration of 50 ms should not be perceived as /short/"
"a duration of 200 ms should not be perceived as /short/"
"a duration of 50 ms should not be perceived as /long/"
"a duration of 200 ms should not be perceived as /long/"

The initial ranking of these constraints is based on the results of distributional learning, so that the listener will at this point already have a reasonable proficiency in mapping auditory duration to phonological length. The L1 model of Authors&SomebodyElse (2003) predicts that the learner will later introduce less-phonetic mapping constraints that map other auditory continua to phonological length, e.g. "an F1 of 300 Hz should not be perceived as /short/", thereby rendering the categories /short/ and /long/ as abstract as the categories /ɪ/ and /i/; we will ignore this later step here (but see §3.5).

After the creation of the /short/–/long/ contrast, the learners can start using the two length categories in their lexical representations. Thus, they will represent Southern /ɪ/ as /i, short/ and Southern /i/ as /i, long/. Once the learners have a correct lexicalization of length, the appropriate mapping from continuous duration to the binary length contrast will be achieved by the Gradual Learning Algorithm. Tableau 9 shows what happens if an intended Southern English |ʃip| 'sheep', which has the representation |ʃiːp| for the learner, is pronounced appropriately with a duration of 104.6 ms, but the learner has a non-optimal constraint ranking that leads her to perceive the vowel as /short/ rather than /long/. She will perceive /ʃip/ and notice a mistake, since the word that she should have recognized is |ʃiːp| 'sheep' rather than |ʃip| 'ship'. She will take action and rerank some constraints with the GLA, thus making it more likely that she will perceive [104.6 ms] as /long/ on the next occasion.

[292 Hz, 104.6 ms] intended = /i, long/	292 Hz not /e/	104.6 ms not /long/	104.6 ms not /short/	292 Hz not /i/
☞ /ʃip/			←*	*
√ /ʃiːp/		*!→		*
/ʃep/	*!		*	
/ʃeːp/	*!	*		

Tableau 9. Acquisition of the categorization of length by a Spanish listener in Southern England.

3.3. L2 simulations

We simulated developmental sequences of typical Spanish learners of English who go to two different English-speaking countries at an age of 200 months. During their childhood in Spain, they will develop the /i/-/e/ distinction shown in Figure 11.

0 months	1 month	10 months	100 months	200 months

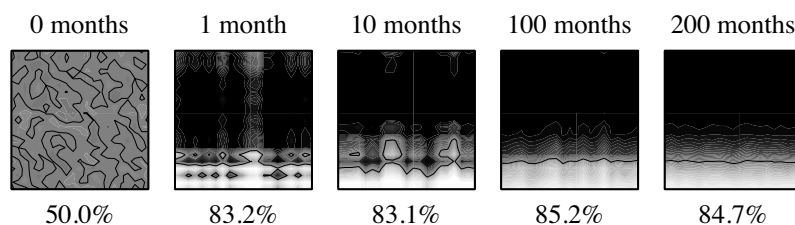| 50.0%	83.2%	83.1%	85.2%	84.7%

Fig. 11. Development of Spanish L1 perception. Black = /i/, white = /e/.

For this Spanish L1 simulation, we started with all F1 and duration constraints ranked at the same height, then applied 1000 inputs per virtual month while using the same evaluation noise and plasticity regime as for the L1 English in Figure 10. The listeners come to rely on the F1 cue only, with the duration constraints staying ranked approximately at their initial height.

The first listener goes to Scotland at an age of 200 months. We assume that she bases her perceptual assimilation on the auditory properties of the average Scottish /ɪ/ and /i/ tokens, so that she will equate Scottish /ɪ/ with her existing category /e/, and Scottish /i/ with her existing category /i/ (Figure 7, left). We also assume that this listener already has correct lexical representations for *ship* and *sheep*. Every virtual month we fed this virtual L2 Scot with 1000 vowels, drawn from the same Gaussian distribution used for the virtual L1 Scot in §3.1. We kept on doing this for 800 virtual months, keeping the evaluation noise constant at 2.0 and the plasticity constant at the low (adult) value of 0.01. The results are in Figure 12.

200 months	201 months	210 months	300 months	1000 months

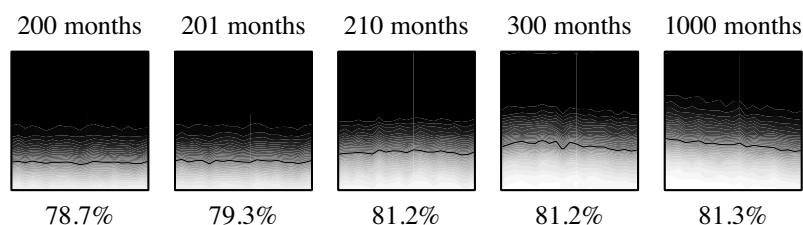| 78.7%	79.3%	81.2%	81.2%	81.3%

Fig. 12. The boundary shift of the simulated Spaniard in Scotland. Black = /i/, white = /e/ (= /ɪ/).

The L2 Scot starts out with a reasonable proficiency of 78.7% (the same boundary gave a fraction correct of 84.7% for the Spanish of Figure 11), then improves it toward an optimal value of 81.3%, by shifting the boundary from 430 to 400 Hz, then tilting it. We see that she becomes as proficient as the L1 Scot in Figure 10, who reached 81.5% correct.[13]

The second listener goes to Southern England at an age of 200 months. We assume that she bases her perceptual assimilation on the auditory properties of the average Southern /ɪ/ and /i/ tokens, so that she will equate both Southern /ɪ/ and Southern /i/ with her existing

---

[13] We did not discuss the representations for the Spanish learners of Scottish English in detail, since the /ɪ/-/i/ contrast does not tell us much about the entire Scottish vowel length system, e.g. that presented in McClure (1977), unlike in the Southern English case. If these learners have no constraints for mapping duration to vowel quality, they will only perform the boundary shift, not tilt the boundary line towards a 10% duration reliance. The data of the real learners in Table 4 does not allow us to decide whether they use duration to this small extent or not at all.

category /i/ (Figure 7, right). The results of 800 virtual months in a virtual Southern environment are in Figure 13.

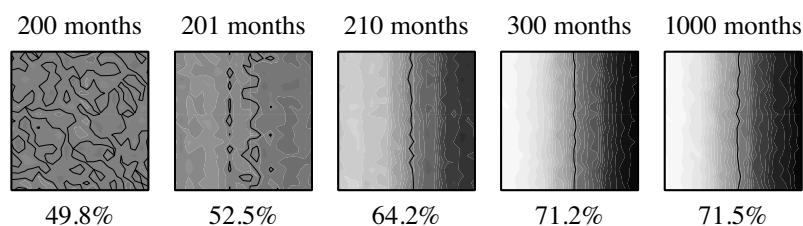| 200 months	201 months	210 months	300 months	1000 months
49.8%	52.5%	64.2%	71.2%	71.5%

Fig. 13. The phonemic separation of the simulated Spaniard in Southern England. Black = /i, long/, white = /i, short/.

The L2 Southerner starts out at chance level (like the infants of Figures 10 and 11; i.e., we assume that this listener has correct lexical representations for *ship* and *sheep* but a very poor categorization, i.e., we again ignore the effects of distributional learning), and improves to 71.5% correct while learning how to map duration to the new length contrast. This performance is rather less than that of the simulated Southerner of §3.1, who has 75.1% correct.

3.4. Comparison of simulated and real L2 listeners, and comparison with L1 learners

The simulated Spaniard in Scotland comes to rely on F1 only, which is what most of the real Scottish-oriented L2 learners do (e.g. subjects **lg** and **mf** in Figure 6). The simulated listener also shows a tiny boundary shift. We cannot detect such a shift between beginning and more advanced real L2 listeners (Figure 6), probably because the L1 Scottish boundary (Figure 4) is very close to the Spanish boundary to start with. This does not imply that developmental boundary shifts do not occur in L2 perception in general; they do, see §4.5.

The simulated Spaniard in Southern England comes to rely on duration only, which is what most of the real Southern-oriented L2 learners do (e.g. subjects **ef** and **of** in Figure 6). This is in stark contrast with what the simulated and real native Southerners do, who rely both on the spectral cue and on the duration cue. As hinted at in §1.6 and confirmed now by the simulations, the cause of the difference between L1 and L2 learners is the L2 learners' strategy of creating a length contrast. The sub-optimal result (71.5% rather than 75.1% correct) is due to the problematic representation of the vowel contrast, namely as /i, short/ versus /i, long/ rather than as the separate symbols /ɪ/ versus /i/ which would allow genuine cue integration. We conclude that our simulations reveal a formal explanation of the attested L2-specific behaviour.

3.5. Where do we go from here?

Our simulations for the Southern-oriented L2 learners (Figure 13) do not lead to a final stage of cue integration. Nevertheless, at least two learners in Figure 6 seem to have reached such a stage. The cause of the exclusive duration reliance in Figure 13 is our exclusive use of 'phonetic' mapping constraints such as "a duration of 110 ms is not /short/", i.e., the only auditory continuum that related to the phonological length feature in the simulations was duration. However, we can imagine that the interlanguage category /short/ could at some point become as abstract as the native English categories /ɪ/ and /i/. As stated in §3.2, the learner will later introduce less-phonetic mapping constraints that e.g. relate the phonological length feature to the auditory continuum of F1. As soon as such abstract constraints (like "an

F1 of 300 Hz is not /short/") become available, learners can start to integrate spectral and duration cues for the length contrast. However, since all 5x2 interlanguage vowels contain one of the feature values /short/ or /long/, a simulation of such a scenario would involve all 13 English vowels and all interlanguage vowels, and is outside the scope of this article. We can predict, though, that since all English high vowels are long, constraints for low F1 values, such as "an F1 of 260 Hz is not /short/", will become high ranked and contribute to a good perception of the /ɪ/–/i/ contrast for highly advanced learners.

4. Discussion: what is transferred and accessed, and what is not?

The formal model for L2 phonemic categorization advanced here bridges the gap between speech perception research and linguistic theory, by applying the explicit hypotheses of Full Transfer and Full Access to L2 speech perception for the first time. We defined Full Transfer as the transfer of L1 categories, L1 perceptual mappings, and L1 blank slates to the initial state of the interlanguage perception grammar, and we defined Full Access as the access to an L1-like category formation device and to an L1-like constraint reranking device. In this section, we will raise the question whether these five ingredients were really attested in the experiments, and to what extent our results contradict or confirm other theories.

4.1. Transfer of lexical elements: reuse of L1 categories in the L2 initial state

An important assumption in our model, following Polivanov (1931) and others, is that starting L2 learners reuse their L1 categories when creating L2 lexical representations.

   To show that our data confirm the assumption of category reuse, we have to consider the results for both groups of learners together. The results for the Scottish-oriented group (native-like discrimination) *could* be explained by the hypothesis that L2 learners are able to tune in immediately to the categories of any language, but this hypothesis would fail to account for the behaviour of the non-native-like Southern-oriented group. The results for the Southern-oriented group (exclusive duration reliance) *could* be explained by the hypothesis that learners start out with no categories at all, and then create new categories, perhaps first on the basis of a universally available dimension like duration, but this hypothesis would fail to account for the very good performance of the Scottish-oriented group. The only hypothesis that can account for both groups at the same time is that the Scottish-oriented group uses two already available categories, whereas the Southern-oriented group starts out with a single already available category and creates a new contrast later on. This confirms the assumption of category reuse, and thereby supports the Full Transfer hypothesis. The idea of category reuse is trivially compatible with Flege's (1995) Speech Learning Model (SLM), which states that an L2 learner has a single phonological space that consists of L1 and L2 categories, where the initial state has L1 categories only.

4.2. Transfer of the grammar: reuse of L1 perceptual mappings in the L2 initial state

Best's (1995) Perceptual Assimilation Model (PAM) states that high-level linguistic experience determines what listeners do when confronted with the cross-linguistic task of having to classify foreign language sounds with L1 categories. This model predicts that Spanish listeners with no experience in English will exhibit two-category assimilation in Scotland and single-category assimilation in Southern England (Figure 7): for a monolingual Spanish listener, typical tokens of Scottish /ɪ/ and /i/ would fall into her /e/ and /i/

categories, respectively, while typical tokens of Southern /ɪ/ and /i/ would all fall into her /i/ category. When combined with the PAM, Full Transfer implies that beginning Spanish learners of English will exhibit two-category assimilation in Scotland and single-category assimilation in Southern England. Since our listening subjects showed later reflexes of both types of perceptual assimilation, our perception experiment turns out to confirm both PAM and Full Transfer.

4.3. Transfer of holes in the grammar: reuse of L1 blank slates in the L2 initial state

Since L1 Spanish has never created discrete categories on the duration continuum, the Spanish perception of the duration continuum is still a blank slate with constraints that handle non-categorizing mappings (Authors & SomebodyElse, 2003). When listening to English, Spanish monolinguals will perceive duration in an acoustic, non-categorizing way (not as a mapping to a single discrete duration category, as Bohn 1995 seems to assume). Under Full Transfer, Spanish learners of Southern English will transfer this non-categorizing mapping to their interlanguage grammar. Our experiment shows indirect evidence for this transfer, because it shows evidence for access to a learning device specific to uncategorized continua, as we show in §4.4. The idea of reuse of blank slates can at least partially account for Flege's (1995) observation that *new sounds,* i.e. sounds in new areas of the phonetic space, are relatively easy to acquire.

4.4. Access to an L1-like acquisition device: new categorization of blank slates

The initial duration continuum for native Southern English infants is a blank slate. The non-categorizing mappings (§4.3) will lead to distributional learning), through which the infants will create two length categories on the basis of their native language input. Under Full Transfer, the initial duration continuum for Spanish learners of Southern English is just such a blank slate. Under Full Access, therefore, they will create two length categories, just like the natives (§1.6, §3.2). Our experiment shows evidence for this: beginners seem to have trouble with the length distinction, while more experienced learners have developed a lexical length contrast. While Bohn (1995) considered this behaviour to be evidence against transfer, our closer formal modelling now actually regards it as evidence in favour of access to an L1-like distributional learning device (and therefore in favour of the transfer of the blank slate).

4.5. Access to an L1-like acquisition device: constraint reranking in the interlanguage

We have been assuming that L2 learners have access to the Gradual Learning Algorithm, which takes the L2 learner through a sequence of constraint rankings in the same way as it does for an L1 learner. In our modelling and simulations, the observational result is boundary shift (§3.3) and the development of cue integration (§3.3, §3.5).

Our experiment does not show direct evidence of boundary shifts in L2 acquisition, since for Spanish learners of Scottish English, the initial position of the /e/-/i/ boundary (Figure 11, right) is already quite close to the position of the Scottish /ɪ/-/i/ boundary (Figure 10), and for Spanish learners of Southern English, the duration boundary that emerged from the simulations (Figure 13) could have been created by distributional learning alone (Authors&SomebodyElse 2003). However, boundary shifts have been attested in other work on L2 speech perception: Caramazza, Yeni-Komshian, Zurif & Carbone (1973) reported that native speakers of French who had begun to acquire English before their seventh birthday exhibited a large shift of the /b/-/p/ boundary; Flege & Eefting (1987) showed that the

perceptual /b/-/p/ boundary of Dutch learners of English depended on the language the learners thought they heard; Authors (2002) showed a similar language mode effect for the /ɛ/-/ɪ/ and /ɪ/-/i/ boundaries of Dutch learners of Spanish. The simulation in Figure 12 shows that our model will have no trouble formalizing boundary-shift effects.

Our experiment shows some evidence of L2 cue integration. A minority of the learners in Figure 6 have diagonal boundaries. In the model of Authors&SomebodyElse (2003), this situation can come about with the introduction of constraints that map F1 values to length categories (§3.5).

4.6. Extending the original definitions of Full Transfer and Full Access to perception

The hypothesis of Full Transfer, as defined by Schwartz & Sprouse (1996), can be regarded as claiming that the L2 learner transfers hidden representations (elements and hierarchies), and their mappings to and from overt forms. For syntax, the hidden representations are the functional categories (and the trees), and the grammar handles the mappings between these and word order. For semantics, the hidden representations are the semantic parts of lexical items, and the semantic categorization system handles the mappings between these and tokens in the real world (e.g., when learning the English word *ship*, Spanish learners may transfer the somewhat smaller semantic extent of the Spanish word *barco*). For phonological perception, we can say that the hidden representations are the phoneme categories (and metrical structures), and the perception grammar handles the mappings between these and auditory events. In §4.1 to §4.3, we have therefore seen evidence for Full Transfer, when applied to phoneme categories.

The hypothesis of Full Access, as defined by Schwartz & Sprouse (1996:41), claims that 'restructuring draws from options of UG'. For phonology, the phonemes should emerge in an L1-like fashion and the mappings should go through developmental stages allowed by Universal Grammar. The typological assumption of OT (Prince & Smolensky, 1993) is that every possible constraint ranking reflects a language allowed by UG. For interlanguage grammars, Full Access means, therefore, that every developmental stage should be identifiable with a constraint ranking. The GLA satisfies this requirement automatically, since it can do nothing besides changing the rankings of the constraints. In §4.4 and §4.5, we have therefore seen evidence for Full Access, when applied to phoneme categories.

5. Conclusions

When learning the English /ɪ/–/i/ contrast, Spanish learners behave differently depending on whether their target dialect is Scottish Standard English or Southern British English. Whereas the learners with a Scottish target behave like the Scottish natives, the learners with a Southern target typically exhibit a pattern of behaviour that is observationally dissimilar to anything that occurs in adult L1 Spanish as well as in adult L1 English and therefore superficially poses a challenge for the Full Transfer hypothesis. We have shown, however, that the formal model for L2 phonemic categorization advanced here successfully accounts for the attested optimal categorization in L1 acquisition as well as for the attested optimal and sub-optimal patterns in L2 acquisition. Our experiment and its modelling and simulations has the combined predictive power of Best's Perceptual Assimilation Model, Flege's Speech Learning Model, and Schwartz & Sprouse's Full Transfer and Full Access hypotheses. Our formalization thereby provides the linguistic mechanism that underlies the generalizations forwarded by several previous models of L2 speech perception.

References

Best, C. (1995). A direct realist view of cross-language speech perception. In W. Strange (ed.), *Speech perception and linguistic experience: Theoretical and methodological issues*. Baltimore: York Press. 171–204.

Best, C., G. McRoberts, & E. Goodell (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Journal of the Acoustical Society of America* 109, 775–794.

Bohn, O.-S. (1995). Cross language speech perception in adults first language transfer doesn't tell it all. *Speech perception and linguistic experience: Theoretical and methodological issues*, edited by Winifred Strange, 279–304. Baltimore: York Press.

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

Bradlow, A. (1995). A comparative study of English and Spanish vowels. *Journal of the Acoustical Society of America* 97, 1916–1924.

Bradlow, A. (1996). A perceptual comparison of the /i/-/e/ and /u/-/o/ contrasts in English and in Spanish: Universal and language-specific aspects. *Phonetica* 53, 55–85.

Broselow, E., S.-I. Chen, & C. Wang (1998). The emergence of the unmarked in second language phonology. *Studies in Second Language Acquisition* 20, 261–280.

Caramazza, A., G. Yeni-Komshian, E. Zurif, and E. Carbone (1973). The acquisition of a new phonological contrast: The case of stop consonants in French-English bilinguals. *Journal of the Acoustical Society of America* 5, 421–428.

Davidson, L. (1997). An Optimality Theoretic approach to second language acquisition. Unpublished Honours Thesis, Brown University.

Eckman, Fred R. (1981). On the naturalness of interlanguage phonological rules. *Language Learning* 31, 195–216.

Flege, J.E. (1987). The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics* 15, 47–65.

Flege, J.E. (1991). The interlingual identification of Spanish and English vowels: orthographic evidence. *Quarterly Journal of Experimental Psychology* 43A, 701–731.

Flege, J.E. (1995). Second language speech learning: Theory, findings and problems. In Winifred Strange (ed.), *Speech perception and linguistic experience: Theoretical and methodological issues*. Baltimore: York Press. 233–277.

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

Flege, J. & W. Eefting (1987). Cross-language switching in stop consonant production and perception by Dutch speakers of English. *Speech Communication* 6, 185–202.

Fox, R.A., J.E. Flege & M.J. Munro (1995). The perception of English and Spanish vowels by native English and Spanish listeners: A multidimensional scaling analysis. *Journal of the Acoustical Society of America* 97, 2540–2550.

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

Gerrits, E. (2001). *The categorisation of speech sounds by adults and children*. Ph.D. dissertation, University of Utrecht.

Gottfried, T. (1984). Effect of consonant context in the perception of French vowels. *Journal of Phonetics* 12, 91–114.

Gottfried, T., & P. Beddor (1988). Perception of temporal and spectral information in French vowels. *Language and Speech* 31, 57–75.

Guion, S., J.E. Flege, R. Akahane-Yamada, & J. Pruitt (2000). An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants. *Journal of the Acoustical Society of America* 107, 2711–2724.

Hancin-Bhatt, B. (1997). Extended full transfer/full access in L2 sound patterns: A prolegomena. In Leather, J. and James, A. (eds.), *New Sounds '97: Proceedings of the Third International Symposium on the Acquisition of Second-Language Speech*. Klagenfurt: University of Klagenfurt, 109–117.

Hancin-Bhatt, B. (2000). Optimality in second language phonology: Codas in Thai ESL. *Second Language Research* 16, 201–232.

Hancin-Bhatt, B., & R. Bhatt (1997a). Optimal L2 syllables: Interactions of transfer and developmental effects. *Studies in Second Language Acquisition* 19, 331–78.

Hancin-Bhatt, B., & R. Bhatt (1997b). Optimality Theory and L2 syllables revisited. Paper presented at New Sounds '97, September 8-11, 1997, Klagenfurt, Austria.

Hayes, R. (2000). Reranking stages in OT analysis of the acquisition of Japanese as a second language. *Carolina Working Papers in Linguistics* 1:1, 1–21.

Hayes, R. (2001). An Optimality-Theoretic account of novel phonetic category formation in second language learning. Ms., University of Arizona.

Hayes, R. (2002). The perception of novel phoneme contrasts in a second language: A development study of native speakers of English learning Japanese singleton and geminate consonant contrasts. In R. Hayes, W.D. Lewis, E.L. O'Bryan & T.S. Zamuner (eds.) *Language in cognitive science* [*Coyote Papers* 12]. University of Arizona Coyote Working Papers. 28–41.

Helmholtz, H. von (1910). *Handbuch der physiologischen Optik*. Vol. 3. Hamburg: Leopold Voss.

Hewlett, N., B. Matthews, & J. Scobbie (1999). Vowel duration in Scottish English speaking children. *Proceedings of the 14th International Congress of Phonetic Sciences* 2157–2160.

Johnson, K., E. Flemming, & R. Wright (1993). The hyperspace effect: Phonetic targets are hyperarticulated. *Language* 69, 505–528.

Labov, W. (1994). *Principles of Linguistic Change. Volume I: Internal Factors*. Blackwell, Oxford.

McClure, J.D. (1977). Vowel duration in a Scottish accent. *Journal of the International Phonetics Association* 7, 10–16.

Morrison, G.S. (2002). Perception of English /i/ and /ɪ/ by Japanese and Spanish Listeners: Longitudinal results. In G.S. Morrison & L. Zsoldos (eds.), *Proceedings of the North West Linguistics Conference 2002*. Burnaby, BC, Canada: Simon Fraser University Linguistics Graduate Student Association. 29–48.

Nittrouer, S. (1992). Age-related differences in perceptual effects of formant transitions within syllables and across syllable boundaries. *Journal of Phonetics* 20, 1–32.

Nittrouer, S. (1996). Discriminability and perceptual weighting of some acoustic cues to speech perception by 3-year-olds. *Journal of Speech and Hearing Research* 39, 278–297.

Nittrouer, S. (2001). Challenging the notion of innate phonetic boundaries. *Journal of the Acoustical Society of America* 110, 1598–1605.

Nittrouer, S., C. Manning, & G. Meyer (1993). The perceptual weighting of acoustic cues changes with linguistic experience. *Journal of the Acoustical Society of America* 94, 1865.

Nittrouer, S. & M.E. Miller (1997). Developmental weighting shifts for noise components of fricative-vowel syllables. *Journal of the Acoustical Society of America* 102, 572–580.

Peterson, G.E., & H.L. Barney (1952). Control methods used in a study of vowels. *Journal of the Acoustical Society of America* 24, 175–184.

Peterson, G.E., & I. Lehiste (1960). Duration of syllable nuclei in English. *Journal of the Acoustical Society of America* 32, 693–703.

Polivanov, E.D. (1931). La perception des sons d'une langue étrangère. *Travaux du Cercle Linguistique de Prague* **4**: 79–96. [English translation: The subjective nature of the perceptions of language sounds. xx. 223–237]

Prince, A., & P. Smolensky (1993). *Optimality Theory: Constraint interaction in generative grammar*. Technical Report TR-2, Rutgers University Center for Cognitive Science.

Ruhm, H.B., E.O. Mencke, B. Milburn, W.A. Cooper, & D.E. Rose (1966). Differential sensitivity to duration of acoustic signals. *Journal of Speech and Hearing Research* 9, 371–384.

Schwartz, B., & R. Sprouse (1996). L2 cognitive states and the Full Transfer/Full Access model. *Second Language Research* 12, 40–72.

Scobbie, James (1998). Interactions between the acquisition of phonetics and phonology. In K. Gruber, D. Higgins, K. Olsen & T. Wysochi (eds.) *Papers from the 34th Annual Meeting of the Chicago Linguistic Society* II. Chicago: Chicago Linguistic Society. 343-358.

Scobbie, J., A. Turk, & N. Hewlett (1999). Morphemes, phonetics and lexical items: The case of the Scottish vowel length rule. *Proceedings of the 14th International Congress of Phonetic Sciences* 1617–1620.

Stevens, S.S., J. Volkmann, & E.B. Newman (1937). A scale for the measurement of the psychological magnitude pitch. *Journal of the Acoustical Society of America* 8, 185–190.

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

Tesar, B., & P. Smolensky (1998). Learnability in Optimality Theory. *Linguistic Inquiry* 29, 229–268.

Tesar, B., & P. Smolensky (2000). *Learnability in Optimality Theory*. Cambridge, Mass.: MIT Press.

Wells, J.C. (1982a). *Accents of English 2. The British Isles*. Cambridge University Press.

Wells, J.C. (1982b). *Accents of English 3. Beyond the British Isles*. Cambridge University Press.