SSLA, **26**, 551–585. Printed in the United States of America.
DOI: 10.1017/S0272263104040021

BRIDGING THE GAP BETWEEN L2 SPEECH PERCEPTION RESEARCH AND PHONOLOGICAL THEORY

Paola Escudero
Utrecht University

Paul Boersma
University of Amsterdam

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 English) 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 second language (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.

Part of the phonology of a language consists of sound distinctions that the speakers perceive and produce. The sound distinctions of a language are signaled by a number of auditory properties (duration, static and dynamic spectral features, periodicity, noise, intensity) that integrate to constitute

We would like to thank the audiences of the 25th Penn Linguistics Colloquium (Philadelphia, 2001), Generative Approaches to Language Acquisition 2001 (Palmela), and EuroSLA 11 (Paderborn, 2001) and of talks in Lisbon, Amsterdam, Edinburgh, Utrecht, and Munich, all in 2001 as well, for their questions and remarks, and Rachel Hayes, Michael Sharwood Smith, and five anonymous *SSLA* reviewers for their helpful comments on earlier versions of this paper. This research was partially sponsored by grant 355-75-003 to Boersma from the Netherlands Organization for Scientific Research (NWO).

Address correspondence to paola.escudero@let.uu.nl or paul.boersma@uva.nl.

phonological contrasts. For instance, the contrast between the English vowels /i/ and /ɪ/ is one of vowel height (or tenseness) as well as length—that is, these vowels differ in spectral features (Peterson & Barney, 1952) as well as in duration (Peterson & Lehiste, 1960); and native speakers (NSs) rely on both of these auditory cues when categorizing these vowels (Bohn & Flege, 1990).

Perceptual cue weighting has crosslinguistic as well as developmental aspects. Crosslinguistically, 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 and Beddor showed 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. As they develop, 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). Because all infants start out with identical perception systems, we expect that the crosslinguistic 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 & Miller, 1997). For instance, Gerrits showed that 4-year-old Dutch children attend to duration cues much more than adult listeners do, for vowels as well as consonants.

Both crosslinguistic 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 NSs of the L2—in production as well as in perception. For instance, Bohn (1995) and Flege, Bohn, and 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 listeners use the two dimensions equally, whereas American English listeners have a preference for the spectral cues. It has also been shown that learners change their cue weighting as their experience with the L2 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 1 month of presence in Canada to a reasonably good reliance on spectrum or duration (but not on both) after 6 months. The question of interest to L2 researchers now is how to explain the perceptual behavior of the learners.

It has long been acknowledged that structural properties of the first language (L1) can be transferred to the interlanguage system. In the realm of phonological perception, Polivanov (1931) gave 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. The formal framework of generative grammar has offered several hypotheses for transfer. The most radical, explicit,

and simple hypothesis is what Schwartz and Sprouse (1996) have called Full Transfer, according to which a beginning L2 learner starts her L2 development by transferring her entire L1 system (except the phonological makeup of lexical entries) to her interlanguage system. It has also long been acknowledged that learners can develop toward a more targetlike performance by accessing L1-like learning mechanisms such as (in phonology) boundary shift, category creation, and the increase of the use of marked structures (e.g., Major, 1987). The most radical, explicit, and simple hypothesis is what Schwartz and Sprouse have called Full Access, according to which an L2 learner subsequently has access to all the principles of Universal Grammar (UG) and her entire language acquisition device, as she had when acquiring her L1.

In the domain of syntactic theory, Schwartz and Sprouse (1996) showed that some apparent counterexamples to the combined Full Transfer–Full Access hypothesis can be reanalyzed as supporting cases if one looks into the data deeply enough or if one's linguistic framework is explicit enough. Such results extend to phonology. Broselow, Chen, and Wang (1998), for instance, argued that if phonological production is modeled within the constraint-based framework of Optimality Theory (OT) of Prince and Smolensky (1993) rather than within a rule-based framework, an apparent counterexample against Full Transfer–Full Access found by Eckman (1981) can be reanalyzed as a case of Full Transfer followed by Full Access. In this and all other OT work in L2 production (Davidson, 1997; Hancin-Bhatt, 1997, 2000; Hancin-Bhatt & Bhatt, 1997; Hayes, 2000), Full Transfer is taken to imply that an L2 learner brings to the learning task the constraint set of her L1 as well as the ranking of these constraints, and 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 Error Driven Constraint Demotion (Tesar, 1995; Tesar & Smolensky, 1998, 2000) or the Gradual Learning Algorithm (Boersma, 1998; Boersma & Hayes, 2001), which were discussed with respect to L2 acquisition by Davidson (1997) and Hayes (2001), respectively. In the present article we argue that these results extend to speech perception. We show that, if we model language-specific perceptual knowledge as an OT grammar, an apparent counterexample against Full Transfer given by Bohn (1995) outside an explicit linguistic framework can be reanalyzed as grammatical transfer from the L1 and subsequent access to L1-like acquisition devices.

Given that 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 behavior 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. It is possible to apply constraint-ranking methods to the modeling of language-specific perceptual knowledge—that is, the listener's knowledge of how to map continuous auditory features to discrete phonolog-

ical surface structures such as segments and syllables. Polivanov (1931), for instance, explained 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 such as these can be straightforwardly formalized as OT perception grammars, as was done by Boersma (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 NSs of Japanese, and found that the NSs show categorical perception, the nonnative speakers show continuous perception, and the learners show a behavior in between these two kinds of perception. Hayes formalized the three kinds of perception with OT analyses and concluded that the OT perception model accounts for the attested fact that the learners' perception changes during development.

The present paper investigates the acquisition of the English /i/-/ɪ/ contrast by L1 Spanish learners, who have difficulty with it in production as well as in perception (Flege et al., 1997). This paper differs from Hayes's in five respects—all of which pose stronger challenges to the OT perception theory. First, for NSs 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's case, and it is not clear from the start that the theory can handle such more complicated cases. Second, we document a typical L2 learning stage that is attested neither in the learner's L1 nor in the L2 and is difficult to label as "intermediate." Third, we consider the acquisition of two closely related L2s—Scottish Standard English and Southern British English—rather than a single L2, and it is not clear from the start that the theory can account for the large differences that we find between the L2 developments of the Spanish learners of these varieties. Fourth, to successfully test the model, these different developmental patterns 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—we have to take into account the generalizations advanced by other empirical researchers on L2 speech perception.

In this article, we show that, for a complicated case with multiple auditory continua and multiple varieties of the L2, the stochastic version of OT (Boersma, 1998), together with the Gradual Learning Algorithm (GLA) that is

associated with it (Boersma & Hayes, 2001), is capable of accounting for the following observed behavior.

1. Both L1 and L2 listeners optimize their perception in accordance with the productions that they encounter—that is, their perception develops in such a way as to cope well with the specific characteristics of the ambient language.
2. L1 listeners arrive at an optimal perception—that is, they achieve the best possible use of the available auditory information.
3. Depending on the target dialect, L2 listeners may reach optimal perception or may manifest suboptimal optimization strategies that are specific to L2 acquisition.

In the present study, these three observed behaviors are demonstrated in experiments involving real listeners, modeled with OT perception grammars, and confirmed in computer simulations.

EXPERIMENT: L1 AND L2 PERCEPTION IN TWO ENGLISH DIALECTS

It is an empirical question whether L2 speakers can learn to perceive nonnative sounds in a way that resembles the behavior of NSs of the L2, 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 show that the perceptual behavior of the native listeners closely follows the relative use of the cues in the production of their dialect, whereas the behavior of the Spanish listeners only partially follows these ambient productions, with many of them behaving in a way that is not found in either of the native groups nor in their L1 Spanish.

Spanish Perception of the English /i/-/ɪ/ Distinction

Several studies have shown that Spanish speakers have problems perceiving certain English vowels. Fox, Flege, and Munro (1995) showed that Spanish listeners are not sensitive to durational information when categorizing vowels and therefore may have difficulty 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) showed that monolingual Spanish listeners associate both English /iː/ and English /ɪ/ with Spanish /i/. 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) showed that some inexperienced Spanish learners of English were 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*, given that they are capable of using these words as labels). Nevertheless, they made 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 paid as much attention to durational as to spectral differences when having to categorize English /i/ and /ɪ/ (spectral cue reliance: 50% and duration: 44%), whereas the English listeners exhibited a strong preference for spectral cues (spectral cues: 88% and duration: 9%).[1] In contradiction to Bohn (1995), however, Flege, Bohn, and Jang (1997, p. 465) 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% for the learners and 88% for the NSs). However, if a measured difference of 47% versus 88% is compatible with the absence of a true difference, the power of the statistical test must be so low that the measured difference is also compatible with a very large true difference (namely, 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 NS perception. Second, it presents contradictory conclusions with respect to the auditory cue weighting in Spanish learners of English. Our experiment attempts to resolve these contradictory findings.

Scottish English and Southern British English Native Production of the Contrast

There is 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. Regarding spectral vowel height, as expressed by the first formant frequency (F1), there is a large height difference between Scottish and Southern English /ɪ/. For instance, Labov (1994, p. 169) showed a Cockney speaker whose /ɪ/ is even higher than her /i/. As for vowel duration, the Scottish vowel length rule (Hewlett, Matthews, & Scobbie, 1999; McClure, 1977; Scobbie, Turk, & Hewlett, 1999) states that in many varieties and styles of Scottish English /i/ and /ɪ/ are equally long (as a transfer from the Scots language), although there could be slight differences between the two before fricatives or voiced consonants. Table 1 shows the F1 and duration values reported for a single Southern English and a single Scottish speaker by Escudero and Boersma (2003), averaged over eight words (produced 50 times each in a carrier sentence) that showed some realistic variation with respect to the voicing of the following consonant and the number of syllables.

Table 1. Duration and F1 for /ɪ/ and /i/ for a Scottish speaker and a Southern English speaker

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

Note. The duration was averaged across four consonantal contexts. The words were *ship*, *sheep*, *lid*, *lead*, *snicker*, *sneaker*, *filling*, and *feeling*.

METHOD[2]

Given that the present 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 L2 and with respect to the learners' degree of experience with the L2. Therefore, we selected 30 Spanish learners of English who differed in experience level and in target dialect (either Scottish or Southern British English). They were 15 women and 15 men, from various regions in Spain and various countries in South America, between 18 and 58 years of age. 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 they participated in the study. To compare their results to those of NS 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, between the ages of 23 and 35) and 21 speakers of Southern British English (10 women and 11 men, between the ages of 19 and 55; all reported to have grown up in the South of England and were judged to have various Southern accents, although 10 of them lived in Edinburgh at the time of the experiment). All of the Spanish and Scottish subjects and 10 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 stimuli for the experiment were isolated synthetic vowels. They were based on the auditory properties of natural exemplars of the vowels /i/ and /ɪ/, which were produced 10 times each by two Scottish English speakers. The average F1 of the naturally produced vowels was 485 Hz for /ɪ/ and 343 Hz for /i/. The average F2 (second formant frequency) was 1890 Hz for /ɪ/ and 2328 Hz for /i/. These values were taken as the basis for the top and bottom edges of the stimulus rectangle (Figure 1 shows the F1 values only). The six vertical

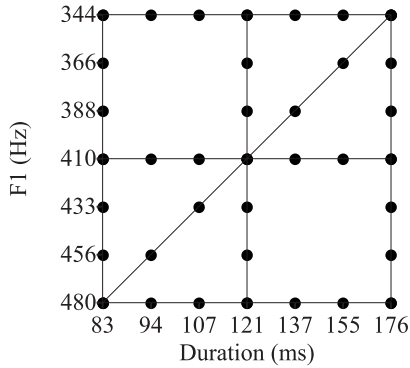

Figure 1. The 37 stimuli.

steps, which led to seven spectrally different stimuli, were equal on the auditorily based mel scale (Stevens, Volkmann, & Newman, 1937), 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 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. In total, 37 vowel stimuli (the points in Figure 1) were created with the Sensyn version of the Klatt parameter synthesizer.

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. 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. For both L1 and L2 perception, we used pictures rather than written words to avoid orthographic effects. There were both verbal and written instructions, which did not use the words *ship* and *sheep* explicitly. The subjects were instructed to guess in case of uncertainty about an answer and to take as much time as needed to make a decision. Every listener heard each of the 37 vowels 10 times. The 370 stimuli were presented to each subject with a different randomization of 10 blocks of 37 trials. After every block, the subjects took a short break.

L1 RESULTS

The results for the two NS groups are shown in Figures 2 and 3. Dark areas indicate a predominance of /i/ responses, light areas a predominance of /ɪ/ responses, and the solid curve is the *boundary line*, which estimates where

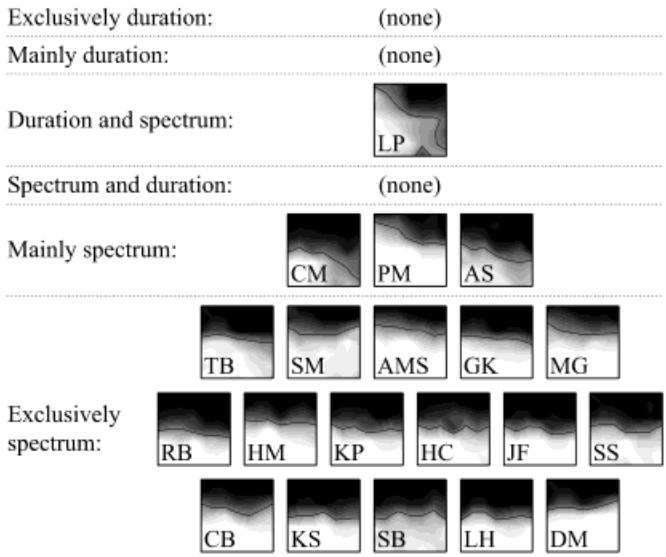

Figure 2. Identification results for each of the 20 Scottish listeners. In each square, duration runs from 83 ms (left) to 176 ms (right), and F1 runs from 480 Hz (bottom) to 344 Hz (top), as in Figure 1.

the subjects were equally likely to respond /ɪ/ and /i/. (To get a continuous representation, the values in the 12 not-measured cells were interpolated from the values in the neighboring cells.) For each subject, 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 crosslinguistic 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 for each subject as the ratio of her duration reliance and her spectral reliance; this ratio can be shown to be a measure of the slope of the listener's boundary line (Escudero & Boersma, 2003). In Figures 2 and 3, 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 and spectrum"; if her ratio is between 1/2 and 1, she relies on "spectrum and 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

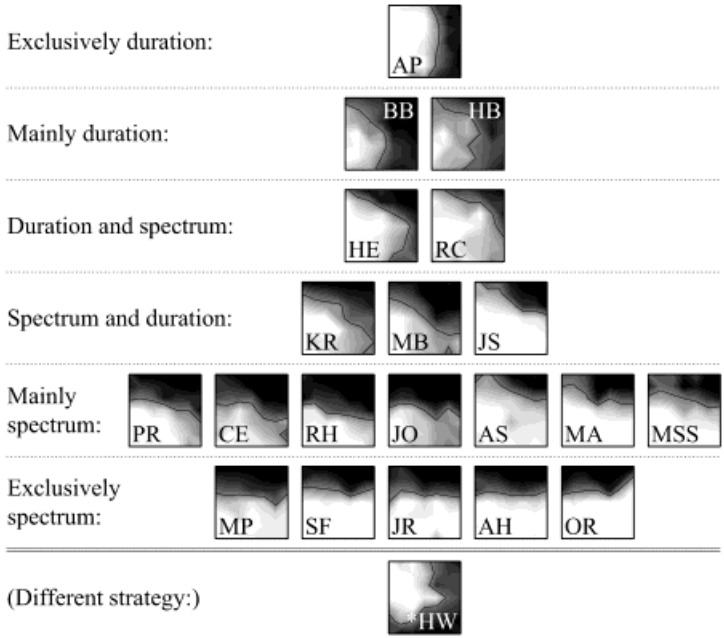

Figure 3. Identification results for the 21 Southern English listeners. Axes as in Figure 2. Subjects HB, RC, KR, MB, RH, JO, MA, MSS, MP, and JR were tested in Edinburgh; the others in Reading.

horizontal), the subject relies "exclusively on spectrum." For the "duration and spectrum" and "spectrum and duration" cases, the boundary is nearly diagonal.

The Scottish listeners showed a clear preference for the spectral cues: Sixteen out of 20 subjects had a reliance ratio of less than 1/4, which means that they relied almost exclusively on the formants. The Southern English listeners showed a different pattern. First, we excluded subject HW from further consideration because she followed an unexpected opposite strategy (i.e., her spectral reliance was statistically reliably negative). Only 5 of the remaining 20 subjects relied almost exclusively on the formants, whereas 14 subjects used a combination of the two cues. One subject relied exclusively on duration.

The results of the Scottish and Southern English listeners indicate that the perceptions of /i/ and /ɪ/ in these two varieties show differences in the same direction as the production (a one-tailed two-sample Kolmogorov-Smirnov test on the first two columns in Table 3 yields $p < .003$).

L2 RESULTS

The same experiment was performed with the group of Spanish learners of English. The results are in Figure 4 and Table 2. A language background ques-

Table 2. Data on Spanish L1 listeners (n = 30)

Subject	Origin	Education	Time in Scotland	Time in England	Time in Ireland	Time in Zimbabwe	Duration reliance (%)	Spectral reliance (%)	Reliance ratio	Cues relied on
BA	S	—	4	0	0	0	94	−17	(−5.53)	Exclusively duration
MAO	S	Dip	0	1	0	0	96	−17	(−5.65)	Exclusively duration
EF	S	EP	0	0	2	0	86	−10	(−8.6)	Exclusively duration
MANL	S	EP	46	0	0	0	90	−7	(−12.9)	Exclusively duration
OF	LA	—	1	108	0	0	99	−7	(−14.1)	Exclusively duration
MVL	LA	—	48	36	0	0	86	−4	(−21.5)	Exclusively duration
PN	S	—	5	13	0	0	99	−4	(−24.8)	Exclusively duration
MC	S	EP	0	0	2	0	79	−3	(−26.3)	Exclusively duration
CT	S	FCE	0	0	6	0	67	9	7.444	Exclusively duration
ARG	S	EP	0	0	0	0	94	20	4.7	Exclusively duration
SND	S	EP	0	0	0	0	89	20	4.45	Exclusively duration
JTN	S	CAE	0	1	0	0	99	27	3.667	Mainly duration
FJRG	S	EP	0	0	0	0	94	27	3.481	Mainly duration
LJ	S	EP	0	6	0	0	70	36	1.944	Duration and spectrum
ABG	S	—	0	2	0	0	40	47	0.851	Spectrum and duration
MCSC	S	EP	0	1	0	0	29	81	0.358	Mainly spectrum
LCHR	S	—	1	1	0	60	14	77	0.182	Exclusively spectrum
LG	LA	—	114	0	0	0	0	96	0	Exclusively spectrum
ADCG	S	—	48	0	0	0	−1	100	(−0.01)	Exclusively spectrum
CC	LA	—	96	36	0	0	−1	66	(−0.02)	Exclusively spectrum
MF	S	—	72	0	0	0	−6	100	(−0.06)	Exclusively spectrum
IR[a]	LA	—	36	0	0	0	19	−90[a]	—	
JR[a]	LA	—	6	312	0	0	13	−100[a]	—	
Beginners:										
MW[a]	LA	—	1	0	0	0	16	−94[a]	—	
AF	S	—	0	0	0	0	87	11	7.909	Exclusively duration
JG	S	—	0	1	0	0	37	77	0.481	Mainly spectrum
AL	S	—	0	0	0	0	13	94	0.138	Exclusively spectrum
JAD	S	—	0	0	0	0	−7	97	(−0.07)	Exclusively spectrum
MT	S	—	0	0	0	0	−10	100	(−0.10)	Exclusively spectrum
DMC[a]	S	—	0	0	0	0	−11	94	(−0.12)	Exclusively spectrum

Note. Origin: S = Spain, LA = Latin America. Education = 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. Time in Scotland, England, Ireland, Zimbabwe = time spent in Scotland, England, Ireland, Zimbabwe, in months (time spent in the United States was not included, because it was never more than a few weeks).
[a] = a reversal of *ship* and *sheep*.

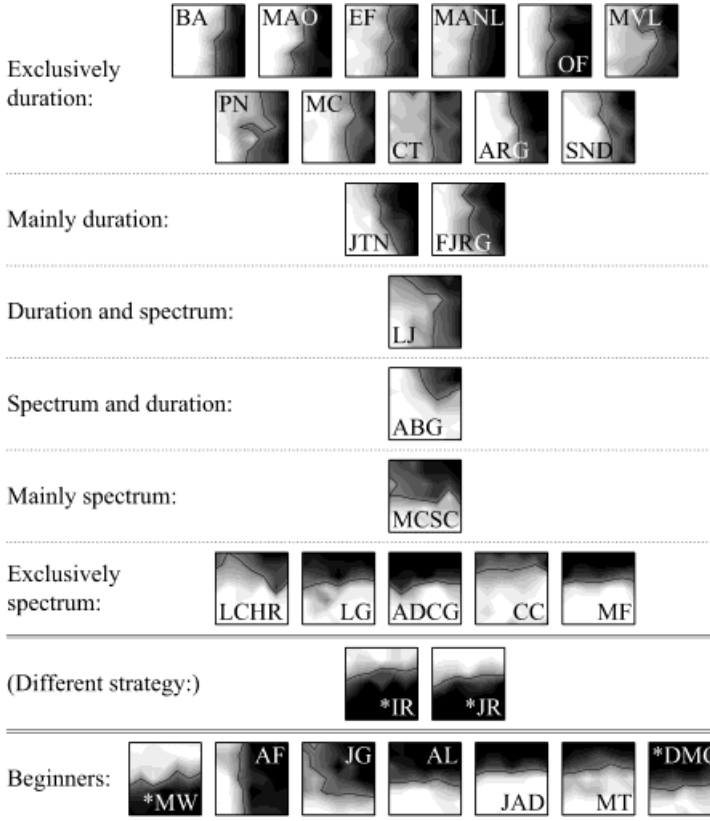

Figure 4. Identification results of 30 Spanish listeners on the English /i/-/ɪ/ contrast. Axes as in Figure 2.

tionnaire showed that the subjects had had 1–15 years of formal English instruction in their home countries and had spent 0–26.5 years in English-speaking countries. All reported still using Spanish, as well as using English with NSs. We labeled seven subjects as beginners—namely, those who had spent less than 2 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* (although they must have learned the orthographies during formal instruction), and they probably created on the fly a strategy on the basis of the only division that their L1 perception allows—namely, the division between Spanish /e/ and /i/. Thus, five of the seven beginners exhibit a horizontal boundary; it is revealing that subject MW testifies of her lexical confusion by consistently assigning 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 1 (which preceded the experiment reported here; see Escudero, 2001,

2002), though the test on the whole square (see Figure 4) had him guessing correctly. Subjects AL, JAD, and MT made the correct guess throughout. Subjects AF and JG showed duration reliance and cue integration, respectively; it is possible that only these two had the correct lexical contrast.

Most of the 23 more advanced subjects used either 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, whereas 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). To check whether the differences in L2 categorization are in the same direction as the differences manifested in the two L2 production environments (the L1 environments in Table 1), we performed a one-tailed Kolmogorov-Smirnov test on the six reliance classes for the 14 Southern English–oriented learners (those with an educational bias or having spent more time in Southern England than in Scotland; i.e., MAO through MCSC, except MVL) versus the 6 Scottish-oriented learners (BA, MVL, LG, ADCG, CC, and MF). The result of $p = .024$ establishes the correlation of L2 perception with the target dialect.

Table 3 compares the cue reliances for the (nonbeginner, nonreversing) Spanish listeners with those of the two groups of native English listeners. More than half of the Spanish listeners with a Southern English target relied mainly on duration, a pattern not found in either of the L1 groups. For what it is worth in this case of a bimodal distribution, two-tailed Kolmogorov-Smirnov tests show that the L2 group has a different average cue reliance ratio both

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

Cue integration	Scottish L1	Southern L1	L2
Exclusively duration	0	1	11
Mainly duration	0	2	2
Duration and spectrum	1	2	1
Spectrum and duration	0	3	1
Mainly spectrum	3	7	1
Exclusively spectrum	16	5	5
Total	20	20	21

from the Scottish L1 group ($p < .001$) and from the Southern English L1 group ($p = .020$). We conclude that, although between-learner differences in L2 categorization are correlated to the target dialect of each learner, many (mostly Southern English–oriented) learners are biased toward a reliance on duration alone that is not found in either of the L1 varieties.

EXPLAINING VOWEL CATEGORIZATION

The primary explanandum about the listening experiments is that both L1 and L2 perception depend on the production environment: The Scottish NSs and the Scottish-oriented L2 learners depended more on the spectral cues than did the Southern English NSs and the Southern English–oriented L2 learners. The secondary explanandum is a divergence between the L2 learners and the NSs: Sixteen out of 21 L2 listeners used only one acoustic cue for categorizing /i/ and /ɪ/, whereas at least the Southern English NSs typically integrated the two cues (14 out of 20). In this section, we argue that speech perception researchers have only partly been able to explain these phenomena, whereas our own explicit linguistic proposal can relate all phenomena to the OT versions of Full Transfer and Full Access.

The issues involved in L2 phonemic categorization have been addressed mainly 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). The two models predict that (and how) L1 linguistic experience determines the behavior of nonnative 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, and Pruitt (2000) concluded that the SLM needs to be extended nontrivially before it can account for early stages in L2 development; and Best, McRoberts, and 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 the PAM's generalizations about the initial state and the SLM's generalizations about later developmental stages.

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 OT (Prince & Smolensky, 1993)—more specifically, its probabilistic version (stochastic OT; Boersma, 1998). Boersma (p. 164) proposed continuous constraint families for mapping auditory continua to discrete "phonetic" categories (e.g., for mapping incoming F1 values to some discrete points along the F1 continuum). However, if phonological categories

are to be arbitrary symbols for purposes of lexical storage, we must assume that any value of *any* auditory continuum could in principle be mapped to *any* phonological category. The general formulation of a simple constraint, then, is "a value *x* on the auditory continuum *y* should not be perceived as the phonological category *z*" (Escudero & Boersma, 2003). 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 such as "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 for modeling the categorization of two English vowels on the basis of two auditory continua.[3] This is one step more complicated than the accounts of one-dimensional categorization presented by Boersma (1998, chapter 8) and Hayes (2001).

FIRST STEP: ENGLISH AND SPANISH L1 VOWEL CATEGORIZATION

To be able to explain the differences between the Scottish and the Southern English NSs, the differences between the Scottish-oriented and the Southern English–oriented Spanish learners of English, and the divergence between the L1 and L2 listeners, we first need a firm understanding of the three relevant L1 developments: How does the perception of the /i/-/ɪ/ contrast develop in the two varieties of English, and how does the perception of the /i/-/e/ contrast develop for Spanish?

We claim that *optimal* vowel categorization involves a dependency of perception on the specific production environment—that is, 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 optimal strategy for minimizing the probability of miscomprehension is to make decisions that lead to *maximum-likelihood behavior* (von Helmholtz, 1910): The optimal listener perceives any incoming auditory event as the phonological category that is most likely to have been intended by the speaker. Figure 5 shows how likely it is for any duration-F1 pair to have been intended as the category /i/, if the distribution of these 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 listener and a Southern English listener are confronted with the same auditory event—for example, [74 ms, 349 Hz], shown twice as a diamond in Figure 5. If both are optimal listeners, the Scottish listener will perceive this auditory event as /i/, given that Scottish *speakers* are

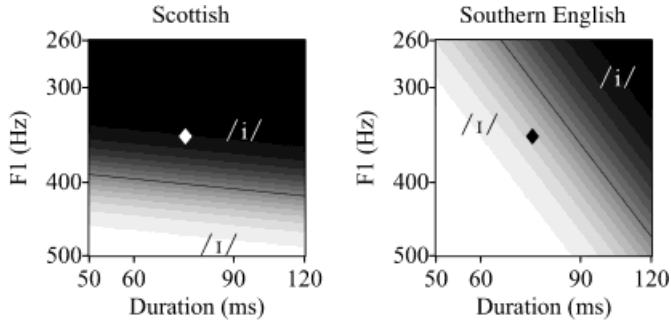

Figure 5. The probability that any given duration-F1 pair was intended as /i/ in a Scottish or Southern English environment. Values range from 0% (white) to 100% (black); the 50% line is also shown. In the pictures, "/i/" and "/ɪ/" depict the average F1 and duration values in the two dialects, taken from Table 1. The diamond is the auditory event [duration = 74 ms, F1 = 349 Hz], discussed in the text.

more likely to intend this auditory event as /i/ than /ɪ/. Likewise, the Southern English listener will best perceive the same event as /ɪ/, given that in her L1 environment this auditory event is more likely to have been intended as /ɪ/ than /i/. More generally, both listeners perceive everything above their own equal-likelihood line (in Figure 5) as /i/, everything below as /ɪ/. The optimal perceiver, therefore, has a category boundary in perception that coincides with the equal-likelihood line in her production environment. When we compare Figure 5 with the results of the real Scottish and Southern English listeners in Figures 2 and 3, we see that the L1 English listeners in our experiment indeed exhibit optimized vowel categorization and integrate the cues to the /i/-/ɪ/ contrast in accord with their own production environment.

Optimal L1 English

In this section, we explain how Scottish and Southern English listeners implement an optimal categorization—that is, how they perceive the diamond in Figure 5 appropriately. The knowledge underlying a Scottish listener's perception of the auditory event [74 ms, 349 Hz] can be represented as a ranking of constraints. One possible ranking with the desired results 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 relevant candidates for the output of the perception grammar. For reasons of space the tableau only shows the four relevant constraints. 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/ (see Table 1). When the auditory event [74 ms, 349 Hz] arrives, the

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

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

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.

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/.

Optimal L1 Spanish

For NSs of Spanish, the most important constraints are those for the spectral features. For the three front vowels, we have three continuous constraint families for F1. Examples of such 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/," and "an F1 of 200 Hz should not be perceived as /a/." Analogously, there are three constraint families for duration.

The constraints must be ranked in a way appropriate for the perception of Spanish—that is, 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 6 shows a possible continuous ranking of the constraints against perceiving /i/, /e/, and /a/ as functions of F1 (modeled analogously to Boersma, 1998, chapter 8).

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

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

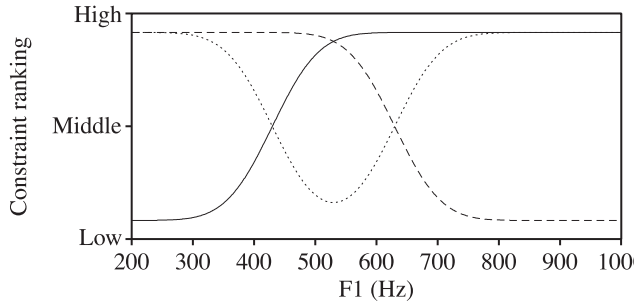

Figure 6. The adult Spanish F1-to-front-vowel perception grammar. The solid curve is the continuous ranking of "do not perceive an F1 of *x* as /i/." The dotted curve is "do not perceive an F1 of *x* as /e/." The dashed curve is "do not perceive an F1 of *x* as /a/."

The perception of Spanish front vowels on the basis of F1 can be derived from Figure 6. For an input of 350 Hz, the figure shows that the highest ranked constraint is "do not perceive an F1 of 350 Hz as /a/." The constraint "do not perceive an F1 of 350 Hz as /e/" is ranked almost as high. The constraint "do not 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/.

Given that 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 6, they are hardly capable of contributing to the determination of the winning candidate.

The Learning Algorithm

We have shown that OT listeners can handle several examples of vowel categorization. However, 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, therefore, 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 (Boersma, Escudero, & Hayes, 2003), then use these categories to create phonological forms in their lexicon (e.g., |ʃɪp|) and construction 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; Boersma, 1998; Boersma & Hayes, 2001) to their perception grammar, driven by recognition in the lexicon.[4]

For L1 acquisition, we begin our modeling at the point where the infant has the lexicon in place. For example, an infant Scottish listener may at some point inadvertently entertain a grammar that would have been appropriate for a Southern English 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 3. However, her environment is Scottish, so this auditory event could well have been related to /ʃip/ rather than to /ʃɪp/. If so, the infant's recognition system detects 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/ because 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 3), the child's GLA changes 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 3). This increases the probability that she will perceive /ʃip/ the next time she hears an F1 of 349 Hz or a duration of 74 ms. The rankings are changed by only a small step along the continuous ranking scale of stochastic OT (e.g., one-thousandth of the high-low distance in Figure 6), 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 similar to those of the adult Scottish listener in Tableau 1.

L1 English and Spanish Simulations

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

Tableau 3. Error-driven learning by the GLA in an OT perception grammar.

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

In our L1 English 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 would have led to the creation of construction constraints like "an F1 of 260 Hz is not /ɪ/," together with a reasonably good initial ranking of these constraints (Boersma, Escudero, & Hayes, 2003). Because category emergence is beyond the scope of the present article, we assume instead, 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 /i/. The virtual listeners are subsequently fed with input-output pairs drawn randomly from Gaussian distributions for the appropriate production environment. We assume that the distributions are centered about the mean F1 and duration values in Table 1. For both vowels and both English varieties, we chose fixed standard deviations of 0.2 octaves for F1 and 0.4 doublings for duration, which are large enough to ensure that a wide range of duration-F1 pairs occurs (these same distributions show up in Figure 5). For computational reasons, the duration continuum was divided into 21 values evenly spaced between 50 and 120 ms along a logarithmic scale, and the F1 continuum was divided up into 21 values evenly spaced between 260 and 500 Hz along a logarithmic scale. The number of relevant constraints, therefore, was 84 (= 21 steps × 2 continua × 2 categories).

Each virtual listener received 1,000 data per virtual month and changed some constraint rankings every time there was a mismatch between her perceived category and the correct category recognized by her lexicon. The size of the *evaluation noise* (the amount of fuzzy ranking associated with Stochastic OT, which is temporarily added to the ranking of each constraint at evaluation time) was held constant at 2.0, and the *plasticity* (the amount by which constraints are moved down or up along 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 7 shows the perceptual performance of the virtual Scot and the virtual Southerner in five stages. As in Figures 2–4, black areas stand for /i/ perceptions, white areas for /ɪ/ perceptions, and the black curve is the 50% category boundary line; as before, the gray areas stand for variable perceptions, which are possible in Stochastic OT as a result of the evaluation noise (Boersma, 1998; Boersma & Hayes, 2001). These pictures were computed by running each of the 21 × 21 = 484 possible auditory events through the perception grammar 1,000 times, keeping the evaluation noise at the same level as during learning. For instance, the token in the center of the square (77 ms, 361 Hz) is perceived by a simulated Southern English 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 slightly below the constraint against perceiving 77 ms as /i/.

In our L1 Spanish simulations, the task of the virtual child was to develop an /i/-/e/ distinction. We started with all F1 and duration constraints ranked

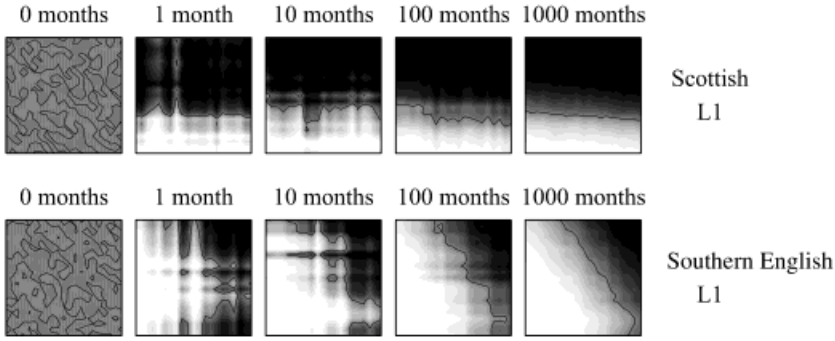

Figure 7. Perceptual development of virtual Scottish (top) and Southern (bottom) listeners. Horizontal axes: duration from 50 ms (left) to 120 ms (right). Vertical axes: F1 from 500 Hz (bottom) to 260 Hz (top). Black = /i/, white = /ɪ/.

at the same height, then applied 1,000 inputs per virtual month and used the same evaluation noise and plasticity regime as for the L1 English. As seen in Figure 8, the listeners come to rely on the F1 cue only, with the duration constraints staying ranked approximately at their initial height (the resulting rankings for the F1-to-vowel mapping are very similar to those in Figure 6).

The locations and slopes of the boundaries in the final stages in Figure 7 compare well with the optimal ones in Figure 5. 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.[5] We have thus accounted for the difference between the two L1 control groups in our experiment and for the final state of L1 Spanish, which is the starting point for the second step.

SECOND STEP: TRANSFER OF STRUCTURES AND PROCESSES

We claim that L2 learners aim at minimizing the probability of perceptual confusion, just like the L1 learners of the previous section. It is advantageous not to start where L1 learners start—namely, without any vowel categories at all.

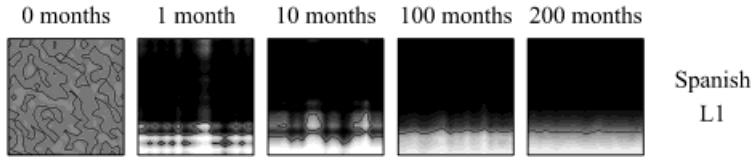

Figure 8. Development of Spanish L1 perception. Black = /i/, white = /e/.

If Spanish learners of English reuse their five native vowel categories /i, e, a, o, u/ in their initial perception of English, they should already be able to handle a five-way contrast, even if this falls short of what is necessary to speak an L2 with 13 vowels proficiently. This reuse of L1 categories is an instance of the transfer of language-specific *structures* (Polivanov, 1931). Additionally, if the beginning Spanish learner of English enters the Scottish production environment, she will probably map /i/ and /ɪ/ to her Spanish /i/ and /e/, respectively, as these are the two categories closest to the average realizations of the vowels in the L2 (from Figure 2, the Scottish boundary is about 400 Hz; according to Bradlow, 1996, the Spanish boundary lies near 430 Hz). This is shown in Figure 9, which also shows that the Spanish learner of the Southern English variety will probably initially map both /i/ and /ɪ/ to her native /i/ category. This reuse of L1 mappings is an instance of the transfer of language-specific *processes*. Both kinds of transfer give L2 learners a head start.

The perceptual behavior of beginning L2 learners can be seen as a case of foreign language speech perception. The PAM by Best (1995) distinguishes the two strategies depicted in Figure 9. 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*.

We now show how a Spanish perception grammar would perceive the average tokens of Scottish and Southern English /ɪ/, whose F1 values and durations are presented in Table 1. Tableaus 4 and 5 are based on the rankings in

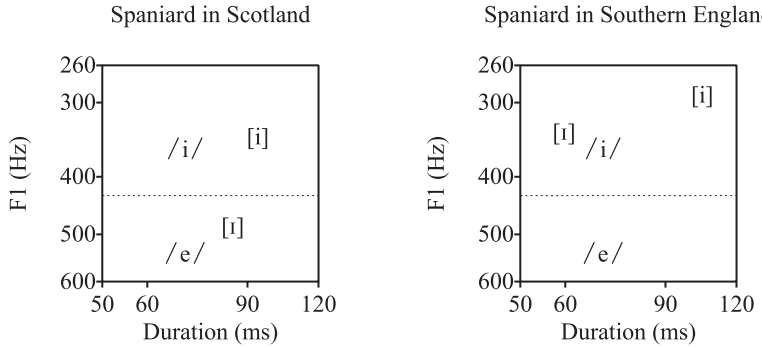

Figure 9. Hypothetical perceptual assimilation by Spanish listeners of different varieties of English. In square brackets: the average realizations of /i/ and /ɪ/ for the environment at hand (taken from the average values in Table 1). Between slashes: the average Spanish vowels (Bradlow, 1996: F1 values 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.

Tableau 4. The perception of the average Scottish /ɪ/ 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/	*!		*	

Figure 6, under the assumption that all duration constraints are ranked at "middle." Thus, some F1 constraints are ranked higher than the duration constraints, and some are ranked lower. Analogously, both the average Scottish /i/ and the average Southern English /i/ are perceived as Spanish /i/. Through Tableaus 4 and 5, Figure 6 provides the explanation of the perceptual assimilation patterns in Figure 9.

Although Best's 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 nonnative contrasts (as suggested by Guion et al., 2000, and Escudero, 2001) and taken to predict the initial state of a beginning L2 learner. Analogously to the usual OT interpretation of Full Transfer, we claim that the behavior 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 4 and 5 as well. The Spanish-L1 speakers in Scotland will perform well because they can use the Spanish categories /i/ and /e/ for storing words with English /i/ and /ɪ/ (i.e., the phonological part of their lexical representation of the word *ship* is |ʃep|). Their only little problem is that their /i/-/ɪ/ boundary is at 430 Hz, not at 400 Hz, as it is for the average native Scot of Figure 2. By contrast, the Spanish-L1 speakers in Southern England are in more trouble because they will use the single Spanish /i/ for storing English words with both /i/ and /ɪ/. Fortunately, as we

Tableau 5. The perception of the average Southern /ɪ/ 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/			*	*

explain in the next section, Full Access allows the learners to overcome entirely or partially these problems of boundary mismatch and single-category assimilation.

THIRD STEP: ACCESS TO L1-LIKE ACQUISITION STRATEGIES

Although the two groups of beginning Spanish learners of English do the best they can given their linguistic experience, their behavior 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 access their L1 acquisition devices to modify their structures (i.e., create more categories or reduce, split, or merge existing categories) or their processes (i.e., shift their category boundaries) into the direction of the L2.

L1-like Boundary Shifts

Spanish learners of Scottish English would have to move their category boundary, perhaps from 430 to 400 Hz, as in Figure 10 (left), which also shows that the centers of the two reused categories have shifted. The availability of this boundary shift is an instance of access to L1-like learning mechanisms.

For Spanish speakers in Scotland, a small mismatch occurs in perception. F1 values around 420 Hz may 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 GLA changes the perception grammar, as in Tableau 6. The learner perceives /ʃip/ but notes that the semantic context requires that she should have perceived /ʃep/ because that is the phonological part of her

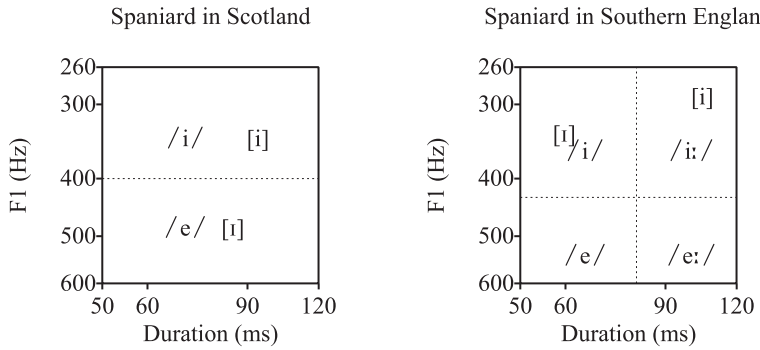

Figure 10. Further L2 perceptual optimization in the two environments. To the left: a shifted boundary in Scotland. To the right: a new length distinction in Southern England, presumably leading to the fourfold *bit-beat-bet-bait* contrast.

Tableau 6. Acquisition of vowel categorization by a Spanish listener in Scotland.

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

lexical representation of the English word *ship*.[6] Tableau 6 shows that she takes action in such a way that she will be more likely to perceive /e/ at the next 420-Hz token. Thus, boundary shifts are handled entirely by the learning algorithm.

L1-like Category Creation

The learning task for Spanish speakers in Southern England is to "notice" that their single /i/ category represents two different phonemes in the L2. This "noticing" (a mechanism for it is described in the next paragraph) could in principle either lead to splitting the /i/ category into two new vowels /i/ and /ɪ/ or to forming a new feature contrast, perhaps a length distinction—that is, a new /short/-/long/ contrast that does not occur in the learners' L1. 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 10 (right) shows the results of applying this strategy.

It remains to be explained why the listeners choose the new length distinction rather than split their /i/ category. Bohn (1995) claimed that it has to do with a universal availability of duration as a source of lexical distinctions, which is called on if the L1 has insufficient spectral distinctions. Bohn tacitly assumed that Spanish speakers have a single duration category, and any explanation for their behavior would have to address the question as to why learners split this single duration category rather than the single high front vowel category. For Bohn, then, the explanation is that there is something universally special about duration. We claim, however, that duration is not special: These learners start not with *one* duration category but with *no* duration category at all, and they use the attested L1 acquisition strategy of distributional learning (Maye, Werker, & Gerken, 2002) to detect two peaks in the duration distribution, which allow them to divide the duration continuum into two categories. 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, which has, to our knowledge, not been proposed as a common L1 learning strategy. Thus, the simple hypothesis of Full Access is compatible with the availability to L2 learners of category formation in general and the formation of a duration contrast in this particular case and at the same time compatible with a general unavailability 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 extending their mastery of English from five vowels to 10. Tentatively, we can say that the resulting assimilation pattern will be as in (1).

(1)

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

This kind of suprasegmental length contrast (attested in our Spanish learners of English for the high front vowels and possibly existing for the other four vowels as well) is not found in the learners' L1, which has no length distinction at all, nor in the L2, in which nearly all vowels (and certainly high front 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, because at first sight one would think that such a theory predicts a learning path that interpolates linearly between the L1 and the L2. However, our theory also crucially relies on the L2 learners' access to the L1-like mechanism of category formation: We propose that the L2 learner starts with a virgin duration continuum (i.e., zero categories rather than one, with no duration-to-category mappings yet) as all infants do, and that she creates two categories as soon as she establishes that the distribution is bimodal (analogously to the explicit L1 OT modeling by Boersma et al., 2003). At that point, the learner adds a binary length contrast to her grammar, by introducing "phonetic" construction constraints such as: "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/," and "a duration of 200 ms should not be perceived as /long/." Distributional learning not only introduces these constraints but also an initial ranking for them, so that the learner starts with a reasonable proficiency in mapping auditory duration to phonological length (as discussed at the end of this section, the learner introduces some more constraints later). After the creation of the /short/-/long/ contrast, the learners can start using the two length categories in their lexical representations. They represent Southern English /ɪ/ as /i, short/ and Southern English /i/ as /i, long/. Once the learners have a correct lexicalization of length, the appropriate mapping from continuous duration to the binary length contrast is achieved by the GLA. Tableau 7 shows what happens if an intended Southern English |ʃip| "sheep," which has the representation |ʃiːp|

Tableau 7. Acquisition of the categorization of length by a Spanish listener in Southern England.

[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/			←*	*
√	/ʃi:p/		*!→		*
	/ʃep/	*!		*	
	/ʃe:p/	*!	*		

for the learner, is pronounced appropriately with a duration of 104.6 ms, but the learner has a nonoptimal constraint ranking that leads her to perceive the vowel as /short/ rather than /long/. She perceives /ʃip/ and notices a mistake because the word that she should have recognized is |ʃi:p| "sheep." She takes action and reranks some constraints with the GLA, thus making it more likely that she will perceive 104.6 ms as /long/ on the next occasion.

L2 Simulations

We simulated developmental sequences of typical Spanish learners of English who have gone to two different English-speaking countries at an age of 200 months. The first listener lives in Scotland. She equates Scottish /ɪ/ with her Spanish /e/, and Scottish /i/ with her Spanish /i/ (Figure 9, left). We assumed that she already has correct lexical representations for *ship* and *sheep*. Every virtual month we fed this virtual L2 Scot with 1,000 vowels, drawn from the same Gaussian distribution used for the virtual L1 Scot. We continued this procedure for 800 virtual months, keeping the evaluation noise constant at 2.0 and the plasticity constant at the low (adult) value of 0.01. Figure 11 shows how the L2 Scot shifts her boundary from 430 to 400 Hz and then tilts it, thus becoming as proficient as the L1 Scot in Figure 7.[7]

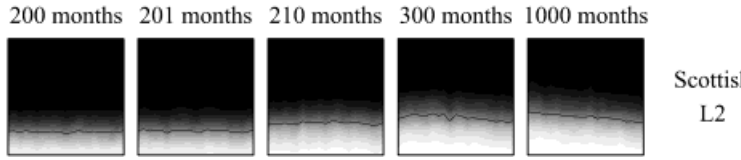

Figure 11. The boundary shift of the simulated Spaniard in Scotland. Black = /i/, white = /e/ (= /ɪ/).

The second listener lives in Southern England. She equates both Southern English /ɪ/ and /i/ with her Spanish /i/ (Figure 9, right). Figure 12 shows how the L2 Southerner starts out at chance level (like the infants of Figures 7 and 8; i.e., we assumed correct lexical representations for *ship* and *sheep* but a very poor categorization) then learns how to map duration to the new length contrast, improving toward the performance of the simulated Southerner in Figure 7 although she continues to ignore the F1 cue.

Comparison of Simulated and Real L2 Listeners and Comparison with L1 Learners

The simulated Spaniard in Scotland comes to rely primarily on F1, like most of the real Scottish-oriented L2 learners (e.g., subjects LG and MF in Figure 4 and Table 2). The simulated listener also shows a tiny boundary shift. We cannot detect such a shift between beginning and more advanced real L2 listeners (Figure 4), probably because the L1 Scottish boundary (Figure 2) 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 (because, in fact, they do occur, as described in the Discussion section). The simulated Spaniard in Southern England comes to rely on duration only, like most of the real Southern English–oriented L2 learners (e.g., subjects EF and OF in Figure 4 and Table 2). This is in stark contrast with the simulated and real native Southerners, who rely on both the spectral and the duration cues; this difference is due to the L2 learners' strategy of creating a length contrast, which leads to a problematic representation of the vowel contrast—namely, as /i, short/ versus /i, long/ rather than as the separate symbols /ɪ/ versus /i/ that would allow genuine cue integration. We conclude that our simulations reveal a formal explanation of the attested L2-specific behavior.

Later Developments

The L1 distributional learning model of Boersma et al. (2003) predicted that learners will initially use a single auditory cue for each phonological contrast in their language. In the case of our simulations for the Southern English–

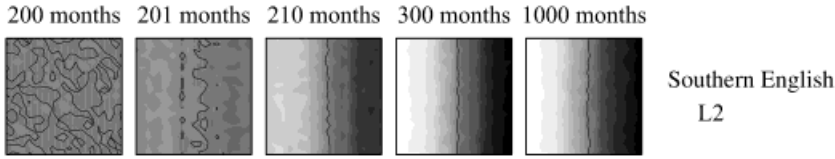

Figure 12. The phonemic separation of the simulated Spaniard in Southern England. Black = /i, long/, white = /i, short/.

oriented L2 learners (Figure 12), the single cue for phonological length is duration; this is formally expressed as our exclusive use of "phonetic" constructions constraints such as "a duration of 110 ms is not /short/." However, the same distributional learning model also predicts that learners will later on introduce less-phonetic construction constraints—in our case, constraints that relate the phonological length feature to the auditory continuum of F1, such as "an F1 of 300 Hz is not /short/." As soon as such constraints become available, the categories /short/ and /long/ have become as abstract as the categories /ɪ/ and /i/, and learners can start to integrate spectral and duration cues for the length contrast. Indeed, at least two learners in Figure 4 seem to have reached such a stage. A computer simulation of such a scenario, however, would involve all 13 English vowels and all interlanguage vowels, given that all 5×2 interlanguage vowels contain one of the feature values /short/ or /long/. We can predict that, because 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 /i/-/ɪ/ contrast for highly advanced learners.

DISCUSSION: WHAT IS TRANSFERRED AND ACCESSED?

The formal model for L2 phonemic categorization advanced here bridges the gap between speech perception research and linguistic theory by applying for the first time the explicit hypotheses of Full Transfer and Full Access (Schwartz & Sprouse, 1996) to L2 speech perception. 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 raise the question as to whether these five ingredients were really attested in the experiments and to what extent our results contradict or confirm other theories.

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 (nativelike 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 behavior of the nonnativelike Southern English–oriented group. The results for the Southern English–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 English–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) 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.

Transfer of the Grammar: Reuse of L1 Perceptual Mappings in the L2 Initial State

Best's (1995) PAM states that high-level linguistic experience determines what listeners do when confronted with the crosslinguistic 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 9): For a monolingual Spanish listener, typical tokens of Scottish /ɪ/ and /i/ would fall into her /e/ and /i/ categories, respectively, whereas typical tokens of Southern English /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. Because our listening subjects showed later reflexes of both types of perceptual assimilation, our perception experiment seems to confirm both the PAM and Full Transfer.

Transfer of Holes in the Grammar: Reuse of L1 Blank Slates in the L2 Initial State

As L1 Spanish has never shown discrete categories on the vowel duration continuum, the Spanish perception of the duration continuum is still a blank slate with constraints that handle noncategorizing mappings (Boersma et al., 2003). When they listen to English, Spanish monolinguals perceive vowel duration in an acoustic, noncategorizing way (not as a mapping to a single discrete duration category, as Bohn, 1995, seemed to have assumed). Under Full Transfer, Spanish learners of Southern English transfer this noncategorizing 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 the next paragraph). The idea of reuse of blank slates can at least partially account for the observation (Flege,

1987; Major, 1987, p. 109) that *new sounds* (i.e., sounds in new areas of the phonetic space) are relatively easy to acquire.

Access to an L1-like Acquisition Device: New Categorization of Blank Slates

The initial vowel duration continuum for native Southern English infants is a blank slate. The noncategorizing mappings lead to distributional learning, through which the infants create two length categories on the basis of their L1 input. Under Full Transfer, the initial vowel duration continuum for Spanish learners of Southern English is just such a blank slate. Under Full Access, therefore, they create two length categories, just like the NSs. Our experiment shows evidence for this: Beginners seem to have trouble with the length distinction, whereas more experienced learners have developed a lexical length contrast. Although Bohn (1995) considered this behavior to be evidence against transfer, our closer formal modeling now actually regards it as evidence in favor of access to an L1-like distributional learning device (and therefore in favor of the transfer of the blank slate).

Access to an L1-like Acquisition Device: Constraint Reranking in the Interlanguage

We have been assuming that L2 learners have access to the GLA, which takes the L2 learner through a sequence of constraint rankings in the same way as it does for an L1 learner. In our modeling and simulations, the observational result is boundary shift and the development of cue integration.

Our experiment does not show direct evidence of boundary shifts in L2 acquisition, given that, for Spanish learners of Scottish English, the initial position of the /e/-/i/ boundary (cf. Figure 11, left) is already quite close to the position of the Scottish /ɪ/-/i/ boundary (cf. Figure 7, top right), and, for Spanish learners of Southern English, the duration boundary that emerged from the simulations (Figure 12) could have been created by distributional learning alone. However, boundary shifts have been attested in other work on L2 speech perception: Caramazza, Yeni-Komshian, Zurif, and Carbone (1973) reported that NSs of French who had begun to acquire English before their seventh birthday exhibited a large shift of the /b/-/p/ boundary. Flege and Eefting (1987) showed that the perceptual /b/-/p/ boundary of Dutch learners of English depended on the language the learners thought they heard. Escudero and Boersma (2002) showed a similar language mode effect for the /e/-/ɪ/ and /ɪ/-/i/ boundaries of Dutch learners of Spanish. The simulation in Figure 11 shows that our model is able to formalize boundary-shift effects.

Our experiment does show some evidence of L1-like L2 cue integration. A minority of the learners in Figure 4 have diagonal boundaries. As previously

noted, the model of Boersma et al. (2003) predicted that this situation is the result of the introduction of constraints that map F1 values to length categories.

Extending the Original Definitions of Full Transfer and Full Access to Perception

The hypothesis of Full Transfer, as defined by Schwartz and 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 equivalent *barco*). For phonological perception, we can say that the hidden representations are the phoneme categories (and metrical structures) and that the perception grammar handles the mappings between these and auditory events. In the previous discussion on *transfer*, we have therefore seen evidence for Full Transfer, when applied to phoneme categories.

The hypothesis of Full Access, as defined by Schwartz and Sprouse (1996, p. 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 UG. The typological assumption of OT (Prince & Smolensky, 1993) is that every possible constraint ranking reflects a language allowed by UG. For OT interlanguage grammars, Full Access means, therefore, that every developmental stage should be identifiable with a constraint ranking. The GLA satisfies this requirement automatically, as it can do nothing besides changing the rankings of the constraints. In the previous discussion on *access*, we have therefore seen evidence for Full Access, when applied to phoneme categories.

CONCLUSION

When learning the English /i/-/ɪ/ 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 NSs, the learners with a Southern English target typically exhibit a pattern of behavior that is observationally dissimilar to anything that occurs in adult L1 Spanish or 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 suboptimal patterns in L2 acquisition. Our exper-

iment and its modeling and simulations have the combined predictive power of Best's PAM (1995), Flege's SLM (1995), and Schwartz and Sprouse's Full Transfer and Full Access hypotheses (1996). Our formalization thereby provides the linguistic mechanism that underlies the generalizations forwarded by several previous models of L2 speech perception.

(*Received 3 March 2004*)

NOTES

1. In the "L1 Results" section, we explain how such percentages are computed.

2. A large part of the experiment described here, including the testing of the Spanish-L1 learners and the Scottish NSs, was reported previously in Escudero (2001, 2002).

3. For the present paper, we could equally well have used positively worded constraints such as "an F1 of 260 Hz should be perceived as /i/," but we happen to know that such constraints do not work for the general case in which multiple auditory continua are mapped to more than two phonological categories.

4. The GLA, a learning algorithm for stochastic OT, was preceded by Error-Driven Constraint Demotion (Tesar, 1995; Tesar & Smolensky, 1998, 2000), a learning algorithm for nonstochastic OT. That learning algorithm would not have worked for our findings, because it is not capable of handling variable mapping—that is, the token [80 ms, 400 Hz] can sometimes represent /ɪ/, sometimes /i/, and Error-Driven Constraint Demotion fails to work if such tokens are presented to it in sequence.

5. Traditionally in the OT literature, grammars map underlying forms to surface forms. Such grammars can be called *production grammars*. Here we have been interested in *perception grammars*, which, with partially different kinds of constraints, map auditory events to phonological structures. An anonymous *SSLA* 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 (Boersma, 1998), in which faithfulness constraints in the speaker's production grammar evaluate the extent to which the listener's perception grammar is able to reconstruct the hidden phonological structure. This ensures that in this model the two grammars do not replicate each other (i.e., perception is primary).

6. It is, of course, not necessary to have actual minimal pairs in the lexicon. Perceiving an intended *shift* as the nonword *sheeft* already suffices for the recognition system to issue a protest, and the same learning step takes place.

7. We did not discuss the representations for the Spanish learners of Scottish English in detail because the /i/-/ɪ/ 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 shift the boundary line and not tilt it toward a 10% duration reliance. The data of the real learners in Table 2 does not allow us to decide whether they use duration to this small extent or not at all.

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* (pp. 171–204). Timonium, MD: York Press.

Best, C., McRoberts, G., & Goodell, E. (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.

Boersma, P. (1998). *Functional phonology* (Doctoral dissertation, University of Amsterdam). The Hague: Holland Academic Graphics.

Boersma, P., Escudero, P., & Hayes, R. (2003). Learning abstract phonological from auditory phonetic categories: An integrated model for the acquisition of language-specific sound categories. *Proceedings of the 15th International Congress of Phonetic Sciences*, 1013–1016.

Boersma, P., & Hayes, B. (2001). Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry*, *21*, 45–86.

Bohn, O.-S. (1995). Cross language speech perception in adults: First language transfer doesn't tell it all. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues* (pp. 279–304). Timonium, MD: York Press.

Bohn, O.-S., & Flege, J. E. (1990). Interlingual identification and the role of foreign language experience in L2 vowel perception. *Applied Psycholinguistics*, *11*, 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., Chen, S.-I., & Wang, C. (1998). The emergence of the unmarked in second language phonology. *Studies in Second Language Acquisition*, *20*, 261–280.

Caramazza, A., Yeni-Komshian, G., Zurif, E., & Carbone, E. (1973). The acquisition of a new phonological contrast: The case of stop consonants in French-English bilinguals. *Journal of the Acoustical Society of America*, *54*, 421–428.

Davidson, L. (1997). An Optimality Theoretic approach to second language acquisition. Unpublished honors thesis, Brown University, Providence, RI.

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

Escudero, P. (2001). The role of the input in the development of L1 and L2 sound contrasts: Language-specific cue weighting for vowels. In A. H.-J. Do, L. Dominguez, & A. Johansen (Eds.), *Proceedings of the 25th annual Boston University Conference on Language Development* (pp. 250–261). Somerville, MA: Cascadilla Press.

Escudero, P. (2002). The perception of English vowel contrasts: Acoustic cue reliance in the development of new contrasts. In J. Leather & A. James (Eds.), *New sounds 2000: Proceedings of the 4th International Symposium on the Acquisition of Second-Language Speech* (pp. 122–131). Klagenfurt, Austria: University of Klagenfurt.

Escudero P., & Boersma, P. (2002). The subset problem in L2 perceptual development: Multiple-category assimilation by Dutch learners of Spanish. In B. Skarabela, S. Fish, & A. H.-J. Do (Eds.), *Proceedings of the 26th annual Boston University Conference on Language Development* (pp. 208–219). Somerville, MA: Cascadilla Press.

Escudero, P., & Boersma, P. (2003). Modeling the perceptual development of phonological contrasts with Optimality Theory and the Gradual Learning Algorithm. In S. Arunachalam, E. Kaiser, & A. Williams (Eds.), *Proceedings of the 25th annual Penn Linguistics Colloquium, Penn Working Papers in Linguistics*, *8*, 71–85.

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 W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues* (pp. 233–277). Timonium, MD: York Press.

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

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

Fox, R. A., Flege, J. E., & Munro, M. J. (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.

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

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

Guion, S., Flege, J. E., Akahane-Yamada, R., & Pruitt, J. (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 J. Leather & A. James (Eds.), *New sounds '97: Proceedings of the 3rd International Symposium on the Acquisition of Second-Language Speech* (pp. 109–117). Klagenfurt, Austria: University of Klagenfurt.

Hancin-Bhatt, B. (2000). Optimality in second language phonology: Codas in Thai ESL. *Second Language Research*, *16*, 201–232.
Hancin-Bhatt, B., & Bhatt, R. (1997). Optimal L2 syllables: Interactions of transfer and developmental effects. *Studies in Second Language Acquisition, 19*, 331–378.
Hayes, R. (2000). Reranking stages in OT analysis of the acquisition of Japanese as a second language. *Carolina Working Papers in Linguistics*, *1*, 1–21.
Hayes, R. (2001). An Optimality-Theoretic account of novel phonetic category formation in second language learning. Unpublished manuscript, University of Arizona, Tucson.
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.), *University of Arizona Coyote Working Papers, 12: Language in Cognitive Science* (pp. 28–41). Tucson: University of Arizona.
Hewlett, N., Matthews, B., & Scobbie, J. (1999). Vowel duration in Scottish English speaking children. *Proceedings of the 14th International Congress of Phonetic Sciences*, 2157–2160.
Labov, W. (1994). *Principles of linguistic change: Vol. 1. Internal factors*. Oxford: Blackwell.
Major, R. C. (1987). A model for interlanguage phonology. In G. Ioup & S. H. Weinberger (Eds.), *Interlanguage phonology: The acquisition of a second language sound system* (pp. 101–124). Rowley, MA: Newbury House.
Maye, J., Werker, J. F., & Gerken, L. A. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*, B101–B111.
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* (pp. 29–48). Burnaby, Canada: Simon Fraser University Linguistics Graduate Student Association.
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., & Miller, M. E. (1997). Developmental weighting shifts for noise components of fricative-vowel syllables. *Journal of the Acoustical Society of America*, *102*, 572–580.
Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of vowels. *Journal of the Acoustical Society of America*, *24*, 175–184.
Peterson, G. E., & Lehiste, I. (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 [The perception of the sounds of a foreign language]. *Travaux du Cercle Linguistique de Prague*, *4*, 79–96.
Prince, A., & Smolensky, P. (1993). *Optimality Theory: Constraint interaction in generative grammar* (Tech. Rep. No. 2). New Brunswick, NJ: Rutgers University Center for Cognitive Science.
Schwartz, B., & Sprouse, R. (1996). L2 cognitive states and the Full Transfer/Full Access model. *Second Language Research*, *12*, 40–72.
Scobbie, J. A. (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 2* (pp. 343–358). Chicago: Chicago Linguistic Society.
Scobbie, J., Turk, A., & Hewlett, N. (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., Volkmann, J., & Newman, E. B. (1937). A scale for the measurement of the psychological magnitude pitch. *Journal of the Acoustical Society of America*, *8*, 185–190.
Tesar, B. (1995). *Computational Optimality Theory*. Unpublished doctoral dissertation, University of Colorado, Boulder.
Tesar, B., & Smolensky, P. (1998). Learnability in Optimality Theory. *Linguistic Inquiry*, *29*, 229–268.
Tesar, B., & Smolensky, P. (2000). *Learnability in Optimality Theory*. Cambridge, MA: MIT Press.
von Helmholtz, H. (1910). *Handbuch der physiologischen Optik* [Handbook of physiological optics]: *Vol. 3*. Hamburg, Germany: Leopold Voss.
Wells, J. C. (1982a). *Accents of English 2: The British Isles*. New York: Cambridge University Press.
Wells, J. C. (1982b). *Accents of English 3: Beyond the British Isles*. New York: Cambridge University Press.