The Subset Problem in L2 Perceptual Development: Multiple-Category Assimilation by Dutch Learners of Spanish

Paola Escudero and Paul Boersma
McGill University / Utrecht University, and University of Amsterdam

Second-language (L2) speech perception research has identified several patterns in the non-native perception of vowel contrasts. The models that account for non-native perception (Perceptual Assimilation Model, Best 1995; and Speech Learning Model, Flege 1995; for a summary of these and other non-native perception models see introduction in Best, McRoberts and Goodell 2001) claim that the attested patterns are a consequence of the influence of the learners' first language (L1). One of the most common patterns reported in the literature, mainly because of its highly problematic nature for learning, is the so-called *single-category assimilation* (Best's term; Flege calls it *perceptual equivalence*), e.g. the perception of English /ɛ/-/æ/, /i/-/ɪ/ and /r/-/l/ by Dutch, Spanish, and Japanese listeners, respectively (Figure 1).

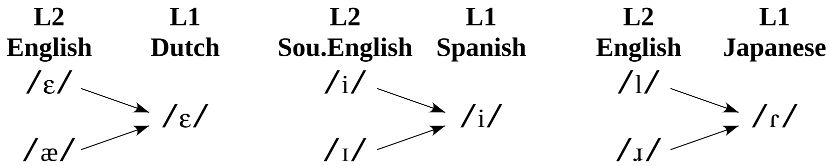

Fig. 1. Three cases of single-category assimilation.

According to Best's and Flege's models, this pattern is manifested as the learners' association of a binary contrast in the L2 with only one segment in their L1. This will cause a very poor category differentiation in L2 perception, if we assume that the initial state of the learner's perception system is a copy of her L1 perception system (an example of Full Transfer; Schwartz and Sprouse 1996). Therefore, we can expect problems with the formation of a lexical contrast, with implications in production. The learner's solution would be to somehow *split* the category to which the contrast has been mapped.

A second pattern identified for non-native perception is the so-called *two-category assimilation* (Best's term; Flege calls it *old contrast*), e.g. the perception of English /p^{h}/-/b̥/ and Spanish /o/-/u/ by Dutch listeners (Figure 2). In opposition to the previous pattern, the learner associates a binary contrast in the L2 with a binary contrast in her L1. According to Best (1995) and Flege (1995), this leads to good category differentiation. Thus, this pattern can be

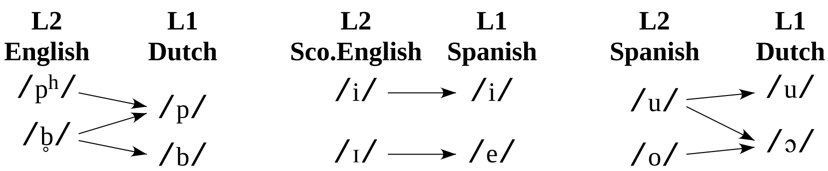

Fig. 2. Three cases of two-category assimilation.

expected not to present large problems for lexicalization (but see Flege 1987 for the problematic nature of this pattern in native-like attainment of production). We believe (again assuming that the L1 perception system is initially copied to the learner's L2) that this pattern can cause a perceptual problem, namely a boundary mismatch in the learner's L2 perception system, leading to problems with lexical access. The learner should solve this problem by *shifting* the boundary between the categories in her L2 to match that of the target language.

The models mentioned above refer not only to the perception of a foreign language, but the effects of the two patterns discussed can also be found in L2 development. Escudero (2001), for instance, reconstructs the patterns in the middle of Figures 1 and 2 as early stages in the development of Spanish learners of Southern British English and Scottish Standard English, respectively, on the basis of the learners' later cue weighting for the /i/-/ɪ/ contrast.

In this paper, we consider another pattern in the perception of non-native contrasts, which we will call *multiple-category assimilation* to comply with Best's terminology. Although this pattern was not addressed by Best (1995) and Flege (1995), and Schmidt (1996) and Strange, Akahane-Yamada, Kubo, Trent and Nishi (2001) just mention it as a possibility, we consider it quite important in the perceptual development of L2 learners. We will show a description of the pattern (§1) and present preliminary evidence for its existence and problematic nature (§2). In §3, we will reveal directly that it causes categorization problems in L2 perceptual development, and expose the most common strategy by which learners solve it.

1. Multiple-category assimilation and the L2 subset problem

Multiple-category assimilation (MCA) involves perceiving a binary contrast in a second language as more than two categories in the first language. For Dutch speakers listening to Spanish, this pattern is likely to occur in the perception of the Spanish front vowels /i/ and /e/, since Dutch has three short front vowels (/i/-/ɪ/-/ɛ/). The Dutch will perceive some tokens of Spanish /i/ as Dutch /i/, but some as /ɪ/, and they will perceive some tokens of Spanish /e/ as Dutch /ɪ/ as well, but others as /ɛ/ (Figure 3). Two other possible examples of MCA are also shown in Figure 3: the perception of the English /p^h/-/b̥/ and /s/-/ʃ/ contrasts as three-way contrasts by Korean and Polish listeners,

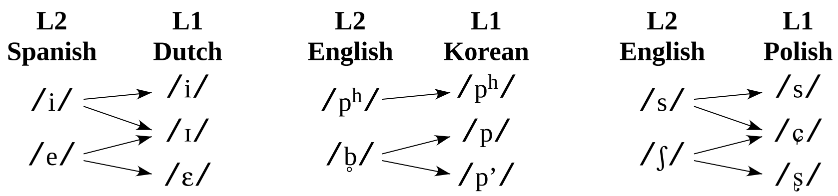

Fig. 3. Three cases of multiple-category assimilation (MCA).

respectively (although Schmidt 1996 shows that only one of the English categories gets assimilated to the middle Korean category).

The reason why the MCA pattern has been neglected in the literature may be that at first sight it does not seem to pose any perceptual problems for L2 learners other than those resembling single-category assimilation (the mapping of some tokens of both Spanish /i/ and /e/ to Dutch /ɪ/) or two-category assimilation (up to two boundary mismatches). Also, one might think that there are few problems in lexicalization. After all, Dutch has as many as 12 primary monophthongs in its vowel inventory (i y u ɪ ʏ ɛ ɔ ɑ aː eː øː oː), whereas Spanish has no more than five (i e a o u). As far as the lexical storage of Spanish morphemes is concerned, it seems to be an advantageous strategy for the Dutch learner of Spanish to re-use five of her already existing vowel categories, perhaps the set /i ɛ ɑ ɔ u/. For learners who follow this strategy, the set of L2 vowels forms a *subset* of the L1 vowel inventory.

But we can identify two possible problems specific to MCA. First, even the learner with the above advantageous strategy has to have *learned* that the Dutch categories /i/ and /ɛ/ exist in Spanish as well, and that the (acoustically intermediate) Dutch category /ɪ/ does *not* exist in Spanish; otherwise, she might have created spurious lexical contrasts (/i/-/ɪ/ or /ɪ/-/e/), with possible repercussions in production. This is a typical example of a *subset problem* in acquisition, which in general is the problem of how the learner can learn on the basis of positive evidence alone that some feature does not exist in the target language. But even if the beginning learner overcomes the lexical subset problem, her "too good" category differentiation leaves her with a *perceptual* subset problem, i.e., even though she 'knows' that she should only perceive /i/ and /ɛ/, she cannot help perceiving many /i/ or /e/ tokens as /ɪ/ (at least if we assume again that the initial state of her perception system is a copy of her L1 perception system). The learner will try to solve the perceptual subset problem by disposing of the extraneous category /ɪ/. She can effectuate this by merging it with either or both of /i/ and /ɛ/, or by moving the two boundaries together, or by deleting it in some other way.

2. Evidence for MCA: front vs. back vowels in Dutch learners of Spanish

Since the MCA pattern goes beyond just presenting a boundary mismatch problem to the learner, we predict that it poses a larger challenge to the learner than two-category assimilation does. As our test case, we chose the perception of Spanish vowels by Dutch learners, since these learners appear to manifest MCA for the front vowels (Fig. 3) and two-category assimilation for the back vowels (Fig. 2). If Dutch learners have poorer front than back vowel categorization in Spanish, this will provide evidence for the problematic nature of MCA as compared to two-category assimilation. Since we also wanted to find out whether any vowel categorization biases change with language experience, we considered learners with various levels of L2 experience. If their bias for front vowel errors decreases with experience, this will suggest that learners change their behaviour in order to solve a problematic MCA pattern during their L2 perceptual development.

The subjects were 38 Dutch learners of Spanish, most of whom were enrolled in the Spanish degree programme at the University of Amsterdam. The learners were divided into three groups on the basis of a language background questionnaire: 11 were judged to be beginners, 18 intermediate, and 9 advanced. Three other subjects turned out to be bilingual Spanish-Dutch. We also tested 11 Dutch-only controls and 44 Spanish-only controls (whose number is so large because their perception might vary according to what part of Spain or Latin America they came from).

In a forced-choice labelling task, the listeners were presented with 25 tokens of each of the five Spanish vowels, embedded between two consonants. These 125 CVC chunks were cut from a Spanish text read aloud by a female Spanish speaker originally from Madrid, now a university teacher of Spanish speaking proficiency in the Netherlands. The stimuli were preceded by a carrier phrase (e.g. *la palabra [kes]* 'the word [kes]'). For reasons that will become clear in §3.1, the embedding consonants were chosen to sound ambiguously Dutch/Spanish. In addition, there were 55 filler stimuli with very Spanish-sounding consonants (e.g. [ror]). The 180 CVCs were presented in blocks of 40 with the option of a short break after each block. The listeners were asked to identify the Spanish vowel in the CVCs by clicking on one of the five Spanish vowels that appeared on a computer screen in Spanish orthography. The task lasted for 7 to 8 minutes.

The results are shown in Figure 4: each of the graphs displays the results for a different listener group. In the graphs, we plotted two types of categorization errors that each listener made: along the horizontal axis, the number of front vowel categorization errors (the number of times the listener clicked /e/ when the stimulus contained an /i/ token or vice versa), and along the vertical axis the number of back vowel categorization errors. Since there were 25 different tokens for each vowel the maximum number for each type of error is 50. The dotted

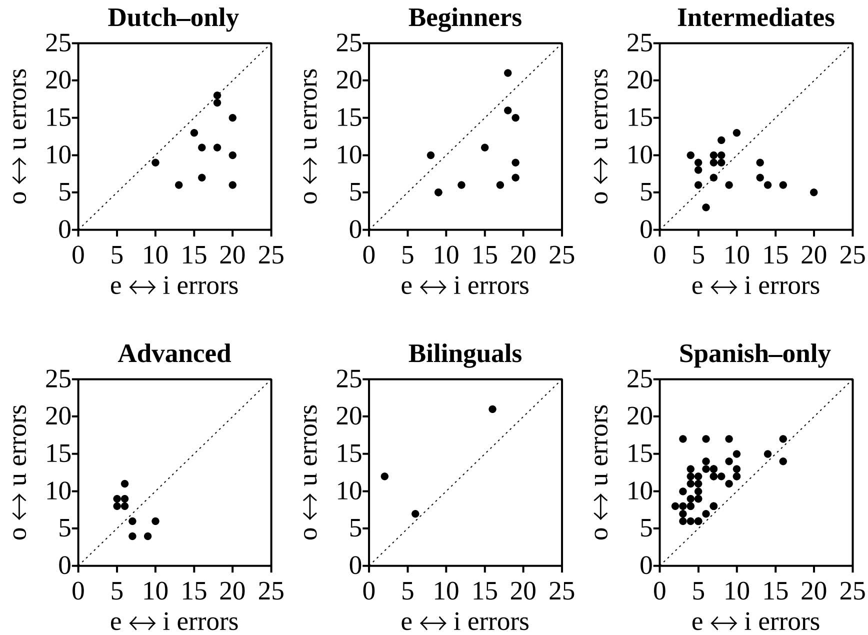

Fig. 4. Identification errors for front vs. back vowels, for 6 listener groups.

diagonal line in each graph shows where a listener would be located if she happened to make equal numbers of front and back vowel categorization errors.

On average, the Dutch-only control group, probably basing themselves on orthography, made more errors with the front than with the back vowels.[1] The beginners showed a comparable bias towards front vowel errors. The intermediate and advanced learner groups show more or less equal numbers of front and back vowel errors. The Spanish-only listeners make more errors with the back than with the front vowels (the fact that many native speakers made a large total number of errors may be due to their region of origin; the five listeners from the Madrid area happened to be in the lower left of the figure, i.e., they made very few errors). Non-parametric statistical tests on the ratio of the number of front vowel errors and the number of back vowel errors show that intermediate and advanced learners have a smaller bias towards front vowel errors than the beginners, i.e., the front error bias decreases with experience level. Also, the Spanish controls turn out to have a smaller front vowel bias than

1. The three listeners plotted at or near the diagonal turned out to be from areas where the local languages (Limburgian, Frisian, Low Saxon) have symmetric vowel systems, i.e., they have an /ɔ ʊ u/ distinction which is often transferred to the local variety of Dutch. We like to think that this is no coincidence.

any of the learner groups, i.e., the average learner of any experience level has not attained native-like perceptual skills.

We observe that our prediction is borne out: Dutch learners of Spanish perform poorer on front than on back vowels; also, learners tend to achieve a better performance and specifically fewer front errors as their experience level increases. These observations provide evidence for the existence of MCA, its problematic nature, and its reduction during L2 development.

3. Direct evidence for the relation of MCA to perceptual proficiency

The evidence provided in §2 for the existence and problematicity of MCA was rather indirect. It might be the case that the front error bias is caused by some other phenomenon that we have not been able to track down. For this reason, the actual design of the experiment did not just contain the identification task described in §2. Rather, it consisted of three separate tasks, of which the one described in §2 was only the third. Each of the three tasks aimed at establishing a specific fact about MCA and its problematic nature:

(1) Multiple-category assimilation exists, i.e., various tokens of Spanish /e/ and /i/ are indeed perceived as the three Dutch categories /ɛ/, /ɪ/, /i/. This was established in a task in which the learners had to classify Spanish vowels into Dutch categories, while thinking they were listening to Dutch.

(2) Learners do something about MCA (showing its problematicity): they avoid perceiving the /ɪ/ category if they listen with their ‘Spanish ears’. This was established by comparing the results of the first task with those of a second task, which was similar to the first except that the learners thought they were listening to Spanish.

(3) The extent to which learners do something about MCA (i.e. the difference between the first and second tasks with respect to the use of the /ɪ/ category) correlates with their proficiency in identifying Spanish vowels, as measured by the third task.

Fact (3), which involves all three tasks, will confirm our main hypothesis, namely that it is indeed MCA that causes the problems that Dutch learners have with Spanish front vowels. We will now describe in detail the three tasks and their results.

3.1. First task: direct evidence for multiple-category assimilation

The first task aimed at showing that learners use all three short Dutch vowel categories /ɛ/, /ɪ/, and /i/, when classifying Spanish vowels. We presented the listeners with the same 25 CVC tokens of each of the five Spanish vowels as described in §2. The stimuli were embedded in a Dutch carrier phrase (*luister naar [kes]* ‘listen to [kes]’), pronounced by the same speaker (a balanced Spanish-Dutch bilingual), and the listeners were (untruthfully) told that the

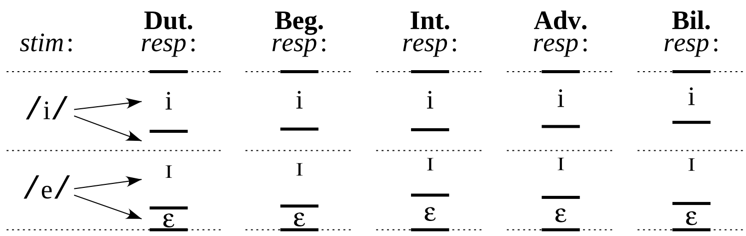

Fig. 5. Average Dutch classification of the 50 Spanish front vowel tokens, when the listeners were told that the language was Dutch.

stimuli had been cut from a Dutch text. To enhance the Dutch ambience, the CVC chunks (for *all* the tasks) had to contain consonants that could have been Dutch, i.e., rhotics, approximants, and some dentals were avoided. Also, we supplemented the 125 Spanish target stimuli with 55 very Dutch-sounding decoys (like [høːs]) cut from a Dutch text read by the same speaker. Although several listeners still reported having heard un-Dutch or even Spanish vowels, this seemed to be the closest we could get towards fooling the listeners into thinking the language they heard was Dutch. The listeners had to choose from the 12 Dutch vowel categories, which were presented orthographically between 'h' and 'k' to avoid the ambiguity problems of Dutch orthography.

Table 1 and Figure 5 show how the five Dutch-speaking groups classified the 25 /e/ and the 25 /i/ tokens. The figure is to be interpreted as follows. The vertical dimension represents a continuum of vowel height (F1). On the left, we see the production of the 50 stimulus tokens. For simplicity, we assume that the 25 /i/ tokens are higher than the 25 /e/ tokens, and that the two groups have height distributions that cause the 50 tokens to form a near-continuum along the height dimension. Thus, the top and bottom dotted lines demarcate the 50 tokens, and the middle dotted line represents the production boundary between the /i/ and the /e/ tokens. On the right, we see the responses for the five groups

Table 1. Average Dutch classification of the 50 Spanish front vowel tokens, when the listeners were told that the language was Dutch. Standard deviations between parentheses.

Group	*N*	Task 1 /i/→/ɪ/	Task 1 /e/→/ɪ/	Task 1 total /ɪ/ use
Dutch-only	11	6.1 (2.4)	18.1 (1.5)	24.2 (2.5)
Beginners	11	6.8 (3.5)	17.5 (3.0)	24.3 (5.3)
Intermediate	18	6.7 (2.3)	14.0 (4.0)	20.7 (3.6)
Advanced	9	7.7 (4.5)	14.8 (5.3)	22.4 (5.5)
Bilingual	3	9.0 (5.2)	16.7 (7.1)	25.7 (9.5)

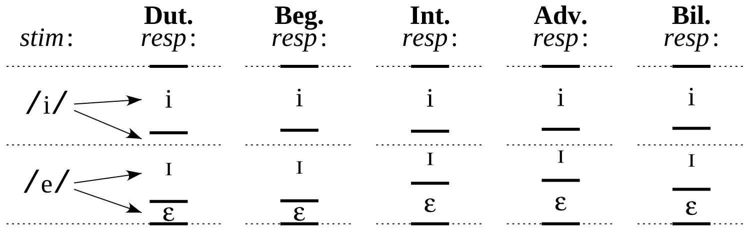

Fig. 6. Average Dutch classification of the 50 Spanish front vowel tokens, when the listeners were told that the language was Spanish.

who knew Dutch. For the Dutch-only group, the average listener classified 6.1 /i/ tokens as /ɪ/, and for simplification we assume that the remaining 18.9 were classified as /i/; to show this, the figure has an /ɪ/-/i/ *boundary line* at a height of 6.1/25 between the dotted line in the middle and the dotted line at the top. The average Dutch-only listener classified 18.1 /e/ tokens as /ɪ/, and for simplification we assume that the remaining 6.9 were classified as /ɛ/, so that the /ɛ/-/ɪ/ boundary line appears at a height of 6.9/25 between the dotted line at the bottom and the dotted line in the middle. In this way, the figure was made to illustrate the relative extent to which the three categories were used for perceiving the 50 Spanish front vowel tokens. We see, for instance, that the Intermediate group uses the /ɪ/ category to a smaller degree than any of the other four groups, but a correlation test does not reveal a reliable pattern of development for the three learner groups.

From Table 1 and Figure 5, we conclude that every group uses all three Dutch categories when perceiving tokens of the two Spanish vowels, i.e., they all show multiple-category assimilation.

3.2. Second task: reduction of MCA in the Spanish perception mode

The second task aimed at showing that learners use the /ɪ/ category less in their Spanish perception than in their Dutch perception. Ten minutes after the first task (in which the subjects filled in the questionnaire), we presented the listeners with the same target stimuli as in the first task, but embedded in a Spanish carrier phrase (*la palabra [kes]* 'the word [kes]'), and supplemented with 55 very Spanish-sounding fillers (like [roɾ]). We told the listeners, now truthfully, that the language they were listening to was Spanish, but we asked them to listen to the stimuli with 'Dutch ears' and to classify them as Dutch vowels. The results for the front vowels are in Table 2 and Figure 6. First, it turns out that there is a reliable effect of experience level on the extent of the use of the /ɪ/ category (for the three learner groups: $r = -0.38$; less than zero; $p < 0.005$). When we compare Figures 5 and 6, we see that the listeners use the /ɪ/ category less when they think the language they hear is Spanish than when they

Table 2. Average Dutch classification of the 50 Spanish front vowel tokens, when the listeners were told that the language was Spanish. Standard deviations between parentheses.

Group	*N*	Task 2 /i/→/ɪ/	Task 2 /e/→/ɪ/	Task 2 total /ɪ/ use	Task1 – Task 2 /ɪ/ reduction
Dutch-only	11	3.9 (1.4)	17.9 (1.4)	21.8 (1.8)	2.4 (2.7)
Beginners	11	4.7 (2.2)	17.7 (3.9)	22.5 (5.1)	1.8 (4.9)
Intermediate	18	4.4 (2.5)	12.1 (4.8)	16.6 (6.0)	4.1 (3.5)
Advanced	9	5.0 (2.8)	11.2 (6.0)	16.2 (6.3)	6.2 (3.2)
Bilingual	3	5.3 (5.1)	14.0 (10.1)	19.3 (13.3)	6.3 (4.5)

think the language they hear is Dutch. Apparently, the learners are unable to comply completely with the explicit task of listening with ‘Dutch ears’. Rather, their different behaviour shows us that they actually use their Spanish ears in the second task. This suggests that there is a low-level language-dependent *perception mode*: people do different things depending on the language that they hear, or think they hear.

Of course, the mere fact that there is a statistically significant difference between the learner groups in the second task, but not in the first task, does not imply that the listeners behaved differently in the two tasks. To assess the difference between the tasks, we have to calculate, for each listener, the degree of /ɪ/ reduction across the two tasks, which is computed by subtracting her use of the /ɪ/ category in the second task from that in the first task. The mean results are in the last column of Table 2, which is computed as the difference of the ‘total /ɪ/ use’ columns of Tables 1 and 2. The degrees of /ɪ/ reduction for the intermediate and advanced groups are reliably different from zero. The fact that these learners use the /ɪ/ category less in the second task than in the first, shows us that if they switch on their Spanish perception mode, they reduce the extent to which they manifest MCA. The extent of this reduction is developmental: for the three learner groups, it correlates with experience level: $r = 0.39$ (greater than zero, $p < 0.005$).

3.3. Third task: perceptual proficiency

The third task was already described in §2. The 125 target stimuli and the 55 fillers were exactly the same as in the second task, but the response categories were the five Spanish vowels. The results for the front vowels are in Table 3 and Figure 7. As a measure of perceptual proficiency, we take the location of the boundary between the /e/ and /i/ categories, e.g., if 8 /e/ tokens were classified as /i/, and 2 /i/ tokens were classified as /e/, the boundary would lie at a height of 6/25 = 0.24 categories below the *optimal* perceptual boundary (which must

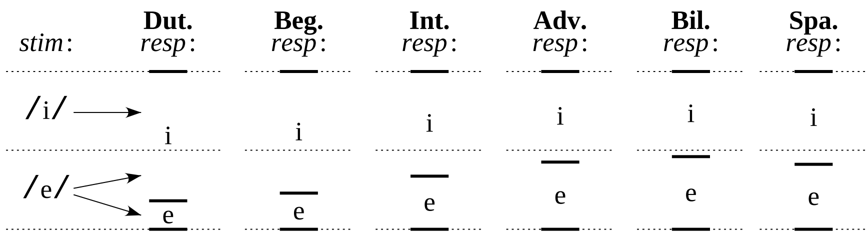

Fig. 7. Average identification of the 50 Spanish front vowel tokens.

be the dotted line in the middle of Figure 7 since it has to match the production boundary). The reason we use this measure rather than an overall error rate, is because the latter is influenced more by factors like attention and non-linguistic individual characteristics.

We see that the location of the boundary becomes more native-like as the learners' experience level increases. For the three learner groups, the boundary mismatch correlates with experience level (r = –0.57, less than zero, $p < 0.0001$).

We can now combine the results of all the three tasks to test our main hypothesis, namely that it is MCA that causes the front vowel problems. This hypothesis will be confirmed if it turns out that having more MCA leads to poorer identification. In Figure 8 we plot the boundary location of the third task (i.e. our measure of perceptual proficiency) as a function of the reduction of /ɪ/ between the first and second tasks (our measure of MCA reduction), for each of the 38 listeners in the three learner groups. Thus, Figure 8 separately shows the values that we had averaged over in the last columns of Tables 1 and 2. The boundary location turns out to correlate strongly with the degree of MCA reduction between the Dutch and Spanish perception modes: r = –0.62 (less than zero, $p < 0.00001$; 95% = –0.78...–0.37). Since the experiment was designed beforehand to test exactly this correlation, we are allowed to interpret this positive result as evidence for the existence of language-dependent low-level

Table 3. Average identification of the 50 Spanish front vowel tokens. Standard deviations between parentheses.

Group	*N*	Task 3 /e/→/i/	Task 3 /i/→/e/	boundary mismatch
Dutch-only	11	16.4 (3.1)	0.4 (0.7)	16.0 (3.2)
Beginners	11	14.2 (5.2)	0.6 (1.1)	13.5 (6.0)
Intermediate	18	8.7 (4.6)	0.5 (1.0)	8.2 (5.0)
Advanced	9	5.2 (2.9)	1.6 (1.7)	3.7 (4.4)
Bilingual	3	5.0 (4.0)	3.0 (3.5)	2.0 (2.0)
Spanish-only	44	5.5 (3.5)	1.0 (0.9)	4.5 (3.9)

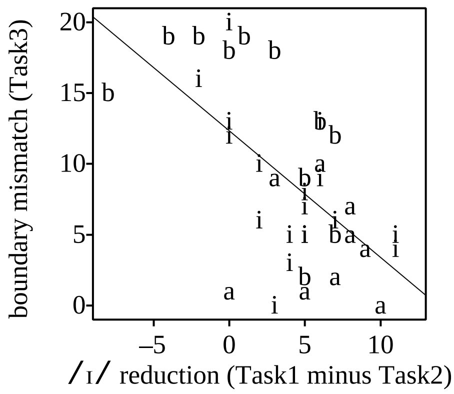

Fig. 8. Perceptual proficiency (0 = optimal, 20 = poor)
as a function of the degree of MCA reduction (0 = none, 10 = strong),
for the 38 learners (b = beginners, i = intermediate, a = advanced).

perception modes. The Spanish perception mode of Dutch learners then gradually develops into a system that accurately places the vowel boundaries, by getting rid of the extraneous /ɪ/ category.

We conclude that the availability of extra categories can render an advantageous one-to-one mapping strategy (i.e., Spanish /i-e/ to Dutch /i-ɛ/) inaccessible to the learner, and that learners take corrective action when confronted with such a situation.

4. Conclusions

Multiple-category assimilation turns out to exist: learners perceive three categories where the target language has only two. While this existence may have been found before, the current paper established the *problematic nature* of MCA for purposes of accurate L2 categorization. We regard the MCA problem as an instance of the subset problem in phonology: the target language contains a subset of the categories of the L1, and the task for the learner is to dispose of the L1 category not shared by the target language. Fortunately, the learner can achieve this because in her L2 perception she can make use of a perception mode dedicated to the L2 (though copied from her L1 perception mode at the beginning), which she can freely change without affecting her L1 perception. The concept of ‘perception mode’ was inspired by our work on Optimality-Theoretic *perception grammars* (Escudero and Boersma 2001), which the learner can modify by the application of a *gradual learning algorithm* (Boersma and Hayes 2001). An application of this linguistic modelling to MCA phenomena will be fruitful, since this modelling is generally capable of coping with overt subset phenomena; we will present computer simulations in later work.

References

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

Best, Catherine, Gerald McRoberts, and Elizabeth Goodell (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Journal of the Acoustical Society of America* **109**: 775–794.

Boersma, Paul, and Bruce Hayes (2001). Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry* **32**: 45–86.

Escudero, Paola (2001). The role of the input in the development of L1 and L2 sound contrasts: Language-specific cue weighting for vowels. In Anna H.-J. Do, Laura Domínguez, and Aimee Johansen (eds.), *Proceedings of the Annual Boston University Conference on Language Development* **25**: 250–261. Somerville, Mass.: Cascadilla Press.

Escudero, Paola, and Paul Boersma (2001). Modelling the perceptual development of phonological contrasts with Optimality Theory and the Gradual Learning Algorithm. Rutgers Optimality Archive **439**, http://roa.rutgers.edu. To appear in *Proceedings of the 25th Penn Linguistics Colloquium*.

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

Schmidt, Anna Marie (1996). Cross-language identification of consonants. Part 1. Korean perception of English. *Journal of the Acoustical Society of America* **99**: 3201–3211.

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

Strange, Winifred, Reiko Akahane-Yamada, Rieko Kubo, Sonja Trent, and Kanae Nishi (2001). Effects of consonantal context on perceptual assimilation of American English vowels by Japanese listeners. *Journal of the Acoustical Society of America* **109**: 1691–1704.