# Learning to perceive a smaller L2 vowel inventory: an Optimality Theory account

**Paul Boersma & Paola Escudero, University of Amsterdam**

**June 30, 2007**

This paper gives an Optimality-Theoretic formalization of several aspects of the acquisition of phonological perception in a second language. The subject matter will be the acquisition of the Spanish vowel system by Dutch learners of Spanish, as evidenced in a listening experiment. Since an explanation of the learners' acquisition path requires knowledge of both the Dutch and the Spanish vowel system, the 12 Dutch and 5 Spanish vowels are presented in Figure 1. Along the vertical axis we find the auditory correlate of perceptual vowel height (first formant, F1), and along the horizontal axis the auditory correlate of perceptual vowel backness (second formant, F2), whose articulatory correlates are tongue backness and lip rounding. A third auditory dimension, duration, is implicit in the length sign (":") used for 4 of the 12 Dutch vowels.

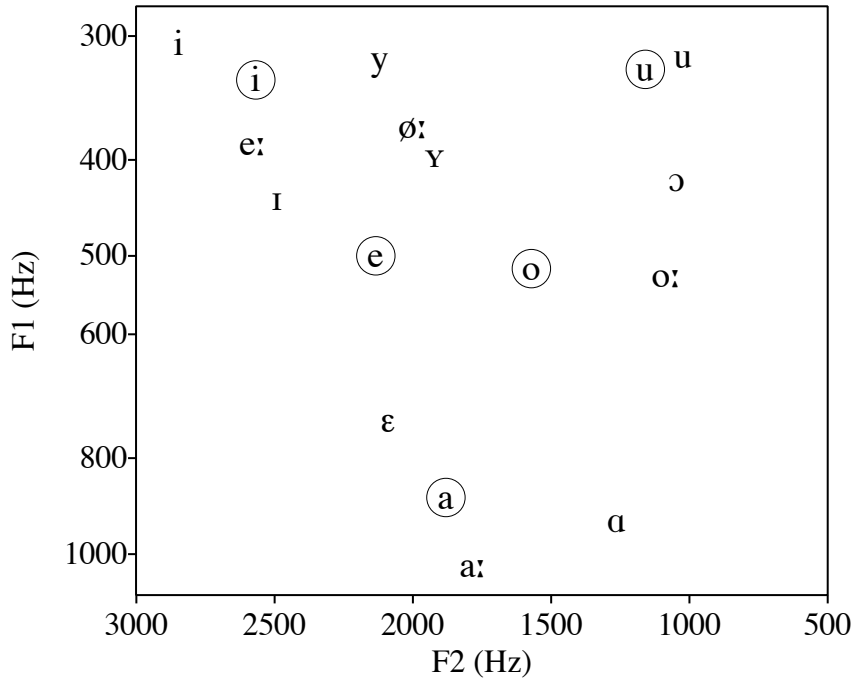


**Fig.1.** The 5 Spanish vowels (circled) amidst the 12 Dutch vowels.

To control for speaker-dependent vocal tract dimensions, we based the two sets of formant values in Figure 1 on the speech of a single speaker, a perfect Spanish-Dutch bilingual (moved to the Netherlands when she was 12, currently a teacher of Spanish speaking proficiency at the University of Amsterdam, with no noticeable foreign accent in either Dutch or Spanish). We see the usual features of the Dutch vowel system: /i/, /y/ and /u/ at the same height, /eː/ and /øː/ at the same height, /ɪ/ and /ʏ/ at the same height, /ɛ/ more open than /ɔ/, /ɑ/ more open than /ɛ/ but somewhat closer than /aː/.

As for most speakers of Dutch, /aː/ is front and /ɑ/ is back. As for many speakers, /ɪ/ and /ʏ/ are a bit lower than /eː/ and /øː/. The height of /ɔ/ shows that this speaker is from one of those large areas that merge the reflexes of both historical /ɔ/ and /ʊ/ into a single relatively high variant at the height of /ɪ/ and /ʏ/ (if this had been true of all speakers of Dutch, a better symbol for the phoneme /ɔ/ would have been /ʊ/). A more idiosyncratic feature of the speaker's regional accent is the low position in the chart of the vowel /oː/, which is due to its large degree of diphthongization (i.e., the three higher mid vowels are phonetically realized by this speaker as [ei], [øy], [ɔu]). As for this speaker's Spanish vowel system, we see that /a/ is rather front, that /e/ and /o/ are not close to any Dutch vowel, and that the extent of the Spanish vowel space is somewhat smaller than that of the Dutch vowel space, with a notable centralization of /o/. The patterns are compatible with what is known about Dutch (Pols, Tromp, and Plomp 1973; Koopmans-Van Beinum 1980), about Spanish (Bradlow 1995, 1996), and about the crosslinguistic correlation between the size of a language's auditory vowel space and the size of its vowel inventory (Liljencrants and Lindblom 1972; Lindblom 1986).

## 1. Ease and difficulty for Dutch learners of Spanish vowels

For Dutch learners of Spanish who want to master the Spanish vowel system, there is something easy as well as something difficult about it. The ease lies in creating lexical representations for Spanish vowels, while the difficulty lies in perception, i.e. in the mapping from raw auditory data to discrete representations that can be used for lexical access.

### 1.1. Easy: lexical symbols for L2 vowels

When native speakers of Dutch learn to use the vowel system of the Spanish language, they seem to have the advantage that the target language has fewer vowels than their native language, so that they have the option of reusing a subset of their native vowel categories for the storage of Spanish lexemes. The phonological representations of entries in the Spanish lexicon can get by with only five vowel categories, which we will denote as $|a|_S$, $|e|_S$, $|i|_S$, $|o|_S$, and $|u|_S$ (in our notations, subscript S is used for structures in the minds of native speakers of Spanish, and underlying forms are given within pipes).[i] Thus, the lexical representation of the word *centrifugado* 'centrifugated' is $|θentɾifuɣaðo|_S$ for native speakers of (European) Spanish. Native speakers of Dutch have to maintain at least 12 vowel categories in their native lexical representations: $|ɑ|_D$, $|aː|_D$, $|ɛ|_D$, $|ɪ|_D$, $|eː|_D$, $|i|_D$, $|ʏ|_D$, $|øː|_D$, $|y|_D$, $|ɔ|_D$, $|oː|_D$, $|u|_D$ (subscript D for structures in the minds of native speakers of Dutch). When learning Spanish, then, they could simply[ii] reuse five of these for representing their L2 Spanish lexemes; no category split, no category creation would be necessary. As we will see when discussing the results of our listening experiment (§1.5), this is what the learners indeed seem to do. The following simplified list shows which Dutch vowels are reused for which Spanish vowels in the interlanguage:

(1) *Identification of lexical symbols for Dutch learners of Spanish*

$|ɑ|_D$ — $|a|_S$
$|ɛ|_D$ — $|e|_S$
$|ɪ|_D$ — $|i|_S$
$|ɔ|_D$ — $|o|_S$
$|u|_D$ — $|u|_S$

Note that this identification does not describe the knowledge of the learners; rather, it is an observation that we as linguists can infer from experimental tasks (as we do in §1.5). The identification in (1) means, for instance, that the Dutch learner's underlying representation of Spanish *centrifugado* is $|θɛntrɪfuɣɑdɔ|_D$. Also note that our use of vowel symbols is not meant to suggest crosslinguistic identity: $|u|_D$ is not a priori more similar to $|u|_S$ than $|ɑ|_D$ is to $|a|_S$.[iii]

### 1.2. Difficult: perceptual boundaries of L2 vowels

While the reuse of existing categories is advantageous in itself, there is an additional gain in the identifications in (1), which are far from arbitrary. This section first shows that these identifications are largely based on language-specific perceived (auditory and structural) similarity, and then shows why such an identification strategy is advantageous.

Typical tokens of an intended native Spanish $|a|_S$ tend to sound like a short somewhat front open vowel, which in a narrow auditory-phonetic transcription is [a] or [a̠]. The spectral quality (F1 and F2) of these tokens is close to that of typical tokens of Dutch $|aː|_D$, which are phonetically realized like the long cardinal IPA open front vowel [aː]; the duration of the Spanish tokens, however, is close to that of typical tokens of Dutch $|ɑ|_D$, which typically sound like the slightly rounded low back vowel [ɑ̟]. Since Dutch listeners, when having to categorize sounds in the [a]-[ɑ̟]-[ɑ̟ː]-[aː] region, weigh the duration cue much higher than the spectral cues (Gerrits 2001: 89), they will classify the Spanish [a]-like tokens as $/ɑ/_D$ rather than as $/aː/_D$.[iv] Another option is to perceive these tokens as $/ɛ/_D$, whose typical realizations in Dutch sound like the cardinal IPA open mid front vowel [ɛ]. In the listening experiment partly discussed below we found that non-Spanish-learning speakers of Dutch perceived Spanish $|a|_S$ as $/ɑ/_D$ 60 percent of the time, as $/ɛ/_D$ 27 percent of the time, and as $/aː/_D$ 4 percent of the time. So it seems that language-specifically perceived similarity, with duration as the main determining cue, largely explains the identifications in (1).[v]

So why would learners choose to base their identifications on perceived similarity, i.e. what advantage does it give them to reuse Dutch categories whose auditory distributions include the most typical tokens of the Spanish correspondents, as in (1)? To answer this, we have to consider what is involved in the listener's *comprehension* task, i.e. her mapping from auditory information to lexical representations that make contact with meaning. In several theories of phonological comprehension (for an overview, see McQueen and Cutler 1997 and McQueen 2005), the process consists of two sequential levels, which can be called *perception* and *recognition*. The ("prelexical") perception process maps auditory to phonological surface representations without accessing the lexicon, and the recognition process maps the phonological

surface representations to underlying forms in the lexicon and is heavily influenced by the semantic and pragmatic context.

(2) *Two-stage comprehension model*

| | auditory representation | perception → | phonological representation | recognition → | lexical representation |
|---|---|---|---|---|---|
| e.g. | [kæso] | | /kɛsɔ/$_D$ | | \|kɑsɔ\|$_D$ 'case' |

The advantage of reusing lexical categories now becomes clear: the learner will exhibit some initial proficiency in her comprehension, at least if she transfers the perception system to her interlanguage system as well. Suppose, for instance, that the learner is in a stage at which she has already correctly stored the Spanish words |kaso|$_S$ 'case' and |keso|$_S$ 'cheese' into her interlanguage lexicon as |kɑsɔ|$_D$ 'case' and |kɛsɔ|$_D$ 'cheese'. A hundred native tokens of an intended |kaso|$_S$ will have a distribution of vowel formants (for the |a|$_S$ part) that is centred around values that are typical of a low front vowel. As suggested above, Dutch monolinguals may hear 60 of these vowel tokens as /ɑ/$_D$, 27 as /ɛ/$_D$. If learners transfer this perception to their interlanguage, they will perceive 60 instances of |kaso|$_S$ as /kɑsɔ/$_D$, 27 as /kɛsɔ/$_D$. In the majority of the cases, then, a beginning learner will perceive /kɑsɔ/$_D$, from which the lexical item |kɑsɔ|$_D$ 'case' can be retrieved quite easily. Thus, comprehension is well served by an initial transfer of native perception (which presupposes an initial transfer of native lexical symbols) to the interlanguage.

But an interlanguage perception system that is identical to the native perception system is not perfect yet. In the example above, 27 percent of intended |kaso|$_S$ tokens, perhaps the most fronted and raised ones, will be perceived as /kɛsɔ/$_D$, from which it is not so easy to retrieve the lexical item |kɑsɔ|$_D$ 'case'.[vi] To improve, the learner will have to learn to perceive tokens in the auditory [æ] region as /ɑ/$_D$ rather than as /ɛ/$_D$ when listening to Spanish. Preferably, though, tokens in that same region of auditory space should continue to be perceived as /ɛ/$_D$ if the learner is listening to Dutch. The following table sums up the ways in which [æ] would then be perceived in the five cases we discussed:

(3) *Five perceptions of the auditory form* [æ]

Monolingual Spanish: [æ] → /a/$_S$
Monolingual Dutch: [æ] → /ɛ/$_D$
Beginning learners when listening to Spanish: [æ] → /ɛ/$_D$ (transfer)
Proficient learners when listening to Spanish : [æ] → /ɑ/$_D$ (native-like)
All learners when listening to Dutch: [æ] → /ɛ/$_D$ (double perception systems)

The situation in (3) would require a duplication of the learner's perception system, where the interlanguage perception system starts out as a clone of the native perception system but subsequently develops towards something more appropriate for the comprehension of the target language, without affecting the L1 perception system (Escudero & Boersma 2002). The experiment described below, in which we show that

Dutch learners of Spanish exhibit different perceptual behaviour when they think they are listening to Dutch than when they think they are listening to Spanish, provides evidence for two separate perception systems in L2 learners.

### 1.3. The listening experiment: method

The method (stimulus material, subjects, tasks) was described before in Escudero & Boersma (2002). We repeat here only what is relevant for the present paper.

**Stimulus material.** The same bilingual speaker as in Figure 1 read aloud a Spanish text, from which we cut 125 CVC (consonant-vowel-consonant) tokens. The consonants were selected in such a way that each of the 125 CVC tokens could pass for a licit Dutch syllable (apart from the vowel).

**Subjects.** Thirty-eight Dutch learners of Spanish performed the three tasks described below. The learners were from various parts of the Netherlands, so that their vowel systems may differ from the one in Figure 1 mainly in the location of /ɔ/ (which for many speakers has [ɔ]- and [ʊ]-like positional variants) and in the location of /oː/ (which for many speakers has the same degree of diphthongization, and the same height, as /eː/ and /øː/). There were two control groups: 11 Dutch non-learners of Spanish performed the first and second tasks only, and 44 native speakers of Spanish performed the third task only.

**First task.** In the first task the subjects were told that they were going to listen to a number of Dutch CVC syllables and had to classify the vowel into the Dutch classes /ɑ/, /aː/, /ɛ/, /ɪ/, /eː/, /i/, /ʏ/, /øː/, /y/, /ɔ/, /oː/, /u/. But what the subjects actually heard was a randomized set of the 125 Spanish tokens. To enhance the Dutch *perception mode*, the tokens were interspersed with 55 CVC tokens that were cut from a Dutch text spoken by the same bilingual speaker; many of these 55 tokens contained very Dutch-sounding vowels and consonants, often corresponding to a recognizable Dutch word, e.g. /ɦøːs/ ‘really’. Also, the 180 CVC tokens were embedded within a Dutch carrier phrase (*luister naar...*).

**Second task.** The second task differed from the first only in the perception mode that we wanted to bring the subjects in. So we told the subjects (correctly, this time) that they were going to listen to Spanish CVC sequences, and we interspersed the 125 CVC tokens (which were the same as in the first task) with 55 very Spanish-sounding tokens (e.g. /roɾ/) and embedded the 180 stimuli within a Spanish carrier phrase (*la palabra...*). Importantly, though, we told the listeners to try to “listen with Dutch ears” to these stimuli and to classify the 180 tokens into the 12 Dutch vowel classes.

**Third task.** The third task differed from the second only in that we told the listeners to listen with Spanish ears and to classify the 180 tokens into the 5 Spanish vowel classes. This task, then, simply tested the learners’ proficiency in the perception of the target language.

### 1.4. The listening experiment: results

When the subjects thought that the language they were hearing was Dutch (Task 1), they responded differently from when they thought the language was Spanish (Task 2): they turned out not to be able to completely “listen with Dutch ears” in Task 2. For details, see Escudero & Boersma (2002, to appear). We now describe the three main differences between the results of the two tasks. In Task 2, the group of 38 listeners

avoided responding with "ɪ". Although most tokens that were scored as "ɪ" in the first task were still scored as "ɪ" in the second (namely 599), many tokens that were scored as "ɪ" in the first task were scored as "i" or "ɛ" in the second (namely, 120 and 101, respectively). The reverse drift was much smaller: the number of tokens that were scored as "i" or "ɛ" in the first task but as "ɪ" in the second were only 27 and 57, respectively. Since the differences between 120 and 27 and between 101 and 57 are significantly greater than zero (see Escudero & Boersma to appear for the statistical tests), we can reliably say that the listener group shied away from the "ɪ" response in the second task. The learners showed an analogous behaviour for "ʏ" responses, which were avoided in the second task, where many of them were replaced with "u" and "ɔ" responses. A third reliable effect was the shift of the "ɑ" response: many tokens that were scored as "ɛ" when the listeners were fooled into thinking the language was Dutch were scored as "ɑ" when the listeners knew it was Spanish, and many tokens that were scored as "ɑ" in the first task were scored as "ɔ" in the second. Finally, the long vowels "aː", "eː", "oː" and "øː" were generally avoided in the responses in Task 2.

The learners showed developmental effects. The degree of "ɪ" avoidance in Task 2 relative to Task 1 correlated with the experience level of the learners (who were divided into 11 beginners, 18 intermediate, and 9 advanced on the basis of an independent language background questionnaire) as well as with the perceptual proficiency level as measured in Task 3 (Escudero & Boersma 2002).

### 1.5. The listening experiment: interpretation

The shift from "ɛ" responses in the first task toward "ɑ" responses in the second shows that the learners reused their Dutch /ɑ/$_{D}$ category for perceiving Spanish /a/$_{S}$. We can explain this shift by assuming that for [æ]-like auditory forms some of the learners follow the mode-dependent strategies predicted in (3) for proficient learners:

(4) *Two separate language modes for a proficient Dutch learner of Spanish*

| Language mode | Token | Perception | Response |
|---|---|---|---|
| Dutch | [æ] | /ɛ/$_{D}$ | "ɛ" |
| Spanish | [æ] | /ɑ/$_{D}$ | "ɑ" |

For the Spanish vowel /i/$_{S}$, which could in principle have been identified with Dutch /ɪ/$_{D}$ or with Dutch /i/$_{D}$, the avoidance of "ɪ" in the second task shows that in fact Spanish /i/$_{S}$ was identified with Dutch /i/$_{D}$. This shows that (1) is correct. The avoidance of the four long vowels in both the first and second tasks confirms the expectation mentioned in §1.2 that duration is a strong auditory cue that can override any spectral similarity.

The developmental effects can be explained by an initial *transfer* of the native perception system to the interlanguage, followed by *lexicon-guided learning*. Thus, the Dutch-appropriate perception of [æ] as /ɛ/$_{D}$ is transferred to the initial state of the learner's interlanguage, so that a beginning Dutch learner of Spanish will perceive [æ] as /ɛ/$_{D}$, regardless of whether she listens to Dutch or to Spanish. When she is listening to Spanish, however, the lexicon will often issue an error message. If the learner perceives an incoming [kæso] as /kɛsɔ/$_{D}$, for instance, higher conceptual processing may force the lexicon to recognize /kɛsɔ/$_{D}$ as |kɑsɔ|$_{D}$ 'case'. If that happens, the lexicon

can 'tell' the perception system to modify itself in such a way that a /kɑsɔ/$_D$ perception becomes more likely in the future (note that the existence of minimal pairs is not required). Both the perception system and lexicon-guided learning are formally modelled in the following sections.

## 2. An explicit phonological model of perception

Perception researchers agree that prelexical perception, i.e. the mapping from auditory to phonological representations, is a language-dependent process for all speakers from about 9 months of age (Werker and Tees 1984; Jusczyk, Cutler, and Redantz 1993; Polka and Werker 1994). This language dependence is enough reason for us as linguists to want to model prelexical perception by linguistic means, e.g. to model it by Optimality-Theoretic constraint ranking, as has been done before by Boersma (1997, 1998, 1999, 2000), Hayes (2001), Escudero and Boersma (2003, 2004), and Pater (2004).[vii] Tesar's (1997, 1998) and Tesar & Smolensky's (2000) Optimality-Theoretic modelling of the process of *robust interpretive parsing*, i.e. a mapping from unanalysed ("overt") sequences of syllables with stress marks to full abstract hierarchical foot structures, can also be seen as a case of Optimality-Theoretic modelling of perception, an idea that was pursued by Apoussidou & Boersma (2003, 2004).[viii]

In our special case of L2 acquisition, perception can depend on the language that learners think they are listening to: the likelihood of mapping [æ] to the Dutch lexical vowel symbol /ɛ/$_D$ depends on whether the learner thinks she is hearing Dutch (more likely) or Spanish (less likely), as we mentioned in §1.4. We therefore model the behaviour of the learner with two separate perception grammars, one for her Dutch perception, which does not change during her learning of Spanish, and one for her Spanish perception, which starts out as a clone of her Dutch perception grammar and subsequently develops towards a more Spanish-appropriate grammar by the lexicon-driven optimization we introduced in §1.5.

### 2.1. Tableaus and constraints that model perception

Optimality-Theoretic perception grammars use the same decision scheme as the more usual Optimality-Theoretic production grammars. Whereas a production grammar takes an underlying lexical representation as its input and yields a pronunciation or surface structure as its output (Prince and Smolensky 1993, McCarthy and Prince 1995), a perception grammar takes an auditory representation as its input and yields a phonological surface structure as its output.

The perceptual process that we restrict ourselves to in this paper is static categorization, where the inputs are static (temporally constant) values of auditory features and the output candidates are language-specific phonological features or phonemes. Escudero & Boersma (2003) proposed that this mapping is evaluated by the negatively formulated constraint template in (5), which directly relates auditory feature values to phonological categories. The reason for its negative formulation will be discussed in §4.5.

(5) *Arbitrary cue constraints*

"A value *x* on the auditory continuum *f* should not be mapped to the phonological category *y*."

For our case, the perception of Dutch and Spanish vowels, the relevant auditory continua are the first formant (F1), the second formant (F2), and duration, and the relevant phonological categories are the 12 Dutch vowel symbols. Examples of the relevant *cue constraints* (the term is by Boersma 2005 and Escudero 2005) are therefore "an F1 of 531 Hz is not /ɔ/$_D$", or "an F2 of 1585 Hz is not /eː/$_D$", or "a duration of 150 ms is not /y/$_D$". We propose that these cue constraints are *arbitrary*, i.e. they exist for any auditory value and any vowel category, regardless of whether that auditory value is a plausible cue for that vowel category. Thus while a typical F1 value for /i/$_D$ is 280 Hz, we indiscriminately allow the presence of constraints like "an F1 of 280 Hz is not /i/$_D$" and "an F1 of 900 Hz is not /i/$_D$". It is the *ranking* of these constraints, not their presence, that determines what auditory values map to what vowel categories. Thus, in order to make it unlikely that an auditory input with an F1 of 900 Hz will ever be perceived as /i/$_D$, the constraint "an F1 of 900 Hz is not /i/$_D$" should be ranked very high, and in order to allow that [i]-like auditory events can be perceived as /i/$_D$ at all, the constraint "an F1 of 280 Hz is not /i/$_D$" should be ranked rather low.

As an example, consider the perception of the typical token of the Spanish vowel |a|$_S$, namely an [a]-like auditory event with an F1 of 877 Hz, an F2 of 1881 Hz, and a duration of 70 ms. In tableau (6) we see that the two spectral cues favour the perception of |aː|$_D$, but that in line with the finding in §1.5 these cues are overridden by the duration constraints, which assert that an overtly short vowel token (e.g. 70 ms long) should not be perceived as the vowel /aː/$_D$.

(6) *Dutch cross-language perception of a typical token of Spanish* |a|$_S$

| [a], i.e. [F1=877, F2=1881, dur=70] | [dur=70] is not /aː/$_D$ | [F1=877] is not /ɛ/$_D$ | [F2=1881] is not /ɑ/$_D$ | [F1=877] is not /aː/$_D$ | [F1=877] is not /ɑ/$_D$ | [F2=1881] is not /ɛ/$_D$, not /aː/$_D$ | [dur=70] is not /ɑ/$_D$, not /ɛ/$_D$ |
|---|---|---|---|---|---|---|---|
| /aː/$_D$ | *! | | | * | | * | |
| ☞ /ɑ/$_D$ | | | * | | * | | * |
| /ɛ/$_D$ | | *! | | | | * | * |

With (6) we can describe the behaviour of the non-Spanish-learning Dutch listeners in the experiment. There are two reasons why the listeners' responses are variable. First, the 25 |a|$_S$ tokens in the experiment were all different, so that some will have been closer to [ɛ], some to [ɑ]. Secondly, listeners are expected to show variable behaviour even for repeated responses to the same token. We model this by using *Stochastic Optimality Theory* (Boersma 1997, 1998; Boersma and Hayes 2001), in which constraints have *ranking values* along a continuous scale and in which some *evaluation noise* is temporarily added to the ranking of a constraint at each evaluation. In tableau (6) this will mean that candidate /ɑ/$_D$ will win most of the time, followed by candidate /ɛ/$_D$.

In general, the candidates in a tableau should be all 12 vowels. Since that would require including all 36 relevant cue constraints, we simplified tableau (6) to include only three candidates, so that we need only consider 9 constraints. The remaining nine candidate vowels can be ruled out by constraints such as "an F1 of 877 Hz is not /i/$_D$" and "an F2 of 1881 Hz is not /ɪ/$_D$", which are probably ranked far above "a duration of 70 ms is not /aː/$_D$", since there were no "i" or "ɪ" responses at all for intended |a|$_S$. Tableau (6) also abstracts away from constraints such as "an F1 of 280 Hz is not /ɔ/$_D$" that refer to auditory feature values that do not occur in the input of this tableau. Such constraints do exist and are ranked along the same continuum as the nine constraints in (6); the constraint "an F1 of 280 Hz is not /ɔ/$_D$" can interact with six of the nine constraints in (6), namely when the input contains a combination of an F1 of 280 Hz with either an F2 of 1881 Hz or a duration of 70 ms.

Since the four long Dutch vowels play no role in the identifications in (1) or in the perception experiment reported in §1.3, we will from now on ignore these long vowels and consider only the eight short vowels as possible candidates. This allows us to ignore the duration constraints and to focus on the spectral cues alone.

### 2.2. Lexicon-driven perceptual learning in Optimality Theory

A tableau is just a *description* of how perception can be modelled in Optimality Theory. A more *explanatory* account involves showing how the ranking of so many constraints can be learned. This section describes Boersma's (1997, 1998) proposal for lexicon-driven optimization of an Optimality-Theoretic perception grammar, as it was first applied to the ranking of arbitrary cue constraints in L1 and L2 acquisition by Escudero & Boersma (2003, 2004).

Throughout our modelling of perception we assume that the learner has already established correct representations in her lexicon. This means that the listener's recognition system (see (2)) can often reconstruct the speaker's intended vowel category, even if the original perception was incorrect. After all, the listener's recognition system will only come up with candidate underlying forms that are actually in the lexicon, and in cases of ambiguity will also be helped by the semantic context (see Boersma 2001 and Escudero 2005:214–236 for Optimality-Theoretic solutions). If the resulting underlying form differs from the perceived surface form, the recognition system can signal to the perception system that the perception has been "incorrect". We will denote such situations by marking the speaker's intention (as recognized by the listener) in the listener's perception tableau with a check mark, as in (7).

(7) *A beginning learner's misperception of a high front token of Spanish* |a|$_S$

| [F1=800, F2=1900] | [F1=800] is not /ɪ/$_D$ | [F2=1900] is not /ɔ/$_D$ | [F1=800] is not /ɔ/$_D$ | [F2=1900] is not /ɑ/$_D$ | [F1=800] is not /ɛ/$_D$ | [F1=800] is not /ɑ/$_D$ | [F2=1900] is not /ɛ/$_D$ |
|---|---|---|---|---|---|---|---|
| √ /ɑ/$_D$ | | | | *!→ | | *→ | |
| ☞ /ɛ/$_D$ | | | | | ←* | | ←* |
| /ɔ/$_D$ | | *! | * | | | | |
| /ɪ/$_D$ | *! | | | | | | |

We can assume that the constraint "an F1 of 800 Hz is not /ɛ/$_D$" in (7) is ranked lower than the constraint "an F1 of 877 Hz is not /ɛ/$_D$" in (6), because 800 Hz is closer to typical F1 values of |ɛ|$_D$ than 877 Hz is. By this lower ranking, the constraint "an F1 of 800 Hz is not /ɛ/$_D$" can be ranked below "an F2 of 1900 Hz is not /ɑ/$_D$", which is of course ranked at nearly the same height as "an F2 of 1881 Hz is not /ɑ/$_D$" in (6). This difference between (6) and (7) now makes /ɛ/$_D$ the winner. However, if the learner's postperceptual recognition tells her she should have perceived /ɑ/$_D$ because the recognized lexeme contains the vowel |ɑ|$_D$, she can mark this candidate in the tableau ("√"), and when she notices that this form is different from her winning candidate /ɛ/$_D$, she can take action by changing her perception system. The changes are depicted in the tableau by arrows: the learner will raise the ranking of the two constraints that prefer the form she considers correct ("←") and lower the ranking of the two constraints that prefer her incorrectly winning candidate ("→"), thus making it more probable that auditory events with an F1 of 800 Hz or an F2 of 1900 Hz will be perceived as /ɑ/$_D$ at future occasions, at least when she is listening to Spanish.

In order to prove that the learning algorithm just described works for Dutch learners of Spanish throughout their L1 and L2 acquisition, we will show two computer simulations. Section 3 will simulate a simplified problem, namely the L1 and L2 acquisition of the mapping from a single auditory continuum (F1) to four vowel heights (exemplified by /ɑ/$_D$, /ɛ/$_D$, /ɪ/$_D$, and /i/$_D$). Section 4 will fully simulate the L1 and L2 acquisition of the mapping from two auditory continua (F1 and F2) to the 12 Dutch vowels and, later, the 5 Spanish vowels of Figure 1.

## 3. One-dimensional vowel loss

We will first simulate the acquisition of a simplified vowel system, one in which a single auditory continuum, namely F1, is mapped to only four vowels. This initial simplification is necessary in order for us to be able to illustrate with explicit graphics how constraint rankings in the perception grammar can lead to an optimal perception in L1 and L2. The two-dimensional case of §4 will then be a straightforward extension.

### 3.1. The L1 language environment

The L1 at hand is a language with only four vowels, simplified Dutch. The vowels carry the familiar labels /ɑ/$_D$, /ɛ/$_D$, /ɪ/$_D$, and /i/$_D$, but they are distinguished only by their F1 values. We assume that the token distributions of the four intended vowels |ɑ|$_D$, |ɛ|$_D$, |ɪ|$_D$, and |i|$_D$ have Gaussian shapes around their mean values along a logarithmic F1 axis, as in Figure 2. The mean values (i.e. the locations of the peaks in Figure 2) are the same as the median F1 values of Figure 1, namely 926, 733, 438, and 305 Hz, and the standard deviation is 0.05 along a base-10 logarithmic scale (i.e. 0.166 octaves). This leads to the curves in Figure 2, where we assume for simplicity that all four vowels occur equally frequently, so that the four peaks are equally high.

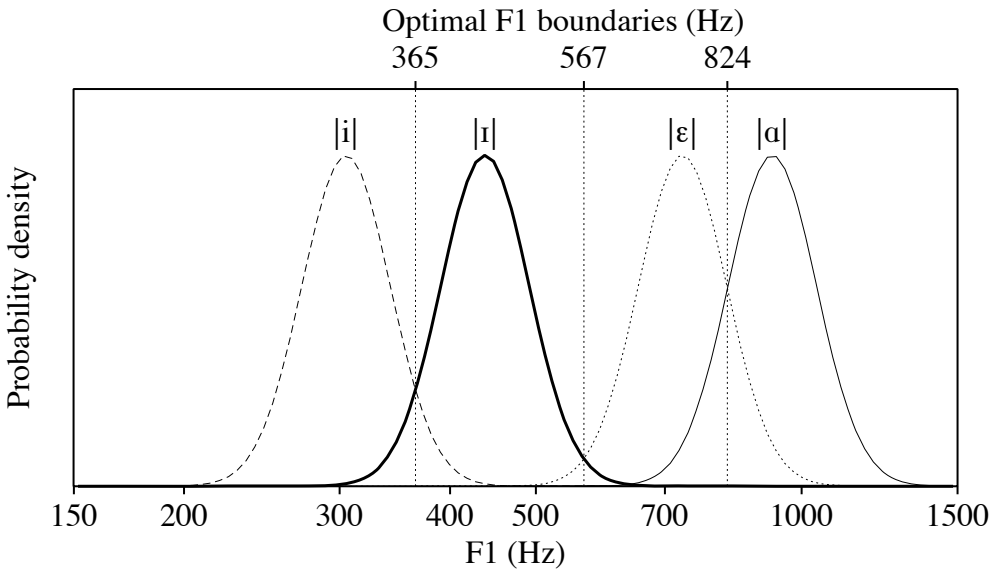


**Fig. 2.** Idealized token distributions for four short Dutch vowels.

### 3.2. Optimal L1 perception

Figure 2, then, describes the distributions of *speakers'* productions of the four intended vowels in a large corpus of one-dimensional Dutch. The task of the *listeners* is to map each incoming F1 value on one of the vowel categories /ɑ/$_D$, /ɛ/$_D$, /ɪ/$_D$, and /i/$_D$, in preparation for subsequent access of a word containing one of the underlying vowels |ɑ|$_D$, |ɛ|$_D$, |ɪ|$_D$, and |i|$_D$. The question now is: what would be an optimal strategy for a listener? We propose that the optimal strategy is to minimize the discrepancy between the perceived vowel and the recognized vowel, i.e. to minimize the number of cases where the listener perceives a certain vowel (e.g. /ɛ/$_D$) but subsequently finds a different vowel (e.g. |ɪ|$_D$) in her lexicon (we call such a situation a *perception error*).

A general strategy that achieves this minimization of the number of perception errors is the *maximum likelihood* strategy (Helmholtz 1910), where the listener perceives any given F1 value as the vowel that was most likely to have been intended by the speaker. In Figure 2 we see that if a listener hears an F1 value of 400 Hz, it is most likely that this was a token of an intended vowel |ɪ|$_D$. We know this because for an F1 of 400 Hz the distribution curve for |ɪ|$_D$ lies above the distribution curves for the other three vowels. In general, any F1 value should be perceived as the vowel whose curve is highest. Which curve is highest in Figure 2 is determined by the three main cutting points of the curves, which lie at 365, 567, and 824 Hz. Given the distributions in Figure 2, then, a maximum-likelihood strategy entails that the listener should perceive all incoming F1 values below 365 Hz as /i/$_D$, all F1 values between 365 and 567 Hz as /ɪ/$_D$, all F1 values between 567 and 824 Hz as /ɛ/$_D$, and all F1 values above 824 Hz as /ɑ/$_D$. If the listener indeed uses these three *optimal boundaries* as her criteria for perception, she will achieve a correctness percentage of 90.5, i.e., of all F1 values that will be drawn according to the distributions of Figure 2 (with equal probabilities for each of the four intended vowels) she will perceive 90.5 percent as the same vowel as she will subsequently find in her lexicon. The remaining 9.5 percent are cases of perception errors, caused by the overlap in the curves of Figure 2 (i.e. in 9.5 percent of the productions an F1 value crosses the boundary with a neighbouring vowel).

The reader will have noticed that our definition of optimal perception (minimizing the number of perception errors) is related to our operationalization of lexicon-driven learning (§2.2), which changes the perception grammar every time the listener makes a perception error. The simulation of the following section will show that lexicon-driven perceptual learning with the GLA indeed leads to optimal boundaries in the listener.

### 3.3. L1 acquisition of the perception of one-dimensional Dutch

In order to be able to do a computer simulation of the F1-only simplified Dutch vowel system, we divide up the F1 continuum between 150 and 1500 Hz in 100 values equally spaced along a logarithmic scale: 152, 155, 159, ..., 1416, 1449, and 1483 Hz. We will assume that only these 100 frequencies are possible incoming F1 values. According to §2.1, we therefore need 400 cue constraints (100 F1 values × 4 vowel categories) that can be formulated like "[F1 = 1416 Hz] is not /ɪ/$_{D}$".[ix]

We assume that in the initial state of our learner all lexical representations are already correct, so that lexicon-driven learning according to tableaus like (7) works flawlessly. We further assume that all 400 cue constraints are initially ranked at the same height, namely at 100.0, so that any F1 value has a probability of 25 percent of being perceived as any of the four vowels. This combination of assumptions is obviously a severe simplification, since a correct lexicalization must depend on a reasonably good perception system, i.e. one whose percentage correct is much higher than 25. Such a reasonably good perception system could be obtained by an Optimality-Theoretic distributional learning method for infants such as the one described by Boersma, Escudero & Hayes (2003), but we will not pursue this here since we are mainly interested in what happens later in life.

We feed our simulated learner with 10,000 F1 values per virtual year, drawn from the distributions in Figure 2 (i.e. more F1 values near the peaks than near the valleys), always telling the learner, as in (7), what would have been the correct perception. Every time there is a mismatch between the perceived vowel and the correct vowel (i.e. the vowel intended by the speaker, as recognized by the listener's lexicon), some rankings change by a small amount, which Stochastic Optimality Theory refers to as the *plasticity* (or *learning step*). The plasticity is 1.0 during the first year, then decreases by a factor of 0.7 every year, ending up as a plasticity of 0.0023 during the 18th virtual year. With a constant evaluation noise of 2.0, this plasticity scheme causes learning to be initially fast but imprecise, and later on slow but accurate.

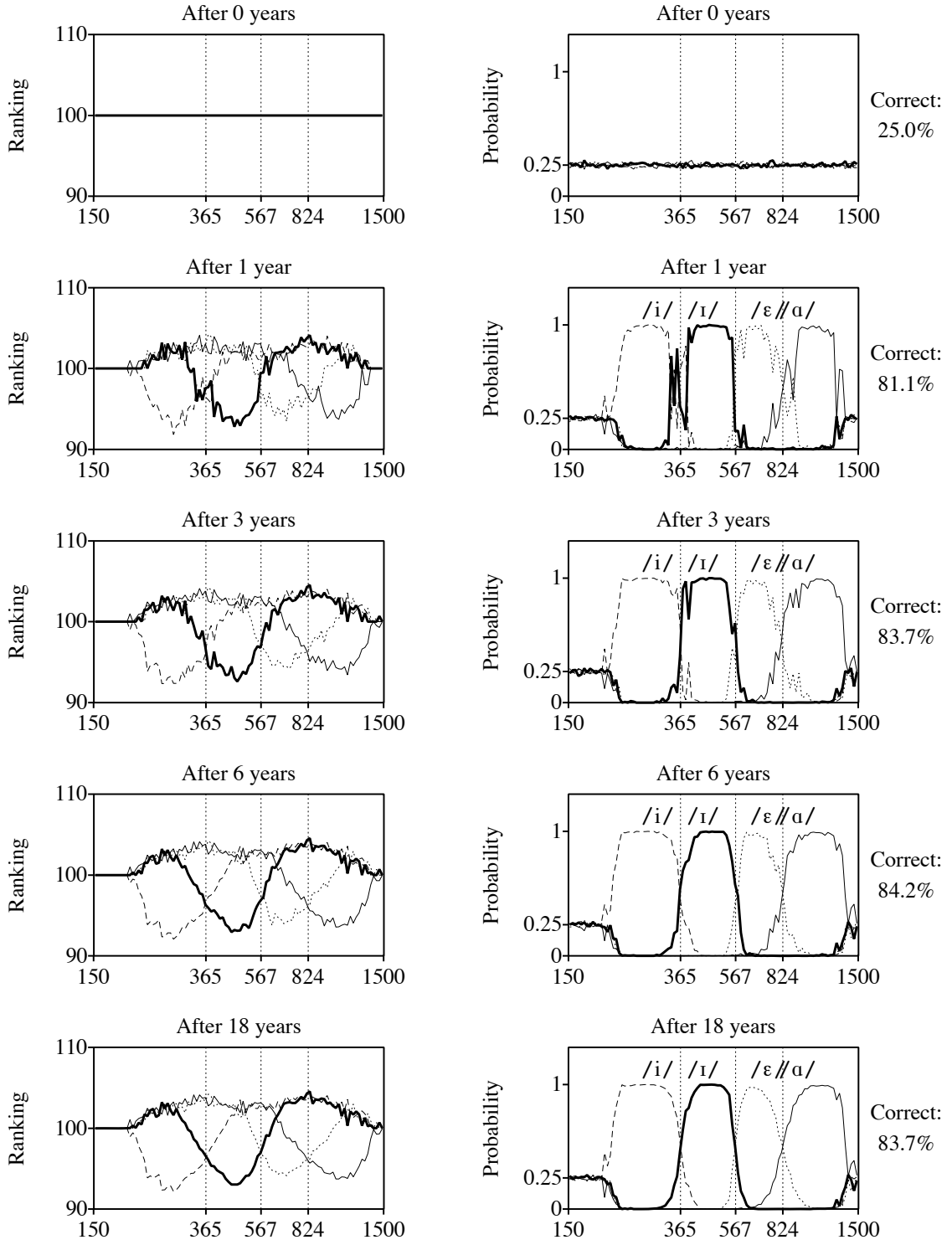


**Fig.3.** Simulated L1 acquisition of Dutch.
Left: the rankings of the four constraint families "[F1=*x*] is not /*vowel*/$_D$".
Right: the identification curves.
Dashed: /i/$_D$; plain thick: /ɪ/$_D$; dotted: /ɛ/$_D$; plain thin: /ɑ/$_D$.

The left side of Figure 3 shows the development of the grammars and is to be interpreted as follows. For every F1 value it is the lowest-ranked constraint that determines into which vowel category the F1 value will most often be classified. For instance, for an F1 of 400 Hz the lowest ranked constraint (the thick curve) is "[F1 =

400 Hz] is not /ɪ/$_D$". Tableau (8) shows that the low ranking of this constraint determines the winning candidate, irrespective of the relative ranking of the other three relevant constraints.

(8) *Perception determined by the lowest curve*

| [F1=400] | [F1=400] is not /ɑ/$_D$ | [F1=400] is not /ɛ/$_D$ | [F1=400] is not /i/$_D$ | [F1=400] is not /ɪ/$_D$ |
|---|---|---|---|---|
| /ɑ/$_D$ | *! | | | |
| /ɛ/$_D$ | | *! | | |
| ☞ /ɪ/$_D$ | | | | * |
| /i/$_D$ | | | *! | |

Every grammar leads to its own perception pattern. In the course of the 18 virtual years we see that the crossing points of the constraint curves come to lie close to the optimal boundaries of 365, 567, and 824 Hz. If a listener with the 18th-year grammar in Figure 3 were to have an evaluation noise to zero, her percentage correct would be about 90.5, just as for the maximum-likelihood listener in §3.2 (the percentage correct can be estimated by running 100,000 F1 values, distributed as in Figure 2, through the grammar and counting the number of correct output vowels). If we assume, however, that the listener has an evaluation noise of 2.0, just as during learning, the percentage correct is a bit lower. It can be shown (Boersma 1997) that in the one-dimensional case the resulting perception grammar is *probability matching*, i.e. the probability of perceiving a certain F1 value as a certain vowel comes to approximate the probability that this F1 value had been intended as that vowel. For instance, we can read off Figure 2 that an F1 value of 400 Hz has 90 percent chance of having been intended as |ɪ|$_D$ and 10 percent chance of having been intended as |i|$_D$. When confronted with an auditory input of 400 Hz, a probability-matching listener will perceive it 90 percent of the time as /ɪ/$_D$ and 10 percent of the time as /i/$_D$. Exactly this is what our learner comes to do, improving her perception of the whole distribution from 25 percent correct to 83.7 percent correct, which is the same value that can be computed from Figure 2.[x] In the rest of this paper we will call probability-matching behaviour "optimal", and forget about maximum-likelihood behaviour, which never occurs in practice anyway.

The right side of Figure 3 shows our virtual listener's *identification curves* (as known from many perception experiments with real listeners), i.e. for each of the four vowels a curve that shows for every F1 value how often that F1 value is perceived as that vowel. These curves are computed by running each of the 100 F1 values through the grammar 1,000 times and counting how often each of the four possible vowels is the winner. The virtual learner grows increasingly confident of her category boundaries, which become optimal for her language environment.

### 3.4. L2 acquisition of the perception of one-dimensional Spanish

After having learned Dutch for 18 years, our virtual learner starts learning Spanish. Our one-dimensional Spanish has the three vowels |a|$_S$, |e|$_S$, and |i|$_S$, whose F1

distributions are centred around the median F1 of Figure 1, again with a logarithmic standard deviation of 0.05. The learner equates the three Spanish vowels with her Dutch categories $|\alpha|_D$, $|\varepsilon|_D$, and $|i|_D$, respectively, as do the real learners of §1.4. Her L2 language environment can thus be described by the curves in Figure 4.

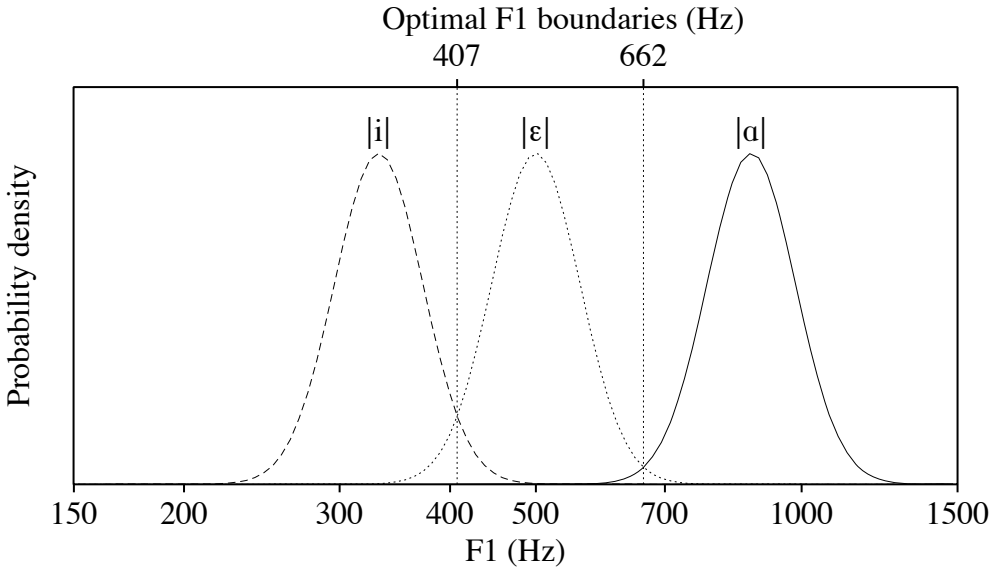


**Fig. 4.** The Spanish vowel environment, with Dutch labels.

The learner's initial interlanguage grammar has to be a copy of her current grammar of Dutch (§1.2), so the picture in the upper left of Figure 5 is identical to the picture in the lower left of Figure 3. Such a grammar handles Spanish better than an infant-like grammar where all constraints are ranked at the same height. Whereas an infant-like grammar (with the four Dutch categories) would score 25 percent correct, the copied Dutch grammar already scores 53.1 percent correct. Nevertheless, this score is far from nativelike, since an adult probability-matching listener of Spanish will achieve 95.5 percent correct (as computed from Figure 4). If she is to gain more accuracy in her L2 environment, our virtual listener will have to learn.

We immerse our virtual learner in a rich Spanish environment where she hears 10,000 vowel tokens a year, as many as during her L1 acquisition. Acknowledging her high motivation, we endow her with a plasticity of 0.01, which is over four times as high as her final L1 plasticity of 0.0023 but of course still only a tiny fraction of her initial L1 plasticity of 1. The development of the virtual L2 learner is shown in Figure 5.

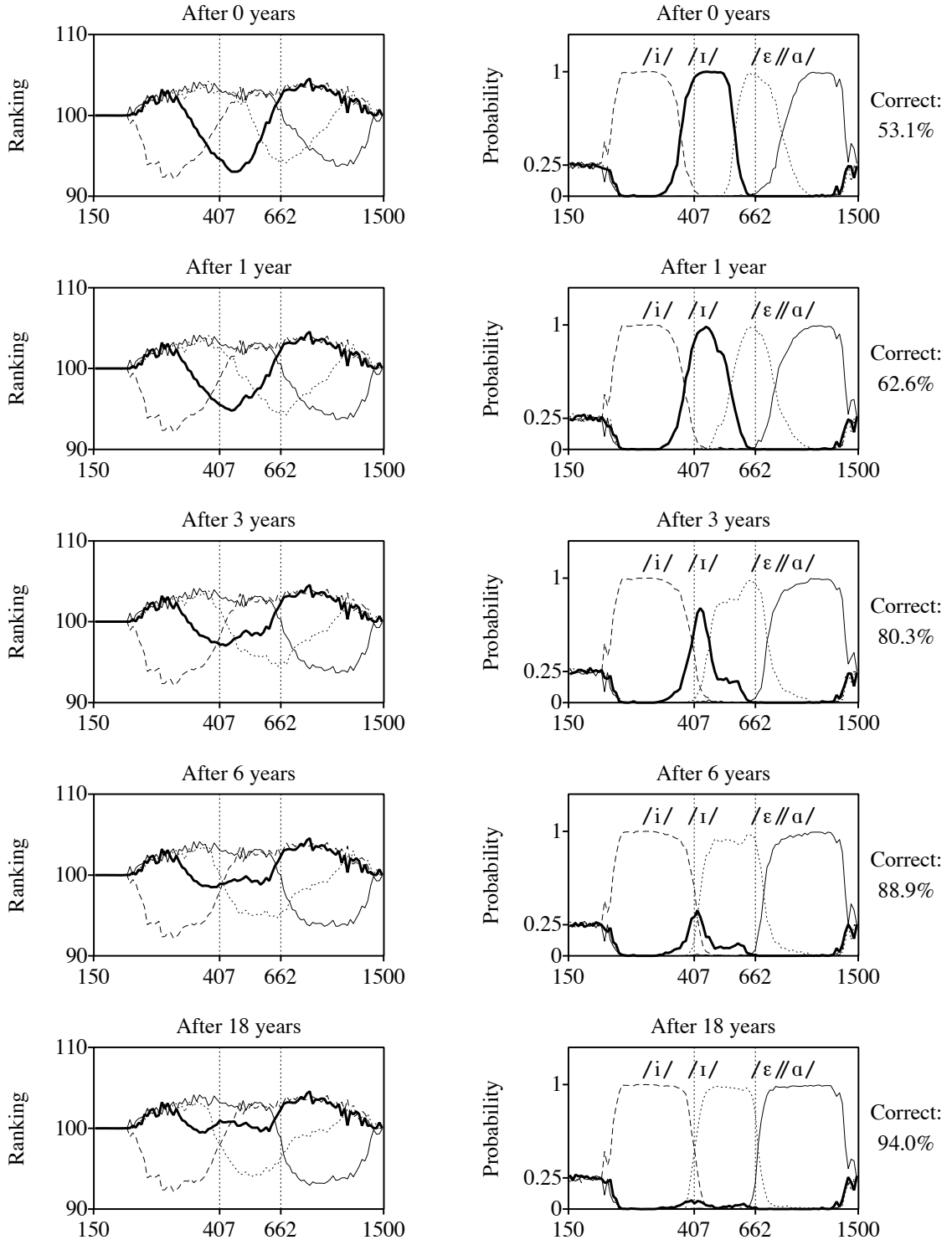


**Fig. 5.** Simulated L2 acquisition of Spanish.
Left: the rankings of the four constraint families "[F1=*x*] is not /*vowel*/$_D$".
Right: the identification curves.
Dashed: /i/$_D$; plain thick: /ɪ/$_D$; dotted: /ɛ/$_D$; plain thin: /ɑ/$_D$.

The main feature of the development is the fall of the /ɪ/$_D$ category. Whenever the learner perceives an incoming F1 value as /ɪ/$_D$, the interlanguage lexicon, which does not contain any instances of |ɪ|$_D$, will tell her that she should have perceived a different vowel, most often /i/$_D$ or /e/$_D$. In all these cases, one of the constraints "[F1=*x*] is not

/ɪ/$_D$" will rise along the ranking scale, thus making it less likely that the next occcurrence of the same F1 value will again be perceived as /ɪ/$_D$.

The learner's proficiency clearly improves, although despite her complete immersion in her L2 environment, despite her raised motivation, and despite her full access to an L1-like learning mechanism (the GLA), she has trouble achieving complete nativelike competence (i.e. 95.5%), even in 18 years. This small failure is mainly due to the plasticity of 0.01, which stresses adultlike precision rather than infantlike learning speed.

## 4. Two-dimensional vowel loss and shift of |ɑ|$_D$

After the oversimplification of §3, our second simulation reflects a more realistic situation, in which two auditory cues, namely both F1 ('height') and F2 ('place'), contribute to the perception of the whole Dutch system of short vowels. We divide both continua in 21 values, as shown in Figure 6. Some height-place combinations cannot occur articulatorily (frog-like sounds in the bottom left) or by definition (the bottom right, where F1 is greater than F2); these are left blank in the figure.

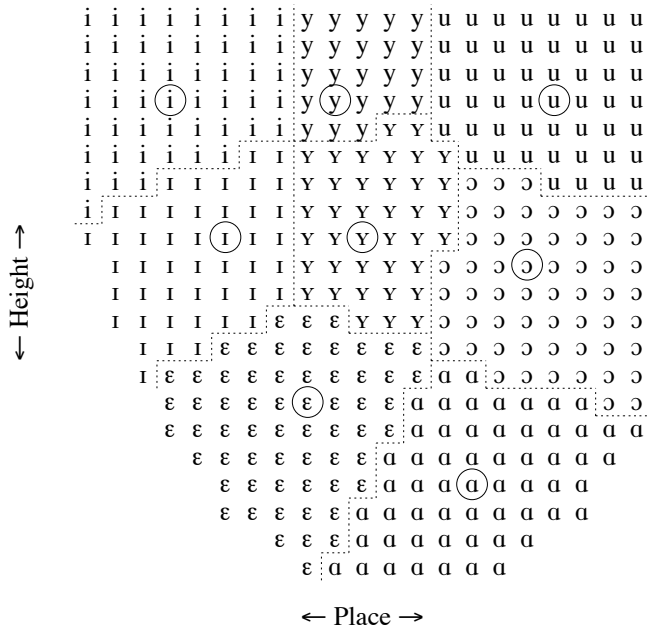


**Fig. 6.** Circles: the centres of the token distributions of the eight short Dutch vowels. Phonetic symbols: the most likely intended vowel for every place-height combination.

### 4.1. The 2-dimensional L1 language environment

Figure 6 summarizes the height and place distributions for native speakers of Dutch. The circles represent the centres of the token distributions of the eight vowels. Their locations are similar to those in Figure 1, but for the purposes of the present section we have made each of them coincide exactly with one of the 21×21 possible height-place values. We assume that the standard deviation of the Gaussian place distribution is 2.0 columns along the horizontal axis, and that the standard deviation of the Gaussian height distribution is 2.0 rows along the vertical axis. We also simplifyingly assume that all short vowels are equally common, except |y|$_D$, which we take to be five times less

common in this simplified Dutch inventory than every other short vowel. Figure 6 then shows for each F1-F2 combination what the most likely intended vowel is. The regions thus attributed to each vowel are delimited by dotted lines in the figure. These "production boundaries" turn out to run at equal distances to the nearest vowels, except for the boundaries around the $|y|_D$ area, which reflect the low token frequency of this vowel.

### 4.2. Optimal 2-dimensional perception

Since Figure 6 shows the most likely intended productions, the production boundaries in this figure must indicate the optimal boundaries for perception as well. We can compute that a probability-matching listener would score 78.2% correct. The following section shows that GLA learners can achieve this optimal perception.

### 4.3. L1 acquisition of the perception of 2-dimensional Dutch

Analogously to §3.3, we feed a virtual Dutch listener 10,000 F1-F2 tokens a year, drawn randomly from the distribution in Figure 6 (i.e. fewer tokens far away from the vowel centres than close to them, and fewer tokens of $|y|_D$ than of every other vowel).

The virtual learner's grammar contains 336 cue constraints (= (21 height values + 21 place values) × 8 vowels), which start out being ranked at the same height. Subsequent learning is performed, as before, via 180,000 tableaus, which in case of a misperception cause a learning step analogous to that in tableau (7). The evaluation noise and plasticity regime are as in §3.3. There is no simple way to show the grammars or identification curves, as there was in the 1-dimensional case of §3.3, but we can compute for every F1-F2 combination what the most likely perceived vowel is, by running each F1-F2 combination through the grammar 1000 times. The results are in Figure 7, which shows the development of the learner's performance. While after one year the "perception boundaries" (the dotted lines that delimit the most-likely-vowel areas) are still rather ragged, after 18 years they are smooth and very close to the production boundaries of Figure 6, leading to fractions correct that compare very well with the optimum reported in §4.2. It turns out that the GLA is indeed capable of creating a stochastic Optimality-Theoretic grammar that exhibits optimal perceptual behaviour.

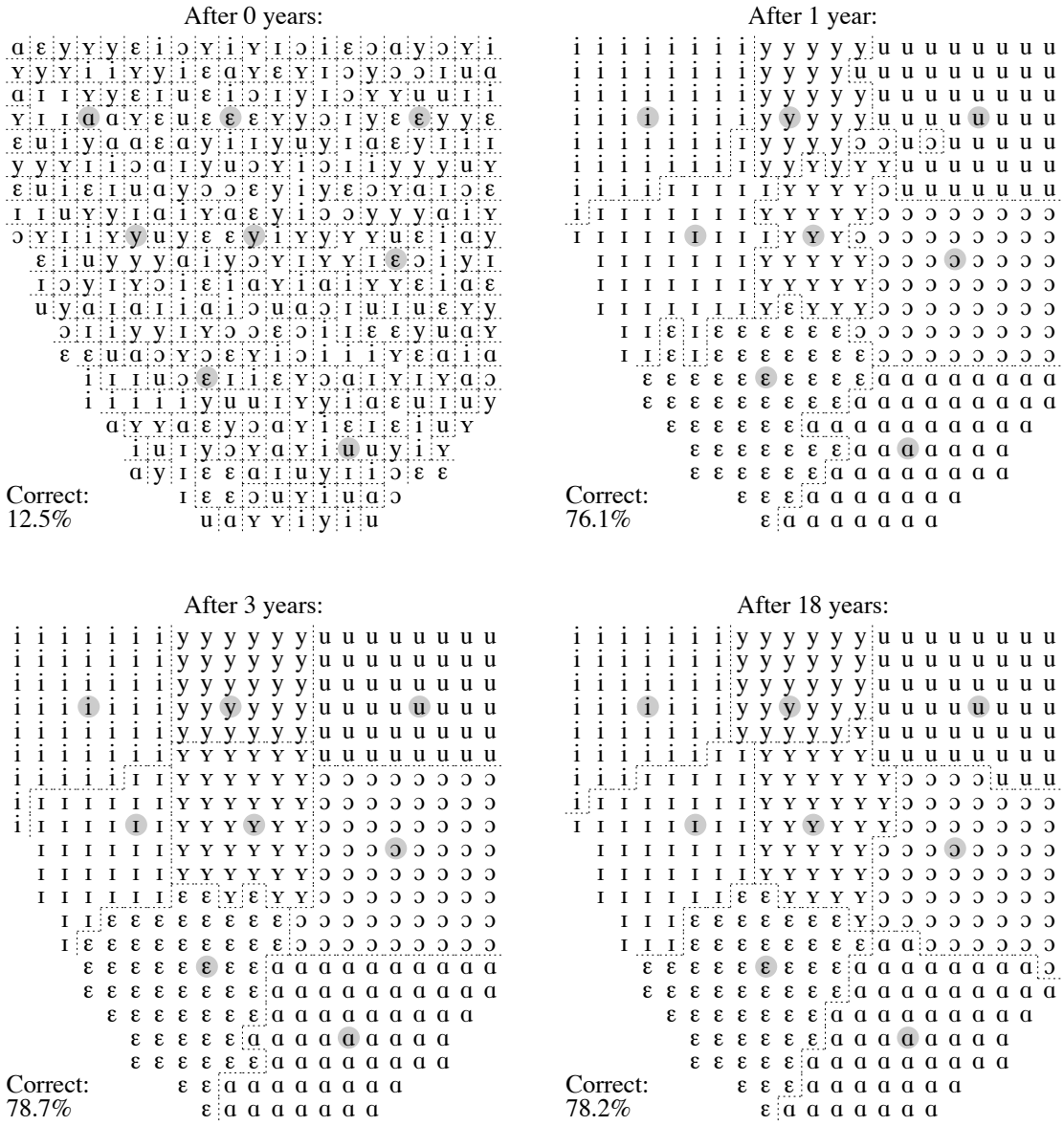


**Fig. 7.** Simulated L1 Dutch vowel classification after 0, 1, 3, and 18 years. Grey disks: the eight Dutch short vowel centres in production.

### 4.4. L2 acquisition of the perception of 2-dimensional Spanish

When the learner moves to Spain, her language environment becomes that of Figure 8, which shows the most likely intended Spanish vowels, under the assumption that the five vowels have equal token frequencies. When the learner copies her Dutch constraint ranking (i.e. the grammar in Figure 7, bottom right) to her Spanish interlanguage grammar, her fraction correct, given the distributions in Figure 8, is 47.6% (cf. 56.6% for the 1-dimensional case of §3.4).

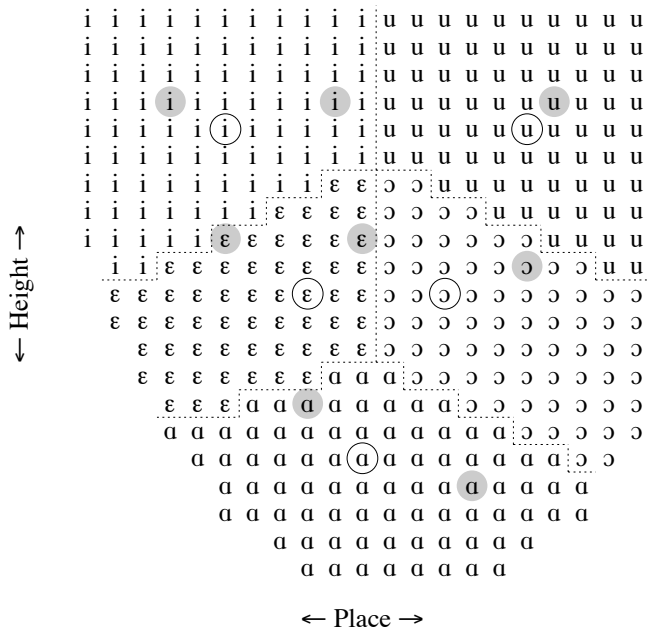


**Fig. 8.** The Spanish vowel environment, with Dutch labels.
Circles: the Spanish vowel centers. Grey disks: Dutch short vowel centres.

As with the 1-dimensional case of §3.4, we immerse the learner in Spanish (10,000 tokens a year, drawn from the distributions in Figure 8, with lexicon-guided correction) with a plasticity of 0.01. The development of classification behaviour is shown in Figure 9. We see that the learner gradually loses her /ɪ/$_{D}$, /ʏ/$_{D}$, and /y/$_{D}$ categories and shifts her /ɑ/$_{D}$ category towards the front, just as the real human subjects did in our listening experiment (/ɪ/$_{D}$ and /y/$_{D}$ never fade entirely, continuing to occupy regions where the Spanish learning environment has offered very few tokens). Nativelike behaviour, which should follow the optimal boundaries in Figure 8 (and reach a fraction correct of 83.7%), is closely approached but never completely attained, mainly as a result of the low plasticity relative to that of infants.

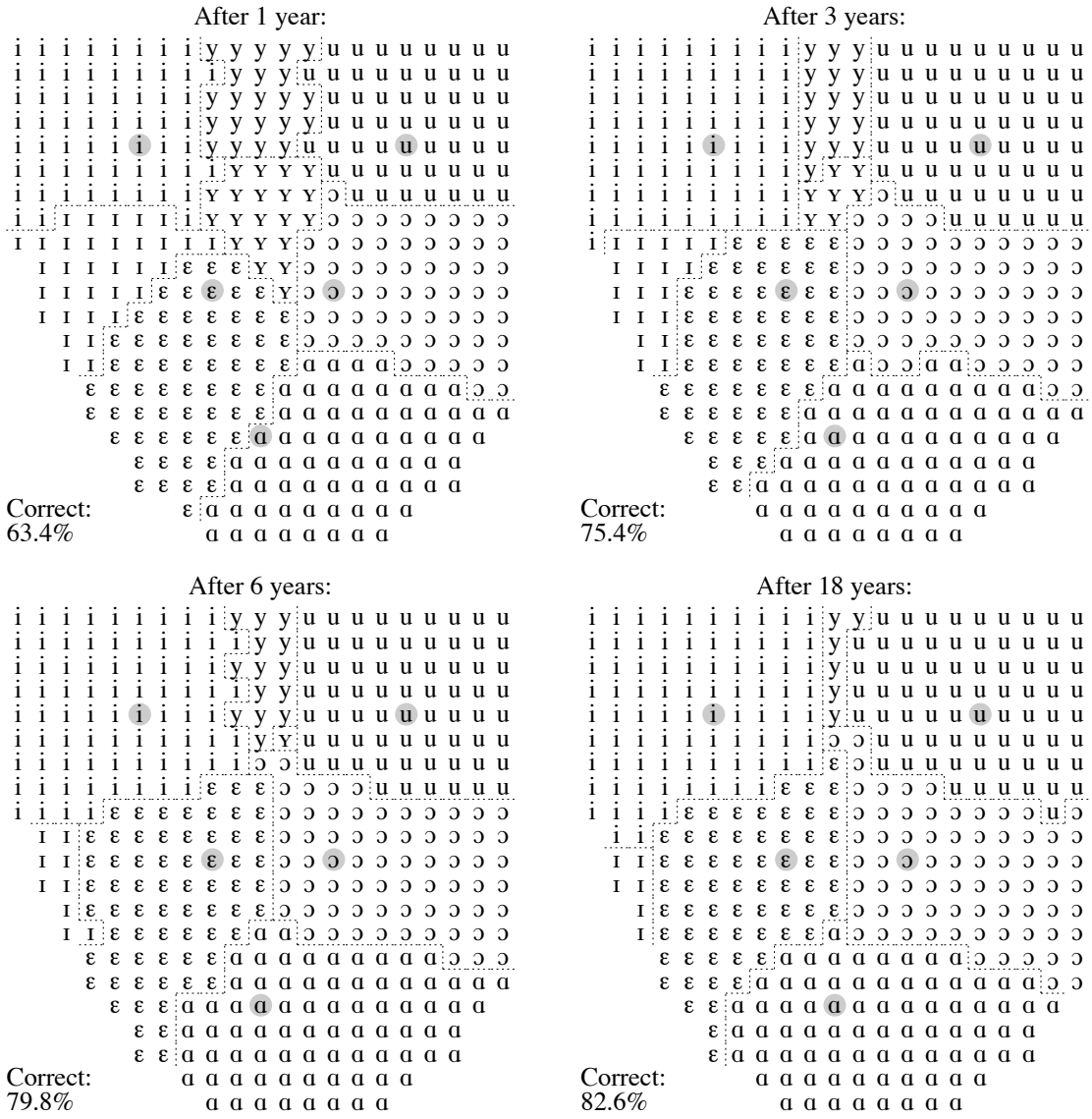


**Fig. 9.** The perception of Spanish by a Dutch learner after 1, 3, 6, and 18 years. Grey disks: the Spanish vowel centres.

### 4.5. The need for negatively formulated cue constraints

In the present paper we have been using cue constraints with negative formulations, such as "an F1 of 400 Hz is not /ɑ/$_D$". Couldn't we just have used positively formulated cue constraints instead, like "an F1 of 400 Hz is /ɑ/$_D$"? There are two cases in which this makes no difference. The first case is that of a single auditory continuum, as in §3: in tableau (8), in which every candidate violates a single constraint, we can simply rank positively formulated constraints in the reverse order of their negatively formulated counterparts, and the outcome will be the same. The second case is that of multiple auditory continua but only two different vowel categories (Escudero & Boersma 2003, 2004): if we have only two categories /A/ and /B/, the constraint "an F1 of 400 Hz is not /A/" is simply equivalent to the constraint "an F1 of 400 Hz is /B/".

But the equivalence does not generalize to cases with two (or more) auditory continua and more than two categories. For instance, an 18-year simulation of the acquisition of L1 Dutch with positively formulated cue constraints leads to a grammar

that exhibits the behaviour in Figure 10, with a fraction correct of 44.9% for the perception of Dutch, an achievement dramatically worse than that of the negatively formulated constraints of Figure 7, which scored 78.2%. In Figure 10, the highest-ranked positively formulated constraint is "[height=6] is $/\varepsilon/_D$"; an entire row of epsilons (the sixth row from below) shows that this constraint has a non-local influence throughout the place continuum. The second-highest constraint is "[place=3] is $/i/_D$"; a complete column of i's (the third column from the left) shows that it has a non-local influence throughout the height continuum.[xi] It can easily be seen that there exists no ranking of these positively formulated constraints that yields a separation into locally confined areas like those that appear in Figure 7: the top-ranked ones always determine the perception of entire rows or columns in the vowel grid.[xii]

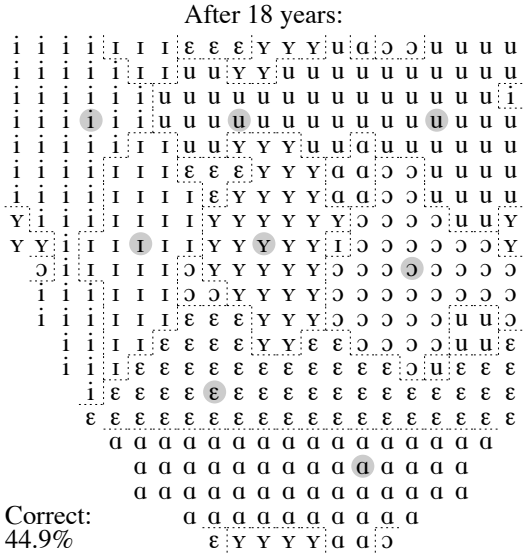


**Fig. 10.** The failure of learning L1 Dutch with positively formulated cue constraints.

## 5. Discussion

Negatively formulated Optimality-Theoretic constraints can handle the categorization of both 1-dimensional and 2-dimensional auditory continua as attested in listening experiments, at least if every category spans a compact local region in the auditory space. Our Optimality-Theoretic perception model shares this property with several connectionist models, starting with the perceptron (Rosenblatt 1962), and with Massaro's (1987) fuzzy logical model of perception. But unlike these other models of perception, it makes a connection with phenomena that phonologists have traditionally been interested in, as witnessed by the perceptual processes that have been modelled in Optimality Theory: the interpretation of metrical feet, which requires structural constraints like IAMBIC and WEIGHT-TO-STRESS (Tesar 1997, 1998; Tesar and Smolensky 2000; Apoussidou and Boersma 2003, 2004); sequential abstraction, which can be handled by the interaction of structural constraints and cue constraints like the Obligatory Contour Principle and the Line Crossing Constraint (Boersma 1998, 2000); the interaction of structural constraints and auditory faithfulness in the categorization of vowel height (Boersma 1998) or consonant length (Hayes 2001); truncation by infants, which requires structural constraints like WORDSIZE (Pater 2004); and ghost segments,

which can be handled by the interaction of structural and cue constraints (Boersma 2005).

The general usefulness of modelling perception in Optimality Theory extends to the specific kinds of cue constraints described here, which are not specific to the task of learning a smaller L2 vowel system. The same kind of constraints have been applied to learning to perceive a *larger* L2 vowel system, i.e. an inventory with new sounds (from Spanish to English: Escudero & Boersma 2004), and to learning an equally large L2 vowel system, i.e. an inventory with similar but non-identical sounds (from Canadian English to Canadian French: Escudero 2005), and they have been combined with auditory-to-auditory constraints in the modelling of L1 category formation (Boersma, Escudero & Hayes 2003).

Optimality-Theoretic accounts of perception and its acquisition thus bridge the gap between phonological theory and the computational modelling of human speech processing.

## References


Apoussidou, Diana and Paul Boersma 2003 The learnability of Latin stress. *Proceedings of the Institute of Phonetic Sciences Amsterdam* 25: 101–148.

Apoussidou, Diana and Paul Boersma 2004 Comparing two Optimality-Theoretic learning algorithms for Latin stress. In Vineeta Chand, Ann Kelleher, Angelo J. Rodríguez and Benjamin Schmeiser (eds.), *Proceedings of the 23rd West Coast Conference of Formal Linguistics*, 29–42. Somerville, MA: Cascadilla.

Boersma, Paul 1997 How we learn variation, optionality, and probability. *Proceedings of the Institute of Phonetic Sciences Amsterdam* 21: 43–58.

Boersma, Paul 1998 *Functional Phonology*. PhD dissertation, University of Amsterdam. The Hague: Holland Academic Graphics.

Boersma, Paul 1999 On the need for a separate perception grammar. Manuscript, University of Amsterdam. [*Rutgers Optimality Archive* 358]

Boersma, Paul 2000 The OCP in the perception grammar. Manuscript, University of Amsterdam. [*Rutgers Optimality Archive* 435]

Boersma, Paul 2001 Phonology-semantics interaction in Optimality Theory, and its acquisition. In Robert Kirchner, Wolf Wikeley, & Joe Pater (eds.), *Papers in Experimental and Theoretical Linguistics*, Volume 6, 24–35. Edmonton: University of Alberta.

Boersma, Paul 2005 Some listener-oriented accounts of hache aspiré in French. *Rutgers Optimality Archive* 730. Revised version to appear in *Lingua*.

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

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

Bradlow, Ann 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, Ellen 2003 Language contact phonology: richness of the stimulus, poverty of the base. In Keir Moulton and Matthew Wolf (eds.), *NELS 34: Proceedings of the 34th Annual Meeting of the North-Eastern Linguistic Society*. Amherst: Graduate Linguistic Student Association of the University of Massachusetts.

Curtin, Suzanne, Toben H. Mintz and M.H. Christiansen 2005 Stress changes the representational landscape: Evidence from word segmentation. *Cognition* 96: 233–262.

Escudero, Paola 2005 *The Attainment of Optimal Perception in Second-Language Acquisition*. Ph.D. dissertation, University of Utrecht. Utrecht: Landelijke Onderzoeksschool Taalwetenschap.

Escudero, Paola and Paul Boersma 2002 The subset problem in L2 perceptual development: Multiple-category assimilation by Dutch learners of Spanish. In Barbora Skarabela, Sarah Fish and Anna H.-J. Do (eds.), *Proceedings of the 26th annual Boston University Conference on Language Development*, 208–219. Somerville, MA: Cascadilla.

Escudero, Paola and Paul Boersma 2003 Modelling the perceptual development of phonological contrasts with Optimality Theory and the Gradual Learning Algorithm. In Sudha Arunachalam, Elsi Kaiser and Alexander Williams (eds.), *Proceedings of the 25th Annual Penn Linguistics Colloquium. Penn Working Papers in Linguistics* 8.1, 71–85.

Escudero, Paola and Paul Boersma 2004 Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition* 26: 551–585.

Escudero, Paola and Paul Boersma to appear *Language modes and perceptual development in Dutch learners of Spanish.*

Gerrits, Ellen 2001 The categorisation of speech sounds by adults and children. PhD dissertation, University of Utrecht.

Hayes, Rachel 2001 An Optimality-Theoretic account of novel phonetic category formation in second language learning. Manuscript, University of Arizona.

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

Jusczyk, Peter W., Anne Cutler and N.J. Redanz 1993 Infants' preference for the predominant stress patterns of English words. *Child Development* 64: 675–687.

Jusczyk, Peter W., Derek M. Houston and M. Newsome 1999 The beginnings of word segmentation in English-learning infants. *Cognitive Psychology* 39: 159-207.

Kenstowicz, Michael 2001 The role of perception in loanword phonology. *Linguistique africaine* 20.

Koopmans-van Beinum, Florien J. 1980 Vowel contrast reduction. An acoustic and perceptual study of Dutch vowels in various speech conditions. PhD dissertation, University of Amsterdam.

Legendre, Géraldine, Yoshiro Miyata, and Paul Smolensky 1990 Harmonic Grammar – a formal multi-level connectionist theory of linguistic well-formedness: theoretical foundations. *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society*, 884–891. Cambridge, MA: Erlbaum.

Liljencrants, Johan and Björn Lindblom 1972 Numerical simulation of vowel quality systems: the role of perceptual contrast. *Language* 48: 839–862.

Lindblom, Björn 1986 Phonetic universals in vowel systems. In John J. Ohala and Jeri J. Jaeger (eds.), *Experimental Phonology*, 13–44. Orlando: Academic Press.

McQueen, James M. and Anne Cutler 1997 Cognitive processes in speech perception. In William J. Hardcastle and John Laver (eds.), *The Handbook of Phonetic Sciences*, 566–585. Oxford: Blackwell.

McQueen, James M. 2005 Speech perception. In K. Lamberts and R. Goldstone (eds.), *The Handbook of Cognition*, 255–275. London: Sage Publications.

Massaro, Dominic William 1987 *Speech Perception by Ear and Eye: A Paradigm for Psychological Inquiry*. Hillsdale: Lawrence Erlbaum.

Pater, Joe 2004 Bridging the gap between perception and production with minimally violable constraints. In René Kager, Joe Pater and Wim Zonneveld (eds.), *Constraints in Phonological Acquisition*, 219–244. Cambridge: Cambridge University Press.

Polka, Linda and Janet F. Werker 1994 Developmental changes in perception of non-native vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance* 20: 421–435.

Polka, Linda, Megha Sundara and Stephanie Blue 2002 The role of language experience in word segmentation: A comparison of English, French, and bilingual infants. Paper presented at the 143rd Meeting of the Acoustical Society of America: Special Session in Memory of Peter Jusczyk, Pittsburgh, Pennsylvania.

Pols, Louis C.W., H.R.C. Tromp and Reinier Plomp 1973 Frequency analysis of Dutch vowels from 50 male speakers. *Journal of the Acoustical Society of America* 53: 1093–1101.

Rosenblatt, Frank 1962 *Principles of Neurodynamics; Perceptrons and the Theory of Brain Mechanisms*. Washington: Spartan Books.

Smolensky, Paul 1996 On the comprehension/production dilemma in child language. *Linguistic Inquiry* 27: 720–731.

Tesar, Bruce 1997 An iterative strategy for learning metrical stress in Optimality Theory. In Elizabeth Hughes, Mary Hughes and Annabel Greenhill (eds.), *Proceedings of the 21st Annual Boston University Conference on Language Development*, 615–626. Somerville, MA: Cascadilla.

Tesar, Bruce 1998 An iterative strategy for language learning. *Lingua* 104: 131–145.

Tesar, Bruce and Paul Smolensky 2000 *Learnability in Optimality Theory*. Cambridge, MA: MIT Press.

Werker, Janet F. and R.C. Tees 1984 Cross-language speech perception: evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development* 7: 49–63.
Yip, Moira 2006 The symbiosis between perception and grammar in loanword phonology. *Lingua* 116: 950–975.

---

[i] We use pipes in order to distinguish underlying forms from phonological surface structures, which are given between /slashes/, and auditory phonetic forms, which are given in an approximate IPA transcription between [square brackets].

[ii] Escudero (2005: 214–236) investigates and models (in Optimality Theory) the possibility that category reuse is not an "easy" instantaneous act that occurs magically at the start of L2 acquisition after all. Escudero proposes instead that category reuse gradually emerges as an automatic result of an initial creation of lexical items with multiple underlying phonological representations and a subsequent reduction of this lexical variability by the process of message-driven learning of recognition.

[iii] Nor *less* similar. A theory of phonology that regards all vowels as a combination of innate (hence crosslinguistically identical) phonological feature values may even consider every vowel at the left in (1) as featurally identical to its counterpart at the right.

[iv] We use slashes ("/") for perceived phonological surface representations. We assume that these representations consist of the same kinds of discrete arbitrary symbols as lexical representations, because the task of the perception process is to turn raw auditory data into discrete representations that are maximally suited for lexical access. See (2) for an explicit model.

[v] Deeper mechanisms than perceived similarity may play a role as well, such as choosing categories that are peripheral in the L1, in order to improve *production* in such a way that other listeners' comprehension improves. This may contribute to linking $|a|_S$ to $|ɑ|_D$ rather than to $|ɛ|_D$. Such a bias towards peripherality also follows automatically (i.e. without goal orientation) from Escudero's (2005: 214–236) model of selecting underlying representations (cf. fn. 2).

[vi] In Optimality-Theoretic terms, having to map a perceived /kɛsɔ/ to an underlying |kɑsɔ| can be said to involve a faithfulness violation in the recognition grammar (Boersma 2001).

[vii] Not included in this list are those who model comprehension as a single mapping in Optimality Theory, namely Smolensky (1996), Kenstowicz (2001), Broselow (2003), and Yip (2006), nor developments more recent than the present paper, such as Boersma (2005) and Escudero (2005).

[viii] We have to point out that Smolensky (p.c.) does not consider perception and robust interpretive parsing to be the same, because our auditory form is more peripheral and continuous than Tesar & Smolensky's overt form, which has already been analysed into discrete syllables. However, we see no reason why the language-specific construction of feet should not be handled in parallel with more peripheral-looking processes like the language-specific mapping from vowel duration to e.g. stress in Italian or to vowel length in Czech. Until there is evidence for prelexical sequential modularity, we will subsume all these processes under the single umbrella of "perception". The literature on the perception of foot structure by infants (e.g. Jusczyk, Houston, and Newsome 1999; Polka, Sundara, and Blue 2002; Curtin, Mintz, and Christiansen 2005) usually talks about "word segmentation", but uses perceptual terminology like "cue weighting".

[ix] A more sophisticated discretization of the F1 continuum, as used by Boersma (1997), would involve taking many more F1 values and allowing the learning algorithm to change the ranking of some neighbouring constraints by a value that decreases with the distance to the incoming F1. This would lead to results similar to those obtained by the simplified discretization of the present paper.

[x] Given a distribution where $p(f, v)$ denotes the probability that a token drawn randomly from the language environment has an F1 of $f$ Hz and was intended as the vowel $v$ (i.e. $\Sigma_{f,v}\, p(f, v) = 1$), the fraction correct for a maximum-likelihood listener can be computed as $\Sigma_f \max_v p(f, v)$, and the fraction correct for a probability-matching listener can be computed as $\Sigma_f (\Sigma_v p(f, v)^2 / \Sigma_v p(f, v))$.

[xi] Computationally inclined readers may wonder why one cannot successively erase lines and columns with identical symbols from Figure 10 until the figure is empty. This is because Figure 10 is based on repeated stochastic evaluations (§4.3), not on a fixed ranking.

[xii] We repeated the same simulations with OT's predecessor Harmonic Grammar (HG; Legendre, Miyata, and Smolensky 1990), where the ranking values are additive weights. With the same type of evaluation noise that turns OT into Stochastic OT, our "Stochastic HG" learners end up with a good separation of the categories, scoring about 78% correct, both for negatively and positively formulated cue constraints. Whether real humans use OT with negative constraints or HG with negative or positive constraints cannot be assessed on the basis of our data or simulations. Biological reality may well be more complex than both OT and HG.