INVENTORIES IN FUNCTIONAL PHONOLOGY*

Paul Boersma

Abstract

Functional Phonology, which makes a principled distinction between articulatory and perceptual representations, features, and constraints, can describe as well as explain the symmetries as well as the gaps in inventories of vowels and consonants. Symmetries are the language-specific results of general human limitations on the acquisition of perceptual categorization and motor skills. Gaps are the results of local hierarchies of articulatory effort, perceptual contrast, and perceptual confusion. There is no need to posit a dedicated inventory grammar: inventories are the automatic result of the constraints and their rankings in the production grammar.

Consider the short-vowel system of Frisian:

$$
\begin{array}{cccc}
i & & y & u \\
& e & \o & o \\
& \varepsilon & & \mathrm{\textturnc} \\
& & a &
\end{array}
\tag{1}
$$

Inventory (1) shows two common properties of inventories:

(a) *Symmetry*. The eight non-low vowels occur in only three heights and three "places"; they are not scattered randomly throughout the space of possible vowels. In §1.1, I will show that this symmetry is real. 'Phonetic' approaches to sound systems like (1) have not taken into account the symmetrizing principles of perceptual categorization and motor learning, though these are general phenomena of human behaviour, not specific to phonology.

(b) *Gaps*. The lower-mid-short-vowel system has a *gap* at /œ/. This must have something to do with the fact, stressed by the trapezoidal shape of (1), that the perceptual front-back contrast is smaller for lower vowels than for higher vowels. 'Phonological' approaches have ignored the explanatory power of the communicatively functional principles that segments will tend to be well contrasting and easy to articulate, although these principles can now easily be expressed as near-universal rankings in a constraint grammar

In Functional Phonology (Boersma 1997a), the effects of perceptual categorization, motor learning, perceptual contrast, and articulatory effort are expressed directly in the grammar. In the current paper, I will show that this approach is capable of representing the symmetries as well as the gaps of segment inventories.

---

1. Phonological and phonetic approaches to inventories

Since this paper is intended for a mixed audience of phonologists and phoneticians, it seems worthwhile to discuss the approaches to the modelling of inventories that have been proposed from each side, and to point out the strengths and weaknesses of both with respect to explanatory power and empirical adequacy. In later sections, I will show that the functional phonological account of inventories combines the strengths of the two sides, without, I hope, copying any of the weaknesses.

As an example, we will make a comparison of how the two approaches have answered the question: "if a language has seven vowels (without differences in length, tone, or phonation), what will they be?"

1.1. The reality of symmetry

The following seven-vowel system is very common (Crothers 1978, Maddieson 1984):

(2)
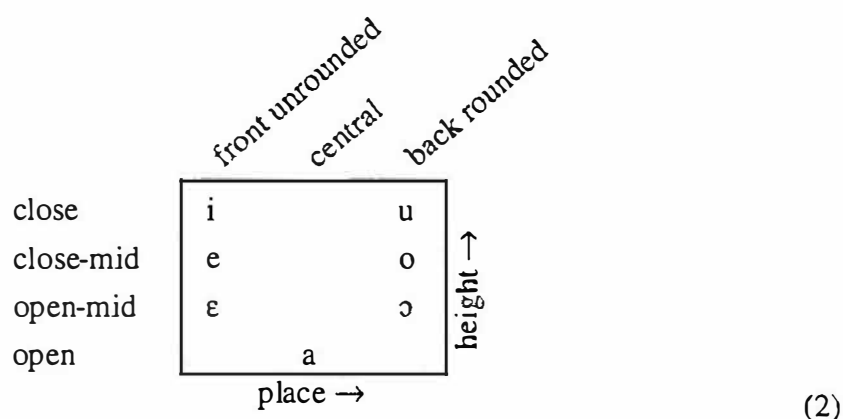
The first thing that meets the eye of the phonologist, is the *symmetry* within this inventory. There are three front unrounded vowels, two close vowels, and so on. Groups like these are called *natural classes*, and they are the most likely candidates to co-occur in synchronic *phonological rules* or historical *sound changes*.

— *Phonological rules.* In northern standard Italian, which has exactly the portrayed vowel structure in stressed syllables (ignoring the diphthongs, and admitting that [a] is rather front), the open-mid vowels merge with the close-mid vowels when these syllables become unstressed as a result of a morphological operation. So /sp'ɛndo/ 'I spend' next to /v'endo/ 'I sell', but /spendi'amo/ 'we spend' next to /vendi'amo/ 'we sell'; likewise /p'ɔrgo/ 'I present' and /s'orgo/ 'I arise', but /pordʒ'amo/ and /sordʒ'amo/. The important thing here is that the back vowel [ɔ] behaves exactly like the front vowel [ɛ], which is the rationale behind suggesting that [ɛ] and [ɔ] constitute a natural class, and the reason why we can talk generalizingly about "open-mid" vowels at all.

— *Sound change.* In some parts of Italy (let's call them area A), /ɛ/ does *not* contrast with /e/ (there is only one mid front vowel). And in area B, /ɔ/ does not contrast with /o/. Now, the areas A and B *are the same*. This means that front and back vowels must have had a common property at the time when the late sound change (merger of lower and higher mid vowels) occurred.

A more striking example of this front/back concerto is found with the sound changes that converted the Latin vowel system into system (2) when the Latin length correlations were lost. Latin long /iː/ and /uː/ became Italian /i/ and /u/ (viːnum →

vino 'wine', luːna → luna 'moon'), while Latin short /i/ and /u/ were both lowered to Italian /e/ and /o/ (fidem → fede 'belief', supra: → sopra 'above'). Latin long /eː/ and /oː/ became Italian /e/ and /o/ (feːtʃiː → fetʃi 'I did', doːnum → dono 'gift')[1], whereas both short /e/ and /o/ were lowered (and often diphthongized) to /(i)ɛ/ and /(u)ɔ/ (pedem → piɛde 'foot', rotam → ruɔta 'wheel'). In all these cases, the height contour of a front vowel was changed in the same way as that of its corresponding back vowel; the symmetry was preserved[2].

– Phonetics. There is nothing mysterious about the common behaviour of vowels at the same height. The symmetry suggests that there must be something similar in vowels of the same height. As this cannot be found in the muscles used for these sounds (genioglossus, lower longitudinals, and risorius for unrounded front vowels; styloglossus and orbicularis oris for rounded back vowels), it must be a *perceptual* similarity. As argued by e.g. Lindau (1975), this perceptual similarity between vowels at equal height is the *first formant* (F_1), the location of the peak in the excitation of the basilar membrane in the human inner ear furthest from the oval window. So, vowels at the same height have equal values for the first formant. Leoni, Cutugno & Savy (1995) measured the acoustic F_1 values (acoustically, formants are measured as peaks in the frequency spectrum) of the Italian seven-stressed-vowel system, and of the five-unstressed-vowel system. /ɛ/ and /ɔ/ have the same F_1, so have /e/ and /o/, and so have /i/ and /u/[3]. Thus, the changes from Latin to Italian can be described as changes in the F_1 *contours* of the vowels or diphthongs.

The degree of symmetry seems to depend on inventory size. While most four-vowel systems (Maddieson 1984) are asymmetric (/a ɛ i u/ instead of /ɛ ɔ i u/), large systems like the 18 long vowels and diphthongs of Geleen Limburgian are very symmetric:

i	y	u		iː	yː	uː	
				iæ	yœ/øœ	oɐ	
e	ø	o		eː	øː	oː	
				ɛi	œy	ɔu	$\downarrow F_1$
ɛ̧	œ̧	ɔ		ɛː	œː	ɔː	
æ		ɑ		æi		ɑu	
					aː		(3)

First note that (apart from a possible emerging split in the opening diphthongs), the 18 long vowels are distributed over only seven distinct F_1 contours.

– Distribution. Sounds of equal length and height form natural classes in Limburgian phonemic distributions:

- [ɛi œy ɔu], [æi ɑu], and [i y u] do not occur before /ʀ/ in the same morph.
- [iæ yœ/øœ oɐ] always carry the acute accent.
- Of the short vowels, only [i u] can occur at the end of a word.

---

[1] Any length in the Italian reflexes is related to Italian stress, not directly to length in Latin.

[2] If the account of the Latin vowel system can be questioned on the ground that we do not know for sure that, say, /iː/ and /i/ had the same quality, the historical relations can be replaced by a comparison of Italian with Sardinian which simply merged Latin /iː/ with /i/, /uː/ with /u/, /eː/ with /e/, and /oː/ with /o/.

[3] Apart from an apparently cross-linguistically fixed and unexplained 30-Hz difference between front and back vowels, with back vowels having the slightly higher F_1.

The point, again, is that the front unrounded, front rounded, and back rounded vowels act in the same way.

– Phonological rules. Sounds of equal length and height also form natural classes in a pervasive phonological rule:

- The Limburgian *umlaut* rule, which is used in the formation of diminutives and in the formation of many plurals, makes the following vowel changes: /u/ → /y/, /o/ → /ø/, /ɔ/ → /œ/, /ɑ/ → /æ/, /uː/ → /yː/, /oɐ/ → /øœ/, /oː/ → /øː/, /ɔu/ → /œy/, and /ɔː/ → /œː/ (also /aː/ → /ɛː/). So, this is an alternation that uses very different tongue-body movements, but keeps vowel height intact.

– Sound change. More proof of the organizational power of vowel height can be seen in sound changes and regional variation. The following examples take the Geleen dialect, which has a very conservative vowel system, as a reference:

- In the Sittard dialect (Dols 1944 [1953]), underlying acute /éː ǿː óː/ (/béːʀ/ 'beer', /zǿːkə/ 'search', /ɣóːt/ 'good') became /éi ǿy óu/ (/béiəʀ/, /zǿykə/, /ɣóut/) and [iæ yœ/øœ uɐ] (/kiæs/ 'cheese', /ɦyœʀə/ 'hear', /ɣʀoɐt/ 'great') became [éː ǿː óː] (/kéːs/, /ɦǿːʀə/, /ɣʀóːt/).
- In the Roermond dialect (Kats 1985), [iæ yœ/øœ uɐ] merges with and into [éː ǿː óː], so that /ɣóːt/ 'good' rhymes with /ɣʀóːt/ 'great'.
- In the Venlo dialect (Peeters 1951), which does not contrast [æ] with [ɛ], [ɛ œ ɔ] are much lower ([ɛ œ ɔ]), [ɛː œː ɔː] are much higher ([e̝ː œ̝ː ɔ̝ː]), and [iæ yœ/øœ oɐ] are [iə yə uə] (with accent contrasts).
- In the Maastricht dialect (Tans 1938), [ɛː œː ɔː] became [eː øː oː], surface acute [íː ýː úː] became [éi ǿy óu], and [iæ yœ/øœ oɐ] became [i y u] (/kis/ 'cheese', /py/ 'paws', /ɣʀut/ 'great').

In all these cases, the three vowel places act in the same way.

The examples discussed above can be multiplied at will for all kinds of languages and features. The symmetry is real, phonetically as well as phonologically. If a front vowel changes its height contour, the corresponding back vowel, if it exists, follows suit in the far majority of cases.

We are in search of a theory that both accounts for symmetry and explains it, i.e. a theory that has both descriptive adequacy and explanatory power.

1.2. The phonological approach to symmetry in rules

To account for symmetry, phonologists take the solution of describing each segment as a bundle of *features*, preferably in such a way that they can describe *both* the inventory *and* the phonological processes in terms of these same features. These features traditionally take on no more than two values, so let us first see how these *binary* features describe the Italian mid-vowel merger rule.

First, vowel height, with its four possible phonetic values in Italian, has to be split up into at least two binary features. Several ways of doing this have been proposed, but let us work here with the labels in the left-hand side of figure (2). For instance, we may use the features [open] and [mid], with the values [+open], [–open], [+mid], and [–mid]. As *place* features, we could use [back] and [round]. So, /e/ would be [–open, +mid, –back, –round]. The mid-vowel merger rule can now be stated as a *feature-changing rule*:

$$\begin{bmatrix} +\text{open} \\ +\text{mid} \end{bmatrix} \rightarrow \begin{bmatrix} -\text{open} \\ +\text{mid} \end{bmatrix} \Big/ [-\text{stress}] \tag{4}$$

which says that if an unstressed segment is [+open] and [+mid], it is *changed* to [–open] and [+mid], without changing any other feature. We could OT-ize this with an equivalent constraint ranking, with faithfulness constraints for each relevant underlying feature value and an ad-hoc phenomenological structural output constraint against open-mid vowels in unstressed position ("*" means "should not occur in the output"):

PARSE (+mid) *Italian mid-vowel merger*

*[+open & +mid / –stress]

PARSE (+open)

(5)
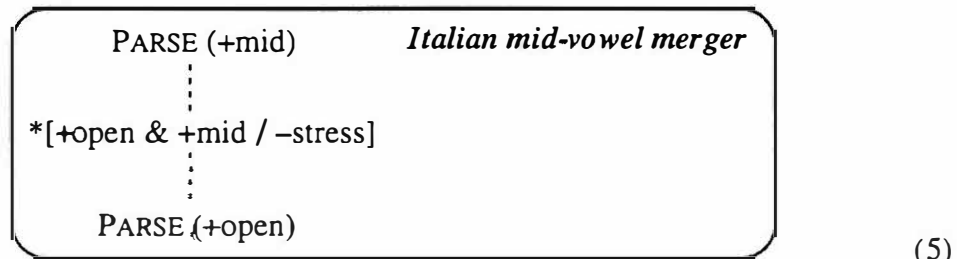
This would correctly account for the surface forms of underlying open-mid vowels:

Input: /ɛ/ unstressed	PARSE (+mid)	*[+open & +mid / –stress]	PARSE (+open)
[ɛ]		*!	
☞ [e]			*
[a]	*!			(6)

The output candidate [ɛ] honours all faithfulness constraints, but violates the constraint against open-mid vowels. In the winning candidate [e], the structural constraint is satisfied at the cost of violating the lower-ranked PARSE constraint that calls for the surfacing of the underlying [+open] specification. The structural constraint could also be satisfied by unparsing the [+mid] specification, giving as our third candidate the open non-mid vowel [a], but that would violate the higher-ranked PARSE constraint that says that underlying mid vowels should surface as mid vowels.

The important asset from formulation (4) or (5) is that it merges the two rules /ɛ/ → [e] and /ɔ/ → [o] into one; it accomplishes this by generalizing over all values for the features [back] and [round].

We can now be more precise about what defines a natural class in a phonological rule. The Italian mid-vowel example gives us five natural classes, defined in various ways:

- The group of segments that undergo the rule must have something in common; this is the *structural description* of the rule: the left-hand side plus the environment clause (unstressed) in (4), or that what is forbidden by the structural constraint in (5): the class of open-mid vowels, consisting of /ɛ/ and /ɔ/.
- The group of segments that are the result of applying the rule must have something in common; this is the right-hand side in (4): the class of close-mid vowels: /e/ and /o/.
- The segment that undergoes the rule must have something in common with the result of the rule, namely, the features not mentioned in (4) or (5). These common properties also define a natural class. So /ɛ/ and /e/ belong to the class of front

unrounded vowels. A natural class found in this way may contain more segments than just these two (in this case, /i/).

- Likewise, /ɔ/ and /o/ belong to the class of back rounded vowels, with /u/.
- Combining the arguments, the four segments /e/, /ɛ/, /o/, and /ɔ/ together must form a subset of yet another natural class, the mid vowels. The rule involves *neutralization* of the class [+mid], as shown by the position of this class at the top of the sandwich in (5) and by the fact that the *merger* of /ɛ/ and /e/ into [e] allows us to rewrite (4) as [+mid] → [−open]. Rule (4) is said to apply *vacuously* to /e/ and /o/: though these segments meet the structural description of the rule, they are not changed by it; in (5), this is reflected by the satisfaction of faithfulness for these underlying segments.

It should be noted that although the proposed account adequately describes the generalizations in the process of mid-vowel merger, it provides no explanation. This is quite apparent in the explicitly arbitrary formulation (4), but the OT version (5) fares no better: while the two PARSE constraints might be seen as natural (but where do the features come from?), the formulation of the structural constraint hides any relations that could explain the behaviour of open mid vowels, such as their relations with their neighbours, or the reason for the dependence of its ranking on stress: though we could imagine a not very active constraint *[+open & +mid / +stress] ranked below PARSE (+open), there is no explanation for why it should rank below *[+open & +mid / −stress].

1.3. The phonological approach to symmetry in inventories

When looking at the Italian inventory (2), we see that the four binary features generate together 16 possible combinations, whereas Italian uses only seven of them, and not a random subset. Now, in Italian, either of the features [back] or [round] is redundant, in the sense that every [+back] vowel is also [+round], and every [+round] vowel is also [+back]. So, as far as the inventory is concerned, we could do without the feature [round] (or [back]) at all. The remaining three features give eight possible combinations, which is enough for Italian. However, if we look at the French inventory:

	front unrounded	front rounded	back rounded
close	i	y	u
close-mid	e	ø	o
open-mid	ɛ	œ	ɔ
open	a		

place →   (height ↑)

(7)

we see that for French we do need all four features [open], [mid], [back] and [round] in the following *complete specification*:

	i	y	u	e	ø	o	ε	œ	ɔ	a
[open]	−	−	−	−	−	−	+	+	+	+
[mid]	−	−	−	+	+	+	+	+	+	−
[round]	−	+	+	−	+	+	−	+	+	−
[back]	−	−	+	−	−	+	−	−	+	−	(8)

Now we take from the theory of *contrastive underspecification* (Clements 1987, Steriade 1987) the idea that the inventory can be described by its *redundancies*:

(a) All back vowels are rounded; and, logically, all unrounded vowels are front.
(b) Low (i.e., open non-mid) vowels are unrounded, so that rounded open vowels must be mid, and rounded non-mid vowels must be non-open.

Instead of the original derivational formulation, it is easier to describe these facts as language-specific *constraints* or *well-formedness conditions* on possible segments, expressible as OT-able output constraints analogous to the structural constraint in (5):

$$*[+back\ \&\ -round]\ \ ;\ \ *[+open\ \&\ -mid\ \&\ +round] \qquad (9a;b)$$
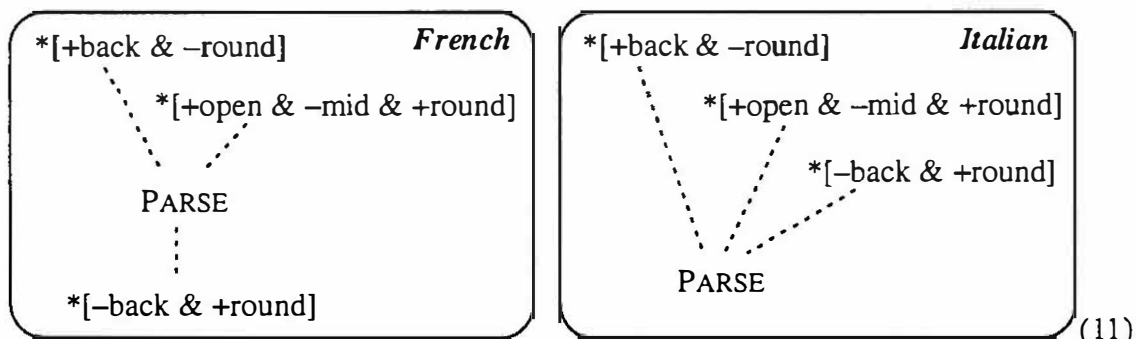
Instead of this bottom-up approach (from segments to constraints), we could also perform the top-down procedure, borrowed from the theory of *radical underspecification* (Archangeli 1984, 1988), of starting with a language-specific feature set, deriving a maximum inventory, and limiting this with universal and/or language-specific constraints. Thus, the four binary features [open], [mid], [round], and [back] yield $2^4 = 16$ possible sounds; of these, the four back unrounded vowels [ɯ], [ɤ], [ʌ], and [ɑ] are ruled out by (9a), and the two low rounded vowels [œ] and [ɒ] by (9b). This leaves exactly the ten attested vowels. So the redundancy constraints form not only *necessary*, but also *sufficient* conditions on possible feature combinations.

According to the principle of the *richness of the base* (Prince & Smolensky 1993), the limitations of each language are caused not by limitations in the lexicon, but by the workings of the constraint system. For inventories this means that the underlying form could contain any universally possible utterance, specified in the usual universal (hybrid) phonological features, and that the grammar filters this into a well-formed utterance. For French, the inventory should *follow* from the set of four features together with the two constraints (9a;b) dominating faithfulness. For instance, a hypothetical underlying /ʌ/ is filtered into a different vowel, most likely [ɔ] or [ε], depending on the exact ranking of the faithfulness constraints:

Input: /ʌ/	*[+back & −round]	PARSE (+back)	PARSE (−round)
[ʌ]	*!		
☞ [ɔ]			*
[ε]		*!		(10)

Collapsing the PARSE constraints, the French vowel system can be described with the grammar (at the left)

French	Italian

*[+back & –round]
 	
    *[+open & –mid & +round]
 	
PARSE
⋮
 	
*[–back & +round]
	*[+back & –round]
	
	    *[+open & –mid & +round]
	
	        *[–back & +round]
	
	PARSE

(11)

We can now describe sound inventories in a generalizing manner: given only four features and two constraints, we can derive a system of ten vowels. Most languages seem to have this kind of economically representable grammars, a phenomenon that we identified earlier as the preference for symmetry. With a basic tenet of radical underspecification theory, it would be good if the constraints proved universal, i.e., if all languages draw from the same set of non-conflicting constraints. The two constraints of the French vowel structure, at least, are widely found throughout the world. In fact, both the Italian and the French system can be described with the same constraint set, but with different rankings, as (11) shows. A simpler representation of Italian, however, without the feature [round], would only have the single dominating constraint *[+open & –mid & +back], if /a/ is considered a front vowel.

Note that the structural filters can still be explained only very indirectly.

1.4. Evaluation of the phonological approach

The most important requirement in the phonological approach is *empirical adequacy*: we aim at a theory that predicts what is possible in language and what is not. The strengths of this approach are reflected in two main points:

(a) *Symmetry*. If the distinctive features for a language have been identified, the principle of *maximum use of available features* guarantees symmetry in inventories.

(b) *No autonomous inventories*. In most phonological theories, inventories are not posited but follow from the grammar. In Radical Underspecification theory (Archangeli 1988), the segments follow from the features, which are used maximally, and from the constraints, which restrict the use of feature combinations. For instance, the default rule [+back] → [+round] was meant to fill in the values of unspecified features, and at the same time this rule was an implicational generalization on segment structure. The optimality-theoretic correlate of this position is the richness of the base, combined with structural constraints.

The supposed connection between inventories and production grammars makes empirical predictions: non-contrastive or redundant feature values are thought to be phonologically inert. Contrastive Underspecification theory (Steriade 1987), for instance, holds that only contrastive features can be transparent to spreading.

The downside of the phonological approach is its lack of *explanatory power*, which the filter constraints of the previous section still share with Chomsky & Halle's (1968) *markedness conventions*, of which they were mere reformulations. The constraints *follow* from the language data, and can, therefore, not *explain* these data; they do not tell us why we should maintain certain features, why these features should be binary, and why we should maintain certain constraints and not others.

– *Features*. The term *distinctive features* suggests that features are chosen on their ability to implement perceptual distinctions. So, vowel height would correspond to the acoustic and perceptual feature of first formant, and the front-back distinction to the second formant. The terms [back] and [round], however, sound as if they refer to articulatory gestures. To quote Hammarström (1973: 161): "[Using] articulatory terms to describe auditory facts (...) may be acceptable for the purpose of many descriptions (as long as one knows what one is doing)". The danger is that if the next generation is not told what they are doing, they will take the articulatory terms at face value. From the functional standpoint, distinctive features can only be perceptual (i.e., auditory and visual) categories, because proprioceptive categories cannot be communicated directly.

– *Binarity*. In (8), the really natural class [+open & –mid] has a more complicated representation than the alleged class [–mid], which would contain /i/ and /a/. This strange situation is a direct result of the obligatory binarity, which breaks up phonetically continuous dimensions.

The real solution to the quantization problem is to let go of binarity as an organizational necessity, and regard vowel height as multi-valued (Ladefoged 1971). Communicatively, the notion of an originally continuous vowel-height feature is not problematic at all: because of innate capacities of human perception, the learner will divide it into a finite number of categories. This number is language-specific; the fact that many features are binary is caused by nothing more than the inability of the listener to distinguish faithfully more than two values of those perceptual features.

– *Constraints*. The largest problem is how to restrict the output of the grammar. In (11), we could indirectly detect two articulatory constraints (against lip rounding and against jaw widening), and two perceptual constraints (favouring maximal distinctivity by requiring back vowels to be round and by preferring that all vowels are either low or high). Radical underspecification theory (Archangeli 1988) tells us that the *default rules* [] → [–round] ("by default, vowels are not rounded") and [+back] → [+round] ("back vowels are rounded") are universal and innate. However, these indirectly stated rules express what we would expect to result from functional considerations: the former is "we'd rather not perform a lip rounding gesture", and the latter is "to implement perceptual backness (low F_2) in a contrast with front vowels, we have to make a tongue-backing gesture as well as a lip-rounding gesture". We would like, therefore, to disentangle these explanations into directly expressed functional constraints. Further, Archangeli admits that language-specific constraints are also needed, but we will see that these can be expressed directly as general functional principles as well.

The *innateness* requirement seems to be connected to the general lack of explicability, although a learnability issue has also often been advanced (for OT: Tesar & Smolensky 1993). From the functional standpoint, we can explain constraints and show that they can be learned (Boersma, to appear), so we need not assume their innateness. Gaps in sound systems, expressed here as arbitrary filters, will be seen to be caused in fact by asymmetries in the human speech production and perception systems.

1.5. The "phonetic" approach to sound inventories

Phonetic attempts to explain sound inventories have used only a few functional principles.

Kawasaki (1982) restricted her explanations to the two perceptual principles of maximization of distinction and salience.

Stevens (1989) tried to explain the commonness of some sounds as the minimization of precision and the simultaneous maximization of acoustical reproducibility.

Liljencrants & Lindblom (1972) investigated how vowel systems would look like if they were built according of the principle of maximum perceptual contrast in a multi-dimensional formant space. They searched for the optimal 7-vowel system by maximizing within a fixed two-dimensional perceptual space the perceptual contrast, which they defined as the sum of inverse-squared distances between all pairs; they based the distance between two vowels on the difference in F_1 and F_2 expressed in Mels. The results were not satisfactory: because they gave equal weight to F_1 and F_2 differences, the simulated systems showed too many place contrasts relative to the number of height contrasts.

Lindblom (1986) did the same by comparing all subsets with seven vowels taken from a fixed set of 19 'possible' vowels, and choosing the subset that has the largest internal perceptual contrast, based on the distance between two vowels in terms of the difference between the excitation patterns that the vowels would give rise to in the inner ear of a listener. This did not solve the F_2 problem.

Ten Bosch (1991) explained vowel systems on the basis of maximal distinctions within an articulatory space bounded by an effort limit based on the distance from the neutral vocal-tract shape. He decided to *fit* the parameter that determines the relative importance of the front-back distinction with respect to the importance of the height distinction, to the data of the languages of the world, assigning a value of 0.3 to the relative importance of the second-formant distance with respect to the first-formant distance.

A similar approach is found in Boë, Perrier, Guérin & Schwartz (1989), Schwartz, Boë, Perrier, Guérin & Escudier (1989), Vallée (1994), Boë, Schwartz & Vallée (1994), Schwartz, Boë & Vallée (1995), and Schwartz, Boë, Vallée & Abry (1997). Their simulations pointed to a value of 0.25 for the weighting of the F_2 distance.

In an attempt to derive, instead of fit, the relative unimportance of place distinctions with respect to height distinctions, Lindblom (1990) suggested that for determining the contrast between two vowels, proprioceptive contrasts in the speaker (jaw height can be felt more accurately than tongue-body place) are equally important as auditory contrasts in the listener. His predicted 'optimal' 7-vowel system was

$$
\begin{array}{llll}
\text{i} & \text{ʉ} & & \text{u} \\
& & \text{ɤ} & \\
\text{ɛ} & & & \\
\text{a} & & \text{ɑ} & \hspace{3cm} (12)
\end{array}
$$

which he considered to be in "extremely close agreement" (p. 79) with the most common 7-vowel systems found in Crothers (1978), which are

$$
\begin{array}{lll} \quad\quad & & \\
\text{i} & \text{i} & \text{u} \\
\text{e} & \text{ə} & \text{o} \\
& \text{a} &
\end{array}
\qquad\qquad
\begin{array}{lll}
\text{i} & & \text{u} \\
\text{e} & & \text{o} \\
\text{ɛ} & & \text{ɔ} \\
& \text{a} & \quad (13a;b)
\end{array}
$$

1.6. Evaluation of the "phonetic" approach

The main problem with a result like (12) is that it is descriptively totally inadequate: it shows no symmetry, no features, no organization. None of these approaches derives the symmetry that is visible in (1). Schwartz et al. (1997) admit that symmetry "does not always emerge from the intrinsic principles of the theory" (p. 261). Indeed, each

of their four proposed six-vowel systems is less symmetrical than any of the four most common six-vowel systems in Maddieson's (1984) database. Basically, the cause of the problem is that the distance function will actually favour an asymmetry of height between front and back vowels, because a difference in F_1 will always contribute positively to the perceptual distance between a pair of vowels.

Also, Lindblom takes *finiteness* for granted, as witnessed by his use of a finite inventory of phonemes. Schwartz et al. (1997) state that "the problem of the finiteness of the number of speech sounds, important from a theoretical point of view, is in fact impossible to address in a technically satisfying way" (p. 265). The Lindblom school appears to consider tone, duration, and voice quality to be independent features, as witnessed by his neglect of these dimensions. Apparently, these three features are tacitly considered "suprasegmental", or better: independent from the other (here: spectral) features; we can call this *autosegmental*. But for large vowel inventories, F_1 is an autosegmental feature like the others; we can see that when we realize that it is an acoustically distinct aspect of vowels, ready to be divided up into a number of perceptual categories by the language learner.

Lindblom himself (1990) tries to tackle the symmetry problem, and boasts of having found self-organization in a hypothetical language consisting of nine CV utterances only. The nine utterances that emerged most often in repeated simulations were rather symmetric together, but these were not simulated as a group. Nevertheless, let's concede that Lindblom's optimization criterion would yield the following very symmetric set of non-low vowels:

$$\begin{array}{ccc} i & y & u \\ e & ø & o \\ ɛ & œ & ɔ \end{array} \qquad (14)$$

Even then, the symmetry would break down if we asked Lindblom's optimization criterion to give us eight instead of nine utterances. Without performing the actual simulation, we can predict that Lindblom's strategy will yield something like:

$$\begin{array}{ccc} i & y & u \\ e & & o \\ ɛ & \cancel{ø} & ɔ \end{array} \qquad (15)$$

because the perceptual space gets narrower as vowel height decreases. In reality, however, we find things like the Frisian short-vowel system (1) without a lowered /ø/, thus retaining four vowel heights. Obviously, it is the features, not the segments, that structure sound systems.

It seems thus impossible to build an algorithm for generating possible sound systems without symmetrizing principles.

So, the phonetic approaches do not perform well on describing symmetry, which we identified in §1.4 as one of the strong points of the phonological approach. The other point was the connection between inventories and the grammar; in all the phonetic approaches, the modelling of inventories is a goal in its own right, and the grammar (natural classes, output constraints) is not even considered.

What, then, could be the strong points of the phonetic approaches?

(c) *Predicting dispersion.* Phonetic principles could explain some of the constraints on the basis of perceptual contrast: if back vowels are round, they are more unlike front vowels than if they are not round; maximizing the perceptual contrast helps the listener to recognize the speaker's message. Further, the vowel bucket is narrower for the low vowels than it is for higher vowels, i.e., the distance between the F_2's of [a] and [ɑ] is much smaller than the distance

between the F_2's of [i] and [u]; this answers the question which of the eight possible feature combinations in Italian should be the most likely candidate for not being found (namely, a low vowel, be it back or front).

(d) *Predicting the gap*. Given three height and place features, the maximal inventory of non-low vowels is (14). The phonetic approaches can answer the question: if this system has one gap, where will it be? The answer is that the gap will be at /œ/, because the front-back distance is smaller there than at the other heights. This simple contrast-based account seems more natural than the awkward feature-cooccurrence constraints of the phonological approach.

(e) *Predicting the arity*. If height and place are multi-valued features, how many values will they have? Specifically, what is the relation between the average number of heights and the average number of places? Unfortunately, the phonetic approaches have not been able to derive this relation, although the general conviction is that it would be possible if we knew enough about the perception of frequency spectra.

Unfortunately, these approaches have not yet been able to measure any phonetic spaces; a problem with one degree of freedom can always be "solved" by fixing one parameter. Unless we accept Lindblom's (1990) proposal for taking into account the speaker's proprioceptive height and place distinctions, the relative importance of the first formant must be sought in its greater loudness with respect to the second formant: the second spectral peak has a larger chance of drowning in the background noise. In Boersma (forthcoming), I computed the distances between the basilar excitation patterns of [a], [i], and [u] in units of just-noticeable differences (jnd), and found that the distance between [i] and [u] was 12 jnd, and the distance between [a] and each high vowel was 18 jnd. This means that a system with four heights is equally well dispersed as a system with three places, namely, with 6 jnd between each pair of neighbours. This would predict that (13a) and (13b) would be equally common inventories, and this seems to be the case.

1.7. "Integrated" approach 1: enhancement

It seems that we will need to combine phonological and phonetic principles if we want to describe and explain inventories at the same time. The example of the rounding of back vowels will make this clear.

In the vowel systems of the languages of the world, most back vowels are round and most rounded vowels are back. The "phonological" approach has not given any explanations for this fact: the correlation between [round] and [back] was viewed as a part of Universal Grammar, hard-wired into the human language faculty. In phonetic terms, however, the explanation of the correlation between [round] and [back] is straightforward. For a maximal perceptual contrast between two places of articulation, a language should have unrounded front vowels (maximum F_2) and rounded back vowels (minimum F_2).

Even in phonetics, however, the necessary distinction between perception and production seems not always to be made. Stevens, Keyser & Kawasaki (1986) speak of the *enhancement* by lip rounding of the perceptual contrast between vowels with high and vowel with low F_2. With a proper division of labour between perception and production, the statement should be altered to: "a maximal F_2 contrast is *implemented* by having a group of vowels with front tongue position and lip spreading, and a group with back tongue position and lip rounding". Rounding, therefore, does not enhance a contrast, but helps to implement it. For why should styloglossus be the *agonist*, and orbicularis oris the *synergist*? The asymmetric interpretation by Stevens et al. of this phenomenon as the enhancement of backness by rounding smacks of a confusion of

the phonological feature [back], which can be used as an arbitrary label for a certain perceptual contrast, with the articulatory gesture of backward tongue-body movement. Apparently acknowledging this problem, Stevens & Keyser (1989) explicitly divide phonological features into primary and secondary features. While this move was in itself data-driven, because partly based on commonness in speech, the notion that frequency of occurrence has a strong correlation with perceptual distinctivity, is indubitable.

1.8. "Integrated" approach 2: inventory constraints

A functionally-oriented Optimality-Theoretic account was given by Flemming (1995), who handles inventories as the result of the interactions between the functional principles of maximizing the number of contrasts and maximizing the auditory distinctiveness of contrasts. These two principles correspond to Passy's (1890) assertion that speakers will try to get their messages across as *quickly* and *clearly* as possible (respectively).

These principles lead to fixed rankings, e.g. for high vowels along a fixed F_2 axis { i y ɨ ɯ u }. First (for the maximization of the rate of information flow), it is more important to maintain *two* contrasts than it is to maintain *three* contrasts:

MAINTAIN 1 F_2 contrast >> MAINTAIN 2 F_2 contrasts >> MAINTAIN 3 F_2 contrasts

(16)

Secondly (for the minimization of confusion), it is less bad to have two vowels at an "auditory distance" of *three* steps along the discretized F_2 axis { i y ɨ ɯ u } than it is to have them at a distance of *two* steps:

$$\text{MINDIST}_{F2} = 1 \gg \text{MINDIST}_{F2} = 2 \gg \text{MINDIST}_{F2} = 3 \gg \text{MINDIST}_{F2} = 4 \quad (17)$$

Interleaving these two constraint families in a dedicated inventory grammar (i.e. a grammar that evaluates inventories directly), we can choose a grammar that gives { i ɨ u } as the best inventory:

	MAINTAIN 1 contrast	MINDIST = 1	MAINTAIN 2 contrasts	MINDIST = 2	MINDIST = 3	MAINTAIN 3 contrasts
i-u			*!			*
☞ i-ɨ-u					*	*
i-ɨ-ɯ-u				*!	*	*
i-y-u				*!		
i-y-ɯ				*!		

(18)

MINDIST is a constraint formulated as the OT optimization criterion itself: "minimize the maximum problem"; therefore, it is probably the surface result of a more primitive constraint system, e.g. with constraints like "distance ≤ 2". In contrast with (18), such a system would rank the inventory { i y u } above { i y ɯ }, because all three pairs would be evaluated, not just the closest pair; this is a desirable property.

The system { i ɨ u } turns out to be better than the Frisian system { i y u }, for every possible ranking of the constraints, as long as the rankings (16) and (17) are kept fixed. The Frisian preference for { i y u } over { i ɨ u } probably has to do with the choice of the gestures that should implement the central F_2 value: either with

frontal tongue-body raising and lip rounding, or with central tongue-body raising. To account for this, constraints against performing the relevant articulatory gestures should be added to the inventory grammar, and Flemming does so in several cases.
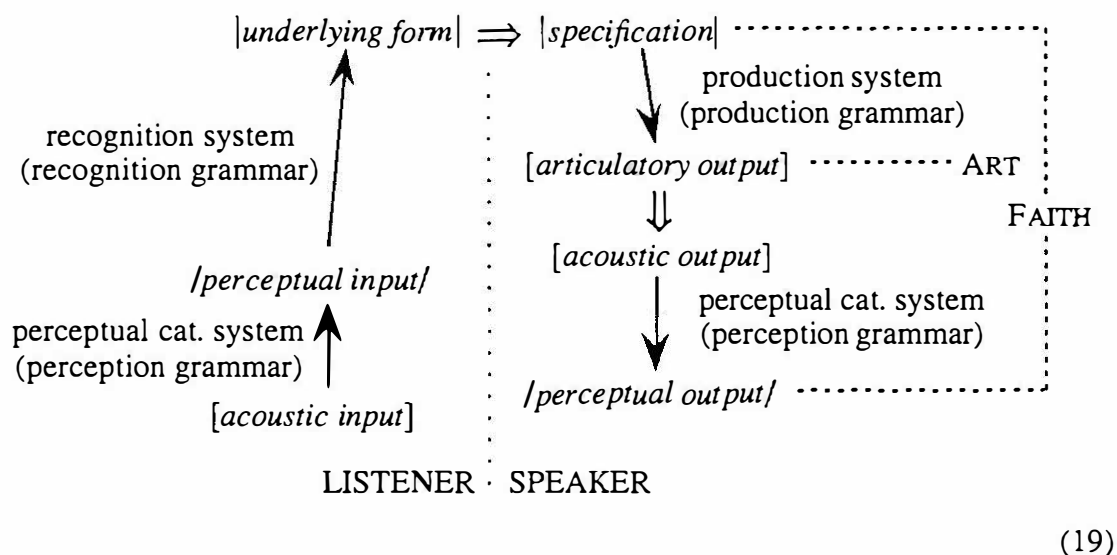
But there is a problem with Flemming's approach, namely that (18) does not represent a production grammar, i.e., it is not a model of how a speaker converts underlying to surface forms: it evaluates inventories instead of output candidates. Flemming gives up a requirement still honoured by the underspecification approaches of §1.3, namely, that inventories are built on the same principles as the grammar. In an OT production grammar, the connection with the inventory can be upheld by the principle of richness of the base; inventory grammars like (18), however, do not explain how a random input is filtered into a well-formed utterance. Thus, while Flemming's approach is more advanced than any of the phonetic approaches discussed earlier, since it combines the notion of sufficient contrast while taking symmetry for granted, and though the notion of the interaction between articulatory effort and perceptual contrast is correct, Flemming's global inventory evaluation procedure is not a model of grammar; it just shows that inventories can be described with strict ranking of principles, just like so many real-life weighings of pros and cons. If, by contrast, the functional principles could be expressed directly in a local production grammar, and this production grammar could derive inventories from richness of the base, a separate global inventory grammar would be superfluous; I will show below how this can be achieved.

1.9. "Integrated" approach 3: local expression of functional principles

The faithfulness and structural constraints of (5) and (11) have direct analogues in functional principles. Structural constraints limit the possible surface structures; the functional principle of minimization of articulatory effort can be expressed in articulatory constraints against the performance of certain gestures. Faithfulness constraints punish any differences between underlying and surface forms; if the two forms are equal, an underlying contrast is still heard on the surface; thus, faithfulness constraints can implement the functional principle of minimization of perceptual confusion in a local manner, without having to compare any forms with all possibly contrasting forms as in Flemming's inventory evaluation procedure.

The implementation of gestural and faithfulness constraints in a theory of grammar requires a principled distinction between articulatory and perceptual features, so we have no hope of translating (11) directly into functional grammars for French and Italian. Instead, we should start a bottom-up procedure from first principles. This will be performed rigorously in the next sections. The resulting theory will combine all the desirable properties that we found in the phonological and phonetic approaches discussed above:

(a) *Symmetry*. Follows from the finiteness of the number of learned perceptual categories and articulatory gestures.

(b) *No autonomous inventories*. Inventory structure follows directly from the constraints in the production and perception grammars, not from a dedicated inventory grammar.

(c) *Predicting dispersion*. Sufficient contrasts emerge from the fact that a listener is also a speaker: local minimization of confusion demands enhancement of contrasts in phonetic implementation.

(d) *Predicting the gap*. The locations of the gaps follow from asymmetries in articulatory effort and perceptual contrast, as these are reflected in the local rankings of gestural, faithfulness, and categorization constraints.

|underlying form| ⟹ |specification| ·······················
                                    ↓ production system
recognition system   (production grammar)
(recognition grammar)   [articulatory output] ··········· ART
                                         ⇓   FAITH
/perceptual input/   [acoustic output]
perceptual cat. system ↑   perceptual cat. system
(perception grammar) ↑   (perception grammar)
      [acoustic input]   ↓
                               /perceptual output/ ·················

                    LISTENER · SPEAKER

(19)

2. Functional Phonology

Functional Phonology makes a principled distinction between articulatory and perceptual representations, features, and constraints.

2.1. Representations and grammars

As illustrated in figure (19), the speaker's *production grammar* handles the evaluation of the entities and relations in the following set of representations:

(1)   The *perceptual specification.* The underlying form of the utterance, specified in terms of perceptual features. The *input* to the OT production grammar.
(2)   The *articulatory implementation.* The surface form of the utterance, in terms of articulatory gestures. In the OT production grammar, many articulatory *candidate* implementations are generated and evaluated directly by gestural constraints, depicted as "ART" in figure (19).
(3)   The *perceptual result.* The surface form in terms of perceptual features. In the production grammar, faithfulness constraints ("FAITH") evaluate the similarity of the perceptual result of each articulatory candidate to the specification.   (20)

The *output* of the OT production grammar is the most harmonic articulatory/perceptual candidate, as defined by the interactions of the gestural and faithfulness constraints.

   The listener's *perception grammar* (perceptual categorization system) maps the acoustic features of an utterance (noise, periodicity, spectrum) onto language-specific numbers of values along language-specific perceptual dimensions. It is used by the listener for the initial categorization of acoustic speech events, and by the speaker to monitor her own output. In this paper, I will make the simplifying assumption that for the listener, perceptual categorization is *followed* by the recognition process.

2.2. Gestures and features

From (1), we see that Frisian speakers must have acquired the articulatory gestures of rounding and spreading of the lips, fronting, backing, and lowering of the tongue body, and lowering of the jaw. Most of the five gestures of the Frisian vowel system

exist in various degrees of distance moved away from the neutral position (to stress the activity of the movement, the list represents each gesture with one of the main muscles involved):

(a) The back of the tongue is raised further for [o] and [ø] (by styloglossus activity) than for [ɔ], and even more so for [u] and [y].
(b) The front of the tongue is raised further for [e] and [ø] (by genioglossus activity) than for [ɛ], and even more so for [i] and [y]
(c) The lips are rounded more strongly for [ø] and [o] (by orbicularis oris activity) than for [ɔ], and even more so for [y] or [u].
(d) The lips are spread more strongly for [e] (by risorius activity) than for [ɛ], and even more so for [i].
(e) The jaw is lowered further (by mylohyoid activity) for [a] than for [ɛ] and [ɔ].

(21)

These things are shown schematically in the upper part of table (22) (the actual numbers are meaningless; they just enumerate locations along a continuous scale):

		i	y	u	e	ø	o	ɛ	ɔ	a
	body: back up	0	0	3	0	0	2	0	1	0
	body: front up	3	3	0	2	2	0	1	0	0
art.	lips: round	0	3	3	0	2	2	0	1	0
	lips: spread	3	0	0	2	0	0	1	0	0
	jaw: down	0	0	0	0	0	0	1	1	2
	height	4	4	4	3	3	3	2	2	1
perc.	place	fr	ce	ba	fr	ce	ba	fr	ba	ce
	(round)	−	+	+	−	+	+	−	+	−

(22)

From (1), we see that Frisian listeners must have acquired four vowel heights. Thus, Frisians have learnt to distinguish four different F_1 "contours" for short vowels. The front-back dimension can be associated with the second main spectral peak: we have the values "maximum F_2, given the value of F_1" (implemented by the acquired gestures of tongue-body fronting and lip spreading), "minimum F_2, given the value of F_1" (implemented by tongue-body backing and lip rounding), and a value in between (implemented by tongue-body fronting and lip rounding). The perceptual features of the nine Frisian short vowels are shown schematically in the lower part of table (22); "fr", "ce", and "ba" stand for [front], [central], and [back], but could also have been named 1, 2, and 3, or [high F_2], [mid F_2], and [low F_2].

Articulatorily or perceptually related segments can form natural classes in phonological processes. Thus, we can talk of the articulatorily defined class of rounded vowels, or of the perceptually defined class of higher mid vowels.

3. Finiteness

The most important thing to be learnt from (1) is the fact that only nine short vowels occur in tens of thousands of words; and this is also the main fact that has to be explained.

3.1. Articulatory constraints and finiteness

The typical articulatory constraint that occurs in the speaker's production grammar is

*GESTURE (g): "an articulatory gesture g is not performed." (23)

The acquisition of motor skills has supplied every speaker with only a finite number of gestures that she can perform. The only "real" gestural constraints that are visible at all in a speaker-oriented grammar and have any claim to psychological reality, are the constraints against the acquired gestures: each of these must be dominated by at least one other constraint, typically a specification-to-perception faithfulness constraint like PARSE, which says that a specified perceptual feature value shall be implemented by *any* gesture.

From the universal descriptive linguistic standpoint, however, there would exist a constraint against every thinkable gesture that humans could learn to perform. Now, most of these universal constraints are undominated and play no role at all; these "virtual" constraints are merely a descriptive device for communication between linguists: they can describe aspects of the learning process and the production of loan words. For instance, the absence in the Dutch speaker's brain of any structures referring to gestures that implement implosives, can be described by an undominated *GESTURE(hyoid: lower) constraint.

Thus, "low-ranked *GESTURE and *COORD constraints determine the language-specific finite set of allowed articulatory features and feature combinations" (Boersma 1997a, p. 42). Therefore, Frisian grammars must simply contain the following dominated constraints: *GESTURE (lips: rounded), *GESTURE (lips: spread), *GESTURE (body: front up), *GESTURE (body: back up), *GESTURE (body: low), and *GESTURE (jaw: low).

3.2. Perceptual constraints and finiteness

The relevant perceptual constraint that occurs in the listener's perception grammar is:

*CATEG (f: v): "the perceptual feature f is not categorized as the value v." (24)

From the linguistic standpoint, there exists a constraint against every thinkable category that humans could learn to perceive. In this sense, the set of categorization constraints is universal, and these constraints are innate in the sense that every normal human child can learn to perceive any category.

However, the acquisition of perceptual classification has supplied every listener with only a finite number of categories that she can perceive. In the grammar of every listener, therefore, most of the universal categorization constraints are undominated and play no role at all; again, these "virtual" constraints are merely a descriptive device for communication between linguists: they can describe aspects of the learning process and the perception of loan words. The only "real" categorization constraints that are visible at all in the listener's perception grammar, are the constraints against the acquired categories: these constraints must be dominated by at least one other constraint, typically the peripheral acoustics-to-perception correspondence constraint PERCEIVE, which says that it is important that an acoustically available feature shall be classified into *any* category.

Thus, "low-ranked *CATEG constraints determine the finite set of allowed perceptual feature values" (Boersma 1997a, p. 50). Thus, Frisian perception grammars must simply contain the following dominated constraints: *CATEG (height: open) (= *CATEG (F_1: maximum)), *CATEG (height: open-mid), *CATEG (height:

close-mid), *CATEG (height: close), *CATEG (place: front) (= *CATEG (F_2: maximum)), *CATEG (place: centre), *CATEG (place: back).

Now we know the causes of the finiteness of segment inventories. Citing Boersma (1997a, p. 51):

The functional view: there are no universal phonological feature values
> "The continuous articulatory and perceptual phonetic spaces are universal, and so are the constraints that are defined on them; the discrete phonological feature values, however, are language-specific, and follow from the selective constraint lowering that is characteristic of the acquisition of coordination and categorization." (25)

An exhaustive use of four vowel heights and three places would lead to a system of twelve vowels, which is three more than Frisian actually has.

The dependence of symmetry on inventory size can be explained with a general property of categorization: the number of perceptual dimensions increases with the number of classes. Speakers of a four-vowel system may recognize the four different excitation patterns associated with /a ɛ i u/; whereas speakers of a 18-vowel system cannot recognize 18 unrelated percepts, but divide up the perceptual space along at least two dimensions: "place" and "height".

3.3. Faithfulness constraints and finiteness

The categorization constraints are not expressed directly in the production grammar. In the production grammar, the categorization is reflected by faithfulness constraints.

An important principle of effective communication is the requirement that specified features are received by the listener. Because the speaker is a listener, too, the correspondence constraint TRANSMIT (Boersma 1997a: §8.2) requires that a specified value (category) of a perceptual feature is heard by the listener as *any* category on that same perceptual tier, and the constraint *REPLACE forbids the two corresponding values to be different. For features with few categories (in this paper, even vowel height will be taken to be such a feature), we can collapse the correspondence and similarity requirements into a single constraint *DELETE or PARSE:

PARSE (f: v): "an underlyingly specified value v of a perceptual feature f appears (is heard) in the surface form". (26)

In Frisian, therefore, we have PARSE constraints for all perceptual categories. These constraints can be abbreviated as PARSE (open), PARSE (open-mid), PARSE (close-mid), PARSE (close), PARSE (front), PARSE (centre), and PARSE (back).

3.4. Gaps and richness of the base

From the functional standpoint, the input to the grammar must be specified in perceptual feature values, i.e. categorizable values of perceptual dimensions specific to the language. For Frisian, this would mean that the input may contain 12 different short vowels, if the categorization of place is independent of height (which is open to doubt, see §5). So:

Richness of the base (functional version):
> "the input may contain any combination of categorizable perceptual features; the combinations that do not occur on the surface are filtered out by the constraint system." (27)

For Frisian, this means that the constraint system will have to explain the gap in the open-mid-vowel system, and the two gaps in the open-vowel system.

4. Local ranking

According to the *local-ranking principle* (Boersma 1997a: §11), gestural and faithfulness constraints can be locally ranked with the functional principles of minimization of articulatory effort and perceptual contrast.
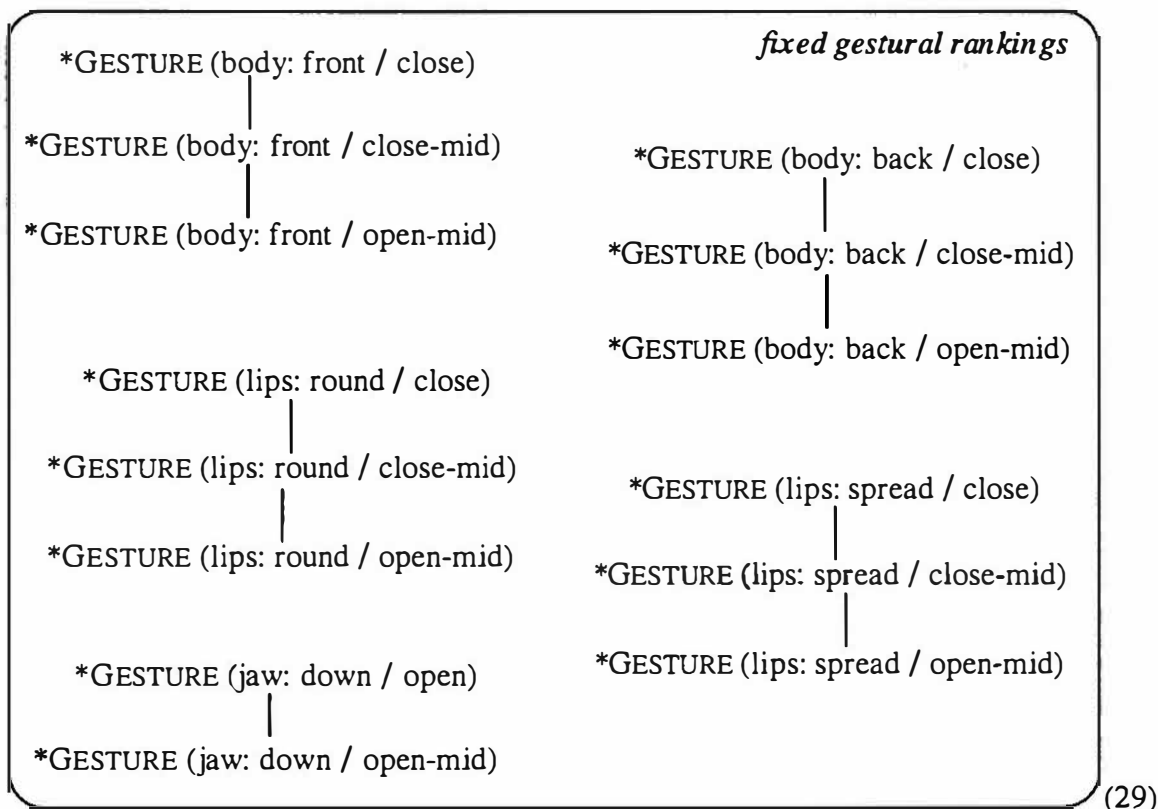
4.1. Local ranking of gestural constraints

The ranking of a gestural constraint may depend on a number of circumstances. These dependences effectively split each *GESTURE constraint into a multidimensionally continuous family:

*GESTURE (a: g / d, v, p, t): "the articulator a does not perform the gesture g along a certain distance d (away from the rest position), and with a certain speed v, reaching a position p for a duration t ." (28)

Basically, articulatory constraints are ranked by *effort*: constraints against gestures that require more effort are universally ranked higher than constraints against easier gestures, but only in the following case: the *same* gesture is more difficult if its distance, speed, duration, *or* precision is greater, *and everything else is kept equal*; this can lead to a fixed ranking of gestural constraints.

With (21), this yields the following fixed distance-based rankings, given the Frisian gesture system:

fixed gestural rankings

*GESTURE (body: front / close)
|
*GESTURE (body: front / close-mid)
|
*GESTURE (body: front / open-mid)

*GESTURE (body: back / close)
|
*GESTURE (body: back / close-mid)
|
*GESTURE (body: back / open-mid)

*GESTURE (lips: round / close)
|
*GESTURE (lips: round / close-mid)
|
*GESTURE (lips: round / open-mid)

*GESTURE (lips: spread / close)
|
*GESTURE (lips: spread / close-mid)
|
*GESTURE (lips: spread / open-mid)

*GESTURE (jaw: down / open)
|
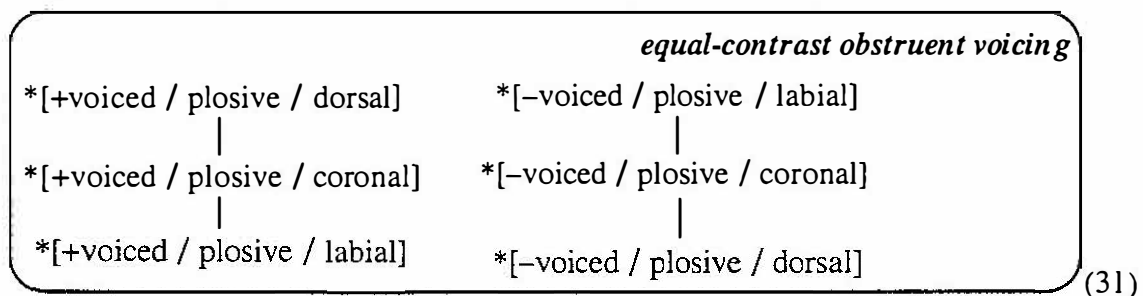*GESTURE (jaw: down / open-mid)

(29)

Since this height-dependent differentiation of the vowel constraints seems to be small once the gestures are mastered, it will be ignored in the rest of this paper. Clearer examples may be found in obstruent voicing and tongue-root inventories.

Many languages with voicing contrasts in obstruents still lack a segment /g/ in their inventory of plosives, i.e., the symmetry is broken by a *gap* at /g/:

$$
\begin{array}{ccc}
\text{(p)} & \text{t} & \text{k} \\
\text{b} & \text{d} & \text{(g)}
\end{array}
\qquad (30)
$$

It is more difficult to maintain a voicing contrast in plosives with a closure close to the larynx, than it is at other places. One of the preconditions for phonation is the presence of a stream of air through the glottis. During the closing interval of plosives, both the nasal and oral pathways are closed, and the flow through the glottis will eventually stop. One of the things that influence the maintenance of the flow is the amount to which the supralaryngeal air will be allowed to expand. For the back closure of [g], the cavities above the glottis are filled earlier with air than in [b] and so voicing will stop earlier in [g] than in [b] because of the more rapid drop in transglottal pressure (Ohala & Riordan 1979).

Thus, a specified degree of voicing is more difficult to maintain for a dorsal plosive than for a labial or coronal plosive. Likewise, a specified degree of voicelessness is more difficult to implement for a labial plosive than for a coronal or dorsal plosive. This leads to a fixed hierarchy of implementation constraints for voiced and voiceless plosives:

$$
\begin{array}{ll}
& \textit{\textbf{equal-contrast obstruent voicing}} \\
*[\text{+voiced / plosive / dorsal}] & *[\text{--voiced / plosive / labial}] \\
\quad\mid & \quad\mid \\
*[\text{+voiced / plosive / coronal}] & *[\text{--voiced / plosive / coronal}] \\
\quad\mid & \quad\mid \\
*[\text{+voiced / plosive / labial}] & *[\text{--voiced / plosive / dorsal}]
\end{array}
\qquad (31)
$$

Because the degrees of voicing and voicelessness were taken constant, we can assume a homogeneous PARSE constraint for the plosive voicing feature values. According to our version of richness of the base, the constraint system should remove an underlying /g/ in a language that lacks [g] at the surface. This will indeed be the outcome if PARSE (±voice) is sandwiched between the coronal and dorsal voicing constraints:

/g/	*[+voice / dorsal]	PARSE (±voice)	*[+voice / coronal]
[g]	*!		
☞ [k]		*	

(32)

Note that with the same hierarchy, coronal voiced plosives surface faithfully:

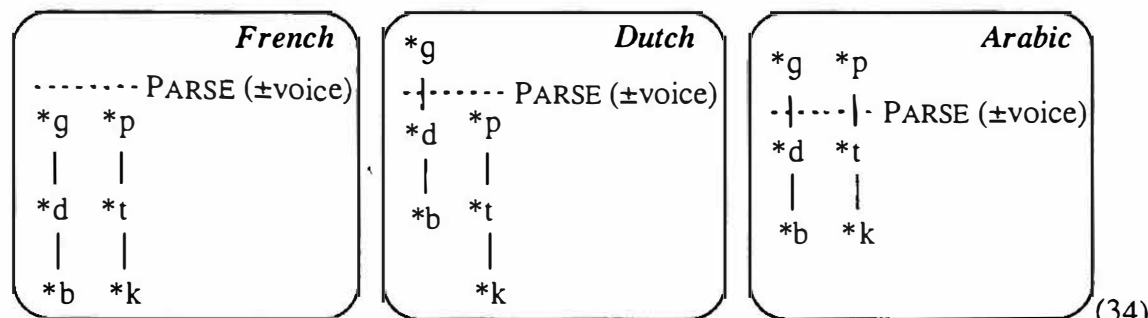| /d/	*[+voice / dorsal]	PARSE (±voice)	*[+voice / coronal]
☞ [d]			*
[t]		*!	

(33)

As the hierarchy for [+voice] is independent from the hierarchy of [−voice] (they use different types of gestures), the three following grammars are some of the possibilities (the gestural constraints are maximally abbreviated; the homogeneous PARSE constraint is shown by a dotted line):

```
┌─────────────────────────┐  ┌─────────────────────────┐  ┌─────────────────────────┐
│              French     │  │ *g              Dutch   │  │ *g  *p          Arabic  │
│  ········ PARSE (±voice)│  │ ·|····· PARSE (±voice)  │  │ ·|····|· PARSE (±voice) │
│  *g  *p                 │  │ *d   *p                 │  │ *d  *t                  │
│  │   │                  │  │  │    │                 │  │  │   │                  │
│  *d  *t                 │  │ *b   *t                 │  │ *b  *k                  │
│  │   │                  │  │      │                  │  │                         │
│  *b  *k                 │  │     *k                  │  │                         │
└─────────────────────────┘  └─────────────────────────┘  └─────────────────────────┘
```
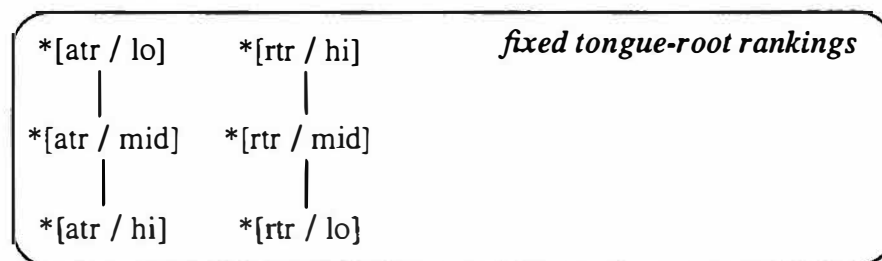(34)

Thus, French shows no gaps, Dutch lacks [g] and Arabic lacks both [p] and [g].

In the realm of vowel inventories, we find analogous rankings in tongue-root systems. If the short-vowel system becomes much larger than the Frisian example of (1), it is probable that speakers construct a third dimension. This is a general property of categorization. If a language has two vowel places (front and back) and more than four segments should be distinguished, the language has the option of dividing the F_1-based height dimension into two new dimensions, say the perceptual correlates of tongue-body (oral) constriction and tongue-root (pharyngeal) constriction, which we shall call [height] and [tr], respectively. Most tongue-root languages (Archangeli & Pulleyblank 1994), have three categories along the height dimension (low, mid, high), and two along the tongue-root dimension (atr, rtr). As is explained in more detail in Boersma (to appear), the following rankings of articulatory effort can be posited:

(a)　The [atr] value is more difficult to implement for lower than for higher vowels.
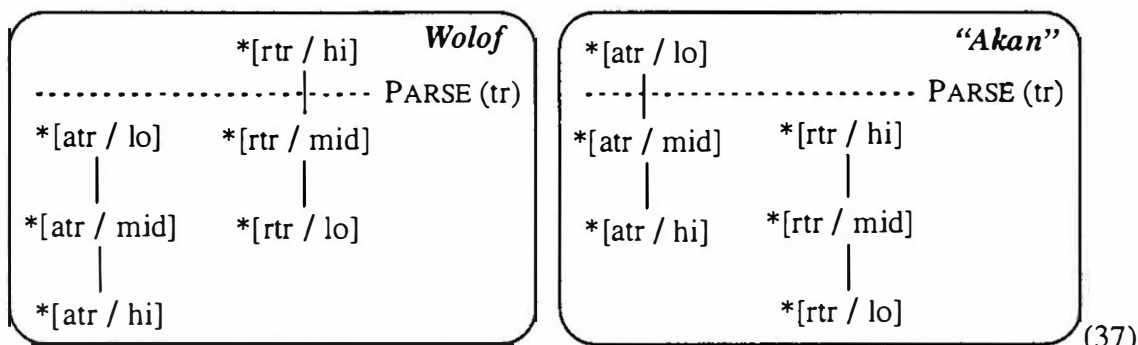(b)　The [rtr] value is more difficult to implement for higher than for lower vowels.

(35)

With the most common categorization, this leads to the following fixed hierarchies of implementation constraints:

```
┌──────────────────────────────────────────────────────────────┐
│  *[atr / lo]      *[rtr / hi]      fixed tongue-root rankings  │
│      │               │                                        │
│  *[atr / mid]     *[rtr / mid]                                │
│      │               │                                        │
│  *[atr / hi]      *[rtr / lo]                                 │
└──────────────────────────────────────────────────────────────┘
```
(36)

These are larger sets of constraints than Pulleyblank's (1994) two *grounding constraints* LO/RTR "if a vowel is low, then it has a retracted tongue root" and HI/ATR "if a vowel is high, it has an advanced tongue root", whose actions are comparable to those of *[atr / lo] and *[rtr / hi], respectively. From the functional standpoint, we should derive, not posit, which of the many possible constraints tend

to be strong and which tend to be weak: *[atr / mid], for instance, also exists although it may be universally lower ranked than *[atr / lo]. Again, we can assume a homogeneous PARSE (tongue root) constraint, because our use of implementation constraints supposes equal tongue-root contrasts for all three heights. Two of the possible grammars are (Archangeli & Pulleyblank 1994):

$$
\begin{array}{cc}
& \text{*[rtr / hi]} \quad \textbf{\textit{Wolof}} \\
\cdots\cdots\cdots\cdots\cdots\cdots\mid\cdots\cdots & \text{PARSE (tr)} \\
\text{*[atr / lo]} \qquad \text{*[rtr / mid]} \\
\mid \qquad\qquad \mid \\
\text{*[atr / mid]} \qquad \text{*[rtr / lo]} \\
\mid \\
\text{*[atr / hi]}
\end{array}
\qquad
\begin{array}{cc}
\text{*[atr / lo]} & \qquad \textbf{\textit{"Akan"}} \\
\cdots\mid\cdots\cdots\cdots\cdots\cdots\cdots & \text{PARSE (tr)} \\
\text{*[atr / mid]} \qquad \text{*[rtr / hi]} \\
\mid \qquad\qquad\quad \mid \\
\text{*[atr / hi]} \qquad \text{*[rtr / mid]} \\
\mid \\
\text{*[rtr / lo]}
\end{array}
\tag{37}
$$

From the set of categorizable front vowels { i, ɪ, e, ɛ, ə, a }, Wolof lacks [ɪ] (if PARSE (height) is ranked high, a hypothetical underlying /ɪ/ would become [i]), and (a hypothetical lexical stratum of) Akan lacks /ə/.

4.2. Local ranking of faithfulness constraints

The ranking of a faithfulness constraint for a particular perceptual feature may depend on the simultaneous presence of other features and on the perceptual events preceding and following that feature:

PARSE (*f*: *v* / *condition* / *environment*): "the value *v* on the perceptual tier *f* in the input is present in the output under a certain *condition* and in a certain *environment*." (38)

Basically, faithfulness constraints are ranked by perceptual *contrast*: constraints that require the faithfulness of strongly distinctive features are ranked higher than constraints for weakly distinctive features, but only in the following case: the *same* replacement is more offensive if the difference between the members of the pair along a certain perceptual dimension is greater, *and everything else is kept equal*; this can lead to a fixed ranking of many pairs of faithfulness constraints.

Along the place dimension, the vowel /i/ has a certain chance, say 10%, of being initially perceived as its perceptual neighbour /y/, In the recognition phase, the listener can correct this misperception, because she has learnt about confusion probabilities (Boersma 1997a, §8.5). Suppose that initial misperceptions are symmetric, i.e., an intended /y/ also has a chance of 10% of being perceived initially as /i/. Thus,

$$P(perc = \text{i} \mid prod = \text{y}) = P(perc = \text{y} \mid prod = \text{i}) = 0.1 \tag{39}$$

If all three high vowels are equally likely to occur in an utterance, the marginal probability of each possible intended production is

$$P(prod = \text{i}) = P(prod = \text{y}) = \tfrac{1}{3} \tag{40}$$

Likewise, the probability that a random utterance is initially categorized as /i/ is

$$P(perc = \text{i}) = \sum_{n=1}^{3} P(perc = \text{i} \mid prod = x_n) \cdot P(prod = x_n) = 90\% \cdot \tfrac{1}{3} + 10\% \cdot \tfrac{1}{3} = \tfrac{1}{3} \tag{41}$$

A table of all these probabilities is

prod↓ perc→	/i/	/y/	/u/	$P(prod = x)$
/i/	0.9	0.1	0	1/3
/y/	0.1	0.8	0.1	1/3
/u/	0	0.1	0.9	1/3
$P(perc = x)$	1/3	1/3	1/3	

(42)

In the recognition phase, the listener can try to reconstruct the speaker's intended utterance by a search for the most likely produced utterance, given the initial perception. For this, she will have to compute the a posteriori probability of every possible produced utterance. For instance, if the listener initially categorizes the utterance, as /i/, the probability that the speaker actually intended /y/ is, with Bayes,

$$P(prod = y \mid perc = i) = \frac{P(perc = i \mid prod = y) \cdot P(prod = y)}{P(perc = i)}$$

(43)

All the a posteriori probabilities are given by

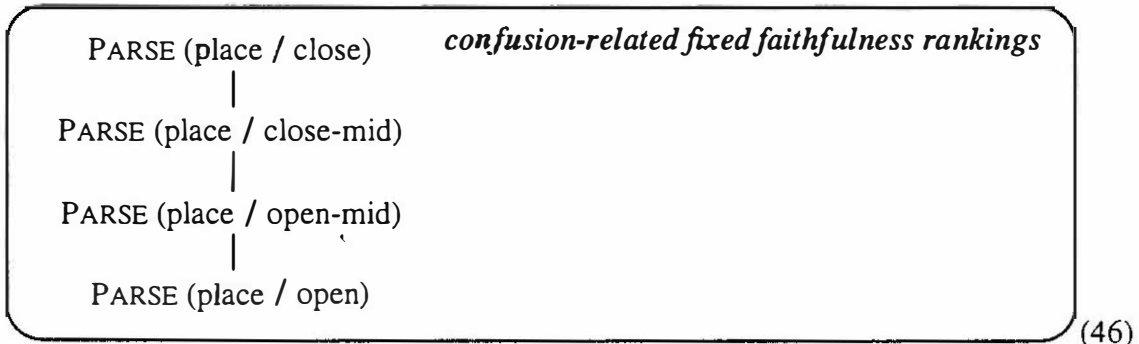| perc↓ prod→	/i/	/y/	/u/
/i/	0.9	0.1	0
/y/	0.1	0.8	0.1
/u/	0	0.1	0.9

(44)

Thus, the probability that a perceived /i/ should be recognized as /y/ is equal to the probability that a perceived /y/ should be recognized as /i/. If we assume that a trained listener is capable of using these numbers in finding the most likely intended utterance (perhaps as a result of the learning algorithm described in Boersma 1997b), we can conclude that it is equally bad for a speaker to pronounce an intended /i/ as [y], as it is for her to pronounce an intended /y/ as [i]: in both cases, the recognition problems for the listener are equally large. Now, because the speaker is also a listener, she can be supposed to "know" this. In a functionally-oriented constraint grammar, this means that the constraints *REPLACE (place: front, central) and *REPLACE (place: central, front) are ranked equally high, or, somewhat loosely, that PARSE (place: x) is ranked equally high for all three place values.

The situation changes if we include the mid vowels in our story. Like the high vowel /i/, the mid vowel /e/ has a certain chance of being perceived as its central counterpart /ø/. But the range of F_2 values decreases as vowels become lower, as illustrated in (1) by the trapezoidal shape of the vowel space. Thus, the confusion probability of /e/ and /ø/ is higher than that of /i/ and /y/, say 20%. The listener has to base her recognition strategy on the following a posteriori probabilities:

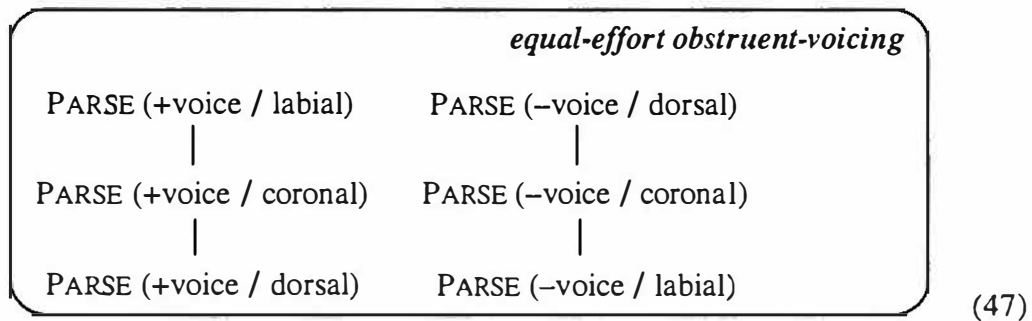| perc↓ prod→	/e/	/ø/	/o/
/e/	0.79	0.20	0.01
/ø/	0.20	0.60	0.20
/o/	0.01	0.20	0.79

(45)

So, under a recognition strategy that maximizes the likelihood of the intended utterance, the chance that the listener successfully corrects a perceived /ø/ into the intended /e/, is larger than the chance that she corrects a perceived /y/ into the intended /i/. This means that a speaker, who knwos this because she is also a listener, can more easily get away with mispronouncing an /e/ as /ø/ than with mispronouncing an /i/ as /y/. Thus, the constraint *REPLACE (place: front, central / close) must outrank *REPLACE (place: front, central / close-mid). Simplifying this with PARSE constraints, we get the following local rankings in the (non-numerical) production grammar:

$$
\begin{array}{c}
\text{PARSE (place / close)} \qquad \textit{confusion-related fixed faithfulness rankings} \\
| \\
\text{PARSE (place / close-mid)} \\
| \\
\text{PARSE (place / open-mid)} \\
| \\
\text{PARSE (place / open)}
\end{array}
$$

(46)

This explains the fact that Frisian shows fewer place contrasts for lower than for higher vowels, but it does not yet explain where the gaps should be.

In our obstruent-voicing example, it will be clear where the gaps are. If the effort that the speaker wants to spend (instead of the perceptual contrast as in §3.1) is taken equal for all three places, the voicing contrast between [g] and [k] will be smaller than the voicing contrast between [d] and [t]. This leads to the following natural constraint ranking:

$$
\begin{array}{cc}
& \textit{equal-effort obstruent-voicing} \\
\text{PARSE (+voice / labial)} & \text{PARSE (−voice / dorsal)} \\
| & | \\
\text{PARSE (+voice / coronal)} & \text{PARSE (−voice / coronal)} \\
| & | \\
\text{PARSE (+voice / dorsal)} & \text{PARSE (−voice / labial)}
\end{array}
$$

(47)

So, keeping the articulatory effort constant, we would have a homogeneous *GESTURE constraint and could get the following constraint interaction:

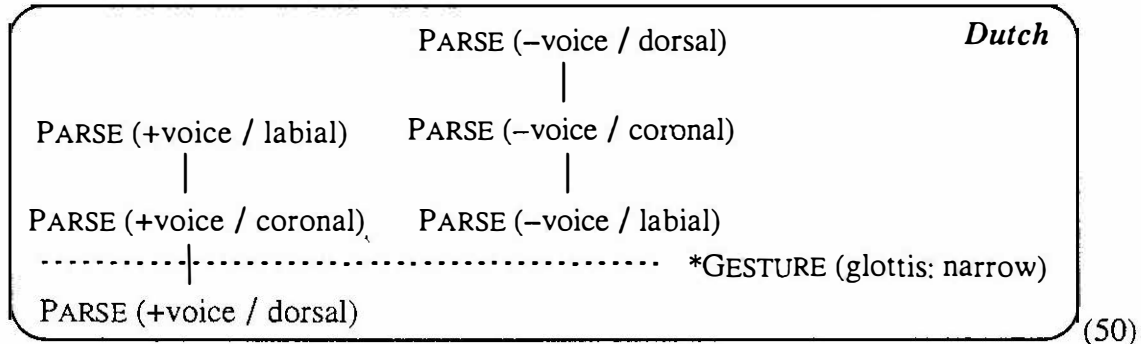| /d/	PARSE (+voice / coronal)	*GESTURE	PARSE (+voice / dorsal)
☞ [d]		*	
[t]	*!		

(48)

Thus, /d/ is parsed faithfully. The dorsal plosive, however, is devoiced:

/g/	PARSE (+voice / coronal)	*GESTURE	PARSE (+voice / dorsal)
[g]		*!	
☞ [k]			*

<div align="right">(49)</div>

The Dutch system could be described as (cf. (34)):

```
                        PARSE (−voice / dorsal)              Dutch
                                  |
  PARSE (+voice / labial)   PARSE (−voice / coronal)
            |                         |
  PARSE (+voice / coronal)   PARSE (−voice / labial)
  ---------+------------------------------------ *GESTURE (glottis: narrow)
  PARSE (+voice / dorsal)
```
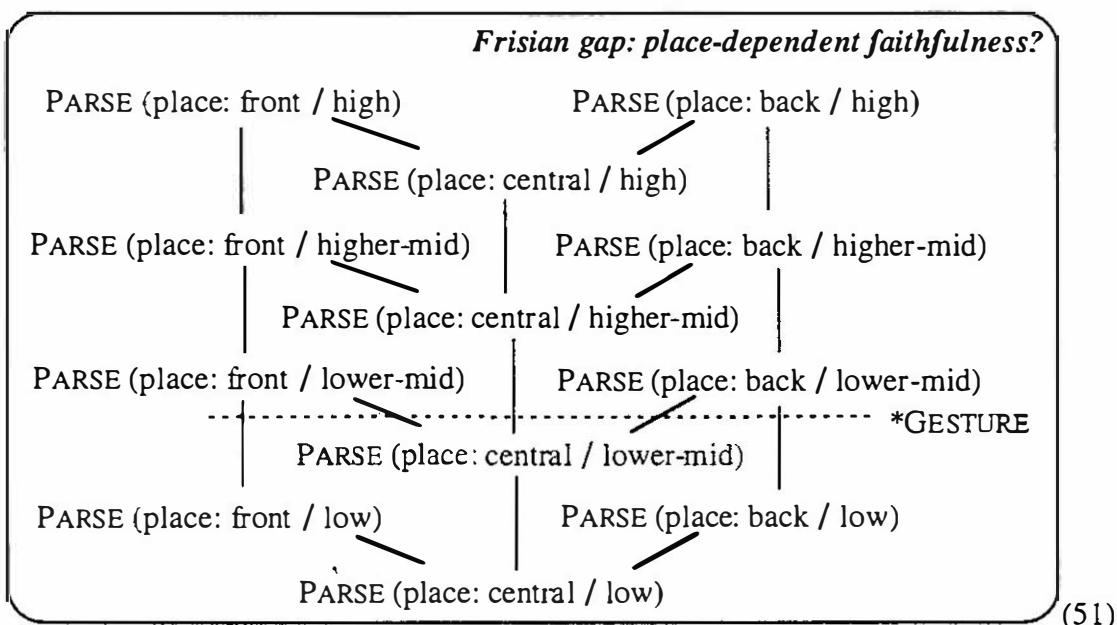
<div align="right">(50)</div>

5. Central gaps

The local rankings of §4 explained why languages tend to have gaps at articulatorily and/or perceptually peripheral locations: the articulatory effort often increases monotonically as we approach a more extreme articulation, and the perceptual contrast often decreases monotonically as a function of another dimension. We will now consider three proposals for the central location of the gap in the Frisian lower-mid vowel system, which has [ɛ] and [ɔ] but lacks [œ].

5.1. An articulatory explanation

The distaste for [œ] could be explained if the effort needed for the rounding gesture is greater than that for the spreading gesture of [ɛ], so that we have the ranking *GESTURE (lips: rounded) >> *GESTURE (lips: spread). There are two problems with this approach. First, this is not a local ranking, because different articulators are involved in lip spreading and rounding. Secondly, the ranking of the PARSE constraints of (46) does not depend on place, so any ranking of the two gestural constraints would treat [ɔ] in the same way as [œ]: either these two sounds are both licensed, or they are both forbidden. The same goes for the fronting gesture of [œ]: if the relevant gestural constraint is ranked higher than PARSE (place / lower-mid), [ɛ] and [œ] are both forbidden; otherwise, they are both allowed. There is no way to derive the correct system with a place-independent PARSE (place).

5.2. A contrast-based explanation

If we make PARSE (place) dependent on place, we may be able to account for the Frisian gap. The following grammar accurately represents the Frisian vowel system:

Frisian gap: place-dependent faithfulness?

PARSE (place: front / high)　　　　PARSE (place: back / high)

　　　PARSE (place: central / high)

PARSE (place: front / higher-mid)　　PARSE (place: back / higher-mid)

　　　PARSE (place: central / higher-mid)

PARSE (place: front / lower-mid)　　PARSE (place: back / lower-mid)

　　　PARSE (place: central / lower-mid)　　　　*GESTURE

PARSE (place: front / low)　　　　PARSE (place: back / low)

　　　PARSE (place: central / low)

(51)

I will now show how the listener's quest for an optimal recognition strategy can give rise to asymmetries in PARSE rankings along a single dimension.

The three place values are not equally well suited for use in a language. Table (42) showed that central values along a perceptual dimension give rise to twice as many confusions as peripheral values. In the history of a language, this could give rise to a pressure towards choosing peripheral values in the process of lexical selection. In our Frisian example, this would take the average confusion between high vowels down from 13.33% in the direction of 10%. For instance, a vocabulary with 40% /i/, 20% /y/, and 40% /u/, would reduce the average confusion probability to 12%, almost half-way the minimum. This lexical shift would reduce the information content per vowel, but not by much: from 1.58 to 1.52 bit. Getting rid of /y/ altogether would reduce the confusion probability to 10%, or 5% after recognition, or 0% after suspension of the central category, but it would also reduce the information content to 1 bit per vowel; this would require a much longer utterance for the same information, violating heavily one of Passy's (1890) functional principles.

In the recognition strategy, the skewed distribution of the place values leads to a shift of the /i/-/y/ discrimination criterion along the continuous F_2 axis in the direction of the centre of the distribution of the /y/ productions (Boersma 1997a: fig. 8.2). If the production distributions are Gaussian, this narrowing of the /y/ category will cause an asymmetry to arise in the confusion probabilities. For instance, the chance that an intended /i/ is categorized as /y/ is 7%, and the chance of an /y/ being categorized as /i/ is 14% (this commonness-related asymmetry is the explanation for the fact that an English intended /θ/ has a larger chance of being perceived as /f/ than the reverse). The perception probabilities of our Frisian example become

$prod\downarrow$　$perc\rightarrow$	/i/	/y/	/u/	$P(prod = x)$
/i/	0.93	0.07	0	0.4
/y/	0.14	0.72	0.14	0.2
/u/	0	0.07	0.93	0.4
$P(perc = x)$	0.4	0.2	0.4	

(52)

The a posteriori probabilities of a certain production given a perceived value are

perc↓ prod→	/i/	/y/	/u/
/i/	0.93	0.07	0
/y/	0.14	0.72	0.14
/u/	0	0.07	0.93

(53)

Thus, an initially perceived /y/ suggests an /i/ recognition candidate more strongly than the reverse. Therefore, it is less bad for recognition to perceive a spurious /y/ than to perceive a spurious /i/. Therefore, it is less bad for the speaker to pronounce an /i/ as [y] than to pronounce an /y/ as [i]. This gives the local ranking *REPLACE (place: front, central) >> *REPLACE (place: central, front), or, more loosely,

PARSE (central)   *commonness-related fixed faithfulness rankings*

PARSE (front)   PARSE (back)

(54)
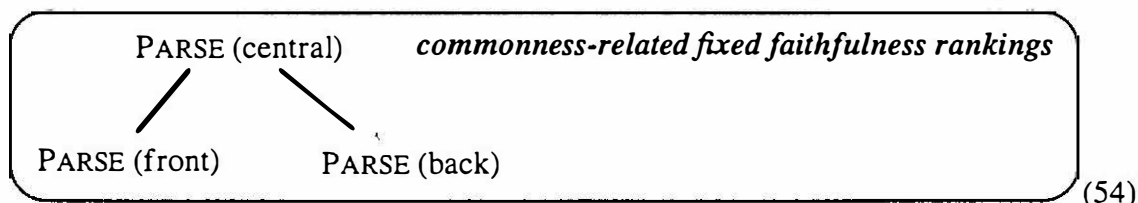
The general empirical prediction from this kind of rankings is that less common perceptual feature values have stronger specifications. For instance, if rounded vowels are less common than unrounded vowels, /i+o/ will have more chance of being assimilated to [yo] than /ye/ to [ie]; if coronals are more common than labials, it is more likely that /n+p/ becomes [mp] than that /m+t/ becomes [nt]; and if nasals are less common than non-nasals, /p+n/ will become [mn] more easily than /m+t/ will become [pt]; no theories of underspecification or privative features are needed to explain these three cross-linguistically well-attested asymmetries.

While the rankings in (54) exhibit a desirable property of phonological processes, they are the reverse of what would be needed to explain the Frisian gap. This becomes dramatically clear when we compare (54) with (51)...

### 5.3. A confusion-based explanation

After discarding two other explanations, we are still left with a plausible explanation for the Frisian gap: the F_2 space for lower-mid vowels is considered too small to easily maintain a three-way contrast. Fewer confusions will arise if the language has an /ɛ/-/ɔ/ contrast than if it has an /œ/-/ɔ/ contrast.

In a production grammar, we could try to describe such a thing by a positive REPLACE (œ, ɛ) constraint (without the asterisk). However, this would effectively introduce an extra level in the phonology! The family *REPLACE (x, y), though formulated as a two-level constraint (a relation between input and output), can actually be seen as an output-only (i.e., one-level) constraint that says "the output should contain no y here". No such move would be possible with a structure-changing positive REPLACE.

Instead of accepting such anti-faithfulness constraints, we should note that the problem of the three-way contrast is in the perception grammar: because of the variation in production and perception, correct categorization is difficult, and not relying on noisy categories will make a better recognition strategy than relying on them. Now suppose that a language has a problematic three-way contrast. The following steps may happen.

Step 1. The middle category gets weaker, i.e. loses some of its lexical occurrences, as described above in §4.2. Variations within and between speakers will lead to random

distributions of the acoustic input to the listener's ear. If the speakers implement three categories with midpoints at [30], [50], and [70] along a perceptual dimension with values from [0] to [100], the inputs to the listener's perception grammar are distributed as follows:
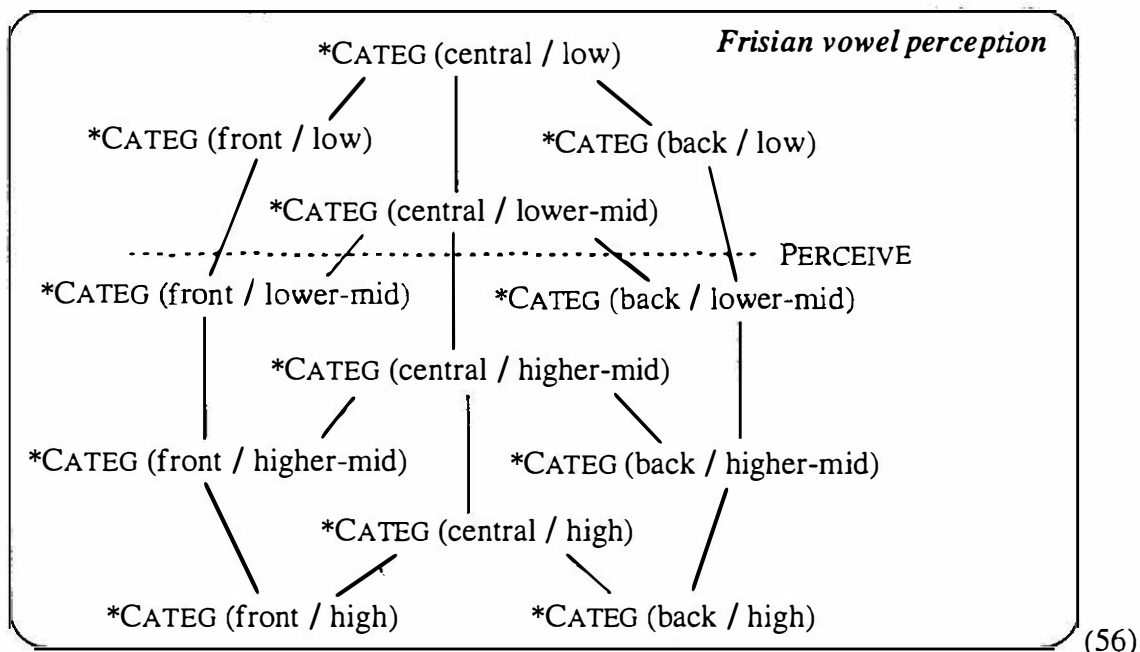
Production distributions of the three categories /30/, /50/, and /70/.

(55)
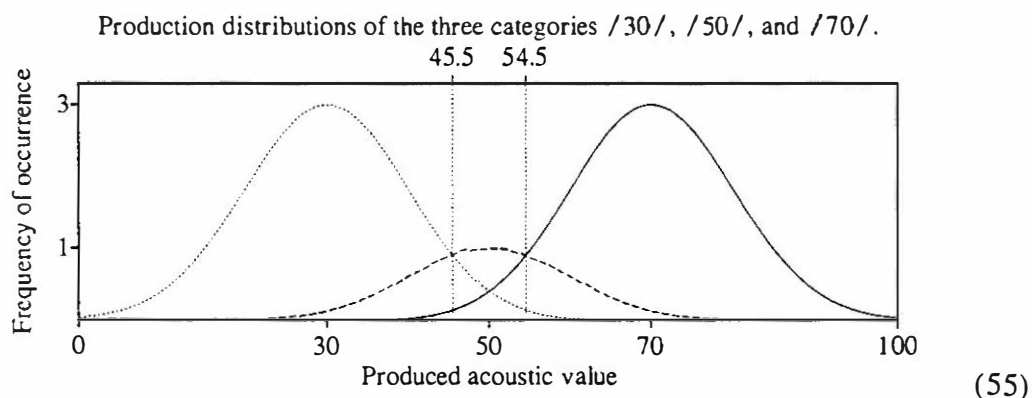
Step 2. The listener will make the fewest mistakes in initial categorization if she uses the criterion of maximum likelihood, i.e., if she chooses the category that maximizes the a posteriori probability (43). For instance, if the acoustic input is [44], an optimal listener will choose the /30/ category because the curve of the distribution of the production of /30/ in figure (55) is above the curve associated with the production of the category /50/, although the value [44] is nearer to the midpoint of the /50/ category than to the midpoint of /30/. Therefore, she will initially categorize all inputs below the criterion [45.5] into the class /30/, all the values between [45.5] and the second criterion [54.5] into the class /50/, and all values above [54.5] into the class /70/. In Boersma (1997b), I showed how an OT listener manages to establish these criteria as a result of an automatic gradual learning process (though she will not actually become a maximum-likelihood listener).

Step 3. If the adjacent categories are close to each other, the criterion shifts can be described as a raising of the *CATEG (central / lower-mid) constraint (Boersma 1997a: fig. 6.4).

Step 4. As the category gets narrower, more utterances of the middle class will be perceived into the neighbouring, broader, classes. Figure (55), for instance, shows that an intended /50/ is perceived as /70/ approximately four times as often as an intended /70/ is perceived as /50/.

Step 5. This will lead to the middle category getting still weaker, i.e., because of the large amount of misperception, the learner will lexicalize many adult /50/ as /70/.

These five steps form a system with positive feedback. Unless checked by requirements of information flow, the process will not stop until all the occurrences of the middle category have vanished, and a newly categorized feature is born. This situation can be described as

The diagram above contains the following constraints arranged hierarchically, titled *Frisian vowel perception*:

*CATEG (central / low)
*CATEG (front / low)
*CATEG (back / low)
*CATEG (central / lower-mid)
*CATEG (front / lower-mid)
*CATEG (back / lower-mid)
PERCEIVE
*CATEG (central / higher-mid)
*CATEG (front / higher-mid)
*CATEG (back / higher-mid)
*CATEG (central / high)
*CATEG (front / high)
*CATEG (back / high)

(56)

The ranking of the four constraints above PERCEIVE means that lower mid vowels cannot be categorized as central, and that the low vowel is not categorized at all along the F_2 dimension. Thus, the gaps in inventory (1) are the result of limitations of categorization, and no constraints against [œ] have to be present in the production grammar, since a hypothetical underlying |œ| could never surface faithfully: even if an underlying |œ| is pronounced as [œ], it will be perceived by the speaker herself as /ɛ/ or /ɔ/; see figure (19) for the role of the perception grammar in the evaluation of faithfulness.

If we vary the ranking of PERCEIVE with respect to the *CATEG (back) family, we see that the following four systems are possible with three heights for non-low vowels:

i		u	i	y	u	i	y	u	i	y	u
e		o	e		o	e	ø	o	e	ø	o
ɛ		ɔ	ɛ		ɔ	ɛ		ɔ	ɛ	œ	ɔ

(57)

Precisely these four systems are fairly common: apparently, grammars are allowed a considerable degree of freedom in ranking the PERCEIVE constraints relative to the *CATEG constraints, but no freedom at all to reverse the universal ranking within the *CATEG (back) family; this gives strong evidence for the local-ranking hypothesis. The first system is more common than the fourth: global (cross-dimensional) contrast measures may predict which of these systems are the most common ones, but cannot preclude any of them beforehand; the local-ranking principle ensures that.

Finally, note that our theory not only tells us *which* sounds there are in an inventory, but also how many, given the number of low *GESTURE and *CATEG constraints; in previous 'phonetic' accounts (§ 1.5-6), this number used to be *posited*.

6. Conclusion

This paper showed that a combination of functional principles, interacting in the production and perception grammars under the regime of Optimality Theory, allows accurate explanation of the symmetries and gaps in vowel and consonant systems.

Symmetry results because the listener interprets a finite number of categories along each of the language's perceptual dimensions, and because the speaker learns a finite number of articulatory tricks and their combinations.

Gaps are explained by local rankings of functional constraints:

(a)   Local rankings of *GESTURE explain articulatorily peripheral gaps.
(b)   Local rankings of PARSE explain perceptually peripheral gaps.
(c)   Local rankings of *CATEG explain perceptually central gaps.

The global optimization criterion of maximal dispersion is a derivative of these local phenomena.

To explain inventories, we need assume no innate features, feature values, or constraints.

7. Postscript: the role of categorization in the production grammar

Superficially, only gestural and faithfulness constraints seem to play a role in the production grammar; categorization constraints are only made explicit in the perception grammar. This would suggest that I invoked the categorization constraints only in order to account for inventory structure, so I would be open to one of the same criticisms that I voiced on Flemming's MINDIST and MAINTAIN-N-CONTRASTS proposals in §1.8.

However, I will show that the categorization constraints actually play an essential role in the production grammar. To see this, we go back to our Italian mid-vowel-merger rule. In stressed position, we have a seven-vowel system, in unstressed position a five-vowel system. This reduced unstressed-vowel system can be caused by a dependence of *CATEG on stress. After all, Italian unstressed vowels are much shorter and less loud than stressed vowels, two properties that make them less resistant against background noise and cause confusion probabilities to be much greater for unstressed than for stressed vowels (with the same distance along the F_1 axis). With the usual dependence of the ranking of *CATEG on confusion probabilities, this may lead to different categorizations of the height dimension for stressed and unstressed vowels. While the four categories for stressed vowels were "low", "lower-mid", "higher-mid", and "high", we are left with three categories for unstressed vowels, which we will call "low", "mid", and high.

Now consider the categorization of the four "stressed" vowel qualities in unstressed position. If the F_1 space is equally large for stressed and unstressed vowels and the categories are equally wide, the F_1 space will be categorized as:

(58)
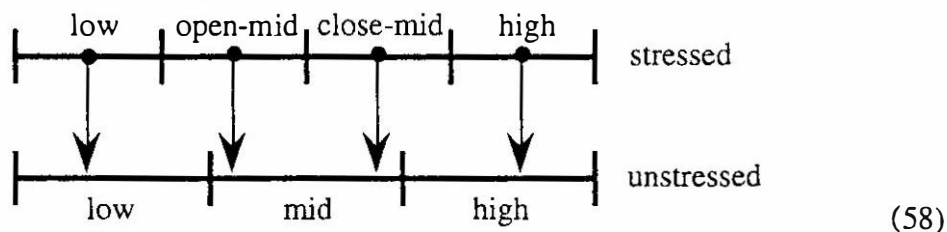
As we see in the figure (suggested by the re-categorization of the midpoints of the "stressed" categories), all of the "low" values map to "low" in the "unstressed" perception grammar, all of "high" maps to "high", and *most* of the open-mid and close-mid realizations will be categorized as "mid". So we see that the names of the three "unstressed" categories are appropriate and that Italian follows the default strategy of the merger of the two central categories: if the speaker pronounces unstressed [a ɛ e i], the listener will initially categorize this as /a "e" "e" i/, where

/"e"/ is a vowel halfway between /a/ and /i/. In finding the underlying form, the listener will have to reconstruct |ɛ| or |e| as appropriate, with the help of the biases of lexical access, syntax, and meaning.

We can now see how the production grammar causes an underlying |ɛ| to be pronounced as [e] in unstressed position:

Input: |ɛ| (unstressed as in |spɛnd+i'amo|)	*REPLACE (open-mid, high)	*REPLACE (open-mid, low)	*REPLACE (open-mid, mid)	*GESTURE (jaw: open)	*GESTURE (jaw: half open)
[a] /a/		*!		*	
[ɛ] /"e"/			*	*!	
☞ [e] /"e"/			*		*
[i] /i/	*!				

(59)

The ranking of the *REPLACE (*underlying category, surface category*) constraints depends on the distance between the midpoints of the underlying and surface categories; as we see in (58), the distance from the "open-mid" underlying category is smallest for the "mid" surface category (1/8 of the scale, as opposed to 5/24 for the "low" surface category). This local ranking of *REPLACE invalidates the candidates [a] and [i][4]. The two remaining candidates [ɛ] and [e] are both perceived as /"e"/, so they both violate the same faithfulness constraint. The buck is passed to the gestural constraints, specifically, to any small differences in jaw-opening effort. As most of the surrounding consonants are usually pronounced with a rather closed jaw, this effort will be larger for [ɛ] than for [e], giving a local *GESTURE ranking that causes the easier [e] candidate to win.

8. References

Archangeli, Diana (1984). *Underspecification in Yawelmani Phonology and Morphology*. Doctoral thesis, MIT, Cambridge. [Garland Press, New York, 1988]

Archangeli, Diana (1988). "Aspects of underspecification theory", *Phonology* 5, 183-207.

Archangeli, Diana and Douglas Pulleyblank (1994). *Grounded Phonology*. MIT Press, Cambridge.

Boë, Louis-Jean, Pascal Perrier, Bernard Guérin, and Jean-Luc Schwartz (1989). "Maximal vowel space", *Proceedings Eurospeech '89* 2, 281-284.

Boë, Louis-Jean, Jean-Luc Schwartz, and Nathalie Vallée (1994). "The prediction of vowel systems: perceptual contrast and stability", in Eric Keller (ed.), *Fundamentals of Speech Synthesis and Speech Recognition*, Wiley, London & New York, 185-213. .

Boersma, Paul (1997a). "The elements of Functional Phonology", ms. University of Amsterdam. [Rutgers Optimality Archive **173**, http://ruccs.rutgers.edu/roa.html]

Boersma, Paul (1997b). "How we learn variation, optionality, and probability", *this volume*, 43-58 [an earlier version is Rutgers Optimality Archive **221**, http://ruccs.rutgers.edu/roa.html]

Boersma, Paul (to appear). "Learning a grammar in Functional Phonology", in Joost Dekkers, Frank van der Leeuw, and Jeroen van de Weijer (eds.), *Optimality Theory: Phonology, Syntax, and Acquisition*.

---

[4] Actually, /a/ and /"e"/ are on opposing sides of the specified value, so local rankability can be questioned: replacement of /ɛ/ with /a/ would also be thinkable (in another language), though Italian follows the globally determined default.

Bosch, Louis ten (1991). *On the Structure of Vowel Systems. Aspects of an extended vowel model using effort and contrast.* Doctoral thesis, Universiteit van Amsterdam.

Chomsky, Noam and Morris Halle (1968). *The Sound Pattern of English.* Harper and Row, New York.

Crothers, John (1978). "Typology and universals of vowel systems", in Joseph H. Greenberg, Charles A. Ferguson, and Edith A. Moravcsik (eds.), *Universals of Human Language*, Stanford University Press, Vol. 2, 93-152.

Dols, Willy (1944). *Sittardse diftongering.* [Alberts' Drukkerijen, Sittard 1953]

Flemming, Edward (1995). *Auditory Representations in Phonology.* Doctoral thesis, University of California, Los Angeles.

Hammarström, Göran (1973). "Generative phonology: a critical appraisal", *Phonetica* 27, 157-184.

Kats, J.C.P. (1985). *Remunjs waordebook.* H. van der Marck, Roermond.

Kawasaki, Haruko (1982). *An Acoustical Basis for Universal Constraints on Sound Sequences.* Doctoral thesis, University of California, Berkeley.

Ladefoged, Peter (1971). *Preliminaries to Linguistic Phonetics.* University of Chicago Press.

Leoni, F.A., F. Cutugno, and R. Savy (1995). "The vowel system of Italian connected speech". *XIIIth International Congress of Phonetic Sciences* 4, 396-399.

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

Lindau, Mona (1975). *[Features] for vowels.* UCLA Working Papers in Phonetics 30.

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

Lindblom, Björn (1990). "Models of phonetic variation and selection", *Phonetic Experimental Research, Institute of Linguistics, University of Stockholm* XI, 65-100.

Maddieson, Ian (1984). *Patterns of Sounds.* Cambridge University Press.

Ohala, John J. and Carol J. Riordan (1979). "Passive vocal tract enlargement during voiced stops", *Journal of the Acoustical Society of America* 65 (S1), S23.

Passy, Paul (1890). *Etude sur les changements phonétiques et leurs caractères généraux.* Librairie Firmin - Didot, Paris.

Peeters, Frans (1951). *Het klankkarakter van het Venloos.* Doctoral thesis Katholieke Universiteit Nijmegen, Centrale Drukkerij, Nijmegen.

Prince, Alan and Paul Smolensky (1993). *Optimality Theory: Constraint Interaction in Generative Grammar.* Rutgers University Center for Cognitive Science Technical Report 2.

Pulleyblank, Douglas (1994). "Neutral vowels in Optimality Theory: a comparison of Yoruba and Wolof", ms. University of British Columbia. [to appear in *Canadian Journal of Linguistics*]

Schwartz, Jean-Luc, Louis-Jean Boë, Pascal Perrier, Bernard Guérin, and Pierre Escudier (1989). "Perceptual contrast and stability in vowel systems: a 3-D simulation study", *Proceedings Eurospeech '89* 2, 63-66.

Schwartz, Jean-Luc, Louis-Jean Boë, and Nathalie Vallée (1995). "Testing the dispersion-focalization theory: phase spaces for vowel systems", *Proceedings XIIIth International Congress of Phonetic Sciences* 1, 412-415.

Schwartz, Jean-Luc, Louis-Jean Boë, Nathalie Vallée, and Christian Abry (1997). "The Dispersion-Focalization Theory of vowel systems", *Journal of Phonetics* 25, 255-286.

Steriade, Donca (1987). "Redundant values", in A. Bosch, B. Need, and E. Schiller (eds.), *Papers from the Parasession on Autosegmental and Metrical Phonology*, Chicago Linguistic Society, 339-362.

Stevens, Kenneth N. (1989). "On the quantal nature of speech", *Journal of Phonetics* 17, 3-45.

Stevens, Kenneth N. and Keyser, Samuel J. (1989). "Primary features and their enhancement in consonants", *Language* 65, 81-106.

Stevens, Kenneth N., Samuel Jay Keyser, and Haruko Kawasaki (1986). "Toward a phonetic and phonological theory of redundant features", in Joseph S. Perkell & Dennis H. Klatt (eds.), *Invariance and Variability in Speech Processes*, Lawrence Erlbaum, Hillsdale, 426-449.

Tans, J.G.H. (1938). *Isoglossen rond Maastricht.* Doctoral thesis, Katholieke Universiteit Nijmegen.

Tesar, Bruce and Paul Smolensky (1993). "The learnability of Optimality Theory: an algorithm and some basic complexity results", ms. Department of Computer Science & Institute of Cognitive Science, University of Colorado at Boulder. [Rutgers Optimality Archive 2]

Vallée, Nathalie (1994). *Systèmes vocaliques: de la typologie aux prédictions.* Doctoral thesis, Institut de la Communication Parlée, Université Stendhal, Grenoble.