Nasal harmony in functional phonology[1]

Paul Boersma, University of Amsterdam, http://www.fon.hum.uva.nl/paul/

April 28, 2000

This paper will show that a theory of phonology that distinguishes between articulatory and perceptual representations and processes (Boersma 1998) accounts for the typological facts of nasal harmony more succinctly and with fewer assumptions for innate substantive devices, than theories that maintain a single kind of phonological features and a single phonological grammar, like the theories applied to nasal harmony by Piggott (1992), Piggott & Van der Hulst (1997), and Walker (1998).

1. Representations: the case of the nasal glottal stop

Nasal harmony is one of the areas of phonology in which articulatory and perceptual representations bear no one-to-one relationship, so that a failure to distinguish between them is bound to lead to confusion and controversy. The single issue discussed most often in the literature is the representation of the sequence [ãʔ̃ã], which can arise from the rightward spreading of nasality from a nasal consonant through a glottal stop, as in Sundanese [nũʔ̃ũs] ‘to dry’ (Cohn 1990: 52). All writers agree that the velum, which must be down during the two instances of [ũ], is also down throughout the glottal stop, but that no nasality is present acoustically during the glottal closure because this closure causes the nasal airflow to fall to zero (Piggott 1992: 39; Cohn 1993: 347; Walker & Pullum 1999: 766). This section will address the articulatory and perceptual representations of the nasal glottal stop, and show why both of these representations are needed in a phonological account of spreading of nasality through glottal stops.

1.1 Articulatory representations

The articulations of [ãʔ̃ã] are shown in (1).

(1) *Articulatory score for* [ãʔ̃ã]

glottis:	adducted	constricted	adducted
velum:	lowered		
lips:	open		
pharynx:	narrow		

[1] An earlier version of this paper was presented at the 4th HIL Phonology Conference, Leiden, January 30, 1999. Thanks to René Kager and Rachel Walker for helpful comments on that version.

In this articulatory score, time runs from left to right, and the tiers are time-aligned with each other. The only two articulatory ***contours*** in this score are the closing and opening of the glottis. These can be regarded as the results of separate closing and opening gestures, or as the results of a composite closing-and-opening gesture. All the other articulators are stationary. The velum stays down, so that the velopharyngeal port stays open, and the supralaryngeal cavities stay in a shape appropriate for [a] (lips open, pharynx narrowed, jaw lowered). The expiratory actions of the lungs are not shown. For practical reasons we can abbreviate articulatory representations as a sequence of IPA symbols within square brackets, thus writing (1) simply as [ãʔ̃ã], with a nasalization symbol above the glottal stop in order to express velum lowering. With an intermediate degree of abbreviation, we could depict this articulation as a tree with links from the feature value [lowered velum] to the three segments involved:

(2) *Articulatory tree for* [ãʔ̃ã]

[*velum*]: lowered

a ʔ a

The dotted line expresses the fact that articulator positions like [lowered velum] must be defined on a continuous stretch in time without any intervening gaps.

1.2 Acoustic representations

The main uncategorized perceptual results of the articulation (1) are summarized in (3).

(3) *Acoustic events as a result of* [ãʔ̃ã]

silence:		+	
place:			gl.bu.
voice:	+		+
F1:	open		open
nasal:	+		+
	ã	_	ˀ ã

Again, time runs from left to right on each tier, and all of these perceptual tiers are time-aligned with each other and with the articulatory score in (1). On the perceptual place tier, we see the abbreviation “gl.bu.”, which stands for a glottal release burst; we can note that [glottal] is a value on the perceptual place tier, just as bilabial, apico-dental, pharyngeal, etc., since all of these are associated with their own acoustic spectral characteristics. “F1” stands for the first formant, which is the acoustic cue that we perceive as vowel height. The feature /nasal/ refers to audible resonances in the nasal tract and is present only during the vowels, not during glottal closure. Below the perceptual tiers, we see an IPA notation of the auditory states and events, where “_” denotes silence and “ˀ” the glottal release burst. The acoustics of (2), then, can be abbreviated with the ***microscopic transcription*** [[ã _ˀã]] (Boersma 1998: 30).

1.3 Perceptual representations

What the listener perceives in [[ã _ˀã]] is not just this sequence of acoustic states and events. She assigns hidden structures to it, like segments, syllables, and feet, in a language-specific way, thus retrieving from the raw acoustic signal a discrete phonological code that allows her to compare the utterance with the entries in her lexicon. One of these abstractions may be the *segmental* level, and we could write the utterance on this level as /ãʔã/, without a nasalization symbol above the glottal stop in order to express the absence of nasality there. As a representation of perceived nasality, however, the shorthand between the slashes is ambiguous: do the two nasalization symbols refer to the same feature value, or to two separate feature values? As an example of the former, the tree in (3) shows a truly ‘linear’ segmental representation:

(4) *Segmental nasality*

```
/nasal/:  +  –  +
          |  |  |
          a  ʔ  a
```

But nothing in what we know about human perception tells us that perceived entities should be continuous: if we look at a car behind a lamp-post, we perceive a single car, not two halves. Likewise, the two nasality cues in [[ã _ˀã]] may well be perceived as a single nasal feature value:

(5) *Suprasegmental nasality*

```
/nasal/:   +  –
          / \/
         a  ʔ  a
```

Which of the two representations (4) or (5) applies in the language at hand, depends on the domain on which nasality tends to be specified. In a language like French, where every vowel can be nasal or non-nasal more or less regardless of the nasality of the adjacent segments, (4) is likely to be the best choice for the listener. In a language like Guaraní, where words have either only nasal or only non-nasal vowels, (5) is more appropriate, because it creates a shorter code.

1.4 Hybrid representations

I think we should stop here and regard (1), (4), and (5) as the representations relevant for phonology. However, generative theories of phonological representations have not stopped here, and have always advocated the existence of a single cognitive feature [nasal], which is supposed to have articulatory as well as perceptual correlates. However, this position becomes problematic in cases where one of these two correlates is absent, as in the case of [ãʔ̃ã] /ãʔã/. An amusing controversy arises in the discussion about whether the glottal stop in [ãʔ̃ã] /ãʔã/ is phonologically nasal or not. Cohn (1993: 349) considers this glottal stop *phonetically nasal*, because the velum is lowered, but *phonologically non-nasal*, since it is *transparent* to nasal spreading, lacking a supralaryngeal node and therefore a nasality node. Piggott (1992: 39), by contrast, calls this glottal stop *phonetically non-nasal*, because there is no nasal airflow, but *phonologically nasal*, because it must be considered a *target* (undergoer) of nasal spreading.

Cohn appears to have an articulatory view of phonetics and a perceptual view of phonology, whereas Piggott seems to have a perceptual view of phonetics and an articulatory view of phonology. The cause of this confusion is that Cohn and Piggott share the standpoint that there must be a single phonological feature [nasal]. As soon as we accept that both the articulatory feature [lowered velum] and the perceptual feature /nasal/ are phonologically active, the controversy vanishes: the glottal stop under discussion is articulatorily ‘nasal’, perceptually ‘non-nasal’, and it is articulatorily an undergoer of ‘nasal’ spreading and perceptually transparent to it. A formal theory about how phonology revolves around these two kinds of representations, is presented in Boersma (1998), and I will presently describe how it works out for the case of nasal harmony.

2 Processes

2.1. The grammar model of functional phonology

Since Cohn and Piggott share the view that phonetic implementation follows the phonology, their opposing views of phonological and phonetic representation return in their views on the derivation of the phonetic form:

(6) *Generative views of spreading nasality through a glottal stop*

	Underlying form:		Phonological form:		Phonetic form:
Cohn:	|naʔa|	→	/nãʔã/	→	[nãʔ̃ã]
Piggott:	|naʔa|	→	/nãʔ̃ã/	→	[nãʔã]

In both of these conflicting derivations, the first arrow denotes the phonological process of nasal spreading, and the second arrow denotes phonetic implementation. If we distinguish between articulation and perception, the derivation becomes very different, the most important difference being the reversal of the order of the phonological and phonetic surface forms:

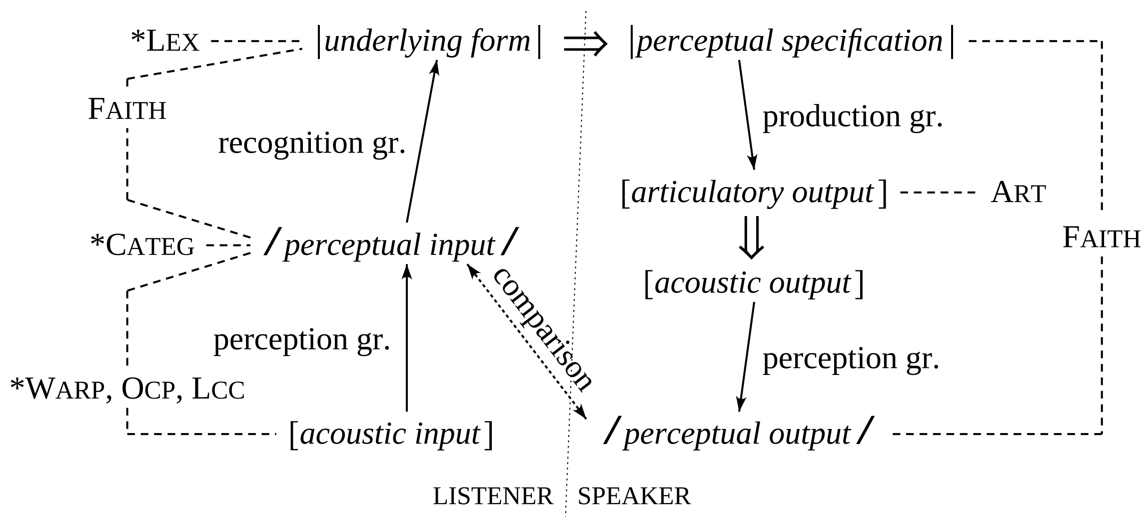
(7) *Functional view of spreading nasality through a glottal stop*

Underlying form:		Articulatory form:		Perceptual form:
|naʔa|	→	[nãʔ̃ã]	→	/nãʔã/

The first arrow denotes the phonology and phonetic implementation, which are not seen as separate modules, and the second arrow denotes the speaker’s perception process, whose task it is to convert a raw articulatory (acoustic, ‘phonetic’) form into a more discrete perceptual (‘phonological’) form. In this ***specification-articulation-perception triad*** (Boersma 1998: ch. 1), the underlying form is a perceptual specification (Saussure 1916), and it is the task of the grammar to choose an implementation that strikes the best balance between the functional principles of ***minimization of articulatory effort***, which favours easy articulatory forms, and ***minimization of perceptual confusion***, which favours perceptual similarity between the underlying specification and the perceptual surface form. The order in (7) expresses the idea that the speech production process, which will have to choose an articulatory implementation, is ultimately perception-oriented, like all human behaviour (Powers 1973).

Figure (8), taken from Boersma (1998: ch. 6), shows that the theory of functional phonology incorporates the specification-articulation-perception triad as the production part of a more complete grammar model that also includes comprehension.

(8) *The grammar model of functional phonology*



The production grammar can be identified with what we know as 'the' grammar from most theories of phonology, but its formalization is quite different (§2.2). The perception grammar occurs twice in (8); its task is to turn raw acoustic forms, which are the automatic results of the articulations of the speaker or others, into more discrete perceptual representations, and it performs sequential abstraction (§2.3). The recognition grammar handles the interaction between phonology and semantics in the process of lexical access (Boersma 1999b).

2.2 The production grammar and its local rankings

The production grammar shown in Figure (8) is modelled as an Optimality-Theoretic grammar consisting of acquired articulatory (ART) and faithfulness (FAITH) constraints. Its input is the perceptual specification; its candidates are articulatory implementations paired with perceptual results:

(9) *Evaluation of articulatory candidates and their perceptual results*

|*spec*|	A	B
☞ [art_1] → /$perc_1$/		*
[art_2] → /$perc_2$/	*!	

Note the intertwining of the production and perception grammars: the arrows in each candidate cell denote the workings of the perception grammar, on which the evaluation procedure in the production grammar can have no influence. The faithfulness constraints, which compare the output of the perception grammar with the underlying specification, are elements of the production grammar.

The production grammar contains a number of ***articulatory constraints*** (depicted as ART in the figure), which evaluate aspects of the innate functional principle of ***minimization of effort***, and work directly on each articulatory output candidate (Boersma 1998: ch. 7). Such a constraint enters the production grammar (initially high-ranked) as soon as the learner discovers the relation between an articulatory gesture and its perceptual result. The most typical example is:

(10) *GESTURE (*articulator*: *gesture* / *distance*, *duration*, *precision*, *velocity*):

“Do not perform a certain *gesture* with a certain *articulator*, along a certain *distance*, for a certain *duration*, and with a certain *precision* and *velocity*.”

Other articulatory constraints militate against the synchronization of two gestures or against the coordination of two simultaneous or sequential gestures.

The ***local-ranking principle*** (Boersma 1998) restricts the typologically possible languages by assuming that pairs of constraints can be ranked in a universal manner if they differ in a single argument or condition, and that they can be ranked in a language-specific manner otherwise. Thus, articulatory constraints can be locally ranked according to articulatory effort, e.g. (10) is ranked higher if the *distance*, *duration*, *precision*, or *velocity* is greater, and everything else stays equal. Otherwise, the rankings are largely language-specific: a *global* measure of articulatory effort (e.g. Boersma 1998: eq. 7.4) can only account for cross-linguistic statistical tendencies.

The production grammar also contains a number of ***faithfulness constraints*** (depicted as FAITH in the figure), which evaluate aspects of the innate functional principle of ***minimization of confusion*** indirectly in an evaluation of aspects of the similarity between the perceptual result of each candidate and the underlying perceptual specification (Boersma 1998: ch. 9). Such a constraint enters the production grammar (initially low-ranked) as soon as the learner’s perception grammar has supplied her with a perceptual category. The most typical example is:

(11) *REPLACE (*feature*: $value_1$, $value_2$ / *condition* / *left-env* _ *right-env*):

“Do not replace a specified value ($value_1$) on a perceptual tier (*feature*) with a different value ($value_2$), under a certain *condition* and in the environment between *left-env* and *right-env*.”

Other faithfulness constraints militate against insertion of surface material (*INSERT) and deletion of underlying material (*DELETE), or against the loss of specified simultaneous and sequential relations between features (*DELETEPATH, *SHIFT).

Faithfulness constraints can be locally ranked according to perceptual confusion, e.g. (11) is ranked higher if $value_1$ and $value_2$ are further apart or if the condition or the environment contribute to a smaller amount of confusion, and everything else stays equal. Otherwise, the rankings are largely language-specific: a *global* measure of perceptual confusion (e.g. Boersma 1998: eq. 4.24) can only account for cross-linguistic statistical tendencies.

As an example, tableau (12) tells us why the ‘linear’ segmental specification |ãʔ̃ã| will probably be implemented as [ãʔã], which the speaker will perceive as the unfaithful form /ãʔã/.

(12) *Realization of an underlying nasal glottal stop*

|ãʔ̃ã|	*REPLACE (nasal: +, –)	*GESTURE (velum: up & down)
☞ [ãʔ̃ã] → /ãʔã/	*	
[ãʔã] → /ãʔã/	*	*!
[aʔa] → /aʔa/	**!*	

In this tableau, we see that both kinds of surface representations are needed in phonology: the faithfulness constraint evaluates the perceptual representations between the slashes (by comparing them to the perceptual specification between the pipes), and the articulatory constraint evaluates the articulatory representations between the square brackets. The second candidate, with a velum raising during the glottal closure, will always lose to the first candidate, regardless of the constraint ranking: we can see that if multiple articulatory candidates yield the same perceptual result, their patterns of faithfulness violations will be identical and their relative harmonicity will be determined solely by the articulatory constraints. The third candidate, with a raised velum throughout, will also lose to the first regardless of the constraint ranking.

Tableau (12) contains an underlying representation that is unlikely (though not impossible) from the point of view of phonological acquisition. Since the learner will hear only the surface form /ãʔã/, she will probable store it in her lexicon as |ãʔã|, at least if nasality is segmental in her language. This is an automatic result of the minimization of faithfulness violations in the recognition grammar (Boersma 1999b), and corresponds to Prince & Smolensky's (1993) idea of Lexicon Optimization. A more common tableau would therefore be (13).

(13) *Realization of an underlyingly non-nasal glottal stop in a nasal environment*

|ãʔã|	*REPLACE (nasal: +, –)	*GESTURE (velum: up & down)
[ãʔã] → /ãʔã/		*!
☞ [ãʔ̃ã] → /ãʔã/		
[aʔa] → /aʔa/	*!*	

This articulatorily perfect result is perfectly faithful to the specification as well.

In many languages, nasality will be autosegmental, so that an underlying representation like (5) is more appropriate than |ãʔã|. In such a case, the tableaus become more complicated (§4).

It is illustrative to compare tableau (12) with how generative phonology would have to handle the underlying form |ãʔ̃ã| if it chooses [ãʔã] as the surface form:

(14) *Realization of an underlyingly non-nasal glottal stop in a nasal environment*

|ãʔ̃ã|	*NASGLOTTALSTOP	IDENT-IO (nasal)
ãʔ̃ã	*!	
☞ ãʔã		*

The faithfulness constraint IDENT-IO (nasal) is the ‘hybrid’ counterpart of *REPLACE (nasal). The problem with tableau (14) is that it needs the markedness constraint *NASGLOTTALSTOP, which has to rule out nasal glottal stops in the surface form. Functionally, it expresses quite indirectly an interaction between articulation and perception (“it is difficult to find an articulation that will lead to a simultaneous perception of nasality and a glottal stop”), whereas in (12) the perceptual and articulatory drives have been separated out. The largest difference for a theory of UG is that the generative approach has to propose large numbers of innate substantive constraints such as *NASGLOTTALSTOP, whereas the functional approach only proposes a few innate templatic constraint families like *REPLACE and *GESTURE, whose substantive content (e.g. the features /nasal/ and [velum]) can be filled in during the acquisition process (Boersma 1998: ch.14; to appear a; to appear b). In the case at hand, *NASGLOTTALSTOP could become superfluous if the ‘hybrid’ surface form is taken as [ãʔ̃ã], as in Walker (1998), but that severs the connection between faithfulness and perception, since IDENT-IO (nasal) will now be satisfied if the velum is lowered.

2.3 The perception grammar and its local and global rankings

In (8), we see that the perception grammar performs several functions (Boersma 1999a): for the speaker, it produces a representation from which she can evaluate faithfulness; for the listener, it produces a perceptual representation of the speech of another person, as an input to the recognition system that will ultimately lead to comprehension; for the learner, it produces perceptual representations of her own speech and of the speech of others, so that the learner can gradually learn to speak in the same way as others do.

The perception grammar is implemented as an Optimality-Theoretic grammar consisting of acquired categorization constraints. Its input is a continuous acoustic signal; its candidates are discrete perceptual representations:

(15) *Evaluation of perceptual candidates*

[[acoustics]]	A	B
☞ $/perc_1/$		*
$/perc_2/$	*!	

The perception grammar generates covert structure: it contains a number of templatic constraint families that help to reduce the raw acoustic material in a language-specific way to a more discrete abstract representation that can be related to the necessarily discrete phonological representations in the lexicon. The lowest-level action of the perception grammar is the conversion of continuous acoustic cues into discrete perceptual feature values with the help of categorization constraints (the *CATEG and *WARP families shown in the figure, see Boersma 1998: ch. 8). For instance, [[m]] will in most languages be mapped on the value /labial/ on the perceptual place tier and on the value /+/ on the perceptual nasality tier. The perception grammar will also generate a ***link of simultaneity*** or ***path*** (in the terminology of Archangeli & Pulleyblank 1994) between these two values, i.e. the value /labial & +/ on the place & nasal tier. Other constraints in the perception grammar control the abstraction of simultaneous and sequential cooccurrence. For instance, [[m]] may be perceived as the single 'segmental' percept /labial nasal/, if the two feature values frequently occur simultaneously in the language at hand. For the subject of nasal harmony, the concept of *sequential* abstraction is more important, and I will spell out its formalization.

A pair of constraint families in the perception grammar together determine the abstraction of sequential acoustic cues into a single percept:

(16) OCP (*f*: *x*; cue_1 | *m* | cue_2)

"A sequence of acoustic cues cue_1 and cue_2 with intervening material *m* is heard as a single value *x* on the perceptual tier *f*."

(17) LCC (*f*: *x*; cue_1 | *m* | cue_2)

"A sequence of acoustic cues cue_1 and cue_2 with intervening material *m* is not heard as a single value *x* on the perceptual tier *f*."

These names are abbreviations of the terms Obligatory Contour Principle and Line Crossing Constraint known from generative phonology as inviolable constraints on representations. In functional phonology, they are violable constraints on the perceptual representation that is derived from an acoustic signal by the perception grammar. For instance, (4) violates OCP (nasal: +; Ṽ | ʔ | Ṽ), and (5) violates LCC (nasal: +; Ṽ | ʔ | Ṽ). The existence of these constraints is a result of general properties of human perception: if we see an object partly obscured by other objects, we can still sometimes perceive the various visible parts together as a single object. Thus, OCP and LCC control the construction of higher-level sequential units such as segments, autosegments and syllables (Boersma 1998: chs. 12, 17; Boersma 1999a).

We can identify some universal local rankings of OCP and LCC:

(18) *Local rankings of OCP*

a. Higher if the sequential combination of cue_1 and cue_2 is more common.
b. Lower if there is more intervening material.

(19) *Local rankings of LCC*

a. Lower if the sequential combination of cue_1 and cue_2 is more common.
b. Higher if there is more intervening material.

For instance, the Portuguese articulation [sɐβɐ̃ũ], which stands for a sequence of contracting lungs, tongue grooving, half-open jaw, lip approximation & opening, a velum lowering gesture, and lip rounding. The acoustic result will be [[sɐβɐ̃ũ]], which stands for sibilant noise, mid-high F1, bilabial place, nasal mid-high F1, nasal low F1. The nasal mid-high F1 and the nasal low F1 tend to cooccur in sequence very often in Portuguese, which, together with the very small amount of intervening material between [[ɐ̃]] and [[ũ]] (namely, none), will probably lead to the perception of [[ɐ̃ũ]] with a single /+/ value on the nasality tier. We can formalize this as in (20).

(20) *A near-universal example of abstraction: nasal diphthongs*

acoustics: [[ɐ̃ũ]]	OCP (nasal: +; Ṽ| | Ṽ)	LCC (nasal: +; Ṽ| | Ṽ)
/*nasal*/: + + | | ɐ u	*!	
☞ /*nasal*/: + ɐ u		*

Thus, /ɐ̃ũ/ will be considered a single nasal unit, and this is why we can say that Portuguese ‘has’ a nasal diphthong. The advantage for the speaker of Portuguese is that she can store the word |sɐβɐ̃ũ| ‘soap’ in her lexicon with a single /+nasal/ value instead of with two.

With slightly more intervening material, perhaps a syllable boundary (which itself is covert structure created by the perception grammar) or a short silence, the result may already be more language-specific, but in the case of rightward spreading of velum lowering through a glottal stop, as in Sundanese, we still expect that the result is perceived with a single /+nasal/ value:

(21) *Perceptual integration of nasality in Sundanese*

acoustics: [[ã _ˀã]]	OCP (nasal: +; Ṽ | _ˀ | Ṽ)	LCC (nasal: +; Ṽ | _ˀ | Ṽ)
/*nasal*/: + – + | | | a ʔ a	*!	
☞ /*nasal*/: + – a ʔ a		*

Unlike the situation in (20), where the name of the LCC constraint feels a little inappropriate, the winning candidate in (21) indeed shows crossing association lines: the non-nasal silence that intervenes between the two nasal cues must be regarded as a /–/ value on the nasal tier.

Sequences with more material between the nasal cues will be less likely to be perceived with single nasality. Articulations that produce a longer non-nasal stretch will be heard either as two separate nasal vowels, or as a single nasal autosegment, mainly depending on whether the language ‘has’ autosegmental nasal harmony or not. For instance, French has nasal and non-nasal vowels in every syllable independently: next to the harmonic |ʃɑ̃sɔ̃| ‘song’ and |pate| ‘pâté’, it has the equally well-formed disharmonic |lapɛ̃| ‘rabbit’ and |mɑ̃to| ‘coat’. As far as the lexicon is concerned, there would be no advantage in storing |ʃɑ̃sɔ̃| with a single /+nasal/, because we would still have to specify whether this feature value applies to the first syllable, or to the second syllable, or to both. On the contrary, perceiving a bimorphemic |mɔ̃+ʃjɛ̃| ‘my dog’ as /mɔ̃ʃjɛ̃/ with a single /+nasal/ value would constitute a faithfulness violation in the production grammar, since one of the two underlying /+nasal/ values would be lost in the perceptual output (see the next paragraph). So in French LCC (nasal) will outrank OCP (nasal) for any intervening consonant (or syllable boundary). In Guaraní, by contrast, which has the harmonic morphemes |tupa| ‘bed’ and |tũpã| ‘god’, but no disharmonic *[tupã] or *[tũpa] (Piggott 1992), it is advantageous for listeners to lexicalize nasality on the morpheme level (or linked only to the final vowel, if there is evidence for leftward spreading), because this economizes on specifying the underlying paths (links) between the nasal values and each syllable. So in Guaraní LCC (nasal) will outrank OCP (nasal) for any intervening consonant, and [tũpã] will be perceived with a single /+nasal/, although it must be implemented with three velar gestures.

The perceptual integration of nasality across syllable boundaries comes at a cost. Consider the concatenation of two underlying morphemes |tãkã| and |tũkã| in a Guaraní-type language. If the result is pronounced [tãkãtũkã], this will necessarily be perceived with a single /+nasal/:

(22) *Necessary perceptual integration across a morpheme boundary*

<table>
<tr><th>acoustics: [tãkãtũkã]</th><th>OCP (nasal: +; Ṽ | C | Ṽ)</th><th>LCC (nasal: +; Ṽ | C | Ṽ)</th></tr>
<tr><td>/nasal/: – + – + – + – +
| | | | | | | |
t a k a t u k a</td><td>*!**</td><td></td></tr>
<tr><td>☞ /nasal/: – – + – –
t a k a t u k a</td><td></td><td>***</td></tr>
<tr><td>/nasal/: – – + – + –
t a k a t u k a</td><td>*!</td><td>**</td></tr>
</table>

Note that there is no ranking of the constraints that will lead to a perception of the third candidate, which would have been optimal for purposes of lexical access. The perception in (22) leads to a faithfulness violation in the production grammar: the underlying form contains two /+nasal/ specifications, the perceptual candidate contains a single /+nasal/ value, so that it violates *DELETE (+nasal). We must expect, then, that a high ranking of *DELETE (+nasal) should be able to force satisfaction of faithfulness at the cost of something else, perhaps the surfacing of an underlying path (link of simultaneity) between [+nas] and a vowel, as in tableau (23).

(23) *Faithfulness violation forces epenthesis*

underlying: |+nas| t a k a + |+nas| t u k a	*DELETE (+nasal)	*DELETEPATH (+nasal)
[tãkãtũkã] → /+nas/ t a k a t u k a	*!	
☞ [tãkatũkã] → /+nas/ /+nas/ t a k a t u k a		*

The first candidate, which has to follow the perception process in (22), merges the two underlying [+nas] specifications in the output, thus violating featural faithfulness once. The second candidate has two separate [+nas] values, which equals the specified number, but has dropped the underlying path between [+nas] and the second vowel. This epenthesis of a [–nas] vowel is not well attested in nasal harmony, but must be expected to be possible, since effects like these are abundant in the case of tone.

The epenthesis in (23) is one of the many possible ‘OCP effects’ in functional phonology (Boersma 1998: ch. 18). It is due to a combination of a high-ranked OCP in the perception grammar, and a resulting violation of a high-ranked faithfulness constraint in the production grammar. Note that if OCP were a constraint in a ‘hybrid’ production grammar, a low ranking of this constraint could have the undesirable result that faithfulness constraints control the creation of covert structure:

(24) *Generative approach to hidden structure*

|+nas| t a k a + |+nas| t u k a	MAX (nas)	MAXLINK (nas)	OCP
[+nas] t a k a t u k a	*!		
[+nas] [+nas] t a k a t u k a		*	
☞ [+nas] [+nas] t a k a t u k a			*

In this example, where subsequent nasal vowel must perhaps be considered adjacent on the level of the syllable head (Piggott & Van der Hulst 1997), high-ranked faithfulness can influence the speaker's interpretation of the surface form. In functional phonology, by contrast, the production grammar has no influence on the mapping from acoustic to perceptual form (Boersma 1999a).

We have started with the perception of nasal diphthongs, and proceeded with the perception of nasality across glottal stops and consonants in general. The question arises whether /–nasal/ gaps in /+nasal/ stretches can be even larger than this, perhaps spanning a syllable, i.e. whether sequences like [pãtikã] are ever perceived with a single /+nasal/. I know of no such instances in nasal harmony, but analogous examples may exist in tongue-root-harmony systems, as in Wolof, where high vowels are transparent to spreading of RTR (Archangeli & Pulleyblank 1994):

(25) *Perception of feature values across syllables*

/RTR//ATR/

t ɛ k k i l ɛ ɛ n

3 Nasal harmony, type A

In this section, I will show that one type of nasal harmony is due to articulatory spreading. Piggott (1992) distinguishes two kinds of nasal-harmony systems, which he calls type A and type B. In type-A nasal harmony, nasality spreads from a nasal segment until the spreading is blocked by a segment that is apparently not compatible with nasality. In Malay, for example, nasality may spread rightward through [j] but not through [k]:

(26) *Nasal spreading in Malay (from the initial consonant to the right)*

[mãj̃ãn] 'stalk'
[mãkan] 'eat'

Thus, /j/ is a ***target*** for nasalization (it's ***nasalizable***), whereas /k/ is a ***blocker*** (it's ***opaque***). The following typology summarizes the possible targets in type-A languages:

(27) *Nasalizable segments (Piggott 1992)*

laryngeals	*glides*	*liquids*	*fricatives*	*plosives*	*language example*
+	–	–	–	–	Sundanese
+	+	–	–	–	Malay, Warao
+	+	+	–	–	Ijo, Urhobo
+	+	+	+	–	Applecross Gaelic

This typology corresponds to the following implicational universals:

(28) *Universals of nasal spreading*

a. If glides can be nasalized, so can vowels and laryngeals.
b. If liquids can be nasalized, so can glides.
c. If fricatives can be nasalized, so can liquids.
d. Plosives cannot be nasalized.

3.1 A segmental functional analysis of type-A nasal harmony

Functionally, the generalization is straightforward. Suppose first that the constraint that is honoured by spreading nasality to the right is *MOVE (velum), i.e. a constraint that aims at postponing the raising gesture of the velum, as an indirect way to minimize the number of raising and lowering gestures of the velum. This gestural definition immediately accounts for the Malay type:

(29) *Glides undergo nasal spreading in Malay*

|maja|	*REPLACE (nas: –, + / liquid)	*MOVE	*REPLACE (nas: –, + / glide)
[mãja] → /mãja/		*!*	
☞ [mãj̃ã] → /mãj̃ã/			*

This account is entirely segmental: the underlying form |maja| has a single |+nasal| segment and, crucially, three segments separately specified for |–nasal|. The result /mãj̃ã/ has four adjacent /+nasal/ values. Only with this segmental approach can we use constraints that refer to complete features, like *REPLACE (nasal: –, +). For a more autosegmental approach, see §3.2. We see, then, that the segment |j|, which is specified as non-nasal, undergoes 'nasal' spreading, i.e., is implemented with a longer velar lowering, which happens to lead to a perception of a nasal glide.

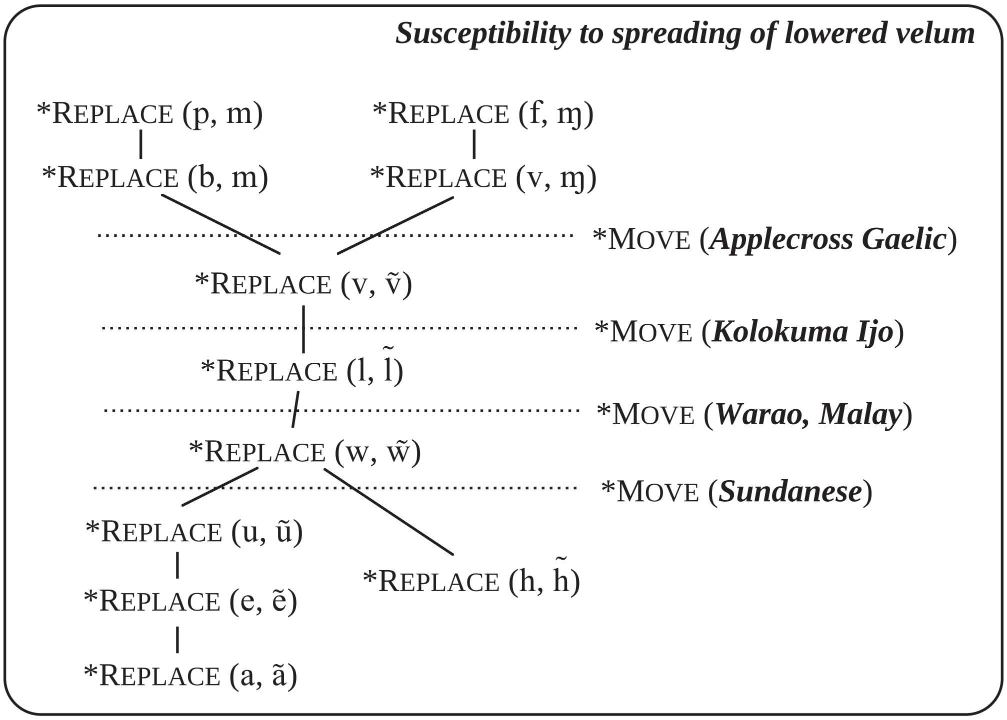
The two *REPLACE constraints in (29) are ranked according to the local-ranking principle, since the closer an oral constriction is, the more its perceptual result is modified by adding a nasal side branch (Schourup 1972: 533). Thus, Malay liquids block nasal spreading:

(30) *Liquids block nasal spreading in Malay*

|mara|	*REPLACE (nas: –, + / liquid)	*MOVE	*REPLACE (nas: –, + / glide)
☞ [mãra] → /mãra/		**	
[mãr̃ã] → /mãr̃ã/	*!		

For purposes of readability, I will abbreviate the two faithfulness constraints in (30) as *REPLACE (l, l̃) and *REPLACE (w, w̃), respectively. The entire fixed confusion-based hierarchy of nasalizability is shown as the *REPLACE constraints connected with solid lines in (31):

(31) *Fixed hierarchy of anti-nasalization faithfulness*



The cases of the obstruents need some comment. If the velum is lowered during a gesture that would otherwise produce a labial plosive, a nasal stop will automatically result. So honouring *MOVE would violate some probably highly ranked faithfulness constraints against deletion of the perception of plosiveness and insertion of the perception of sonorancy:

(32) *Plosives block nasal spreading in Malay*

|maka|	*REPLACE (k, ŋ)	*MOVE
☞ [high velum etc.] → /mãka/		*
[low velum etc.] → /mãŋã/	*!	

The fact that these faithfulness (correspondence) constraints are ranked so high, can be attributed to the strong perceptual repercussions of their violation, according to any reasonable global measure of perceptual distance. For instance, *REPLACE (k, ŋ) can be seen as a shorthand for the conjunction of *DELETE (plosive) & *INSERT (sonorant).[2] For fricatives, a tableau analogous to (32) can be drawn for most languages:

[2] By virtue of enabling additive ranking, a phonological theory that allows conjunction of faithfulness constraints or of gestural constraints (but no mixes) is more directly functional than a theory without this possibility.

(33) *Fricatives block nasal spreading in most languages*

|masa|	*REPLACE (s, n)	*MOVE
☞ [high velum etc.] → /mãsa/		*
[low velum etc.] → /mãnã/	*!	

Since we have no way of telling whether the deletion of frication is worse than the deletion of plosiveness, the top pairs of constraints are not universally ranked with respect to each other. Note that the candidate set in (33) is restricted in such a way that a nasalized fricative cannot occur as a perceptual output. This is because speakers of most languages would not know how to produce such a sound, so that the relevant articulation does not show up as a candidate at all. In Applecross Gaelic, people are reported to be able to produce it, so their tableau is like:

(34) *Nasalized fricatives reported for Applecross Gaelic*

|masa|	*REPLACE (s, n)	*MOVE	*REPLACE (s, s̃)	*GESTURE (special trick)
[high velum etc.] → /mãsa/		*!		
[low velum etc.] → /mãŋã/	*!			
☞ [special trick] → /mãs̃ã/			*	*

So (31) shows a four-way typology, based on the ranking of *MOVE with respect to the fixed hierarchy. The typology also seems to predict the existence of languages that show nasalization of plosives and/or fricatives (as nasal stops), though the number of these languages may be very low because of considerations of global measures of distinctivity, which may force the language-specific rankings of *REPLACE (p, m) and *MOVE (velum) to be drawn from distributions (“windows”) that hardly overlap. Nevertheless, final plosives in Sanskrit become nasal when followed by a nasal consonant, and in Lewis Gaelic (e.g. Ladefoged, Ladefoged, Turk, Hind & Skilton 1997), plosives but not fricatives nasalize after the masculine article |ən|: |ən+palə| ‘the town’ → /əmalə/; |ən+kʰaːr| ‘the car’ → /əŋhaːr/; |ən+fɛr| ‘the man’ → /əɱfɛr/.

3.2 An autosegmental functional analysis of type-A nasal harmony

An autosegmental approach might be more realistic than the segmental approach of §3.1. The two Malay forms in (26) are probably perceived as in (35):

(35) *Autosegmental perception of nasals in Malay*

/*nasal*/: + +
m a j a n

/*nasal*/: + – +
m a k a n

The /mãj̃ãn/ form is particularly interesting (for a similar type, see Piggott 1992: 59). Since the nasalization of /ãj̃ã/ is a result of rightward spreading from /m/, and not of a leftward spreading from /n/, it is plausible that /ãj̃ã/ is perceived in the same nasal stretch as its left neighbour, but in a different stretch from its right neighbour. The Malay perception grammar can achieve this by having a high-ranked OCP (+nas), but an even higher ranked LCC (+nas; V | | nasal stop). This move ensures that both underlying |+nasal| specifications in |majan| surface faithfully, thus satisfying *DELETE (+nasal). We see here another case of the difference between articulation and perception: /mãj̃ãn/ has two perceptual /+nasal/ values, although it is implemented with a lowered velum throughout; this is the reverse case of /tũpã/ in Guaraní, which has a single perceptual /+nasal/ value but is implemented with two separate velum lowering gestures.

Tableau (29) now becomes:

(36) *Glides undergo nasal spreading in Malay*

underlying: |*nasal*|: + – (m a j a)	*DELETE (+nas)	*INSERTPATH (+nas / liquid)	*MOVE	*INSERTPATH (+nas / glide)
[mãja] → /*nasal*/: + – (m a j a)			*!*	
☞ [mãj̃ã] → /*nasal*/: + (m a j a)				*
[baja] → /*nasal*/: – (b a j a)	*!			

This analysis tacitly assumes a general low ranking of faithfulness for the negative value of /nasal/, i.e. for the constraints *DELETE (–nasal) and *DELETEPATH (–nasal). This low ranking means that we could have removed all occurrences of /–nasal/ from (36), thus essentially obtaining an analysis in terms of a privative feature /nasal/. In either case, the autosegmental approach leads to an analysis equivalent to the segmental analysis of §3.1.

The conclusion, then, must be that any account of type-A nasal harmony expressed into functionally rankable directly functional constraints, is observationally, descriptively, and explanatorily adequate, since it accounts for the data, predicts the typology, and needs no assumptions and principles except those rooted in general properties of motor behaviour and perception. I will now look at four alternative analyses.

3.3 Perception-based spreading in type-A languages?

As an alternative functional analysis of type-A nasal harmony, we might propose that the rightward nasal spreading is not caused by postponing an articulatory gesture, but by honouring a faithfulness constraint, say MAXIMUM (nas), which aims at maximizing the duration of the perception of nasality. But we will have to rule out spreading through plosives in Malay. A discontinuous sequences such as [mãkã] can be ruled out in either of two ways: first, by a strong constraint *INSERT (nas) against the insertion of nasality; second, by an articulatory constraint *GESTURE (velum) against velum movement.

The first solution, with *INSERT (nas), only works if [mãkã] is perceived with two separate instances of nasality:

(37) *Faithfulness-only account of Malay*

|nas| / m a k a	*INSERT (nas)	*DELETE (plosive)	MAXIMUM (nas)
☞ [mãka] → /nas/ m a k a			**
[mãŋã] → /nas/ m a ŋ a		*!	
[mãkã] → /nas//nas/ m a k a	*!		*

But this solution is problematic, since perceptually the glottal stop is as much of a plosive as /k/ is, so that if [mãkã] is perceived with two nasals, [mãʔã] should also be perceived with two nasals, but then the correct candidate [mãʔ͡ã] /mãʔã/ would violate *INSERT (nas) and lose.

The second solution, with *GESTURE (velum), also works if [mãkã] is perceived with a single instance of nasality:

(38) *Malay with non-directional articulatory constraints*

|nas| / m a k a	*DELETE (plosive)	*GESTURE (velum)	MAXIMUM (nas)
☞ [mãka] → /nas/ m a k a			**
[mãŋã] → /nas/ m a ŋ a	*!		
[mãkã] → /nas/ m a k a		*!	*

But this solution is problematic, too. For an underlying form |makan|, it predicts [mãkãn] rather than the correct [mãkan]:

(39) *Malay with non-directional articulatory constraints*

<table>
<tr><th>|nas| |nas|
m a k a n</th><th>*DELETE (plosive)</th><th>*GESTURE (velum)</th><th>MAXIMUM (nas)</th></tr>
<tr><td>[mãkan] → /nas//nas/
m a k a n</td><td></td><td>*</td><td>**!</td></tr>
<tr><td>[mãŋãn] → /nas/
m a ŋ a n</td><td>*!</td><td></td><td></td></tr>
<tr><td>*☞* [mãkãn] → /nas//nas/
m a k a n</td><td></td><td>*</td><td>*</td></tr>
</table>

The problem here is that the final [n] forces velum lowering in the first, correct, candidate, too, giving the final vowel of [mãkãn] an articulatorily free ride. Since this articulatory licensing of perceptual nasalization does not seem to occur in this example, the conclusion must be that the spreading in type-A languages must be due to a spreading of the velum lowering gesture in the leftward or rightward direction, or both.

3.4 Walker's (1998) approach to type A

Walker (1998) proposed a family of SPREAD constraints (analogously to *MOVE), with an explicit definition in terms of the number of nasal association lines. Walker expresses the nasalizability hierarchy with cooccurrence constraints for 'hybrid' features:

(40) *Walker's hierarchy of nasalizability*

*NASOBSSTOP >> *NASFRICATIVE >> *NASLIQUID >>
>> *NASGLIDE >> *NASVOWEL >> *NASSONSTOP

These are constraints in the style of the ***grounding conditions*** of Archangeli & Pulleyblank (1994). Such a constraint is thought to have become an innate element of Universal Grammar during the course of evolution, as a result of the selection pressure associated with the interaction between functional principles. In a functional theory of phonology, which expresses function directly, these indirectly functional constraints should be superfluous.

And indeed, some of Walker's constraints have no correlate in a functional account. Walker needs the structural constraint *NASOBSSTOP, which is a filter against the cooccurrence of [+nas], [–cont], and [–son] (Walker 1998: 36), in order to rule out the unpronounceable nasalized labial plosives, which would otherwise be the winner:

(41) *The need for superfluous cooccurrence constraints*

maka	IDENT-IO (±sonorant)	*NASOBSSTOP	SPREAD (nasal)	*NASVOWEL
maka			***!	
☞ mãka			**	*
mãk̃ã		*!		**
mãŋã	*!			**

In the functional account of (32), a candidate perceived as /mãk̃ã/ can never occur, simply because no articulation can produce it. This means that if we distinguish between articulation and perception in the production grammar, several phonetically impossible combinations do not have to be stated as inviolable constraints in the grammar, as they have to in a grammar with hybrid representations.

The second problem is that the hierarchy of structural constraints (40) does not generate the typology by itself. To rule out [mãŋã] for |maka|, Walker (1998: 113) still needs faithfulness constraints like IDENT-IO (±sonorant) and IDENT-❁O (±voiced), which are comparable to *REPLACE (k, ŋ). Thus, Walker's approach needs a hierarchy of structural constraints as well as a hierarchy of faithfulness constraints (i.e. other IDENT constraints are ranked lower), whereas the functional approach needs nothing more than a confusion-based hierarchy of faithfulness.

3.5 Sonority hierarchy and type A

The hierarchies in (31) and (40) are reminiscent of the ***sonority hierarchy***, and indeed the sonority scale has come up in at least one account of nasalizability (Gnanadesikan 1995). This is natural from a generative point of view, since the sonority hierarchy is a very good candidate for an innate phonological device, since nearly all languages will then use it for syllabification. However, the position of /h/ in the hierarchy is problematic, as Gnanadesikan notes. In the nasalizability hierarchy, faithfulness for /h/ is ranked on a height comparable to that of vowels, because nasalizing this sound will not strongly change its main perceptual features (noise and spectrum). In hierarchies for syllabification, on the other hand, /h/ will pattern with the other fricatives /f/ and /s/ in its preference for the syllable margin, which is again only natural since the sound is voiceless (Boersma 1998: 455). Gnanadesikan gives an example of a two-year-old child, who pronounces |bilˈou| 'below' as [fibˈou], copying the initial obstruent to replace the sonorant onset of the stressed syllable, but pronounces |bihˈajnd| 'behind' as [fihˈajn], *not* copying the initial obstruent to replace the apparently non-sonorant onset of the stressed syllable.

We must conclude that the plausibly innate device of the sonority hierarchy has an exception in the direction of immediate functionality, and is not an arbitrary universal. This is a strong argument against substantive innateness in phonology, especially if the exception does not play a role in many languages, since this would leave only a small number of generations to have selected the presumably innate exception.

3.6 Piggott’s (1992) account of type A

Piggott’s (1992) account for the nasal-spreading typology (38) proposes some problem-specific innate principles for UG:

(42) *Piggott’s principles of nasal harmony (simplified)*

a. “The class of blockers must constitute a natural class with the nasal consonants.”
Nasals are ***stops***, so one of those classes must be the class of stops: /m/, /n/, /p/, /t/, which accounts for the blockers in Applecross Gaelic.
Nasals are also ***consonantal***, so depending on whether glides are consonantal, we have the classes /m/, /n/, /p/, /t/, /f/, /s/, /l/, /r/ (Warao) and /m/, /n/, /p/, /t/, /f/, /s/, /l/, /r/, /j/, /w/ (Sundanese).
And nasals are ***sonorant***, so we would expect the class /m/, /n/, /l/, /r/, /j/, /w/, i.e. a language in which obstruents are targets, but sonorants block!

b. “The class of blockers must not be limited to sonorants.”
This exception rules out the third possibility in (42a).

c. “There is a natural class called ***non-approximant consonants***.”
This class consists of /m/, /n/, /p/, /t/, /f/, /s/, accounting for Ijo.

While (42a) sounds like a general principle that could find application in other areas, the UG principles (42b) and (42c) are obviously specific to the problem of type-A nasal harmony. Since these principles are of advantage to only a very small minority of languages, they are very unlikely to have had any chance of emerging by a rich enough selection during the course of evolution (a few hundred generations).

4 Nasal harmony, type B

In this section, I will show that by contrast with type-A articulatory spreading, type-B nasal harmony is due to perceptually-based spreading. Type-B nasal-harmony languages (Piggott 1992) are characterized by morpheme-level or word-level specifications for |+nasal| or |–nasal|, and most segments surface differently in nasal and non-nasal morphemes:

(43) *Type-B nasality contrasts (Southern Barasano)*

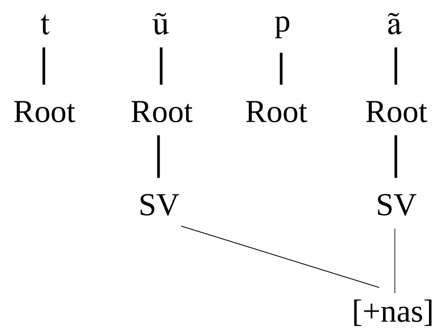
In |–nasal| morphemes:	In |+nasal| morphemes:
a, u	ã, ũ
w, j	w̃, j̃
l, r	l̃, r̃
ᵐb	m
s	s
t, k	t, k

Other type-B languages have comparable systems. Tuyuca (discussed at length by Walker 1998) has /b/ instead of /ᵐb/.

4.1 Transparency of plosives

One of the conspicuous properties of type-B nasal-harmony languages is the transparency of plosives to nasal spreading. So, in Guaraní we have [tupa] ‘bed’ and [tũpã] ‘god’, but no *[tupã] or *[tũpa] (Piggott 1992). In Piggott’s analysis, nasality is spread from right to left across all segments that have a Spontaneous Voicing (SV) node, i.e. all sonorants:

(44) *Piggott’s spreading along the Spontaneous Voicing tier*

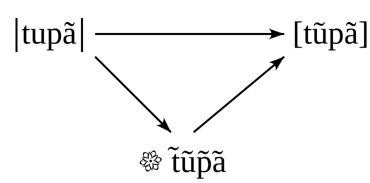


Since the two vowels are adjacent on the SV tier, Piggott’s analysis has the desirable property of ***locality*** in spreading processes.[3]

Piggott & Van der Hulst (1997) reanalyse the process as spreading on the syllable level: a nasalized vowel, being the head of its syllable, makes nasalization a property of the syllable, and this then spreads to adjacent syllables, nasalizing all the sonorants in every affected syllable. This move allows Piggott & Van der Hulst to account for more facts than Piggott (1992), such as the fact that all sonorants in syllables with nasalized vowels are nasalized themselves, and the similarity with vowel-harmony processes. Again, however, the locality requirement has informed the search for a higher structure in which nasalization is continuous.

Walker (1998: 43) also explicitly wants to honour the locality requirement, invoking the line-crossing constraint as an inviolable well-formedness condition on (hybrid) phonological representations. Her analysis is stated in terms of the ***sympathy*** device introduced by McCarthy (1998):

(45) *Walker’s analysis of transparency to leftward spreading*



In this view, the surface form [tũpã] is derived from the underlying form |tupã| as well as from a sympathetic (❀) form /t̃ũp̃ã/, which is itself derived from the underlying form by maximal spreading, but which contains two unpronounceable segments.

Both Piggott’s and Walker’s theories work. However, a theory that distinguishes between articulation and perception in phonology must maintain that feature geometries are illusions

---

[3] As a detail, we may note that the always recalcitrant segment /h/ has no SV node, so that Piggott’s analysis would predict that it is transparent, not nasalizable.

evoked in the linguist who advocates hybrid representations, and that these illusions will evaporate if the correct distinctions are made (Boersma 1998: 22, 442). I will show that if we separate articulation and perception, we do not need Piggott’s feature geometry or Walker’s sympathy approach.

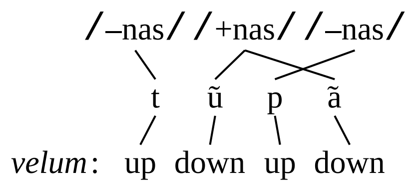
Because of the morpheme-level specification, which successfully nasalizes at least all vowels, sequences of two nasal vowels with an intervening obstruent are very common in these languages. For this reason, OCP for nasality will outrank the line-crossing constraint:

(46) *Perceiving nasality across a plosive*

acoustics: [tũpã]	OCP (nas: +; Ṽ | plosive | Ṽ)	LCC (nas: +; Ṽ | plosive | Ṽ)
perception: /–nas/ /+nas/ /–nas/ /+nas/ t ũ p ã	*!	
☞ perception: /–nas/ /+nas/ /–nas/ t ũ p ã		*

Thus, the two nasalized vowels are perceived with a single value [+nasal] on the perceptual nasality tier, despite the intervening plosive. The articulation, on the other hand, cannot be regarded as continuous, since the velum has to go up and down for the labial plosive. This combination of discontinuous articulation and perceptual unity can de pictured as:

(47) *Asymmetry between articulation and perception*



Note that (47) does not imply that /+nas/ precedes the second /–nas/: if we see a car standing behind a lamppost, do we see it to the left or to the right of the lamppost? Obviously both. Likewise, /+nas/ in (47) both precedes and follows the second /–nas/, contra Archangeli & Pulleyblank (1994: 37), who invoke the assumption of strict precedence of feature values within tiers in an attempt to derive the inviolability of the line-crossing constraint (more precisely, ‘gapped’ configurations) from more basic principles. The universality of the locality condition is an illusion brought about by the ubiquity of articulation-based spreading, which is never discontinuous (because that would violate its very purpose of gesture reduction), and by the rarity of a high OCP across salient intervening material. However, a high rate of cooccurrence of nasality in adjacent syllables, as in type-B nasal harmony languages, will shift the balance in the direction of perceptual unity. At least if speech perception is like other kinds of perception.

4.2 Why all sonorants are nasalizable in type-B languages

Another conspicuous property of type-B nasal-harmony languages as that they all nasalize their glides and liquids. A hierarchy of nasalizability, similar to the one for type-A languages, does not appear to exist. I will discuss one possible cause.

Type-A languages seem to be like most languages in that they have to deal with constraints against the replacement of the trill |r| with a nasalized trill /r̃/. If the constraint refers to the difference between the two, and the difference is in the nasality, we must conclude that the underlying form contains a |–nasal| specification (or, equivalently, lacks a |nasal| specification).

This will be different in type-B languages, which make a point of applying nasality to the morpheme or word level. If nasality is suprasegmental, segments are less likely to be specified for nasality themselves. So the perceptual specification of the segment will not contain any specification for |–nasal|. The only relevant specification for |r| is |trill|, and both /r/ and /r̃/ honour *DELETE (trill). So the relevant specifications for all segments are:

(48) *Type-B nasality contrasts (Southern Barasano)*

In |–nasal| morphemes:	In |+nasal| morphemes:	
a, u	ã, ũ	|low/high vowel, +son|
w, j	w̃, j̃	|back/front glide, +son|
l, r	l̃, r̃	|lat/trill, +son|
$^{\mathrm{m}}$b	m	|stop, +son|[4] (following Piggott 1992)
s	s	|fricative, –son|
t, k	t, k	|plosive, –son|

This means that the surface forms reflect the faithfulness constraints in the following way:

(49) *Faithfulness handling*

a. All the specified features surface faithfully in oral as well as in nasal words.
b. In a nasal context, the obstruents violate *DELETEPATH (nasal). Since there is a partially nasal segment (/$^{\mathrm{m}}$b/), I'll speak of *DELETEPATH (half a nasal) instead, and assign two violations for entirely non-nasal obstruents in a nasal environment.
b. In a non-nasal context, the sonorant stops introduce partial nasality, so they violate *INSERT (half a nasal). Any entirely nasal segment would violate it twice.

Here are the tableaus for all segments:

(50) *Nasalizing a liquid (or vowel or glide)*

|ara + *nasal*|	*DELETE (*any feature*)	*DELETEPATH (half a nasal)	*INSERT (half a nasal)	*GESTURE (velum)
☞ [ãr̃ã] → /ãr̃ã/				
[ãrã] → /ãrã/		*!*		**

[4] For Tuyuca, with a /b/-/m/ contrast, the specification would be |stop, +voi|, which gives simpler tableaus.

(51) *Nasalizing a sonorant stop*

|a[+son,stop]a + *nasal*|	*DELETE (*any feature*)	*DELETEPATH (half a nasal)	*INSERT (half a nasal)	*GESTURE (velum)
☞ [ãmã] → /ãmã/				
[ã^{m}bã] → /ã^{m}bã/		*!		**
[ãbã] → /ãbã/	*!	**		**

(52) *Oralizing a sonorant stop*

|a[+son,stop]a|	*DELETE (*any feature*)	*DELETEPATH (half a nasal)	*INSERT (half a nasal)	*GESTURE (velum)
[ama] → /ama/			**!	**
☞ [a^{m}ba] → /a^{m}ba/			*	**
[aba] → /aba/	*!			

(53) *Nasalizing a plosive (or fricative)*

|a[–son,plos]a + *nasal*|	*DELETE (*any feature*)	*DELETEPATH (half a nasal)	*INSERT (half a nasal)	*GESTURE (velum)
[apa] → /apa/		**		
☞ [ama] → /ama/	*!*			**

The ranking can be summarized as *DELETE (segmental feature) >> FAITH (nasal) >> *GESTURE (velum). The first ranking seems quite natural, because nasality will always be realized on the vowels, so that it is not very important to have it realized on the consonants as well. The second ranking expresses the idea that velar gestures do not play any role in the phonology. In this respect, type-B languages differ completely from type-A languages.

To sum up: in type-A languages, glides, liquids, fricatives, and plosives *don't* want to be nasalized, but sometimes they are forced to be (by articulatory spreading); in type-B languages, these segments *do* want to be nasalized, but sometimes they cannot be (because of fricativity and plosivity faithfulness). For the liquids, which are in the middle of the hierarchy, these opposing desires must lead to nasalization being much more common in type-B than in type-A languages. This is borne out: liquids show blocking behaviour in many type-A languages, because a nasalized liquid sounds differently from a non-nasalized liquid, and they are nasalized in all type-B languages, because a nasalized liquid is still a liquid.

4.3 Walker's generalization

The functional approach appears to have missed a possible unification of type-A and type-B languages, discovered by Walker (1998), who observed that the typology of type-A languages lacks a type with nasalized plosives, and that type-B languages actually fill this gap by allowing nasalized plosives in their sympathetic candidates. The problem with this approach becomes apparent when we consider the only point in which the two overlap: the case of the fricatives. In those few type-A languages where nasality spreads through fricatives, these fricatives are *undergoers*, i.e. they are nasalized themselves. In type-B languages, fricatives are *transparent*. This difference corresponds exactly with the basic difference that we found between the two types: type-A languages have articulatory spreading, i.e. spreading of a lowered velum, and type-B languages have perceptual spreading, i.e. spreading of perceptual nasality even if it involves additional velum effort.

4.4 An independent need for nasal cooccurrence constraints?

It might be argued that constraints like *NASLIQUID, which were used by Walker (1998) but which were shown to be superfluous in a functional approach, are independently needed in phonological descriptions, because they express the fact that most languages tend to have non-nasal liquids like /r/ but not nasalized liquids like /r̃/. With the axiom of Richness of the Base (Prince & Smolensky 1993), any underlying |r̃| should be converted to [r] by the grammar, presumably with the help of a markedness constraint like *NASLIQUID. But in functional phonology, segmental restrictions are not a task of the production grammar. In a language without surfacing nasalized liquids, the learner will never perceive a nasalized liquid, so that she will rarely be forced to construct a nasalized liquid in a lexical form. This poorness of the base is one of the three automatic restrictions in the functional grammar model (§4.5). If a language does have surface nasalized fricatives, the perception grammar will be happen to handle them, and their cross-linguistic markedness will be irrelevant during production.

I still have to explain why languages tend not to have nasalized liquids. If the grammar does not handle this, what does? Suppose that a language has nasalized liquids but no non-nasalized liquids, and that it also has plain nasal stops. If there is a little variation in the nasality of the liquids in this language, the nasalized liquids will tend to become replaced by non-nasal liquids, because this would improve the perceptual distinctivity within the phoneme inventory (Boersma 1998: ch. 17). Also, the large confusion probability between the nasalized liquids and the nasal stops will cause many learners to merge some of the nasal and nasalized phonemes into larger, easier to distinguish, categories. Gradually, the language will come to be built around an average balance between perceptual confusion and articulatory effort.

4.5 Three indirect restrictions on perceptual surface forms

The functional grammar model leads to three automatic restrictions not known in generative phonology.

First, the acoustics is an automatic result of the articulation, so that the perception tends not to contain any features that nobody can produce. Hence, the functional approach has no need for *NASOBSSTOP.

Second, all candidates are learned articulations. For instance, we do not usually generate candidates with nasal fricatives, so that *NASFRICATIVE is not needed to rule them out.

Third, the lexical forms have restrictions approximately identical to those of perception. So we do not have to account for what would happen to an underlying |tũpa| in Guaraní, since the lexicalization process will have led to morpheme-level nasality in all lexical items. This levies our responsibilities with respect to Richness of the Base. The main case in which the workings of Richness of the Base become apparent, is the borrowing of foreign words. For example, Desano borrowed the Portuguese word [sɐβɐ̃ũ] 'soap' as [sabo], and [ʒoɐ̃ũ] 'John' as [ɲũ] (Kaye 1971). An OT approach along the lines of Prince & Smolensky (1993) would assume that the underlying forms are the disharmonic |sɐβɐ̃u| and |ʒoɐ̃ũ|, and have the grammar convert these to the harmonic [sabo] and [ɲũ]. However, there is no reason not to assume that these words were actually perceived harmonically and that they were therefore lexicalized as |sabo| and |ɲũ|.

4.6 Acquisition

Some of this reasoning may give the impression of circularity. Type-B languages are characterized by a high nasality OCP, caused by a high correlation of nasality in consecutive syllables, and this high correlation of nasality again stems from the specification of nasality in the lexicon, which is on the level of the morpheme or linked to a fixed single vowel for every morpheme. And this morpheme-level lexical specification has again been caused by the unified perception of nasals across consonants, caused by a high nasality OCP.

But this circularity was to be expected, since the high OCP and the nasality correlation strengthen and feed each other. Imagine a language like French, with nasality specified for every vowel separately. Now, some cause (perhaps type-A spreading), introduced in this language, may raise the correlation between the nasality values of the vowels in consecutive syllables. If this correlation exceeds a certain threshold, a new learner is likely to introduce a high nasality OCP in her perception grammar. This will have a large effect on her lexicon: while the previous generation had nasality specifications for every vowel separately, the new generation will have morpheme-level nasality specifications for harmonic morphemes and have to mark the disharmonic morphemes as lexical exceptions. If the learner regularizes some of the exceptions, she will produce more harmonic morphemes than she had heard in her environment. As each new generation reduces the number of exceptions, as is common in language change, the disharmonic forms will gradually die out. While this is happening, every next generation will have more reason to posit a high OCP. The result is a stable type-B nasal harmony language.

Conclusion

A functional account of multisegmental nasality leads us to identify two types: the articulatory spreading of velum lowering, which occurs in Piggott's type-A languages, and the perceptual harmony of nasality on the morpheme or word level, which occurs in Piggott's type-B languages.

Generative accounts of nasal harmony have to take recourse to ad-hoc natural classes, exceptions to exceptions, grammaticization of constraints against unproducable perceptual output, functional exceptions to innate hierarchies, feature geometry, and multi-level OT. If all these things were really needed, UG would be full of substantive phonological detail. However, the functional approach to phonology can account for the facts of nasal harmony without assuming anything but general properties of human motor behaviour and perception. This is compatible with the view that the phonological part of the innate language device does not contain much more than: the cognitive abilities of categorization, abstraction, wild generalization, and extrapolation; the storage, retrieval, and access of arbitrary symbols; a stochastic constraint grammar; a gradual learning algorithm; laziness; the desire to understand others; and the desire to make oneself understood.

References

Archangeli, Diana & Douglas Pulleyblank (1994). *Grounded phonology*. Cambridge: MIT Press.

Boersma, Paul (1998). *Functional phonology: Formalizing the interactions between articulatory and perceptual drives* [*LOT International Series* **11**]. PhD dissertation, University of Amsterdam. The Hague: Holland Academic Graphics. [http://www.fon.hum.uva.nl/paul/diss/]

Boersma, Paul (1999a). *On the need for a separate perception grammar*. Ms. [Rutgers Optimality Archive **358**, http://ruccs.rutgers.edu/roa.html]

Boersma, Paul (1999b). 'Phonology-semantics interaction in OT, and its acquisition.' In Robert Kirchner, Wolf Wikeley, & Joe Pater (eds.): *Papers in Experimental and Theoretical Linguistics*. Vol. 6. Edmonton: University of Alberta. [Rutgers Optimality Archive **369**, http://ruccs.rutgers.edu/roa.html]

Cohn, Abigail (1990). *Phonetic and phonological rules of nasalization.* Doctoral thesis, UCLA. [*UCLA Working Papers in Phonetics* **76**]

Cohn, Abigail (1993). 'The status of nasalized continuants.' In Marie K. Huffman & Rena Krakow (eds.), *Nasals, nasalization, and the velum. Phonetics and phonology*, Vol. 5. San Diego: Academic Press. 329–367.

Gnanadesikan, Amalia (1995). *Markedness and faithfulness constraints in child phonology*. Manuscript, University of Massachusetts. [Rutgers Optimality Archive **67**, http://ruccs.rutgers. edu/roa.html]

Kaye, Jonathan (1971). 'Nasal harmony in Desano.' *Linguistic Inquiry* **2**: 37–56.

Ladefoged, Peter, Jenny Ladefoged, Alice Turk, Kevin Hind, and St. John Skilton (1997). 'Phonetic structures of Scottish Gaelic.' *UCLA Working Papers in Phonetics* **95**: 114–153.

McCarthy, John J. (1998). *Sympathy and phonological opacity*. Ms. University of Massachusetts, Amherst. [Rutgers Optimality Archive **252**, http://ruccs.rutgers. edu/roa.html]

Piggott, Glyne (1992). 'Variability in feature dependency: the case of nasality.' *Natural Language and Linguistic Theory* **10**: 33-78.

Piggott, Glyne & Harry van der Hulst (1997). 'Locality and the nature of nasal harmony.' *Lingua* **103**: 85–112.

Powers, William T. (1973). *Behavior: The control of perception*. Chicago: Aldine.

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

Saussure, Ferdinand de (1916). *Cours de linguistique générale*. Edited by Charles Bally & Albert Sechehaye in collaboration with Albert Riedlinger. Paris: Payot & C^{ie}. [2nd edition, 1922]

Schourup, L. (1972). 'Characteristics of vowel nasalization.' *Papers in Linguistics* **5**: 530–548.

Walker, Rachel (1998). *Nasalization, neutral segments, and opacity effects*. Doctoral thesis, University of California, Santa Cruz.

Walker, Rachel & Geoffrey K. Pullum (1999). 'Possible and impossible segments.' *Language* **75**: 764–780.