On the need for a separate perception grammar

Paul Boersma
University of Amsterdam
paul.boersma@hum.uva.nl, http://www.fon.hum.uva.nl/paul/
October 4, 1999

In this article, I will adduce evidence for the existence of a separate grammar for phonological perception, by showing that this would allow us to solve some old paradoxes and that it would help us create a more economical theory of phonology.

1 Articulatory and perceptual representations in phonology

The functional hypothesis for linguistics (Passy 1891) holds that languages are built according to functional principles of efficient communication. For speaking, these functional principles are minimization of articulatory effort and minimization of perceptual confusion. It seems plausible, then, that phonology should distinguish between articulatory and perceptual representations, so that these functional principles can be evaluated within their own natural spaces (Boersma 1989). Within the framework of Optimality Theory (Prince and Smolensky 1993), this idea has been pursued in various forms by Flemming (1995), Jun (1995), Steriade (1995, 1996), Hayes (1996), Boersma (1998), and Kirchner (1998). The form that I will defend here is Functional Phonology (Boersma 1998), which proposes a rigorous division of labour not only between the constraints that implement the various functional principles, but also between the descriptions of the processes of production and comprehension.

The grammar model of functional phonology is summarized in Fig. 1; the right-hand side models the speaker, the left-hand side the listener.

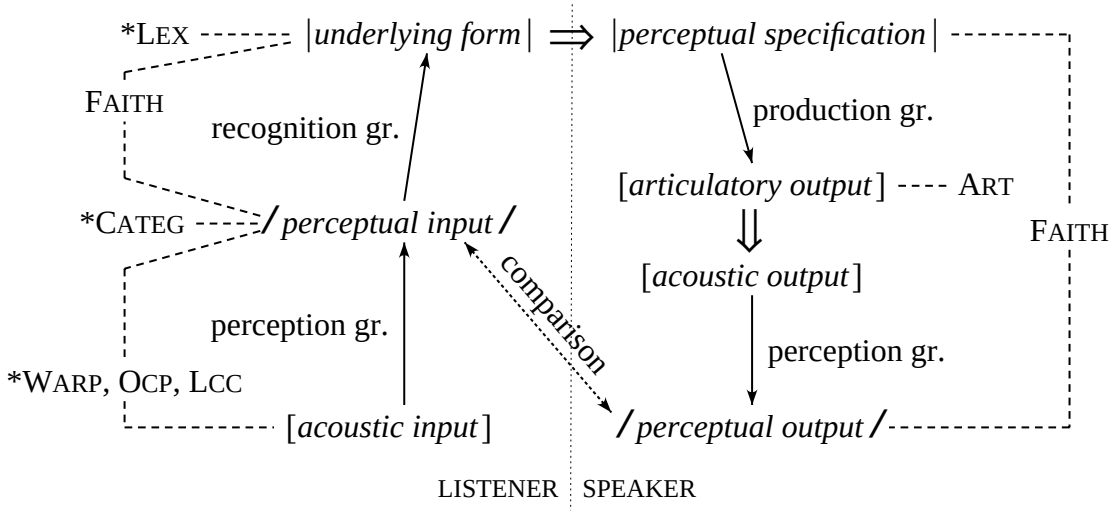


Figure 1
Grammar model of functional phonology

We see four processing systems in Fig. 1:

(1) *Four processing systems in functional phonology*

a. The speaker uses her **production grammar** to map underlying forms to phonetic forms. This is the grammar that is modelled in most theories of phonology.
b. The ***perception grammar*** occurs twice in Fig. 1. It maps phonetic forms to phonological surface structures. The speaker uses it for her own speech, the listener for the speech of others.
c. The listener uses her ***recognition grammar*** to map phonological surface structures to underlying forms in the lexicon.
d. The ***comparison module*** compares a learner's own output forms with adult surface forms. If there are differences between the two, this module changes the learner's grammars.

The model in Fig. 1 exhibits no more than a single articulatory representation:

(2) *One articulatory representation in phonology*

a. The ***articulatory output*** of the production grammar is a set of timed articulatory gestures (muscle positions and tensions, and their simultaneous and sequential coordinations and abstractions).

Further, the model shows no fewer than six perceptual representations:

(3) *Six perceptual representations in phonology*

a. The ***perceptual specification*** of the utterance contains the forms as stored in the lexicon, expressed in terms of discrete perceptual features (periodicity, spectrum, loudness) and their simultaneous and sequential relations and abstractions.
b. The ***acoustic output*** is the automatic physical or peripheral auditory result of the articulatory output.
c. The ***perceptual output*** is the acoustic output as perceived by the speaker herself.
d. The ***acoustic input*** is the physical or peripheral auditory signal received by the listener.
e. The ***perceptual input*** is the acoustic input as perceived by the listener.
f. The ***underlying form*** is what the listener finds in the lexicon. It equals the perceptual specification of the speaker (Saussure 1916: 99).

Finally, Fig. 1 shows a number of constraint sets, which form the ingredients of the three grammars, all of which can be modelled as Optimality-Theoretic constraint grammars. Some of the constraint sets (ART, *CATEG, *LEX) evaluate representations directly, others (FAITH, *WARP) evaluate perceptual similarities between representations. In the roles that these constraints have to perform in phonology, we again see a skewing towards the perceptual side:

(4) *The single role of articulation in phonology*

a. In the speaker's production grammar, ***gestural*** constraints (ART) evaluate aspects of articulatory effort (§3.2)

(5) *The five roles of perception in phonology*

a. In the speaker's production grammar, ***faithfulness*** constraints (FAITH) indirectly evaluate aspects of perceptual confusion by evaluating aspects of perceptual similarity between the specification and the perceptual output (§3.2).
b. In the speaker's and listener's perception grammars, ***categorization*** constraints (*CATEG and *WARP) indirectly evaluate aspects of perceptual confusion by handling the discretization of acoustic continua into language-specific perceptual feature values (§2.2).
c. In the speaker's and listener's perception grammars, ***sequential abstraction*** constraints (OCP, LCC) determine the grouping of temporally related representations into larger units (§7).
d. In the listener's recognition grammar, faithfulness constraints (FAITH) evaluate aspects of perceptual similarity between the perceptual input and the underlying form (§5). It is possible that these constraints are the mirror images of the faithfulness constraints in the production grammar, and ranked in the same order.
e. In the comparison module, a ***gradual learning algorithm*** will take action if the learner's and adult forms are different (§4). Unlike the three grammars, this fourth processing system is not itself modelled as an Optimality-Theoretic grammar; instead, it is capable of changing the three grammars.

(6) *The role of semantics in phonology*

a. In the listener's recognition grammar, ***lexical access*** constraints (*LEX) handle the dependence of recognition on the semantic context (§5).

I will start and finish the remainder of this article with a discussion of the perception grammar. After a short review of the simplest activity of the perception grammar, namely categorization (§2), we are ready for a fundamental discussion of the role of perception in the production grammar (§3). I will then briefly touch upon the formalization of learning (§4) and of the recognition grammar (§5), before addressing the issue of empiricism with respect to substantive content in phonology (§6), which will allow us to turn to the more complicated activity of the perception grammar, namely sequential abstraction (§7).

2 The perception grammar, part one: perceptual categorization

2.1 The task of the perception grammar

What we perceive of a speech utterance in a communicative setting tends to be quite different from what is acoustically there. The acoustic speech signal contains phenomena that the peripheral auditory mechanism will identify as continua of loudness, periodicity, noise, and frequency spectra, and as temporal relations between these. The phonological surface structure that we perceive, however, is a much more structured representation: it may contain sequential tiers of discrete perceptual feature values (voicing, tone, vowel height, nasality, place, sonorance, frication) and their immediate simultaneous and sequential combinations (segments), as well as larger structures such as syllables and feet. Since all these structures are not directly observable, they can be called *covert* or *hidden* surface structures. It is the

task of the language-specific perception grammar to construct them from the overt acoustic signal.

2.2 Categorization into discrete perceptual feature values

The perception grammar classifies continuous auditory inputs into a finite number of discrete categories, by means of an interaction between several constraint families (Boersma 1998: 161–172, 336–343, 375–379). Each candidate is evaluated for its conformance to the language-specific set of phonological feature values (*CATEG) and for its closeness to the acoustic event (*WARP).

I will give only a brief example here, taken from Boersma (1998: 165). Suppose that a language entertains three vowel heights, and that these heights are associated with first formant values of 260, 470, and 740 Hz, i.e. high, mid, and low vowels, respectively (ignoring speaker normalization, which is also a task of the perception grammar). This can be expressed in the perception grammar as a low ranking of the *CATEG constraints that militate against mapping acoustic inputs to the perceptual classes /260 Hz/, /470 Hz/, and /740 Hz/, and a high ranking of the infinite number of *CATEG constraints for any other first formant class:

(7) *CATEG (vowel height: *x*)
"do not categorize into the vowel height class of *x* Hz."

Thus, *CATEG is ranked low if *x* is 260, 470, or 740 Hz, and high for all other values of *x*. Now suppose the listener hears another speaker produce a first formant of 530 Hz. By what we know about the perception of vowel height (e.g. Fry, Abramson, Eimas & Liberman 1962), the listener will probably classify this into the 470 Hz category (i.e. perceive it as a mid vowel), because that is by far the nearest category, being only 60 Hz away from the acoustic input. This phenomenon can be described in the perception grammar with a family of *WARP constraints, which militate against large discrepancies between the acoustic form and the perceptual form:

(8) *WARP (F_1: *x*; vowel height: *y*)
"do not categorize an acoustic F_1 of *x* Hz into a vowel height class of *y* Hz."

These constraints are ranked by their distance to the height category, e.g., there is a universal ranking *WARP ([530 Hz]; /470 Hz/) >> *WARP ([520 Hz]; /470 Hz/). This fixed ranking causes (or is caused by) the preference for the perception into nearby categories. The mapping of an acoustic [530 Hz] into a perceptual /470 Hz/ can now be summarized in the following tableau (only a few constraints are shown):

(9) *Perception of an acoustic form as the nearest perceptual feature value*

[530]	*CATEG (other)	*WARP ([530], /740/)	*WARP ([530], /470/)	*CATEG(/260/), *CATEG(/470/), *CATEG(/740/)
☞ /470/			*	*
/530/	*!			
/740/		*!		*

This is the simplest case. Other interesting things can happen; for instance, if *WARP is ranked higher than *CATEG(other) for distances over 100 Hz, an input of [600 Hz] will be perceived as /600 Hz/, perhaps creating a new weak category /lower mid vowel/.

2.3 Perception of feature combinations

Perceptual feature values as discussed in §2.2 are the simplest instance of language-specific hidden structure in the phonological surface form. A little more complicated is the perception of simultaneous and sequential combinations of feature values as a single, more abstract, percept. Such perceptual aggregation becomes useful if the composing feature values co-occur frequently.

For instance, in languages where the perceptual feature values /+nasal/ and /labial place/ are often simultaneously combined, it becomes advantageous for the listener to perceive the combination as a single percept /m/. Likewise, in a language where /mid vowel height/ and /front vowel place/ (the latter equivalent to the value “high” on the second formant tier) often co-occur, a single percept /e/ will emerge. One of the advantages of perceptual abstraction is seen in the lexicon, where words that contain this sound can now be represented more economically. Of course, this comes at the cost of maintaining an extra perceptual category, and it is the perception grammar that finds the optimal balance between these conflicting demands.

Perceptual integration of simultaneous combinations will thus lead to effects of segmental organization. Likewise, integration of sequential combinations will lead to the construction of larger temporal perceptual units, such as long tones (perhaps across intervening plosives), word-level nasality, segments (again), geminates, NC clusters, syllables, and feet, all of which must be considered language-specific if they are to be constructed by a general perception grammar. The formalization of sequential abstraction, which involves constraints appropriately labelled with the traditional terms OCP and LCC, will be deferred to §7, after a defence of the empiricist nature of phonological substance.

2.4 What the perception grammar is *not* about

It is probably appropriate here to prevent two possible misconceptions about what the perception grammar is.

First, the perception grammar is not about *audibility*. The fact that it maps the acoustic form [530 Hz] on the perceptual category /470 Hz/ does not mean that the listener cannot

hear the difference between the acoustic forms [530 Hz] and [470 Hz]. In fact, listeners can often identify neighbouring dialects by small acoustic differences between vowel realizations. What the perception grammar *is* about, is the balance that must be struck between the loss of phonetic information and the miscategorizations that will occur when there are too many categories. As a historical sound change, category merger has the functional virtue of reducing misunderstanding, especially when speakers from two dialects with shifted vowel systems are mixed. In this way, for instance, standard Dutch shows the effects of the mix of southern and western speakers, whereas nearly all of the more localized dialects retain larger vowel systems (Weijnen 1991).

Second, the perception grammar is not about *reporting*. It is characteristic of partially automatic behaviour such as language that people are not able to report on the details of what is going on inside. In fact, it is unusual for a listener to be able to consciously identify the stress patterns and intonation contours of a real-life utterance, whereas her correct interpretation of its semantics and pragmatics proves that her perception grammar does identify the stress patterns and intonation contours correctly.

Thus, the output of the perception grammar contains in some respects more, in other respects less information than listeners are aware of, and the linguist can only indirectly fathom its workings. For instance, the regional covariation of the degree of diphthongization and the height of Dutch /eː/ and /oː/ indicates that these vowels are not just perceived as separate segments, but that the first formant must have been autosegmental at some level of perceptual representation, for many speakers throughout the ages. In general, just as phonological phenomena have always been taken to supply evidence for underlying and other hidden structures, it is the same phonological phenomena that must lead the linguist to propose the constraints of the perception grammar and their rankings.

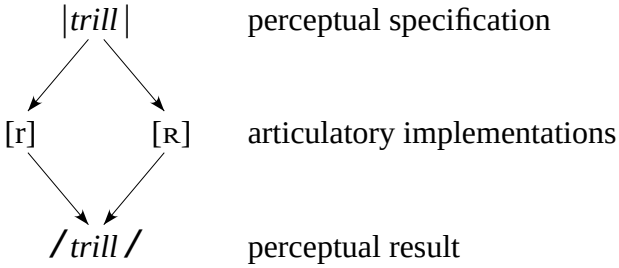
3 The role of the perception grammar in production

3.1 Speech production is perception-oriented

A simple example will show that the specification of an utterance is cast in perceptual, not in articulatory terms.

The Dutch sound written as <r> can be realized as an alveolar trill or as a uvular trill. These realizations are articulatorily quite distinct, but perceptually rather similar. Apparently, the perceptual specification for this sound is that it should be perceived as vibrating, and both articulatory implementations lead to perceptual results that are faithful to this specification:

(10) *Perception orientation in the Dutch trill*



This convergence on a single perceptual result, which is compared with a perceptual specification, is typical of the general *perceptual control loop* of human motor behaviour (Powers 1973). This loop is not specific to language, since it also controls, for instance, the way in which we travel from A to B in the presence of obstacles. Neither is this loop, if applied to language, specific to speech: in Sign Language of the Netherlands, the perceptual specifications include hand positions and finger orientation, and any articulatory implementation (angles of joints) that achieves this will produce a faithful perceptual result.

3.2 Formalization of the production grammar

The perceptual result is not always identical to the perceptual specification, as it was in (10). Rather, it tends to be the result of a negotiation between the constraints that aim to minimize articulatory effort and the constraints that aim to maximize similarity of the perceptual result to the specification.

The prototypical articulatory constraint is

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

"do not let a certain *articulator* perform a certain *gesture*, 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. According to the *local-ranking principle* (Boersma 1998), *GESTURE can be locally ranked according to articulatory effort, e.g. (11) 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 could only account for cross-linguistic statistical tendencies.

The prototypical faithfulness constraint in the production grammar is

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

"do not replace, on a certain perceptual tier (*feature*), a specified value ($value_1$) 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 and deletion of underlying material, or against the loss of specified simultaneous and sequential relations between features. Faithfulness constraints can be locally ranked according to perceptual confusion, e.g. (12) is ranked higher if $value_1$ and $value_2$ are further apart or if the condition or the environment contributes to a smaller amount of confusion, and everything else stays equal. Otherwise, the rankings are largely language-specific: again, a global measure of perceptual confusion could only account for cross-linguistic statistical tendencies.

As an example, consider the case of nasal place assimilation, in which nasal consonants, but not plosives, assimilate to any following consonant (Boersma 1998: 224). First, we note that the articulatory gain of assimilation of nasals (|an+pa| → [ampa]) is equal to the articulatory gain of assimilation of plosives (|at+ma| → [apma]) since in both cases the speaker economizes on a tongue-tip opening-and-closing gesture, as the lip gesture that replaces it will be shared with the following consonant:

(13) *Articulatory scores for* [anpa] *and* [ampa]

a. [anpa]

tongue tip:	open		closed			open		
lips:	open			closed			open	
velum:	closed		open			closed		
acoustics:	a	ã	n	n͡m	_	[p]	a	

b. [ampa]

tongue tip:	open							
lips:	open		closed				open	
velum:	closed		open			closed		
acoustics:	a	ã	m		_	[p]	a	

In the microscopic acoustic transcription below each score, we find the automatic acoustic result of the articulation; [n͡m] stands for a combined nasal (acoustically coronal, visually labial), [_] for a silence, and [^{p}] for a labial release burst. Thus, the main articulatory difference between the non-assimilating candidates ([anpa] and [atma]) and the corresponding assimilating candidates ([ampa] and [apma], respectively) lies in the violation or satisfaction of the constraint *GESTURE (tongue tip: close & open).

If it is not in the articulation, the difference between nasals and plosives must be in the ranking of the faithfulness constraints: since the plosives /p/ and /t/ are easier to distinguish than the nasals /m/ and /n/, the constraint against implementing |t| as something perceived as /p/ must be universally ranked higher than the constraint against implementing |n| as something perceived as /m/; this is expressed as the universal ranking *REPLACE (place: cor, lab / plosive) >> *REPLACE (place: cor, lab / nasal). The following tableaus show how this works out if the gestural constraint happens to be ranked in between these faithfulness constraints:

(14) *Nasals undergo place assimilation*

|an+pa|	*REPLACE (place: cor, lab / plosive)	*GESTURE (tongue tip: close & open)	*REPLACE (place: cor, lab / nasal)
[anpa] → /anpa/		*!	
☞ [ampa] → /ampa/			*

(15) *Plosives are immune to place assimilation*

|at+ma|	*REPLACE (place: cor, lab / plosive)	*GESTURE (tongue tip: close & open)	*REPLACE (place: cor, lab / nasal)
☞ [atma] → /atma/		*	
[apma] → /apma/	*!		

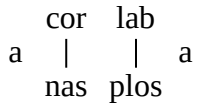
Given the fixed ranking of the *REPLACE constraints, we predict a three-way typology: apart from the language type in (14) and (15), in which only nasals assimilate, there must be a language type where *GESTURE is ranked at the bottom (nothing assimilates) and a language type where *GESTURE is ranked on top (plosives as well as nasals assimilate). This corresponds to Mohanan's (1993) implicational universal, which says that if plosives assimilate, then nasals will assimilate too (if everything else, including their place of articulation, is equal).

Since the articulatory and faithfulness constraints both reside in the production grammar, the Optimality-Theoretic tableaus that formalize their interaction are a little different from the traditional tableaus that have candidates with single 'hybrid' representations. In (14) and (15), we see pairs of representations in the candidate cells. The first member of each pair, written between square brackets, is the articulatory output candidate (phonetic form), which is evaluated by the gestural constraint. The second member of each pair, written between slashes, is the perceptual output (phonological form), whose similarity to the perceptual specification (the form written between pipes in the top left cell) is evaluated by the faithfulness constraints. Finally, the arrow that connects the two members of each pair stands for the perception grammar, which maps the automatic acoustic result of the phonetic form to the phonological surface structure.

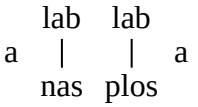
The similarity of the trancriptions for the articulatory and perceptual outputs in (14) and (15) is deceptive: both kinds of notations are just convenient shorthand. Thus, [anpa] stands for a combination of tongue, lip, and velum movements superposed on expiration and glottal adduction, as in (13), whereas /anpa/ stands for perceived voicing, nasality, and coronal and labial place:

(16) *The meaning of the perceptual shorthands* /anpa/ *and* /ampa/

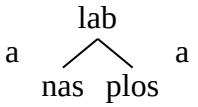
a. /anpa/:



b. /ampa/:



c. /ampa/:



Note that there are no one-to-one relationships between articulation and perception: a perceivable nasality requires not just a velum lowering, but also an airstream and preferably voicing. The distinction between articulation and perception will become more explicit in §7, where we look at perceptual abstraction.

Regarding the difference between (16b) and (16c): by assuming the non-autosegmental representation in (16b), we have been able to state the faithfulness violation in (14) simply as a replacement of /cor/ by /lab/, as we can easily see when comparing (16a) with (16b). However, if we assume instead the autosegmental representation in (16c), the difference between it and (16a) becomes more complex: the violated faithfulness constraints are now *DELETE (place: cor) and *INSERTPATH (place & nasal: lab & +). The constraints that handle the difference between (16b) and (16c) will be thoroughly discussed in §7.

3.3 Paradox solved: structuralism versus generativism

I will show that the functional model of the production grammar (Fig. 1) combines the virtues of the structuralist and generative grammar models, without copying the disadvantages of either. From Fig. 1, we can distil the following linear view of the production grammar:

(17) *The functional view of the grammar*

|underlying form|	→	[phonetic form]	→	/phonological form/
(discrete)		(continuous)		(discrete)
(perceptual specification)		(articulatory output)		(perceptual output)

We thus have a continuous representation in between two discrete representations. The observability of the phonetic form makes it necessary that the two arrows represent separate processing systems: the production grammar and the perception grammar. The ordering in (17) is natural only if we make a principled distinction between articulatory and perceptual representations, and indeed we see that former grammar models, which did not make this distinction, proposed a different order of representations.

First, there is the structuralist grammar model, which had the phonetic and phonological representations in reverse order:

(18) *The structuralist view of the grammar*

|underlying form|	→	/phonological form/	→	[phonetic form]
(discrete)		(discrete)		(continuous)
(morphophonemes)		(autonomous phonemes)		(allophones)

The main argument for the discrete phonological surface structure in (18) was that it could handle the perception of *sameness* (Bloomfield 1933: 128; Hockett 1965: 194). A similar view is nowadays the prevailing view of postlexical phonology, where the second arrow is dubbed *phonetic implementation* and rarely found worthy of investigation.

Second, there is the (early) generative grammar model (Halle 1959, Chomsky 1964, Postal 1968, Chomsky & Halle 1968), in which the underlying form was mapped by a single grammar to an observable phonetic form:

(19) *The generative view of the grammar*

|underlying form|	→	[phonetic form]
(discrete)		(continuous)

In comparison with the structuralist model, the intermediate representation had vanished. The structuralists had maintained that the phonemic form contains all and only the contrastive phonemes. This led Halle (1959) to argue that according to this view, Russian voice assimilation would be phonemic in t→d / _ b, because /t/ and /d/ are contrastive phonemes of Russian, and that it would be allophonic in tʃ→dʒ / _ b, because there is no voiced counterpart to the /tʃ/ phoneme in Russian. Thus, the arguably unitary phenomenon of voice assimilation would have to be divided among two processing systems. Apparently, the intermediate level was an artificial structure invented by the linguist, and the generative phonologists dispensed with it. However, the improvement came at a cost: the generative

model could not express sameness any longer. It is probably for that reason that phonologists have largely returned to the structuralist view of a separation between the phonological and the phonetic component of the grammar.

There is a way out of this predicament. The functional grammar model (17) combines the advantage of having phonology and phonetics together in a single grammar of parallel evaluation without intermediate levels, with the advantage of being able to express sameness on the level of the perceptual output. The structuralists and generativists shared the assumption that *if* a phonological surface structure existed, it would have to be intermediate between the underlying form and the phonetic form. The functional phonology model belies this assumption by putting the phonological form at the end.

3.4 A grammar of performance: solving Hale & Reiss' tooth-loss paradox

Hale & Reiss (1998) present several arguments against collapsing the phonological and phonetic modules. One of these arguments is about the impossibility of sudden changes in the grammar as a result of sudden changes in the body. I will show that the argument runs moot if we distinguish the roles of articulation and perception.

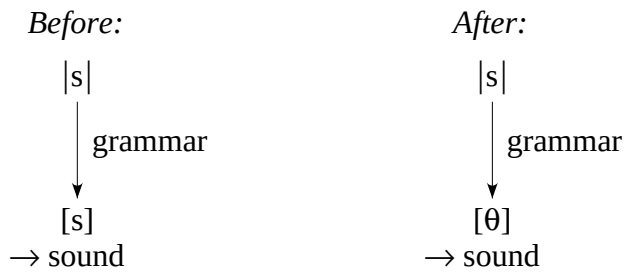
On April 27, 1998, Charles Reiss wrote to the Optimality List:

> "My reading of the generative linguistics literature is that it is about knowledge states, not behavior. If I don't start flossing, all my teeth may fall out — my pronunciation will change, but my phonology won't."

Hale & Reiss use arguments like these to discredit the theory that a grammar could describe, for instance, the performance difficulties that cause children to pronounce their utterances with poor faithfulness to the adult forms. This would mean that a lot of the acquisition literature (Smith 1973, Gnanadesikan 1995, Smolensky 1996) must be wrong.

As an example, I will discuss the pronunciation of |s|, Most speakers implement the perceptual specification |voiceless high-frequency sibilant noise| with the help of a tongue-grooving gesture that allows them to direct a jet of air along the ridges of some medial teeth. If the medial teeth fall out, the ridges will no longer contribute to the loudness of the noise, and the perceptual result will be a non-sibilant fricative, perhaps classified as /θ/. In derivational theories with hybrid phonological representations, this would be described in the style of Smith (1973) as the addition of a rule /s/ → /θ/:

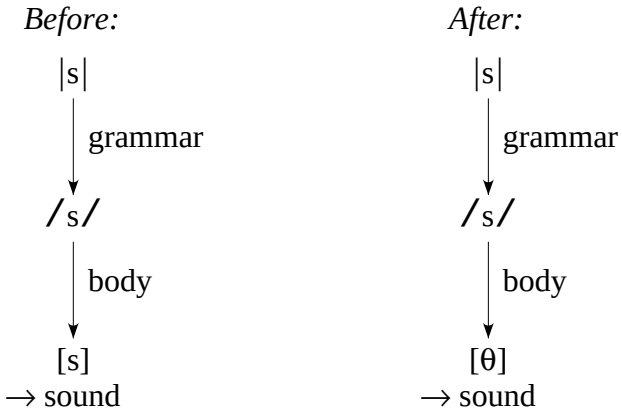
(20) *Generative tooth loss*



If the tooth loss is instantaneous, so is the grammar change. But the grammar is about knowledge states, which should not change suddenly as a result of a performance problem. Hale & Reiss correctly conclude that a monostratal grammar model cannot work. Their

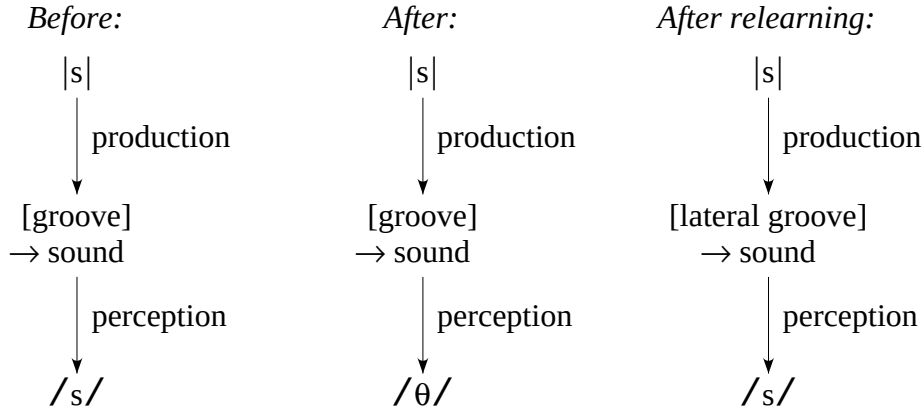
solution is a model in which the phonological component (“the grammar”) precedes the performance system (“the body”):

(21) *Structuralist tooth loss*



The alternative solution, of course, is to distinguish between articulatory and perceptual representations, and propose the functional grammar model:

(22) *Functional tooth loss*



We see that the production grammar does not change immediately when the teeth are lost. The perception grammar does not change either; its output has changed only because the sound has changed. Thus, by placing the sound in the middle instead of at the end, we can collapse phonology and phonetics, competence and performance, grammar and body.

A note must be made here about what the production grammar does. It is seen to *determine* the articulatory output, though it *controls* the perceptual output. This means that the grammar may actually change as the ultimate result of the loss of the medial teeth, for instance when the speaker learns that she can produce a sibilant sound by deflecting a jet of air in a lateral direction, so that it hits the ridges of some of the remaining teeth. This situation, which is seen on the right in (24), is taken up next.

4 The role of perception in learning the production grammar

4.1 The gradual learning algorithm

As we saw above, the functional model of phonology restores the role of *sameness* in the grammar. One application of this concept is found in the Gradual Learning Algorithm (Boersma 1997, 1998, to appear a; Boersma & Hayes 1999), which allows the learner to change her production grammar on the basis of data from her language environment. This algorithm is *error-driven*, which means that the learner's comparison module (Fig. 1) will change her grammar only if her own perceptual output is not the *same* as her perception of the corresponding adult form. As an example, tableau (23) shows an overly faithful stage in the acquisition of nasal place assimilation, which is typical for four-year-old children (for Spanish: Hernández-Chávez, Vogel & Clumeck 1975):

(23) *Detecting an adult faithfulness violation in nasal place assimilation*

underlying form = |an#pa| adult surface form = /ampa/	*REPLACE (place: cor, lab / plosive)	*REPLACE (place: cor, lab / nasal)	*GESTURE (tongue tip: close & open)
☞ [anpa] → /anpa/			←*
√ [ampa] → /ampa/		*!→	

When concatenating the morphemes \|an\| and \|pa\|, the learner produces something that she perceives as /anpa/, whereas an adult produces a form that the learner perceives as /ampa/. The learner will assume that the adult form is correct (√) and, noticing that the two forms are not the same, conclude that her own winner is incorrect (*☞*). Our error-driven learner will then take action by raising the rankings of all constraints violated in her incorrect form and lowering the rankings of all constraints violated in the correct form, by a small amount (*plasticity*) along a continuous scale of constraint ranking. These changes are indicated by the arrows in the violation cells. After the learner has processed a fair number of learning data, the constraints will be ranked in the adult order. This algorithm is guaranteed to converge to a grammar equivalent to the adult's, and leads, if combined with noisy evaluation, to realistic learning curves (Boersma 1998: 284, Boersma & Levelt 1999).

4.2 What is being compared?

The comparison module does not just compare phonemic representations. Since we know that children learn to replicate the adult system to a much finer degree of phonetic detail, more concrete levels of representation must be available to the comparison module as well. At the other end of the spectrum, children must learn to correct lexical stresses on a more abstract level of representation. Thus, several degrees of abstraction have to be available to the learner simultaneously. In §7.7, we will see an example of this simultaneity. As for now, we will first discuss an example of the learning of a single feature value.

4.3 Relearning after tooth loss

We return to our tooth loss example of §3.4. The normal production of |s| can be described with the ranking of an appropriate gestural constraint below an appropriate faithfulness constraint:

(24) *The normal situation*

|s|	*DELETE (noise: sibilant)	*GESTURE (tongue: groove)
☞ [groove] → /s/		*
[no groove] → /θ/	*!	

After the loss of the medial teeth, the grammar has not changed, though if both the grooving and the non-grooving gesture now result in the same perception /θ/, the speaker may decide to do without the complicated grooving gesture:

(25) *Articulatory output change without a change in the grammar*

|s|	*DELETE (noise: sibilant)	*GESTURE (tongue: groove)
[groove] → /θ/	*	*!
☞ [no groove] → /θ/	*	

Thus, the violation marks have changed, and so has the articulatory output. For this articulatory change to happen, the speaker had to learn that her grooving gesture now violates *DELETE (noise sibilant), i.e., her knowledge of the articulation-to-perception map had to be updated. Still, however, the production grammar (i.e. the constraint ranking) has not changed.

Next, the speaker will notice a discrepancy between her own speech and what is the norm in her environment: her own output lacks the perceptual feature /sibilant/, although she perceives that other speakers do produce this feature. Here is where learning comes in. As soon as the unfortunate speaker learns that a laterally directed groove can compensate for her loss, a relevant gestural constraint enters the grammar from above, i.e. initially ranked at the top (Boersma 1998: 280, to appear a):

(26) *Learning to compensate for tooth loss*

specification: |s| norm: /s/	*GESTURE (tongue: lateral groove)	*DELETE (noise: sibilant)	*GESTURE (tongue: groove)
[medial groove] → /θ/		*	*!
☞ [non-groove] → /θ/		←*	
√ [lateral groove] → /s/	*!→		

Gradually, the two moving constraints will reverse their rankings, and the speaker will produce the normal sibilant fricative again. Note that, analogously to what was said in §2.4 about the perception grammar, tableau (26) is not about pronounceability: the presence of the [lateral groove] articulation in (26) already means that it is pronounceable; a low ranking will additionally be needed in order to ensure that the speaker considers it worthwhile to produce this in a communicative situation as well.

5 The role of perception in the recognition grammar

As we see from Fig. 1, the recognition grammar performs more or less the reverse mapping from the production grammar.

Smolensky (1996) has proposed that we can use a single grammar for production and comprehension. His example is the English word |kæt| 'cat'. In production (i.e. in going from the underlying to the hybrid surface form), a young child may pronounce this as [ta] because of the high ranking of some structural constraints like *CODA:

(27) *Unfaithful production*

|kæt|	*CODA	FAITH
[kæt]	*!	
☞ [ta]		**

In comprehension (i.e. in going from the hybrid surface form to the underlying form), the same grammar will still map the adult form [kæt] on the correct underlying form:

(28) *Faithful comprehension*

[kæt]	*CODA	FAITH
☞ |kæt|	*	
|skæti|	*	*!*

This works because the structural constraints now evaluate the *input*, so that in comprehension they assign the same number of violations to all candidates and cannot contribute to determining the winning underlying form. Thus, tableaus like (28) always select the most faithful candidate.

But that is also the drawback of Smolensky's approach: it does not work in cases where the correct underlying form exhibits faithfulness violations. Consider the case of final devoicing in Dutch, where |ʀɑt| 'rat' and |ʀɑd| 'wheel' fall together on the surface:

(29) *Final devoicing in Dutch*

|ʀɑt| 'rat' → [ʀɑt] (cf. [ʀɑtə] 'rats')
|ʀɑd| 'wheel' → [ʀɑt] (cf. [ʀaːdərə] 'wheels')

If a tableau can only consider the phonology, the faithful candidate always wins, even if the semantic context would require ‘wheel’:

(30) *Failure to recognize* |ʀɑd| *‘wheel’*

[ʀɑt]	*VOICEDCODA	FAITH
☞ |ʀɑt|		
|ʀɑd|		*!

In order to be able to recognize ‘wheel’, we should take into account the semantic context, and we will do this by introducing a lexical-access contraint (Boersma, to appear b):

(31) *LEX (|*underlying form*| ‘*concept*’ / *context*)

“Do not recognize the sign |*underlying form*| ‘*concept*’ in the given semantic *context*.”

These constraints can be locally ranked according to semantic context and to token frequency. Thus the constraint *LEX (|ʀɑt| ‘rat’ / ‘turn’) militates against recognizing the sign |ʀɑt| ‘rat’ in the context ‘turn’, and it will be ranked higher than *LEX (|ʀɑt| ‘rat’ / ‘gnaw’). Likewise, *LEX (|ʀɑd| ‘wheel’ / ‘turn’) will be ranked higher than *LEX (|ʊil| ‘wheel’ / ‘turn’), because |ʊil| is a more common word for ‘wheel’ than |ʀɑd| is. In tableau (32), we see how the functional-style recognition grammar manages:

(32) *Success in recognizing* |ʀɑd| *‘wheel’*

perc. input = /ʀɑt/	*LEX (|ʀɑt| ‘rat’ / ‘turn’)	*GESTURE (obstruent voicing)	*REPLACE (height etc.)	*LEX (|ʀɑd| ‘wheel’ / ‘turn’)	*LEX (|ʊil| ‘wheel’ / ‘turn’)	*REPLACE (–voice, +voice)
|ʀɑt| ‘rat’	*!					
☞ |ʀɑd| ‘wheel’				*		*
|ʊil| ‘wheel’			*!**		*	

Note the interaction between phonological and semantic constraints: the third candidate is ruled out on the basis of its lack of phonological similarity to the perceptual input, and the first candidate is ruled out on the basis of its distance to the semantic context.

As far as Smolensky’s proposal of grammar sharing is concerned, we can note (a) that lexical-access constraints evaluate the *underlying form*, so that in the production grammar they assign the same number of violations to all candidates and cannot contribute to determining the winning articulatory form; and (b), that the faithfulness constraints that are ranked low in the production grammar will tend to correspond to reverse counterparts that are ranked low in the recognition grammar, since the listener must undo the same faithfulness violations that the speaker produces. So it may still be the case that the faithfulness contraints and their relative rankings are shared between the two grammars.

6 Function and arbitrariness, innateness and empiricism

Saying that languages are organized according to functionally optimizing principles is not the same as saying that their organization must be *directly* functional, i.e., is not innate. After all, evolution caused many advantageous properties to become innate, and those connected with phonology include at least the following:

(33) *What's innate in phonology*

a. *Peripherals* of speech production (versatile tongue and larynx) and auditory perception (spectrum, periodicity, noisiness, temporal coincidence and ordering, and intensity).
b. Cognitive capabilities: *categorization* of perceived entities into classes of partial equivalence, *abstraction* of simultaneity relations and sequential relations into higher-level constructs, wild *generalization* and extrapolation, and the manipulation of *arbitrary symbols* (storage, retrieval, access).
c. *Decision making*: stochastic constraint grammars and a gradual learning algorithm.
d. *Functional drives*: the desire to understand and make oneself understood, and laziness.

Some of these properties, including some that may have arisen as a side effect of the evolution of language, have clear *general* evolutionary advantages; others properties are more specific to the language faculty; however, none of these properties refers to specific substance in phonology, i.e. the specific features and structures that have been proposed in the generative literature. And indeed, I have made several empiricist suggestions above: §2.3 suggested that perceptual feature values can be merged on the basis of what is functionally advantageous in the language at hand, and §4.3 suggested that new articulatory gestures can be introduced into the grammar as a result of performance problems. The claim must be that at least the *substance* of phonology must be the result of a general learning mechanism (which itself, of course, must be innate):

(34) *Learned phonological entities*

a. *Phonological features*. Sign languages use different features than spoken languages.
b. *Phonological feature values*. Languages tend to divide up the vowel height dimension into equally sized categories. There is no evidence that four-height systems are extended three-height systems, or that three-height systems are really four-height systems with a gap.
c. *Prosodic constituents*. Regarding the controversies about the mora, the syllable, and higher prosodic constituents, it seems safe to say that languages build parochial sequential structures (§7).

If these entities are not innate, then any devices, parameters, or constraints that directly refer to specific substance cannot be innate either. Instead, they must be regarded as the results of interactions between more fundamental innate principles:

(35) *Non-elements of phonological theory*

a. *Autosegmental spreading*. In |an+pa| → /ampa/, the spreading of the lip gesture is forced by a high-ranked faithfulness constraint for consonantal nasality. Without spreading, the deletion of the tongue tip gesture would give rise to /aãpa/, which does not faithfully render consonantal nasality.

b. *OCP effects*. In English |kɪs+(ɪ)z| → [kɪsɪz] 'kisses', the epenthesis is forced by a high-ranked faithfulness constraint for sibilancy. Without epenthesis, sequential abstraction (§7) would cause the candidate [kɪsz] to be heard as /kɪs/, in which one of the underlying sibilants does not surface.

c. *Feature geometry*. The prime evidence for the 'place node' (McCarthy 1988) is the fact that nasal place assimilation tends to occur before labials, coronals, and dorsals, and not before pharyngeals and laryngeals. However, this grouping must already be expected if faithfulness for consonantality is simply ranked high (Boersma 1998: 442): changing /n/ to /m/ or /ŋ/ preserves consonantality, whereas changing it to a nasalized pharyngeal does not.

d. *Phonetically grounded constraints* (Archangeli & Pulleyblank 1994): the usual practice in explaining nasal place assimilation (e.g. Padgett 1995) is to attribute this phenomenon directly to a specific structural constraint like "change the place of a nasal" that outranks the general faithfulness constraint "don't change place". But typologically more correct is to regard it as an interaction between the more fundamental "don't move the tongue tip" and "don't change the place of a nasal" (Boersma, to appear d).

e. *Syllable constraints like *CODA* (Prince & Smolensky 1993). |kalamitɛit| is scanted as [kaː laː mi tɛit], not as [kaːl ʔaːm ʔit ʔɛit], because the prevocalic acoustic cues here are better than the postvocalic cues, and, moreover, the second candidate violates *INSERT (glottal stop).

A related empirically testable claim is that all universal phonology must be directly functional and that all arbitrary phonology must be language-specific, i.e. that **there are no arbitrary substantive universals in phonology**. This is a rather strong claim, since it is immediately falsified as soon as anyone comes up with a substantive universal that has no direct functional explanation. The evidence to date, however, shows that the reverse falsification is more likely to occur: all proposed substantive phonological universals seem to have directly functional exceptions, i.e., phonology seems to be functionally perfect. Such detailed exceptions are not expected for innate properties. Apart from the examples in (35), a notable example is found in the *sonority hierarchy*, which is slightly different for syllabification (/h/, being voiceless, prefers margins) than for susceptibility to spreading of nasality (/h/, not being perceptually affected by nasality, does not block), so that it seems that not even the sonority hierarchy, which is at work in most languages, is a good candidate for an arbitrary universal (Boersma 1998: 455).

To sum up, the ingredients of a functional theory of phonology are:

(36) *Ingredients of a functional theory of phonology*

a. functional principles expressed directly as OT-type constraints;

b. arbitrary facts of the language, plus generalizations, also expressed as constraints.

I will not have much to say here about the second ingredient; it is the one that is generally ignored by every universalist theory of phonology. The subject matter of the first ingredient is comparable to that of autosegmental phonology and feature geometry.

7 The perception grammar, part two: sequential abstraction

In this section, I will show how the perception grammar handles the integration of sequences into larger aggregates. This involves a constraint that could be called MERGE, but in order to connect to the existing literature, I will call it OCP. I will first show that a logical contradiction arises if, as several people have proposed, OCP is seen as a violable constraint in the production grammar. I then conclude that OCP must reside in the perception grammar, and show how it formalizes OCP effects.

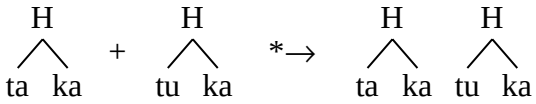
7.1 The OCP in generative phonology

The Obligatory Contour Principle (OCP) was originally introduced in autosegmental phonology as an inviolable constraint on representations. It says "adjacent identical elements are forbidden" (McCarthy 1988). This means, for instance, that the tones in the phonetic form [jévésè] are never represented as HHL, but always as HL:

(37) *The OCP as an inviolable constraint on possible representations*

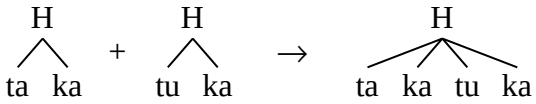
One of the effects of this OCP is the merger of two adjacent identical elements. Consider two morphemes that surface as [táká] and [túká]. Underlyingly, they both carry a single high tone: |H-taka| and |H-tuka|. Now concatenate the two, giving an underlying form |H-taka + H-tuka|. The OCP says that the result cannot be the simple concatenation:

(38) *Impossible effect of concatenation*



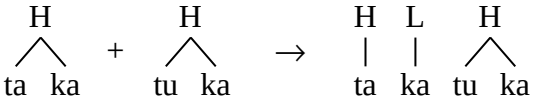
If the phonetic form is simply [tákátúká], it must be represented with a single H:

(39) *Merger as an OCP effect*



The drawback of this common merger is that one of the two underlyingly present high tones does not reach the surface. In some languages, therefore, the result will be the epenthesized form [tákàtúká], with a HLH sequence:

(40) *Epenthesis as an OCP effect*



The intervening low tone causes satisfaction of the OCP, because it causes the two high tones to be non-adjacent. The advantage is that both underlying tones are present on the surface, but the drawback is that the surface contains a non-underlying low tone.

The forms in (39) and (40) are two ways to satisfy the OCP. In Optimality Theory, the constraint OCP has been proposed as being one of the many constraints in a grammar consisting of strictly ranked constraints (Myers 1994, Urbanczyk 1995):

(41) *The OCP as a production constraint*

|H-taka + H-tuka|	OCP	*DELETE (tone: H)	*INSERT (tone: L)
H H ta ka tu ka	*!		
H ta ka tu ka		*!	
☞ H L H ta ka tu ka			*

This neatly shows how the language ranks the disadvantages of the various solutions.

7.2 The problem with the OCP in the production grammar

I will show that the above Optimality-Theoretic account is incompatible with the following assumption, which is tacitly shared by most phonologists:

(42) *The structuralist assumption*

> "Within a given language, every phonetic output form has only one phonological surface representation."

This non-neutralizing property of phonetic implementation has been the main criterion for identifying the intermediate representation in the structuralist grammar model (18):

\|underlying form\| → /phonological surface form/ → [phonetic form]

As before, the first arrow is "phonology", the second arrow "phonetic implementation".

In Optimality Theory, however, an OCP in the production grammar should be *violable* like all constraints. This means that it must be logically possible that OCP is ranked below the tonal faithfulness constraints. In this case, OCP would allow two high-toned morphemes to concatenate without change:

(43) *OCP allows identical adjacent elements*

|H-ta + H-pa|	*DELETE (tone: H)	OCP
☞ H H | | ta pa		*
H ◿◺ ta pa	*!	

Thus, if OCP is violable, the grammar must allow adjacent identical elements. However, a monomorphemic high-toned morpheme would surface unchanged:

(44) *OCP disallows identical adjacent elements*

|H-tapa|	*DELETE (tone: H)	OCP
H H | | ta pa		*!
☞ H ◿◺ ta pa		

This would mean that in one and the same language, the phonetic form [tápá] can have two different phonological surface representations, depending on the underlying form. This neutralization violates the structuralist assumption. Therefore, the existence of a violable OCP in the production grammar is incompatible with that assumption.

7.3 The solution to the OCP problem: it belongs in the perception grammar

The problem identified in the previous section can be solved by reversing the order of the two surface representations with respect to the structuralist grammar model, so that the production grammar looks like (17):

\|underlying form\| → [phonetic form] → /phonological surface form/

The first arrow is "phonology & phonetics", the second arrow "perception grammar". Note that this satisfies the structuralist assumption (42) *trivially*: every phonetic form can yield only one phonological surface form; this is guaranteed by the direction of the second arrow, which is reversed with respect to the structuralist model.

The OCP, then, must be a constraint in the perception grammar, since it takes part in the building of covert structure, e.g. in deciding whether a phonetic sequence of high tones on two consecutive syllables must phonologically be regarded as one or two H values on the perceptual tone tier. This perceptually abstracting OCP can be formalized as follows (Boersma 1998: 241):

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

"A sequence of two acoustic cues cue_1 and cue_2 is perceived as a single value *x* on the perceptual tier *f*, **despite** the presence of some intervening material *m*."

In our example, the relevant constraint is OCP (tone: H; V́ | $_\sigma$][$_\sigma$ | V́). This constraint says that two phonetic high tones should be perceived as a single H value on the perceptual tone tier, if no more than a syllable boundary intervenes between the two (the classification of a certain fundamental frequency as a "phonetic" high tone must also be handled by the perception grammar, as must the assignment of syllable boundaries).

In determining whether perceptual aggregation will occur or not, OCP must compete with a constraint that *dis*favours abstraction. This constraint is LCC (Line Crossing Constraint), which is the other traditional representational constraint of autosegmental phonology (Boersma 1998: 241):

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

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

Like most of the constraints inroduced earlier, OCP and LCC can be locally ranked. The OCP constraint is ranked higher if the sequential combination of cue_1 and cue_2 is more common, and it is ranked lower if there is more intervening material. The reverse correlations hold for the LCC.

Now, depending on the relative ranking of OCP (tone: H; V́ | $_\sigma$][$_\sigma$ | V́) with respect to LCC (tone: H; V́ | $_\sigma$][$_\sigma$ | V́), the perception grammar will map the high tones of [tápá] either on a single perceptual H, or on two:

(46) a. OCP (tone) >> LCC (tone):

tone: H –
/ \ (Mende)
t a p a

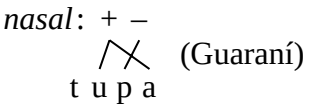
b. LCC (tone) >> OCP (tone):

tone: H – H
| | | (Chinese)
t a p a

The assignment of a high OCP to Mende is based on the fact that most tone rules in Mende regard tone as autosegmental (Leben 1973); note that a low-ranked line crossing constraint is violated, since we find a consonant or syllable boundary in between the phonetic high tones. The assignment of a high LCC to Chinese is based on the fact that each syllable in Chinese carries its own tone contour, so that it is advantageous for a listener not to collapse a sequence of two high-toned syllables only to find out that this merger has to be subsequently undone by the recognition grammar. The distinction between these languages is thus correlated to the commonness of co-occurrence of the two cues: in Chinese, a sequence of two high-toned syllables is as coincidental as any other sequence of tone contours, whereas in Mende such a sequence would often be a sign of a single lexical tone.

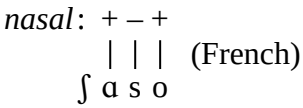
This discussion does not apply to tone alone. Other possibly autosegmental features, like nasality, behave quite analogously:

(46) a. OCP (nasal) >> LCC (nasal):



(Guaraní)

b. LCC (nasal) >> OCP (nasal):



(French)

In Guaraní, which has morpheme-level nasality and a high correlation of nasality values in consecutive syllables, the form [tũpã] 'god' will be perceived with a single nasal (Boersma, to appear c); in French, where the nasality of adjacent syllables is not related, the form [ʃɑ̃sõ] 'song' will be perceived with two separate nasals.

7.4 The perception of place

The examples of autosegmental nasality and tone are already on a fairly abstract level, i.e., the intervening material must be expressed in elements that are themselves fairly abstract covert structures like non-nasal consonants or even syllable boundaries. The OCP is also at work, however, in the integration of more concrete acoustic material such as place cues.

Consider the utterance normally transcribed as [apa]. In almost all languages this will be perceived as a sequence of three segments, i.e. as /apa/. The middle segment can be described as /voiceless labial plosive/. However, the perceived labiality must come from two labial place cues (a formant transition from the preceding vowel and a release burst into the following vowel), which are separated by a stretch of silence, i.e., the microscopic acoustic transcription of [apa] is [[a pˀ _ ᵖ a]], where [pˀ] is the transition (i.e. an unreleased labial stop), [_] is the silence, and [ᵖ] is the labial release burst. In order to perceive the two labial cues as a single /labial/ value on the perceptual place tier, the listener must have a high-ranked OCP (place: labial; transition | silence | burst); otherwise, she would perceive the two cues as two separate values on the place tier:

(47) *Integration of place cues in intervocalic short plosives*

acoustics: [[a pˀ _ ᵖ a]]	OCP (place: labial; transition | silence | burst)	LCC (place: labial; transition | silence | burst)
place: lab – lab a ˀ _ ᵖ a	*!	
☞ *place*: lab – a ˀ _ ᵖ a		*

Note that the winning candidate indeed shows crossing association lines: the silence that intervenes between the two labial cues must be regarded as a null value on the perceptual place tier.

In short plosives in intervocalic position, the integration of the two place cues is probably nearly universal, because short intervocalic plosives tend to be abundant in nearly every language in which they occur at all. A less universal integration will be found in cases where there is more intervening material or a less common co-occurrence of the cues. Starting from the short plosive, the next simplest case is the long plosive, which simply has more

intervening material, namely a longer silence, between the two place cues. In this case, we may expect languages to behave differently, depending on the commonness of co-occurrence:

(48) a. OCP (place; | _ː |) >> LCC (place; | _ː |):

place: lab –
a ⁿ _ː ᵖ a (Italian)

b. LCC (place; | _ː |) >> OCP (place; | _ː |):

place: lab – lab
a ⁿ _ː ᵖ a (English)

The notation OCP (place; | _ː |) abbreviates OCP (place: lab; pⁿ | _ː | ᵖ), where [_ː] stands for the long silence. The assignment of a high OCP to Italian is based on the common occurrence of tautomorphemic geminates in this language; the assignment of a high LCC to English is based on the fact that a phonetic geminate in this language is even less common than most other consonant sequences, and always signals a morpheme boundary, so that it is advantageous for a listener to send a couple of separate labial values to the recognition grammar.

We can go one step further. In many languages, a homorganic nasal-plosive sequence like [ŋk] is very common, so it may be perceived with a single velar place value. Quite probably, [ŋsk] will be perceived with two separate velar place values because of its expected relative rarity and the larger amount of intervening material on the perceptual place tier, namely a rather loud alveolar noise. The perception grammar of such a language must have a high OCP (place) for intervening short silences (as in [[ŋ_ᵏ]]), and a high LCC (place) for intervening sibilants (as in [[ŋs_ᵏ]]):

(49) *Nasal-plosive place integration across an intervening short silence*

[[ŋ_ᵏ]]	LCC (pl: velar; side | s_ | bu)	OCP (pl: velar; side | _ | bu)	LCC (pl: velar; side | _ | bu)	OCP (pl: velar; side | s_ | bu)
vel – vel ŋ _ k		*!		
☞ vel – ŋ _ k			*	

(50) *Nasal-plosive place separation across an intervening sibilant*

[[ŋs_ᵏ]]	LCC (pl: velar; side | s_ | bu)	OCP (pl: velar; side | _ | bu)	LCC (pl: velar; side | _ | bu)	OCP (pl: velar; side | s_ | bu)
☞ vel alv vel ŋ s _ k				*
vel alv ŋ s _ k	*!			

In these tableaus, “side” stands for the oral side-branch resonance during velum lowering, and “pl” and “bu” abbreviate “place” and “burst”. A case in which the rankings in (49) and (50) are crucial is diminutive formation in Limburgian (Boersma 1998: 242, 432). In this language, the diminutive morpheme is expressed as lenition + umlaut + tone shift + |kən|. When applied to the stem |dɛŋ̄g| ‘thing’, the first three parts of this morpheme leave the concatenation |déŋ+kən| as the specification relevant to our purposes. The surface form, however, is /déŋskə(n)/:

(51) *Epenthesis of* /s/ *in Limburgian*

underlying: vel | ŋ + vel | k	*DELETE (place: velar)	*INSERT (noise: sibilant)
[déŋkə] → vel ⋀ ŋ k	*!	
☞ [déŋskə] → vel alv vel \ | / ŋ s k		*

In this tableau, I have assumed that the perception grammar ultimately maps place values on segments (i.e. labelled timing slots), so that we see basically segmental representations appear in the perceptual output candidates. However, the /ŋk/ candidate, though consisting of two segments, contains a single velar place value. As in (39), the primary effect of the OCP constraint on the simple concatenation is a faithfulness violation in the production grammar. The observable traditional ‘OCP effect’, however, is the epenthesis of /s/ (cf. (40)). We can see that both OCP-LCC rankings in (49) and (50) are crucial: if LCC had been ranked high for intervening silences, the first articulatory candidate in (51) would have given rise to a perceptual form with two separate velar place values, thus incurring no marks for *DELETE and becoming the winner; likewise, if OCP had been ranked high for intervening sibilants, the second candidate would have violated *DELETE as well, so that, again, the first candidate would have won.

7.5 What’s a segment?

A theory that denies the innateness of phonological substance cannot view the segment as a cross-linguistic concept. Abstract units will have to emerge in a language-specific way as a result of general simultaneous and sequential abstraction in perception.

But languages do have similar types of abstraction. First, there is simultaneous abstraction (§2.3): |m| is not only the feature [labial], the feature [nasal], and the *path* [labial & nasal], but also the higher-level construct [labial nasal], if these features commonly co-occur in the language. Second, there is sequential abstraction: while [[a pˀ _ ᵖ a]] is usually perceived as /apa/, languages perceive [[a pˀ _ːᵖ a]] variably as /apːa/ or /appa/, and the same a priori ambiguity exists for whether [ts] and [mp] are basically perceived as two /p/-

like units or as one; things like [msp], with even more intervening material, will generally be perceived as two separate labials.

Thus, simultaneous and sequential abstraction lead to properties commonly associated with *root nodes* and *timing slots*, respectively. Both of these have traditionally been correlated with the notion of segment.

7.6 What's a mora?

The mora is a label used variably for language-specific sequences. Proposals for bimoraicity typically depend on what is functional in the language at hand. When the mora is proposed to account for syllable weight, every VC sequence, if any, will count as bimoraic, regardless of the nature of the C (McCarthy & Prince 1986, Hayes 1989). When the mora is proposed as a tone-bearing unit, however, V+sonorant sequences tend to count as bimoraic but V+obstruent sequences tend to be monomoraic (Lithuanian: Kenstowicz 1971; Limburgian: Hermans 1984; but even obstruents can bear, though not show, tone in Ga'anda: Kenstowicz 1994: 364).

7.7 What's a syllable?

The syllable may be the most frequently discussed linguistic construct, and this alone is witness of its likely non-universality. Take the Dutch word /mˈɛləkˌ/ 'milk'. Linguists and non-linguists alike may come up with three different syllabifications. Booij (1995) consistently writes this type of word with a single syllable /mɛlk/, probably because the schwa may be absent if the word is affixed, as in /mˈɛl(ə)kən/ 'to milk' or /mˌɛlkəʀˈɛi/ 'dairy farm'. However, if we count the number of local maxima of sonority, we get two syllables: /mɛ.ləkˌ/, and that is also the subdivision of the falling minor third used when calling: [mɛ́ː ləːk] 'come home I've got milk for you / finally give me my milk'. And if we count released consonants, we get three of them: /mɛ.lə.k/ (e.g. Kaye, Lowenstamm & Vergnaud 1985).

Let us pursue Dutch syllabification further. Coronals seem to be special in this language in that they can occur in places where labials and dorsals cannot, namely in the margins of syllables with consonant clusters. But as the following list shows, the specialness of the coronal sibilant fricative is different from that of the coronal stops:

(52) *Different kinds of coronal specialness in Dutch*

a. Sibilants occur at either end:	pɑks	spɑk	smɑk
b. Plosives occur only finally:	pɑkt	*tpɑk	*tmɑk

Thus, both kinds of coronals can occur at the end of a coda cluster. Historically, this came about by two processes of merger of words ending in -CC or -CəC. If the second C was a labial or dorsal, both endings neutralized to -CəC, and if the second C was a coronal, both endings neutralized to -CC:

(53) *The history of final clusters in Dutch*

Labial or dorsal ending: -CVC]				*Coronal ending*: -CC			
ɦɛnəp	→	ɦɛnəp	‘hemp’	lɪxt	→	lɪxt	‘light’
mɔnək	→	mɔnək	‘monk’	lʊft	→	lʏxt	‘air’
mɛlək	→	mɛlək	‘milk’	bɛst	→	bɛst	‘best’
aːrd+əx	→	aːʀd+əx	‘nice’	ɦoːvət	→	ɦoːft	‘head’
bɛrx	→	bɛʀəx	‘mountain’	ɑmbət	→	ɑmt	‘profession’
ɦɛlp	→	ɦɛləp	‘help’	loːp+ət	→	loːp+t	‘walk-3SG’
ʊɔlf	→	ʊɔləf	‘wolf’	maːk+ədə	→	maːk+tə	‘made-1SG’
kɑlk	→	kɑlək	‘chalk’	zɪŋ+ət	→	zɪŋ+t	‘sings’

As can be seen, the process applied within as well as across morphemes.

In initial position, the generalization is different: coronal sibilants are allowed at the margin, but coronal stops are not (except before an approximant). The correct generalization for coronal stops, including nasals, in initial and final clusters is that they occur in the *second* position in each cluster:

(57) *Symmetry for stop clusters in onsets and codas*

kni ‘knee’	zɪŋt ‘sings’	Allowed: non-coronal stop followed by coronal stop.
*tmi	*zɪnk	Not allowed: coronal stop followed by non-coronal stop.

As we see, clusters of a nasal and a non-nasal stop obey the sonority hierarchy, but in each licit form (/kni/ and /zɪŋt/), it is the coronal that comes second. The explanation starts from the observation that coronal ballistic stop closures are faster than labial or dorsal ones. In order to hear the releases of both stops, |atk| must therefore be rendered as [[a _ ^{t} _ _ ^{k}]], and |akt| must be rendered as [[a _ _ ^{k} _ ^{t}]]. The overall durations of these two forms are the same, but the |kt| cluster will be perceptually more coherent than |tk|. To see how this perceptual coherence could have played a role in syllabification, we consider six possible levels of perceptual abstraction, labelling each with an appropriate arbitrary symbol σ_1...σ_6 and with a mnemonic though unsubstantial “level” name:

(58) *Possible Dutch levels of perceptual abstraction*

σ_1:	m ^{m} ɛ l^{ˈ} l ^{l} ə k^{ˈ} _ ^{k} _ ^{t}	“acoustic” cue level
σ_2:	m ɛ l ə k t	“segment” level
σ_3:	mɛ lə k t	“release” level (government phonology)
σ_4:	mɛ lə kt	“cluster” level
σ_5:	mɛ ləkt	“phonetic syllable” level (sonority pattern)
σ_6:	mɛlkt	“abstract syllable” level

The “acoustic” level contains transitions, states, releases, and silences, all of which may be phonologically relevant, especially the releases (Saussure 1916:79–95, Steriade 1993).

Evidence could be adduced for each of the levels in (58). The point here is that they are arbitrary language-specific constructs, and that they may exist all at the same time. The formalization of the construction of these levels has to run via the OCP and LCC constraints again. As a function of level of abstraction, these constraints can once more be locally ranked: OCP (σ_{i+1}; *cue*$_1$ | *m* | *cue*$_2$) >> OCP (σ_i; *cue*$_1$ | *m* | *cue*$_2$), i.e., two cues with given

intervening material *m* have a larger chance to sit together in a high-level construct than to sit together in a low-level construct.

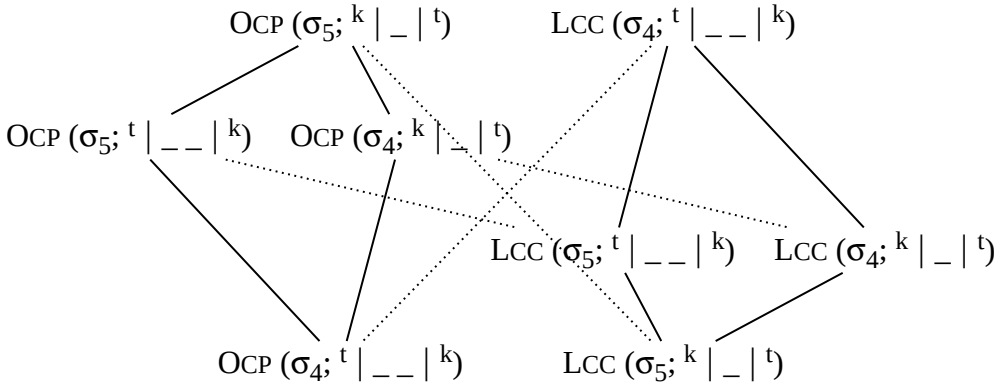
I propose, then, that at a certain point during the history of Dutch, the final clusters in /mɔn(ə)k/ ‘monk’ and /maːk(ə)t/ ‘makes’ were perceived differently on the σ_4 level:

(59) *Possible earlier Dutch levels of perceptual abstraction*

	^{n} ə _ _ ^{k}	^{k} ə _ ^{t}	acoustic level
σ_2:	m ɔ n ə k	m aː k ə t	“segment” level
σ_3:	mɔ nə k	maː kə t	“release” level
σ_4:	mɔ nə k	maː kət	“cluster” level
σ_5:	mɔ nək	maː kət	“phonetic syllable” level (= σ_6)

The difference between the two σ_4 representations is a consequence of the higher ranking of OCP (σ_4) for [[^{k} ə _ ^{t}]] than for [[^{n} ə _ _ ^{k}]], which has more intervening material (a longer silence). Therefore, the two releases in [[^{n} ə _ _ ^{k}]] will get integrated only at the fifth level, where they are combined into the same “syllable”. This state of affairs is depicted in the following tableau:

(60) *Different integration for dorsal-final and coronal-final clusters*



In this tableau, fixed rankings are shown by solid lines, and language-specific rankings by dotted lines. The ultimate result of the difference in the σ_4 representations was that a certain generation of speakers identified the perceived final cluster in /maːkət/ with other clusters constructed at the σ_4 level, like those in /lɪxt/, which had no schwa (but still three “releases”). This identification led to the new generalizing form [maːkt].

The explanation of the role of the coronals in Dutch syllable structure can now be summarized as follows:

(61) *Coronals in Dutch syllables*

a. /s/ can be put anywhere, because it needs no adjacent vowels for its acoustic cues: it is truly self-sounding, like a vowel.
b. /t/ and /n/ can be put in the second position of consonant clusters, because they have fast releases.

So we see that the explanations for the fricative and the stops are unrelated. In fact, these explanations can be traced further back to unrelated physiological properties:

(62) *Unrelated causes of coronal specialness*

a. Sibilant noise requires a sharp ridge to jet air along (§3.4); the tongue blade happens to be located behind such a ridge, as long as the medial teeth are still there.

b. The tongue blade happens to make ballistic stop closures faster than the lips or dorsum.

The two kinds of specialness are not only unrelated, they are also formalized in different grammars. For coronal fricatives, the production grammar contains high faithfulness for sibilancy in any position. For coronal stops, the perception grammar contains a high cluster-OCP for [kt], and a low cluster-OCP for [nk].

Some of the mystery of the special status of coronals (Paradis & Prunet 1991) seems to vanish if we distinguish between production and perception.

8 Conclusion

By distinguishing between production and perception in phonological representations as well as processing systems, we can solve some phonological paradoxes and mysteries, without needing to assume any innate substance.

References

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

Bloomfield, Leonard (1933). *Language*. British edition, 1935. London: George Allen & Unwin.

Boersma, Paul (1989). Modelling the distribution of consonant inventories by taking a functionalist approach to sound change. *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam* **13**. 107–123.

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

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.

Boersma, Paul (to appear a). 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*. Oxford: Oxford University Press.

Boersma, Paul (to appear b). On the need for a separate recognition grammar. In Robert Kirchner, Wolf Wikeley, & Joe Pater (eds.) *Papers in Experimental and Theoretical Linguistics.* Vol. 6. Edmonton: University of Alberta.

Boersma, Paul (to appear c). *Nasal harmony in functional phonology*. Presented at HILP 4, January 1999, Leiden.

Boersma, Paul (to appear d). *Typology and acquisition in functional phonology*. Presented at Utrecht Phonology Workshop, June 1998.

Boersma, Paul & Bruce Hayes (1999). *Empirical tests of the Gradual Learning Algorithm*. Ms. University of Amsterdam and UCLA. [Rutgers Optimality Archive **348**, http://ruccs.rutgers.edu/roa.html]

Boersma, Paul & Clara Levelt (1999). Gradual constraint-ranking learning algorithm predicts acquisition order. To appear in *Proceedings of Child Language Research Forum* **30**. Stanford, Calif.: CSLI.

Booij, Geert E. (1995). *The phonology of Dutch*. Oxford: Oxford University Press.

Chomsky, Noam (1964). *Current issues in linguistic theory*. The Hague: Mouton.

Chomsky, Noam & Morris Halle (1968). *The sound pattern of English*. New York: Harper and Row.

Flemming, Edward (1995). *Auditory representations in phonology*. PhD dissertation, UCLA.

Fry, D.B., Arthur S. Abramson, Peter D. Eimas & Alvin M. Liberman (1962). The identification and discrimination of synthetic vowels. *Language and Speech* **5**. 171–179.

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

Hale, Mark & Charles Reiss (1998). Formal and empirical arguments concerning phonological acquisition. *Linguistic Inquiry* **29**. 656–683.

Halle, Morris (1959). *The sound pattern of Russian*. The Hague: Mouton.

Hayes, Bruce (1989). Compensatory lengthening in moraic phonology. *Linguistic Inquiry* **20**. 253–306.

Hayes, Bruce (1996). Phonetically driven phonology: the role of Optimality Theory and inductive grounding. To appear in *Proceedings of the 1996 Milwaukee Conference on Formalism and Functionalism in Linguistics*. [Rutgers Optimality Archive **158**, http://ruccs.rutgers.edu/roa.html]

Hermans, Ben (1984). Het Limburgs en het Litouws als metrisch gebonden toontalen. *GLOT*. 48–70.

Hernández-Chávez, Eduardo, Irene Vogel & Harold Clumeck (1975). Rules, constraints and the simplicity criterion: An analysis based on the acquisition of nasals in Chicano Spanish. In Charles A. Ferguson, Larry M. Hyman & John J. Ohala (eds.) *Nasálfest*. Stanford University. 231–248.

Hockett, Charles. F. (1965). Sound change. *Language* **41**. 185–205.

Jun, Jongho (1995). Place assimilation as the result of conflicting perceptual and articulatory constraints. *West Coast Conference of Formal Linguistics* **14**. 221–237.

Kaye, Jonathan, Jean Lowenstamm & Jean-Roger Vergnaud (1985). The internal structure of phonological elements: A theory of charm and government. *Phonology* **2**. 305–328.

Kenstowicz, Michael (1971). *Lithuanian phonology*. PhD dissertation, University of Illinois.

Kenstowicz, Michael (1994). *Phonology in generative grammar*. Oxford: Blackwell.

Kirchner, Robert (1998). *Lenition in phonetically-based Optimality Theory.* PhD dissertation, UCLA.

Leben, William (1973). *Suprasegmental phonology*. PhD dissertation, MIT, Cambridge Mass. [New York: Garland Press, 1980]

McCarthy, John (1988): Feature geometry and dependency: a review. *Phonetica* **45**: 84–108.

McCarthy, John & Alan Prince (1986). Prosodic morphology. Ms. University of Massachusetts and Brandeis.

Mohanan, K.P. (1993). Fields of attraction in phonology. In John A. Goldsmith (ed.) *The last phonological rule: Reflections on constraints and derivations.* Oxford: Blackwell. 24–69.

Myers, J. Scott (1994): *OCP effects in Optimality Theory*. [Rutgers Optimality Archive **6**, http://ruccs.rutgers.edu/roa.html]

Padgett, Jaye (1995). Partial class behavior and nasal place assimilation. *Proceedings of the Arizona Phonology Conference: Workshop on Features in Optimality Theory*. Coyote Working Papers, Univ. of Arizona, Tucson. [Rutgers Optimality Archive **113**, http://ruccs.rutgers.edu/roa.html]

Paradis, Carole & Jean-François Prunet (eds.) (1991). *The special status of coronals: Internal and external evidence*. San Diego: Academic Press.

Passy, Paul (1891). *Etude sur les changements phonétiques et leurs caractères généraux*. Librairie Firmin - Didot, Paris.

Postal, Paul M. (1968). *Aspects of phonological theory*. New York: Harper and Row.

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]

Smith, Neilson V. (1973). *The acquisition of phonology: A case study*. Cambridge: Cambridge University Press.

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

Steriade, Donca (1993). Closure, release, and nasal contours. In Marie Huffman & Rena Krakow (eds.) *Nasals, nasalization, and the velum*. Academic Press, New York. 401–470.

Steriade, Donca (1995): *Positional neutralization*. Ms. UCLA.

Steriade, Donca (1996). *Licensing laryngeal features*. Ms. UCLA.

Urbanczyk, Suzanne (1995). *Double reduplications in parallel.* [Rutgers Optimality Archive **73**, http://ruccs.rutgers.edu/roa.html]

Weijnen, Anton (1991). *Vergelijkende klankleer van de Nederlandse dialecten*. The Hague: SDU Uitgeverij.