Institute of Phonetic Sciences,
University of Amsterdam,
Proceedings 25 (2003), 31-42.

OVERT FORMS AND THE CONTROL OF COMPREHENSION*

Paul Boersma

Abstract

This paper shows that the commonly held serial view of the incorporation of overt forms in the grammar (e.g. Hayes 1996 for phonology, and Legendre, Smolensky & Wilson 1998 for syntax) is inconsistent with the even more commonly held view that if two distinct underlying forms are pronounced identically, at least one of them must violate faithfulness. By contrast, *perceptual control grammars* (Boersma 1998 for phonology, and Jäger 2002 for syntax) turn out to be consistent with this view of faithfulness.

1 Introduction

Optimality Theory claims to have replaced serial derivation with parallel evaluation. But when considering the inclusion of phonetic detail into the theory, most researchers revert to a serial view. For instance, Hayes (1996) admits: "Following Pierrehumbert (1980) and Keating (1985), I assume that there is also a phonetic component in the grammar, which computes physical outcomes from surface phonological representations. It, too, I think, is Optimality-theoretic [...]". This testimony can be abbreviated as in (1), in which the arrows denote language-specific mappings, which can presumably be modelled as Optimality-Theoretic grammars (I will use the subscripts *u*, *s*, and *a* for underlying, surface, and articulatory forms, respectively).

(1) *The serial view of production in phonology*
$[\textit{underlying form}]_u \rightarrow [\textit{surface form}]_s \rightarrow [\textit{articulatory form}]_a$

This is the prevailing view among phonologists who think that phonetic implementation should be modelled in the grammar at all. Syntacticians are a bit more than phonologists inclined to work with three representations, and a serial view of the grammar, as in (2), tends to be implicit in GB-style OT syntax (e.g. Legendre, Smolensky & Wilson 1998).

(2) *The serial view of production in syntax*
$[\textit{target logical form}]_T \rightarrow [\textit{logical form}]_L \rightarrow [\textit{phonetic form}]_P$

---

* This paper appeared in Jennifer Spenader, Anders Eriksson & Östen Dahl (eds.), *Proceedings of the Stockholm Workshop on Variation within Optimality Theory*, April 26–27, 2003, Department of Linguistics, Stockholm University, pp. 47–56 (available from http://roa.rutgers.edu).

In this paper, I will show that the serial view contradicts the very reason why OT-ists work with faithfulness constraints, which is summarized in (3).

(3) *The legitimacy of faithfulness*
If two different underlying forms are pronounced identically, at least one of their surface forms must violate a faithfulness constraint.

This axiom expresses the intuition that the way to formalize neutralization in OT is by punishing it with a faithfulness violation. I will assume the correctness of this assumption, because without it, faithfulness constraints would lose their indirect functional grounding.

If our interpretation of faithfulness is correct but incompatible with the serial view of the production grammar, it is the serial view that will have to go. I will replace it with (4).

(4) *The perceptual control view of the production grammar*
phonology: $[underlying]_u \rightarrow ([articulatory]_a \Rightarrow [auditory]_o \rightarrow [surface]_s)$
syntax: $[target]_T \rightarrow ([phonetic]_P \rightarrow [logical]_L)$

This perceptual control view reverts the order of all forms except the underlying form. The single arrows on the right stand for the reconstruction that the *listener* will be able to carry out on the message, and faithfulness constraints will be interpreted as evaluating (the speaker's view of) the extent to which the listener can reconstruct the message intended by the speaker. These recovery processes are language-specific and will therefore be modelled with Optimality-Theoretic grammars; the double arrow represents a language-independent process that therefore does not have to be modelled as a grammar.

Sections 2 to 5 will show how exactly the serial view goes wrong. Sections 6 to 8 will show that the control view does meet the legitimacy of faithfulness, and that it is the most natural view of OT production grammars that involve more than two representations.

2 Two representations, non-serial: McCarthy & Prince (1995)

Those versions of OT that work with only two representations have no fear of needing serial derivation. In Correspondence Theory (McCarthy & Prince 1995), the two representations are called *input* and *output*, but once one works with more than two representations, or studies both production and comprehension, such process-dependent labels are not sufficient, so I will instead use the more explicit traditional terms *underlying form* (UF) and *surface form* (SF). Tableau (5) shows how this version of OT models production.

(5) *McCarthy & Prince's formalization of production*

$[underlying]_u$	STRUCT$_s$	FAITH$_{us}$
$[surface_1]_s$		
$[surface_2]_s$		
$[surface_3]_s$		

Like the representations, the constraints are labelled with *u* and *s* in order to make explicit what representations they evaluate. Thus, the structural constraints, abbreviated here as STRUCT$_s$, evaluate aspects of the surface candidates only, while the faithfulness constraints, abbreviated here as FAITH$_{us}$, evaluate aspects of the similarity between the underlying form and the surface candidates (the order of STRUCT$_s$ and FAITH$_{us}$ in this schematic tableau has no relation to their relative ranking). An analogous tableau can be drawn for syntactic production with two representations (Legendre, Smolensky & Wilson 1998), in which the input is a *target form* (TF) and the output a *logical form* (LF). Such a tableau maps a [*target*]$_T$ to one of a number of candidates [*logical$_i$*]$_L$ via an evaluation of structural constraints at LF (STRUCT$_L$) and faithfulness constraints between TF and LF (FAITH$_{TL}$).

The two two-representation grammar models of production are summarized in (6).

(6) *Production models with two representations*
phonology: [*underlying*]$_u$ → [*surface*]$_s$
syntax: [*target*]$_T$ → [*logical*]$_L$

While I will need to modify the number of representations later on, I will assume that the faithfulness relation is defined correctly here. What is more, when introducing a third and fourth representation I will continue to assume that SF is *defined* as the form whose similarity to UF is evaluated by faithfulness constraints. This definition allows us to derive from (3) an important intermediate result, formulated in (7).

(7) *The locus of neutralization*
If two different underlying forms are pronounced identically, this neutralization must occur somewhere in the mapping from underlying form to surface form.

We can see that this must be true by arguing that if the neutralization took place outside the path by which UF is mapped to SF, faithfulness constraints would not be able to evaluate it, hence (3) would be violated. There is, however, a small catch to this reasoning, as will become clear in the following section.

3 Three representations, non-serial: Tesar & Smolensky (2000)

The need for a third representation in phonology stems from the fact that language-learning children do not hear fully structured surface forms in their environment. Instead, they hear unstructured *overt forms*. For instance, when confronted with a sequence of three syllables, the second of which is stressed, they initially hear the overt form [σ σ́ σ]$_o$ and have to *learn to construct* one of the surface forms [(σ σ́) σ]$_s$ or [σ (σ́ σ)]$_s$, depending on whether their ambient language has iambic or trochaic feet. For this reason, Tesar & Smolensky (2000) propose a grammar model with three forms and two processes. Both mappings in (8) are language-specific, and they are handled by a single Optimality-Theoretic grammar.

(8) *Tesar & Smolensky's grammar model*
production: [*underlying form*]$_u$ → [*full structural description*]$_s$
interpretation: [*overt form*]$_o$ → [*full structural description*]$_s$

The non-seriality of this grammar model relies heavily on *containment*, i.e., both the overt form and the underlying form are contained in the full structural description, see (9).

(9) *Non-serial grammar model with containment*

production: $[underlying]_u \rightarrow [full\ description]_s \Rightarrow [overt]_o$

e.g. $[\sigma\ \sigma\ \sigma]_u \rightarrow [(\sigma\ \acute{\sigma})\ \sigma]_s \Rightarrow [\sigma\ \acute{\sigma}\ \sigma]_o$

and $[\text{taːg}+\varnothing]_u \rightarrow [\text{taːg}_{\langle\text{voi}\rangle}+\varnothing]_s \Rightarrow [\text{taːk}]_o$

comprehension: $[overt]_o \rightarrow [full\ description]_s \Rightarrow [underlying]_u$

e.g. $[\sigma\ \acute{\sigma}\ \sigma]_o \rightarrow [(\sigma\ \acute{\sigma})\ \sigma]_s \Rightarrow [\sigma\ \sigma\ \sigma]_u$

and $[\text{taːk}]_o \rightarrow [\text{taːg}_{\langle\text{voi}\rangle}+\varnothing]_s \Rightarrow [\text{taːg}+\varnothing]_u$

The second example in (9) is the nominative singular of the German word $[\text{taːg}]_u$ 'day'. The phonological part of the case ending is the null morpheme $[\varnothing]_u$. The word is pronounced with final devoicing and with aspiration of the initial voiceless plosive, i.e. as $[\text{t}^\text{h}\text{aːk}]_o$ (for the difference between this overt form and the one given by Tesar & Smolensky, i.e. $[\text{taːk}]_o$, see below). The two double arrows in (9) are simple mechanical mappings. First, the mapping from the surface form to the overt form is mechanical, as summarized in (10).

(10) *Extracting the overt form from the full structural description*

a. Delete hidden material such as parentheses, morphological boundaries, and null morphemes: $[(]_s \Rightarrow []_o$, $[)]_s \Rightarrow []_o$, $[+]_s \Rightarrow []_o$, $[\varnothing]_s \Rightarrow []_o$

b. Interpret the insertion and deletion marks: $[\text{g}_{\langle\text{voi}\rangle}]_s \Rightarrow [\text{k}]_o$

The mapping from the surface form to the underlying form is equally mechanical, as summarized in (11).[1]

(11) *Extracting the underlying form from the full structural description*

a. Delete metrical parentheses and stress marks: $[(]_s \Rightarrow []_u$, $[)]_s \Rightarrow []_u$, $[\acute{\sigma}]_s \Rightarrow [\sigma]_u$

b. Delete the insertion and deletion marks: $[\text{g}_{\langle\text{voi}\rangle}]_s \Rightarrow [\text{g}]_u$

We can now see that (7) does not necessarily follow from (3). Consider the German underlying forms $[\text{raːd}+\varnothing]_u$ 'wheel-NOMSG' and $[\text{raːt}+\varnothing]_u$ 'advice-NOMSG', both of which are pronounced $[\text{ʁa̠ːt}]_o$, i.e., the final obstruent voicing contrast is neutralized. The full structural descriptions are $[\text{raːd}_{\langle\text{voi}\rangle}+\varnothing]_s$ and $[\text{raːt}+\varnothing]_s$, respectively. In the style of the containment faithfulness constraints of Prince & Smolensky (1993), the first of these forms violates PARSE (voi), while the second violates no faithfulness constraints at all. This means that metarule (3) is satisfied. But metarule (7) is not: the two surface forms have different structures, so the neutralization must take place in the mapping from SF to OF, i.e. in the steps $[\text{d}_{\langle\text{voi}\rangle}]_s \Rightarrow [\text{t}]_o$ and $[\text{t}]_s \Rightarrow [\text{t}]_o$. In other words, the neutralization takes place *after* it has been evaluated by the faithfulness constraints. To prevent this counter-intuitive situation, one would have to introduce the separate metarule in (12).

(12) *The anti-diacritical metarule*

Processes are evaluated where they are implemented.

If this metarule is assumed, (7) does follow from (3). We must note that (12) is incompatible with the containment view of the surface form: in order to prevent neutralization from being implemented after its evaluation, surface forms should contain $[\text{t}]_s$ rather than $[\text{d}_{\langle\text{voi}\rangle}]_s$, and if morpheme boundaries and null morphemes are subject to faithfulness as well, surface forms should not contain any instances of $[+]_s$ or $[\varnothing]_s$ either. This idea was implemented in later developments of Optimality Theory, as described in the next section.

---

[1] Tesar & Smolensky (2000: 79) actually give $[\text{taːg}_{\langle\text{voi}\rangle}]_s$ rather than $[\text{taːg}_{\langle\text{voi}\rangle}+\varnothing]_s$ for the full structural description, thereby violating containment.

4 Three representations, serial: Correspondence Theory with overt forms

Correspondence Theory (McCarthy & Prince 1995) is the OT dialect that assumes the anti-diacritical metarule (12). The surface form no longer contains insertion or deletion symbols or morphological information. This does not mean that all hidden material is erased: metrical structure is traditonally kept, since it is often hard to imagine how stress assignment can be handled without reference to hidden foot structure. The grammar model now turns into (13).

(13) *Serial grammar model with correspondence*
production: $[\textit{underlying}]_u \rightarrow [\textit{surface}]_s \rightarrow [\textit{overt}]_o$
e.g. $[\sigma\,\sigma\,\sigma]_u \rightarrow [(\sigma\,\acute{\sigma})\,\sigma]_s \rightarrow [\sigma\,\acute{\sigma}\,\sigma]_o$
and $[\text{ta:g+}\varnothing]_u \rightarrow [\text{ta:k}]_s \rightarrow [\text{t}^\text{h}\underline{\text{a}}\text{:k}]_o$
comprehension: $[\textit{overt}]_o \rightarrow [\textit{surface}]_s \rightarrow [\textit{underlying}]_u$
e.g. $[\sigma\,\acute{\sigma}\,\sigma]_o \rightarrow [(\sigma\,\acute{\sigma})\,\sigma]_s \rightarrow [\sigma\,\sigma\,\sigma]_u$
and $[\text{t}^\text{h}\underline{\text{a}}\text{:k}]_o \rightarrow [\text{ta:k}]_s \rightarrow [\text{ta:g+}\varnothing]_u$

The overt form is represented here with aspiration, unlike in (9), since German-learning children cannot a priori decide whether German aspiration is allophonic or not; for their part, they may well be learning a language with an underlying triple contrast between voiced, voiceless, and aspirated plosives, in which case aspiration is crucial. This general criticism of the view in (9) renders the SF-to-OF mappings in (9) and (13) non-mechanical. The change in SF between (9) and (13) renders the SF-to-UF mappings in comprehension non-mechanical as well, since the surface form $[\text{ra:t}]_s$ should now be mapped to either $[\text{ra:d+}\varnothing]_u$ or $[\text{ra:t+}\varnothing]_u$, probably depending on the semantic and pragmatic context. Both the SF→OF and SF→UF mappings have now become non-trivial, so that both the production and the comprehension process must be regarded as consisting of two serially ordered subprocesses. For production, we can identify these processes as *phonology* and *phonetic implementation*, and for comprehension, they are *perception* and *recognition.*

If the order of the subprocesses in the production model in (13) is correct, (7) reduces to the very simple statement in (14).

(14) *The non-neutralization of phonetic implementation*
The mapping from surface to overt form does not neutralize.

Despite its simplicity, (14) turns out to be extremely difficult to enforce, because it conflicts with the requirement that faithfulness constraints should be able to evaluate the UF-SF similarity. To see this, we consider two extreme interpretations of what a surface form is.

The first possible interpretation for SF is that it is a rather abstract form consisting of the same kind of discrete elements as UF. Under such an interpretation, the SF in (13) is $[\text{ta:k}]_s$, and its similarity to UF is easy to evaluate: it violates IDENT_{us} (voi) because the underlying segment $[\text{g}]_u$ is voiced and its corresponding surface segment $[\text{k}]_s$ is not; the remaining parts of the underlying form, $[\text{t}]_u$ and $[\text{a:}]_u$, surface perfectly. While faithfulness constraints work well under this interpretation of SF, the non-neutralization of the SF→OF mapping cannot be guaranteed: who can tell whether the aspiration of $[\text{t}]_s$ causes neutralization or not? Presumably it does not in German, but consider a couple of allophonic rules in Sanskrit and Japanese. In Sanskrit, an underlying $[\text{s}]_u$ surfaces as $[\text{h}]_o$ utterance-finally. Since the voiceless $*[\text{h}]_u$ is not a possible lexical segment in Sanskrit, this must be regarded as an allophonic rule, hence $[\text{s}]_u \rightarrow [\text{s}]_s \rightarrow [\text{h}]_o$. However, an underlying $[\text{r}]_u$ surfaces as $[\text{h}]_o$ utterance-finally as well, hence $[\mathrm{r}]_u \rightarrow [\mathrm{r}]_s \rightarrow [\mathrm{h}]_o$. But this is impossible, because it would mean that both $[\mathrm{s}]_s$ and $[\mathrm{r}]_s$ neutralize into $[\mathrm{h}]_o$ during phonetic implementation. A similar case occurs in Japanese, where $[\mathrm{z}]_u$ turns into the allophonic affricate $[\mathrm{d̠ʑ}]_o$ before $[\mathrm{i}]_s$, hence $[\mathrm{z+i}]_u \rightarrow [\mathrm{zi}]_s \rightarrow [\mathrm{d̠ʑi}]_o$, but $[\mathrm{d}]_u$ undergoes the same change, hence $[\mathrm{d+i}]_u \rightarrow [\mathrm{di}]_s \rightarrow [\mathrm{d̠ʑi}]_o$, again showing neutralization in phonetic implementation. These two cases of neutralization would leave faithfulness constraints powerless: despite the neutralization of $[\mathrm{s}]_u$ and $[\mathrm{r}]_u$ in Sanskrit, or $[\mathrm{z+i}]_u$ and $[\mathrm{d+i}]_u$ in Japanese, no faithfulness constraints are violated, since the surface forms are identical to the underlying forms. To be true, this situation could be patched up: unnatural derivations like $[\mathrm{r}]_u \rightarrow [\mathrm{s}]_s \rightarrow [\mathrm{h}]_o$ and $[\mathrm{d+i}]_u \rightarrow [\mathrm{zi}]_s \rightarrow [\mathrm{d̠ʑi}]_o$ would do the trick of violating faithfulness by moving the neutralization to the UF→SF mapping, but the complication of the additional two unnatural changes (r→s and d→z) is something most phonologists nowadays would prefer to avoid. Precisely this type of complications was the reason for Halle (1959) to propose that an intermediate form (SF) does not exist. This is the standpoint taken by Chomsky & Halle (1968), according to whom the grammar maps UF to OF via a potentially large number of intermediate representations, none of which has any special status. Chomsky & Halle can be regarded as taking the opposite viewpoint from the abstract-SF viewpoint discussed above: for them, SF is the same as OF, and it is maximally rich. Such a situation does work fine for the requirement of non-neutralization of phonetic implementation, but a phonetically rich SF cannot be used by faithfulness constraints. There is no simple way in which the similarity of a discrete UF with a phonetically detailed SF could be evaluated: does $[\mathrm{t^h a̠ːk}]_s$ violate DEP (aspiration) or not? If faithfulness constraints are to have any meaning at all, the underlying and surface forms should be *commensurable*, i.e., they should consist of the same kind of elements.

It seems that we have too many requirements for SF. For commensurability with UF, SF should be maximally abstract, but in order to make sure that the faithfulness constraints capture all cases of neutralization, SF should be maximally rich. This is probably why a worked-out serial theory of the production grammar, as summarized in (13), has never been proposed. While the issues tackled in the Correspondence Theory literature can often bear agnosticism with respect to the problems with serialism, phonetically-oriented dialects of OT cannot get by without facing these problems, as I will discuss in the following section.

5 Phonetic detail, serial

Phonetically inspired theories of phonology have to make a principled distinction between two overt forms: an articulatory form and an auditory form (Boersma 1989, Flemming 1995, Steriade 1995, Hayes 1996, Kirchner 1998). It is natural to assume that the speaker will produce an articulatory form and that the listener will start from an auditory form. The serial grammar model of (13) will turn into (15), although none of the works cited makes this proposal more explicit than the footnote from Hayes (1996) that I quoted in the Introduction above. I will label articulatory forms with *a*, and continue to label auditory forms with *o*.

(15) *Serial grammar model with phonetic detail*

production: [*underlying*]$_u$ → [*surface*]$_s$ → [*articulatory*]$_a$

e.g. [σ σ σ]$_u$ → [(σ σ́) σ]$_s$ → [σ σ́ σ]$_a$

and [taːg+∅]$_u$ → [taːk]$_s$ → [tʰa̠ːk]$_a$

comprehension: [*auditory*]$_o$ → [*surface*]$_s$ → [*underlying*]$_u$

e.g. [σ σ́ σ]$_o$ → [(σ σ́) σ]$_s$ → [σ σ σ]$_u$

and [tʰa̠ːk]$_o$ → [taːk]$_s$ → [taːg+∅]$_u$

In (15), I have regarded the commensurability requirement as more important than the non-neutralization requirement. After all, one could still require that the phonetic implementation subprocess is non-neutralizing, perhaps by a smart technical invention. But that is not how I will handle the problem, because one can observe here a *conspiracy*: the technical details of a formalization of phonetic implementation would have to conspire in such a way that it does not map two distinct SFs to the same OF. As we learned from Prince & Smolensky (1993), whenever there seems to be a conspiracy there must be something wrong with the theory.

6 Phonetic detail, non-serial

I propose that the thing that is wrong with the theory in (15) is the serial UF→SF→AF mapping, and more in particular the supposedly non-neutralizing SF→AF mapping. We can observe that there is nothing wrong with the *reverse* mapping, OF→SF, which occurs in (15) as well. For instance, the OF→SF mapping is typically neutralizing, as can be expected from any mapping without conspiring requirements. Thus, the continuous detailed auditory form [tʰa̠ːk]$_o$ will be perceived as the segment sequence [taːk]$_s$, but [tʰɑ̟ːk]$_o$ will also be perceived as [taːk]$_s$, since German allows some variation in the place of the long low vowel. Some things nearby will be perceived differently: both [da̠ːk]$_o$ and [ta̠ːk]$_o$ will be perceived as [daːk]$_s$ because German usually devoices its initial 'voiced' plosives, and both [tʰaɐ̯k]$_o$ and [tʰɐːk]$_o$ will be perceived as the segment sequence [tark]$_s$ because German [r]$_u$ is vocalized as a lower mid central vowel when appearing in the coda of a syllable, often influencing the preceding vowel. From the literature, we know that OT grammars typically cause some cases of neutralization to occur. It is natural, therefore, to model the OF→SF mapping in OT (as a *perception grammar*, Boersma 1998), but it is unnatural to try to model SF→AF in OT.

If phonetic implementation cannot be modelled in OT, and it is still language-specific (as the examples show), the question remains whether it should be modelled at all. I propose that it should not. Instead, the reverse mapping, OF→SF, which is needed in comprehension anyway, should take its place. We obtain the grammar model in (16).

(16) *Perceptual control view of phonological production*

production: [*underlying*]$_u$ → ([*articulatory*]$_a$ ⇒ [*auditory*]$_o$ → [*surface*]$_s$)

e.g. [taːg+∅]$_u$ → ([tʰa̠ːk]$_a$ ⇒ [tʰa̠ːk]$_o$ → [taːk]$_s$)

comprehension: [*auditory*]$_o$ → [*surface*]$_s$ → [*underlying*]$_u$

e.g. [tʰa̠ːk]$_o$ → [taːk]$_s$ → [taːg+∅]$_u$

The second single arrow after 'production' is not phonetic implementation, but its reverse, namely perception. The idea is that the speaker chooses an articulation (AF) whose auditory result (OF) will be perceived by the listener as a form (SF) that is as similar as possible to the speaker's intended message (UF), given the articulatory constraints. In other words, the objective of the speaker is to *control* the listener's perception, in the same sense in which Powers (1973) argued that *all* behaviour serves the control of perception. The double arrow in (16) is the mapping from articulatory form to auditory form; this is a language-independent mapping that involves physical (acoustical) and physiological transmissions.

The grammar model in (16) satisfies all three requirements (3), (7), and (12). If two different UFs are pronounced in the same way, i.e., if they have identical articulatory and auditory forms, the corresponding SFs will be identical as well; the direction of the arrows ensures this, since an OT grammar will always yield the same output for the same input as long as the ranking of the constraints does not change; hence, (7) is satisfied. Metarule (3) is then also satisfied, because a single SF cannot be identical to two different UFs at the same time. Metarule (12) has become irrelevant, since diacritics cannot pass from UF to AF, let alone to SF (though it is not impossible that the perception process *constructs* some default morphological information, e.g. that the SF in (16) is really $[\text{taːk+}\varnothing]_{\text{s}}$).

The interpretation of what a faithfulness constraint is, has changed now: faithfulness constraints evaluate (the speaker's view of) the extent to which the listener will be able to reconstruct the intended message without lexical access. The interpretation of what phonetic implementation is, has also changed: phonetic implementation does not exist as a module of the grammar. Analogously to (16), (17) proposes a control grammar model for syntax.

(17) *The control view of syntactic production*

production: $[\textit{target}]_{\text{T}} \rightarrow ([\textit{phonetic}]_{\text{P}} \rightarrow [\textit{logical}]_{\text{L}})$

comprehension: $[\textit{phonetic}]_{\text{P}} \rightarrow [\textit{logical}]_{\text{L}} \rightarrow [\textit{target}]_{\text{T}}$

7 The control view of the candidate generator

The parentheses around AF⇒OF→SF in (16) mean that the production grammar has to find the optimal triplet of AF-OF-SF combinations. In the same production grammar, constraints on articulatory effort evaluate the articulatory form (AF), structural constraints evaluate the surface form (SF), and faithfulness constraints evaluate the similarity of the surface form to the underlying form (UF). Instead of (5), tableaus will look like (18).

(18) *The control view of a production tableau*

$[\textit{underlying}]_{\text{u}}$	ART_{a}	STRUCT_{s}	FAITH_{us}
$[\textit{art}_1]_{\text{a}} \Rightarrow [\textit{aud}_1]_{\text{o}} \rightarrow [\textit{surf}_1]_{\text{s}}$			
$[\textit{art}_2]_{\text{a}} \Rightarrow [\textit{aud}_2]_{\text{o}} \rightarrow [\textit{surf}_2]_{\text{s}}$			
$[\textit{art}_3]_{\text{a}} \Rightarrow [\textit{aud}_3]_{\text{o}} \rightarrow [\textit{surf}_3]_{\text{s}}$			

The single arrow in each cell means that SF has to be computed from OF in a language-specific way, without reference to UF. This makes it impossible to have two candidates in which the auditory forms are identical but the surface forms are not.

Tableau (19) shows how the German neutralization example works in this model.

(19) *The control view of neutralization*

[raːd+∅]$_u$	NoFinalVoicedObstruent$_a$	Ident$_{us}$ (voi)
[ʁa̠ːd]$_a$ ⇒ [ʁa̠ːd]$_o$ → [raːd]$_s$	*!	
☞ [ʁa̠ːt]$_a$ ⇒ [ʁa̠ːt]$_o$ → [raːt]$_s$		*

In such simple cases, the control view works similarly to Correspondence Theory. The Sanskrit case of multiple sources for the [h]$_o$ allophone is more interesting. Consider the UF [maːt̪ar]$_u$ 'mother', which is pronounced as [maːt̪əh]$_a$. The question is to what extent the listener can reconstruct the underlying form from the auditory form [maːt̪əh]$_o$. Since all overt instances of [ə]$_o$ derive from an underlying [a]$_u$ (throughout Sanskrit phonology this vowel acts as the short counterpart to [aː]$_u$), the listener will have no problems in perceiving [ə]$_o$ as [a]$_s$. The case is more difficult for [h]$_o$. Since the lexicon does not contain any instances of voiceless [h]$_u$, there is no point in perceiving [h]$_o$ as [h]$_s$. On average, the listener will do better in reconstructing intended messages if she notes that the great majority of instances of [h]$_o$ in Sanskrit derive from an underlying [s]$_u$ (final [r]$_u$ is far less common). The tableau in (20) shows how the listener will therefore perceive [maːt̪əh]$_o$ as [maːt̪as]$_s$.

(20) *The perception of an overt voiceless glottal fricative in Sanskrit*

[maːt̪əh]$_o$	*[ə]$_s$	*[h]$_s$	[h]$_o$ is not [k]$_s$	[ə]$_o$ is not [i]$_s$	[h]$_o$ is not [r]$_s$	[h]$_o$ is not [s]$_s$	[ə]$_o$ is not [a]$_s$
[maːt̪əh]$_s$	*!	*					
[maːt̪ar]$_s$					*!		*
☞ [maːt̪as]$_s$						*	*
[maːt̪is]$_s$				*!		*	
[maːt̪ak]$_s$			*!				*

We see that the perception process can be modelled in OT quite well. The constraints in (20) have been modelled in the style of Escudero & Boersma (2001). The constraints against perceiving [ə]$_o$ as anything but [a]$_s$ or against perceiving [h]$_o$ as anything but [s]$_s$ must be ranked high. In particular, it must be worse to perceive [h]$_o$ as [r]$_s$ than to perceive it as [s]$_s$. Escudero & Boersma show that such rankings automatically emerge during lexicon-driven acquisition as a result of different likelihoods, i.e., for the overt form [h]$_o$ the candidate [s]$_s$ is more likely to be 'correct' than the candidate [r]$_s$, since the learner is more likely to find [s]$_u$ than [r]$_u$ in her lexicon afterwards during recognition. Finally, the constraints *[ə]$_s$ and *[h]$_s$ must be ranked high, since such structures do not occur in the lexicon (alternatively, the candidate generator might not generate candidates with such structures in the first place, in which case we could do without these constraints).

We can now construct the production tableau for [maːt̪ar]$_u$, as in (21). For brevity, the two overt forms (articulatory and auditory) have been collapsed into one, labelled *ao*.

(21) *The control view of neutralization into a distant allophone*

$[\text{maːt̪ar}]_u$	NOFINALRHOTIC$_a$	IDENT$_{us}$ (son)
$[\text{maːt̪ər}]_{ao} \rightarrow [\text{maːt̪ar}]_s$	*!	
☞ $[\text{maːt̪əh}]_{ao} \rightarrow [\text{maːt̪as}]_s$		*

Since it is optimal for the listener to map $[\text{maːt̪əh}]_o$ to $[\text{maːt̪as}]_s$, there is no candidate like $[\text{maːt̪əh}]_{ao} \rightarrow [\text{maːt̪ar}]_s$. Thus, a given AF can never appear twice in the same tableau. In the formulation by Jäger (2002) for syntax, all candidates in production tableaus must be 'hearer-optimal'. This is crucial in this case, since if we had been allowed to include the candidate $[\text{maːt̪əh}]_{ao} \rightarrow [\text{maːt̪ar}]_s$, it would have become the winning candidate since it violates none of the relevant constraints. In the same vein, two of the six candidates in tableau 15 of Legendre, Smolensky & Wilson (1998) would not be generated in a control view of a syntactic production grammar, since their phonetic forms are identical to those of two hearer-optimal candidates (this *might* help solving one of the problems that they note...).

Interestingly, we see in (21) that the intermediate representation in the $[\text{r}]_u \rightarrow [\text{s}]_s \rightarrow [\text{h}]_o$ mapping discredited in §4 now reappears in the mapping $[\text{r}]_u \rightarrow ([\text{h}]_o \rightarrow [\text{s}]_s)$. In the present case, however, the occurrence of $[\text{s}]_s$ is not inspired by a metalinguistic need to prevent neutralization in phonetic implementation, but by the most sensible guess for Sanskrit listeners.

8 How control grammars incorporate phonetic detail

Since the control view of the production grammar does not allow a separate component for phonetic implementation, it remains to be shown how it is capable of expressing language-specific needs for certain phonetic details. As an example, tableau (22) shows how the aspiration in the initial plosive in the German $[\text{taːt}+\varnothing]_s$ 'deed-NOMSG' comes about.

(22) *The control view of the implementation of phonetic detail*

$[\text{taːt}+\varnothing]_u$	IDENT$_{us}$ (voi / 96%)	IDENT$_{us}$ (voi / 80%)	IDENT$_{us}$ (voi / 20%)	*ASP$_a$	*LAX$_a$	IDENT$_{us}$ (voi / 4%)
☞ $[\text{t}^\text{h}\text{a̠ːt}]_{ao} \rightarrow$ 95% $[\text{taːt}]_s$, 5% $[\text{daːt}]_s$				*		*
$[\text{ta̠ːt}]_{ao} \rightarrow$ 40% $[\text{taːt}]_s$, 60% $[\text{daːt}]_s$			*!			*
$[\text{d̥a̠ːt}]_{ao} \rightarrow$ 10% $[\text{taːt}]_s$, 90% $[\text{daːt}]_s$		*!	*			*
$[\text{da̠ːt}]_{ao} \rightarrow$ 2% $[\text{taːt}]_s$, 98% $[\text{daːt}]_s$	*!	*	*		*	*

If constraints are ranked along a continuous scale, and some noise is added to the rankings at evaluation time (Boersma & Hayes 2001), the output of the perception grammar will vary from evaluation to evaluation. Hence, each of the four candidates has certain probabilities of being perceived as $[\text{taːt}]_s$ and as $[\text{daːt}]_s$. For instance, the voiceless unaspirated articulation $[\text{ta̠ːt}]_a$ is ambiguously perceived as $[\text{taːt}]_s$ 40% of the time, as $[\text{daːt}]_s$ 60% of the time. I assume that the speaker knows these percentages (to compute them, she could run $[\text{ta̠ːt}]_o$ through her perception grammar a number of times) and that the production grammar contains constraints that refer to them. For instance, [ta̱ːt]$_{ao}$ violates IDENT$_{us}$ (voi / 20%) because the probability that this candidate is perceived as the faithfulness-violating [daːt]$_{s}$ is more than 20%. Since it is worse to violate IDENT (voi) 80% of the time than it is to violate it only 20% of the time, the tableau exemplifies a fixed ranking by confusion probability. The tableau also contains a couple of articulatory constraints, which express the idea that it costs some effort to either aspirate a plosive, as in [t^{h}]$_{a}$, or to render it fully voiced, as in [d]$_{a}$.

(23) *The control view of the implementation of phonetic detail*

[dax+∅]$_{u}$	IDENT$_{us}$ (voi / 96%)	IDENT$_{us}$ (voi / 80%)	IDENT$_{us}$ (voi / 20%)	*ASP$_{a}$	*LAX$_{a}$	IDENT$_{us}$ (voi / 4%)
[t^{h}a̱x]$_{ao}$ → 95% [tax]$_{s}$, 5% [dax]$_{s}$		*!	*	*		*
[ta̱x]$_{ao}$ → 40% [tax]$_{s}$, 60% [dax]$_{s}$			*!			*
☞ [d̥a̱x]$_{ao}$ → 10% [tax]$_{s}$, 90% [dax]$_{s}$						*
[da̱x]$_{ao}$ → 2% [tax]$_{s}$, 98% [dax]$_{s}$					*!	

The same ranking explains the pronunciation of [d]$_{u}$ as lenis voiceless, exemplified in tableau (23) for the underlying form [dax]$_{u}$ 'roof'. In this case, the candidate that would serve the listener best (namely [da̱x]$_{ao}$) fails to win, because the speaker does not bother to trade the articulatory gain of not performing the obstruent voicing gestures for an only slightly lower probability of confusion.

9 Conclusion

Unlike theories that propose a serial modularity of phonology and phonetic implementation, the perceptual control view of Optimality-Theoretic production grammars allows us to use faithfulness constraints for the purpose that they were designed for (including the evaluation of neutralization) and in the way they were defined by Correspondence Theory (namely as evaluating two commensurable discrete representations), while at the same time it allows us to explain the details of continuous phonetic implementation.

References

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 (1998): *Functional phonology*. PhD thesis, University of Amsterdam. The Hague: Holland Academic Graphics.

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

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

Escudero, Paola, & Paul Boersma (2001): "Modelling the perceptual development of phonological contrasts with Optimality Theory and the Gradual Learning Algorithm." *Rutgers Optimality Archive* **439**, http://roa.rutgers.edu. Appeared in Sudha Arunachalam, Elsi Kaiser & Alexander Williams (eds., 2003), *Proceedings of the 25th Annual Penn Linguistics Colloquium. Penn Working Papers in Linguistics* **8.1**: 71–85.

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

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

Hayes, Bruce (1996): “Phonetically-driven phonology: The role of Optimality Theory and inductive grounding.” *Rutgers Optimality Archive* **158**, http://roa.rutgers.edu. Appeared in Michael Darnell, Edith Moravcsik, Michael Noonan, Frederick Newmeyer & Kathleen Wheatley (eds., 1999), *Functionalism and formalism in linguistics*, Vol. I: *General papers*, 243–285. Amsterdam: John Benjamins.

Jäger, Gerhard (2002): “Learning constraint sub-hierarchies: The Bidirectional Gradual Learning Algorithm.” *Rutgers Optimality Archive* **544**, http://roa.rutgers.edu. To appear in Henk Zeevat & Reinhard Blutner (eds.), *Optimality Theory and pragmatics*. Palgrave Macmillan.

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

Legendre, Géraldine, Paul Smolensky, & Colin Wilson (1998): “When is less more? Faithfulness and minimal links in wh-chains.” In Pilar Barbosa, Danny Fox, Paul Hagstrom, Martha McGinnis & David Pesetsky (eds.), *Is the best good enough?: optimality and competition in syntax*, 249–289. Cambridge, Mass.: MIT Press.

McCarthy, John, & Alan Prince (1995): “Faithfulness and reduplicative identity.” In Jill Beckman, Laura Walsh Dickey & Suzanne Urbanczyk (eds.), *Papers in Optimality Theory. University of Massachusetts Occasional Papers* **18**, 249–384. Amherst, Mass.: Graduate Linguistic Student Association.

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

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

Steriade, Donca (1995): “Positional neutralization.” Unfinished ms, Department of Linguistics, UCLA.

Tesar, Bruce, & Paul Smolensky (2000): *Learnability in Optimality Theory*. Cambridge, Mass.: MIT Press.