Polivanov’s idea of phonology in perception

Paul Boersma, June 6, 2006

In his discussion of the perception of sounds in a foreign language, Polivanov (1931) proposes an account in terms of inviolable structural constraints and violable cue constraints. The present paper shows that Polivanov’s proposal can be formulated in all details with the decision mechanism of Optimality Theory (Prince & Smolensky 1993) and fits well within Boersma’s (2005) model of bidirectional phonology and phonetics.

1 What is perception?

In general, perception is the mapping from raw sensory data to more abstract mental representations, or any step therein. In phonology, the perception task for the listener is to map a raw continuous auditory representation (AudF) to a discrete phonological surface structure (SF). This task corresponds to what phoneticians in the lab call an *identification* task.

It is useful to point out to what kind of perception I am *not* referring here. If a listener identifies two different auditory forms as the same phonological structure, I will say that these two forms are ‘perceived’ as the same structure. If I say that two auditory forms are perceived as the same phonological structure, I do not mean to say that the listener cannot hear them apart. Listeners can often discriminate sounds that they would classify as the same phoneme. Phoneticians in the lab call this a *discrimination* task. The discriminability of two auditory forms is partly determined by their auditory distance, partly by whether they are classified as the same phonological category in their language: from 9 months of age, human listeners in whose language a certain pair of auditory tokens belongs to two different categories are better at discriminating them than are listeners in whose language this same pair of auditory tokens belongs to a single category (for an overview, see Jusczyk 1997). Thus, the discrimination task measures a partly universal, partly language-specific degree of perceptability of a contrast, whereas the identification task measures what the listener regards as the speaker’s most likely intended phonological surface structure. The two tasks, then, are different, and since the goal of speech comprehension is to reconstruct the speaker’s intended message, I will ignore the extralinguistic discrimination task and use the term ‘perception’ only for the linguistic perception process, which can be equated with the phonetician’s identification task in the lab. Other possible terms for the same thing are *prelexical perception* and *phonetic parsing*.

2 Modelling robust language-specific perception in OT

To model perception in OT, we must have constraints that evaluate the output (SF) and constraints that evaluate the mapping between the input (AudF) and the output, as in (1).

(1) *Perception constraints*

a. *STRUCT$_S$: structural constraints on the output (SF);
b. CUE$_{AudS}$: cue constraints that relate input (AudF) and output (SF).

The *STRUCT$_S$ constraints are the same ones as in production (Prince & Smolensky 1993), where they interact with faithfulness constraints. The CUE$_{AudS}$ constraints cannot really be called faithfulness constraints, since they compare two incommensurable kinds of representations: the auditory form, which consists of universally available continuous formants, pitches, noises and durations, and the phonological surface form, which consists of language-specific abstract discrete structures. These constraints have been called *mapping* constraints or *construction* constraints, but the most appropriate name is probably *cue constraints* (Boersma 2005a, Escudero 2005) since just as the word *faithfulness* the term *cue* implies a relation between two representations ("an SF can be *faithful to* a UF"; "an AudF can be a *cue for* an SF"). The cue constraints that have been proposed in the OT literature are OCP$_{AudS}$ (Boersma 1998) and the generalized categorization constraint family "[*x*]$_{Aud}$ is not perceived as /*y*/$_S$" (Escudero & Boersma 2003, 2004; Boersma & Escudero 2004; Escudero 2005).

3 Polivanov: Japanese learners of Russian

The first example of constraint ranking in perception was provided by Polivanov (1931), who observed that Japanese learners of Russian perceive the Russian pronunciation [tak] (which reflects the underlying form |tak| 'so') as the Japanese phonological surface structure /.ta.ku./. The present section translates this into OT.

Consider the auditory form [^{t}a_^{k}]. As you can see in a spectrogram when you say [tak], this sound consists of a high-frequency noise burst ([^{t}]), followed by a periodic sound with formants around 1000 Hz ([a]), followed by a silence ([_]), followed by a burst with a peak around 2500 Hz ([^{k}]). A listener of Russian will have to map this to the phonological form /.tak./, which is a single syllable (syllables are delimited by periods here) that consists of an ordered sequence of three of the 40 (or so) Russian phonemes ('ordered' because /.kat./ would mean something else). The Russian listener can subsequently easily look this up in her lexicon and finds |tak| 'so', a common interjection expressing agreement. How can we model the Russian perception of [a], or equivalently, [periodic, sonorant, F1 = 800 Hz], i.e. a periodic sonorant sound with a first formant of, say, 800 Hz? Simply like the following tableau:

(1) *Russian perception of [a]*

[periodic, sonorant, F1 = 800 Hz]	*/t/ [periodic]	*/b/ [sonorant]	*/i/ [F1=800Hz]	*/e/ [F1=800Hz]	*/a/ [F1=800Hz]
☞ /a/					*
/e/				*!	
/i/			*!		
/b/		*!			
/t/	*!				

Russian has vowel phonemes like /a/, /e/ and /i/, periodic (i.e. voiced) non-sonorant consonant phonemes like /b/, and non-periodic (i.e. voiceless) phonemes like /t/. When hearing [a], the listener will have to choose from among at least these 5 sounds. Because the sound is periodic, the speaker cannot have intended to say /t/. This is such an important restriction (*constraint*) that I put it on top (i.e. in the left column). The *candidate* perception /t/ thus *violates* the constraint "a periodic (voiced) auditory form cannot be /t/" (abbreviated as */t/ [periodic]). This violation is marked in the first column by an asterisk ("*"). Because this constraint is so high-ranked, its violation immediately rules out the /t/ candidate. In order words, this violation is *crucial*, and we denote that with an exclamation mark ("!").

The second candidate that can be ruled out is /b/, because a sonorant auditory form cannot refer to a plosive. This is the second column. Regarding only the top two constraints, all vowels are still good candidates, because all vowels are periodic and sonorant. We then look at the formant information. The phoneme /i/ typically comes with an F1 of 300 Hz, /e/ perhaps with 500 Hz, and /a/ perhaps with 750 Hz. If you hear an F1 of 800 Hz, it must be very unlikely that the speaker has intended to put an underlying |i| into your head. That is the third column. It must also be slightly unlikely that the speaker's intention was |e|. That is the fourth column. There is still a difference between 750 and 800 Hz, but this difference is not so bad, so the fifth constraint is probably really low-ranked. The remaining candidate is /a/; it violates only the fifth constraint, and this violation does not rule out /a/ (since there are no other candidates left), hence no exclamation mark in this column.

This is all the theoretical machinery we need for Optimality-Theoretic modelling of perception.

Now consider the auditory form [^{t}a_^{k}] again. We saw how Russians would perceive it, but how would Japanese perceive it? Japanese words cannot have a plosive at the end of a syllable (i.e. in *coda*). A Japanese listener probably takes that into account when hearing [tak], so the perception /.tak./ is unlikely. So what will a Japanese learner of Russian do when first hearing a Russian say the utterance [^{t}a_^{k}]?

If the candidate perception /.tak./ is out of the question, perhaps the Japanese listener ignores the [^{k}] release burst and decides to perceive just /.ta./? Or perhaps the Japanese listener hears the [^{k}] release burst and decides that the speaker intended a /k/, which must then have been followed by a vowel, so that some more candidate structures are /.ta.ko./ and /.ta.ku./?

To start to get at an answer, consider what Japanese sounds like. Short high vowels that are not adjacent to a voiced consonant tend to be pronounced voiceless. Thus, the word |káku| is usually pronounced [^{k}á_^{k}u̥]. Such a devoiced vowel will often lose all of its auditory cues, if there is even a slight background noise. So the auditory form is often not much more than [^{k}á_^{k}]. Thus, Japanese listeners are used to interpreting a silence, i.e. the auditory form [ ], as the vowel /u/. They will perceive the Russian [^{t}a_^{k}] as /.ta.ku./. Tableau 2 shows the candidates that I have been discussing, and the reasons why three of them are ruled out.

(2) *Japanese foreign-language perception of Russian*

[^{t}a_^{k}]	NoPlosiveCoda$_S$	*/ / [^{k}]	*/o/ [ ]	*/u/ [ ]
/.tak./	*!			
/.ta./		*!		
☞ /.ta.ku./				*
/.ta.ko./			*!	

This Japanese behaviour when confronted with foreign codas generalizes to silences next to voiced consonants, e.g. Japanese have been reported not to hear the distinction between [ebzo] and [ebuzo] at all, interpreting both as /.e.bu.zo./ (Dupoux, Kakehi, Hirose, Pallier, Fitneva & Mehler 1999). It is the cause behind Japanese loanword adaptations, such as /.e.ki.su.to.ra./ for the European word *extra*.

The phenomenon in tableau (2) underlines the language-specificity of perception, because native listeners of Russian will perceive the same auditory form [^{t}a_^{k}] as the surface structure /.tak./. In tableaus like (2), such an outcome can be achieved by a much lower ranking of NoPlosiveCoda$_S$. The language-specificity of perception, then, corresponds to the freedom that every language possesses to rank the constraints in their own order.

Apropos /.e.ki.su.to.ra./. Why would the Japanese not have borrowed this as /.e.ki.su.tu.ra./, with this less audible /u/ vowel? The answer is that Japanese does have syllables that sound like [tu] (or [du]). They do have syllables that sound like [tsu] (and [dzu]), but apparently they do not like to perceive [t] as /tu/, which they would have to pronounce as [tsu]. So they take /to/ instead, despite its 'full' vowel. But it is a compromise.

The auditory form under discussion is [‿^{d}rama], where the funny symbol in the beginning stands for the sound of voicing with your mouth closed, and the superscript d stands for the alveolar plosive burst.

A Russian listener would perceive this auditory form as the phonological structure /.dra.ma./. A Japanese listener will not perceive it as /.dɾa.ma./, because that form contains a syllable onset that consists of two consonants (something that phonologists call a *complex onset*), and such structures are forbidden in Japanese. The candidate /.dɾa.ma./ therefore violates a structural constraint at Surface Form, say */.CC/ ("no complex onsets"). Tableau (3) makes this explicit.

(3) *Japanese foreign-language perception of Russian*

[̰{*alv*,*burst*}ɾama]	*/.CC/	*/ / [*burst*]	*/du/	*/*vel*/ [*alv*]	*/*fric*/ [*burst*]	*/o/ [ ]	*/u/ [ ]
/.dɾa.ma./	*!						
/.ɾa.ma./		*!					
/.du.ɾa.ma./			*!				*
/.gu.ɾa.ma./				*!			*
/.zu.ɾa.ma./					*!		*
☞ /.do.ɾa.ma./						*	

One way to satisfy the Japanese onset constraint is to perceive [̰^{d}rama] as /.ɾa.ma./, which does not have a complex onset. This would involve throwing away some positive auditory cues, namely the voicing murmur and the alveolar burst. As in the case of [^{t}a_^{k}], Japanese listeners seem not to like throwing away positive cues, i.e. a constraint like */ /[*burst*] is ranked high. This takes care of candidate 2.

The third option is to perceive /.du.ɾa.ma./, hallucinating a /u/ analogously to the /.ta.ku./ case. But Japanese does not allow the structure /du/ on the surface. This is what the third constraint expresses.

The fourth option is to perceive /.gu.ɾa.ma./. This has the allowed sequence /gu/. But this velar candidate ignores the cues for alveolar place.

The fifth option is to perceive /.zu.ɾa.ma./, a phontactically allowed sequence that would be pronounced as [dzuɾama]. This does honour the alveolar place cue but ignores the auditory cue for plosiveness (namely the burst), positing instead a fricative. Because this candidate is more or less possible (according to Polivanov), we must conclude that the alveolar place cue is more important than the plosiveness cue. This is an example of *cue weighting*. The tableau shows this as a fixed ranking of the fourth and fifth constraints.

The sixth option is to perceive /.do.ɾa.ma./. This honours all the place and manner cues for /d/ but has the drawback of hallucinating the full vowel /o/ rather than the half-vowel /u/. It wins because there is no better option.

Please note that the ranking of the constraints in tableau (2) still occurs in tableau (3). This has to be. A single constraint ranking (i.e. a single grammar) has to account for all the forms in the language.

Polivanov suggested that some speakers might choose the fifth candidate. Such speakers would have the ranking in tableau (4).

(4) *Japanese foreign-language perception of Russian*

[‿{*alv*,*burst*}rama]	*/.CC/	*/ / [*burst*]	*/du/	*/*vel*/ [*alv*]	*/o/ [ ]	*/*fric*/ [*burst*]	*/u/ [ ]
/.dɾa.ma./	*!						
/.ɾa.ma./		*!					
/.du.ɾa.ma./			*!				*
/.gu.ɾa.ma./				*!			*
☞ /.zu.ɾa.ma./						*	*
/.do.ɾa.ma./					*!		

Polivanov says that in this variation two constraints compete. They are the fifth and sixth constraints in tableau (4). There is a way to express this variation in a single tableau. In tableau (5), the two constraints are ranked at the same height. This is to be interpreted in the following way: when the tableau is evaluated (i.e. when the listener hears [‿^{d}rama]), the listener perceives /.zu.ɾa.ma./ in 50 percent of the cases, and /.do.ɾa.ma./ in the remaining 50 percent of the cases. Hence the two pointing fingers.

(5) *Two optimal candidates*

[‿{*alv*,*burst*}rama]	*/.CC/	*/ / [*burst*]	*/du/	*/*vel*/ [*alv*]	*/*fric*/ [*burst*]	*/o/ [ ]	*/u/ [ ]
/.dɾa.ma./$_{S}$	*!						
/.ɾa.ma./$_{S}$		*!					
/.du.ɾa.ma./$_{S}$			*!				*
/.gu.ɾa.ma./$_{S}$				*!			*
☞ /.zu.ɾa.ma./$_{S}$					*		*
☞ /.do.ɾa.ma./$_{S}$						*	

It is not yet fully correct. The same ranking does not work well for *extra*. The part [^{t}ra] tends to be perceived as /.to.ɾa./. The trouble is to rule out the candidate /.tu.ɾa./, which is what many phonologists regard as the phonological structure behind the Japanese pronunciation [tsuɾa]. Thus, the impossible thing is not the structure /tu/, but the pronunciation [tsu]. The constraint "*/tu/", analogous to the third constraint in tableaus (3) to (5) will therefore not work.

The solution must lie in the fact that the naked auditory release burst [t], without affrication, cannot be a good representative of the structure /tu/, which must be pronounced with a full affricate [ts]. So we get tableau (6).

(6) *Japanese foreign-language perception of Russian*

[{*alv*,*burst*, –*affr*}ra]	*/.CC/	*/ / [*burst*]	*/tu/ [–*affr*]	*/*vel*/ [*alv*]	*/*fric*/ [*burst*]	*/o/ [ ]	*/u/ [ ]
/.tɾa./	*!						
/.ɾa./		*!					
/.tu.ɾa./			*!				*
/.ku.ɾa./				*!			*
/.su.ɾa./					*!		*
☞ /.to.ɾa./						*	

I am still not entirely satisfied, because I would like to attribute the insertion of /o/ after both /t/ and /d/ to the same cause.

4 Robust perception: Richness of the Base is in comprehension

The *robust perception* mentioned in §3 is related to two concepts that have been proposed earlier in OT. First there is *richness of the base* (Prince & Smolensky 1993), according to which inputs (to production) can be anything: even hypothetical underlying forms that do not actually occur in the lexicon of the language at hand will be converted by the grammar (constraint ranking) to well-formed surface structures. In the perception case, richness of the base resides in the auditory form, which is the input to perception and can be anything: even auditory events that do not normally occur in the listener's language environment will be converted by the grammar to (more or less) well-formed surface structures. Since we refer to this as robust perception, we should perhaps rename Prince & Smolensky's version of richness of the base to *robust production*, to make its orientation explicit. The second concept related to robust perception is *robust interpretive parsing* (Tesar & Smolensky 1998, 2000), according to which the listener succeeds in making sense of any *overt form* (in Tesar & Smolensky example a sequence of syllables marked for stress) by converting it to a sensible surface structure (in Tesar & Smolensky's example a sequence of feet with head syllables), even if the listener's grammar could never generate such a structure in production. To the extent that Tesar & Smolensky's interpretive parsing can be equated with what others call perception, the concepts of robust perception and robust interpretive parsing are not just related, but identical (a difference between them will be discussed later). I will now make plausible that the two concepts can indeed be equated.

As an example of language-dependent interpretive parsing, Tesar (1997) mentions the 'overt form' [σ ˈσ σ], which is a sequence of three syllables of which the middle one is stressed. The task of the listener is to map this overt form to a more abstract metrical structure. According to Tesar, the overt form [σ ˈσ σ] will be interpreted as the foot structure /(σ ˈσ) σ/ in a left-aligning iambic language, and to /σ (ˈσ σ)/ in a right-aligning trochaic language, depending on the language-specific ranking of the structural (metrical) constraints. This looks straightforwardly like what I have defined as perception. Although Tesar (and Smolensky) never draw a tableau that has the overt form as its input and the interpreted structure as its output (all of their tableaus include the winning candidates in *production*), such a tableau can be drawn easily, as here in tableaus (7a) and (7b), which use a subset of Tesar's constraints.

(7a) *Metrical perception in a left-aligning iambic language*

[σ ˈσ σ]	FEETLEFT$_S$	IAMBIC$_S$	TROCHAIC$_S$	FEETRIGHT$_S$
☞ /(σ ˈσ) σ/			*	*
/σ (ˈσ σ)/	*!	*		

(7b) *Metrical perception in a right-aligning trochaic language*

[σ ˈσ σ]	FEETRIGHT$_S$	TROCHAIC$_S$	IAMBIC$_S$	FEETLEFT$_S$
/(σ ˈσ) σ/	*!	*		
☞ /σ (ˈσ σ)/			*	*

While Tesar & Smolensky's surface structures are uncontroversially the same kind of thing as the output of perception in my earlier perception tableaus, the same cannot be immediately claimed about the overt forms. In (7) they are labelled as 'auditory'. But are they really? After all, the form [σ ˈσ σ] already consists of syllables, which are language-dependent higher-level structures, and my use of the discrete IPA stress symbol already abstracts away from the continuous auditory correlates of stress such as intensity, pitch, and duration. But I want to assert that the foot structures in the output candidates of (7) are even more abstract and high-level than this overt form. What we see in (7), then, is a step on the way from the universal auditory form to the language-specific phonological surface structure. Thus, tableaus (7) represent a step in the perception process. Now, I do not mean to imply that the perception process consists of a sequence of steps. The mapping from auditory cues to segments, from segments to syllables, and from syllables to feet could well be done in parallel. In that case, the mapping from segment to syllable could well depend on the foot structure that the listener has to create at the same time. I assume that, indeed, the various facets of perception work in parallel in much the same way as the various facets of production work in parallel in most published OT analyses. And since in OT analyses of production one can find mappings at various levels of abstraction, I take the liberty of doing the same for perception and declare tableaus (7) as perception tableaus, thus identifying Tesar & Smolensky's interpretive parsing with the perception process.

The grammatical framework by Tesar & Smolensky is less restrictive than that by Polivanov. Whereas Polivanov assumes that structural constraints are in GEN (inviolable) and cue constraints in CON (violable), Tesar & Smolensky follow the usual Optimality-Theoretical standpoint that structural constraints are violable, i.e. reside in CON. This violability is exemplified in tableaus (7) and I will assume that it is correct. In other words, phonotactic constraints can conflict with each other in perception, in which case their relative ranking becomes crucial.

The robustness of the perception process has already been illustrated with the Japanese perception of a foreign [tak]. Tesar & Smolensky's robustness point applies to first-language acquisition, and specifically to their proposal that a speaker/listener uses the same constraint ranking in production as in perception. A child learning the left-aligning iambic language of tableau (6), for instance, may have at a certain point during her acquisition period the grammar FEETLEFT$_S$ >> TROCHAIC$_S$ >> IAMBIC$_S$ >> FEETRIGHT$_S$. This left-aligning trochaic grammar is incorrect, since it causes an underlying |σ σ σ| to be produced as /(ˈσ σ) σ/$_S$. When such a child hears the correct overt form [σ ˈσ σ], however, she will interpret it as /(σ ˈσ) σ/, which can easily be seen by reversing the two foot-form constraints in (6). Since the child's robust perception can make sense of a form that she would never produce herself, the child is able to notice the discrepancy between the two forms /(ˈσ σ) σ/ and /(σ ˈσ) σ/, and can take action, perhaps by reversing the ranking of TROCHAIC$_S$ >> IAMBIC$_S$ in her grammar. Thus, Tesar & Smolensky's point is that robustness helps learning. In sum, we conclude that the robustness of the perception process proposed in this section helps in the acquisition of a first and second language and in loanword adaptation.

5 More examples of perception in OT

This section reviews some more examples of how perception has been formalized in Optimality Theory.

5.1 Autosegmental constraints on tone

An early example of a structural constraint in phonology is the Obligatory Contour Principle (Leben 1973, Goldsmith 1976). In theories of suprasegmental tone, this constraint militates against the occurrence of two identical tones in a row. Meyers (1997) investigated the OCP as a constraint in OT. In Boersma (1998, 2000) the OCP was interpreted as the counterpart of the Line Crossing Condition, in the sense that many structures that violate the OCP do not violate the LCC and vice versa (in this respect, the two constraints are similar to pairs like ALIGNFEETLEFT and ALIGNFEETRIGHT, or IAMBIC and TROCHAIC). The explicit definitions of the two constraints are given in (8).

(8) *Autosegmental constraints*

a. OCP$_S$ (*feature value*, *material*): the surface form cannot contain two instances of *feature value* if not more than a certain amount of *material* intervenes;
b. LCC$_S$ (*feature value*, *material*): a single instance of *feature value* in the surface form cannot span across a certain amount of *material*.

These definitions are different from those in Boersma (1998), where these constraints were cue constraints. The current definition is closer to what phonologists are used to (e.g. Myers 1997). Tableaus (9) and (10) show examples from Boersma (2000). In both cases the auditory input consists of two syllables with high level tones (denoted here with acute symbols), but the perceived surface structure depends on the language at hand.

(9) *Shona perception of a suprasyllabic high tone*

[ɓáŋgá]	$[\acute{\sigma}] \rightarrow / \begin{array}{c}\mathrm{H}\\ \mid \\ \sigma\end{array} /$	OCP$_S$(H,]$_\sigma$)	LCC$_S$(H,]$_\sigma$)
☞ $/ \begin{array}{c}\mathrm{H}\\ \diagup\ \diagdown\\ \text{ɓ a ŋ g a}\end{array} /$			*
$/ \begin{array}{ccccc}\mathrm{H} & & & & \mathrm{H}\\ \mid & & & & \mid\\ \text{ɓ} & \text{a} & \text{ŋ} & \text{g} & \text{a}\end{array} /$		*!	
$/ \begin{array}{ccccc}\mathrm{H} & & & & \\ \mid & & & & \\ \text{ɓ} & \text{a} & \text{ŋ} & \text{g} & \text{a}\end{array} /$	*!		

(10) *Mandarin perception of a sequence of syllabic high tones*

[ʂ̹áfá]	$[\acute{\sigma}] \rightarrow / \begin{array}{c}\mathrm{H}\\ \mid \\ \sigma\end{array} /$	LCC$_S$(H,]$_\sigma$)	OCP$_S$(H,]$_\sigma$)
$/ \begin{array}{c}\mathrm{H}\\ \diagup\ \diagdown\\ \text{ʂ̹ a f a}\end{array} /$		*!	
☞ $/ \begin{array}{cccc}\mathrm{H} & & & \mathrm{H}\\ \mid & & & \mid\\ \text{ʂ̹} & \text{a} & \text{f} & \text{a}\end{array} /$			*
$/ \begin{array}{cccc}\mathrm{H} & & & \\ \mid & & & \\ \text{ʂ̹} & \text{a} & \text{f} & \text{a}\end{array} /$	*!		

In Shona, a sequence of two high-toned syllables is interpreted on the phonological surface level as having a single high tone (H). Tableau (9) describes in detail how a word with such a sequence is perceived. The auditory form of the word 'knife' is [ɓáŋgá]. The third candidate in (9) is ruled out because there is a cue constraint that says that an auditorily high-toned syllable has to be perceived as a syllable that is linked to an H tone in the (more abstract) phonological structure. The third candidate violates this constraint because the second syllable is auditorily high but not linked to an H in the full structure (the third candidate would be the appropriate structure for the auditory form [ɓáŋgà] instead). The second candidate is ruled out because it has two H tones that are separated by no more than a syllable boundary. This form then violates the tone-specific OCP constraint that says that two H tones cannot be separated by a syllable boundary only. The first form, with a single H tone, then wins, although it violates the generalized line-crossing constraint that says that two H tones cannot be separated by a syllable boundary or more. In Mandarin Chinese, exemplified in tableau (10) with the word 'sofa', the situation is the reverse: an H tone cannot be shared by consecutive syllables, since every syllable is specified separately for one of the four possible tones of this language.

5.2. Autosegmental constraints on nasality

What can be done for tone can be done for any feature that is suprasegmental in one language but segmental in the other. Tableaus (11) and (12), again from Boersma (2000), show examples for nasality.

(11) *Guaraní perception of suprasyllabic nasality*

[tũpã]	$[\tilde{V}] \rightarrow / \genfrac{}{}{0pt}{}{N}{V} /$	OCP$_S$(nas,]$_\sigma$)	LCC$_S$(nas,]$_\sigma$)
☞ N / t u p a /			*
N N / t u p a /		*!	
N / t u p a /	*!		

(12) *French perception of segmental nasality*

[ʃɑ̃sɔ̃]	$[\tilde{V}] \rightarrow / \genfrac{}{}{0pt}{}{N}{V} /$	LCC$_S$(nas,]$_\sigma$)	OCP$_S$(nas,]$_\sigma$)
N / ʃ ɑ s ɔ /		*!	
☞ N N / ʃ ɑ s ɔ /			*
N / ʃ ɑ s ɔ /	*!		

In Guaraní, nasality is assigned at the word level: there are words pronounced as [tũpã] 'God' and [tupa] 'bed', but no words pronounced as *[tũpa] or *[tupã]. The usual view (e.g. Piggott 1992, Walker 1998) is that the form [tũpã] has to be interpreted as having a single nasality (N) value. Tableau (11) formalizes this as a high ranking of the OCP for nasality in Guaraní. In French, the nasality of consecutive vowels is uncorrelated, since there are words pronounced as [ʃɑ̃sɔ̃] 'song', [lapɛ̃] 'rabbit', [ʃapo] 'hat', and [pɔ̃so] 'poppy'. This means that nasality has to be stored separately with every vowel in the lexicon. If perception is to be aimed at maximally facilitating lexical access (Boersma 2000), French perception must map the two nasalized vowels in [ʃɑ̃sɔ̃] to two different [nas] feature values in the phonological surface structure, as in tableau (12).

5.3. Loanword adaptation

We are now ready to discuss the subject of loanword adaptation. There has been much controversy as to whether loanword adaptation is due to 'perception' or to 'phonology'. But in an OT account of perception, in which phonological (structural) constraints influence the perception process, there is no dichotomy. Tableaus (13) and (14) give the example (from Boersma [2000] 2003: 32) of the adaptation of the Portuguese auditory forms [ʒwẽw̃] ‘John’ and [sɐbẽw̃] ‘soap’ by speakers of Desano (Kaye 1971), another nasal harmony language. The structural constraints */C–nas V+nas/ and */σ–nas σ+nas/ militate against nasal disharmony within and across syllables, respectively, and the cue constraints [V±nas] → /V±nas/ and [C±nas] → /C±nas/ express the favoured interpretation of nasality cues for vowels and consonants, respectively.

(13) *Desano adaptation of Portuguese*

<table>
<tr><th>[ʒwẽw̃]</th><th>*/C–nas V+nas/</th><th>*/σ–nas σ+nas/</th><th>[V±nas]
→ /V±nas/</th><th>[C±nas]
→ /C±nas/</th></tr>
<tr><td>N
|
ʒ u</td><td>*!</td><td></td><td></td><td></td></tr>
<tr><td>☞ N
/\
ɲ u</td><td></td><td></td><td></td><td>*</td></tr>
<tr><td>ʒ u</td><td></td><td></td><td>*!</td><td></td></tr>
</table>

(14) *Desano adaptation of Portuguese*

<table>
<tr><th>[sɐbẽw̃]</th><th>*/C–nas V+nas/</th><th>*/σ–nas σ+nas/</th><th>[V±nas]
→ /V±nas/</th><th>[C±nas]
→ /C±nas/</th></tr>
<tr><td>N
|
s a b o</td><td>*!</td><td></td><td></td><td></td></tr>
<tr><td>N
/\
s a m o</td><td></td><td>*!</td><td></td><td>*</td></tr>
<tr><td>N
//\\
n a m o</td><td></td><td></td><td>*</td><td>*!*</td></tr>
<tr><td>☞ s a b o</td><td></td><td></td><td>*</td><td></td></tr>
</table>

Since Polivanov (1931), then, foreign-language perception and loanword adaptation have been seen by some to involve an interaction between language-specific cue constraints, which partly reflect auditory closeness, and language-specific structural constraints. This is phonology and perception at the same time.

5.4. Arbitrary relations between auditory and surface forms

The cue constraints in (9) to (14) look a bit like faithfulness constraints, e.g. “if there are nasality and vowel cues in the input, the output must have nasality linked to a vowel”. Such simplifying formulations disguise what is really going on, namely a partly arbitrary relation between auditory input and phonological output. The arbitrariness becomes especially visible if we consider cases of *cue integration*. Tableaus (15) and (16), from Escudero & Boersma (2004), give examples of the integration of auditory vowel height (first formant, F1) and auditory duration into the single contrast between the English vowels /i/ and /ɪ/.

(15) *Perception of an auditory event in Scottish English*

[74 ms, 349 Hz]	*/ɪ/ [349 Hz]	*/i/ [74 ms]	*/ɪ/ [74 ms]	*/i/ [349 Hz]
/ɪ/	*!		*	
☞ /i/		*		*

(16) *Perception of the same auditory event in Southern British English*

[74 ms, 349 Hz]	*/i/ [349 Hz]	*/i/ [74 ms]	*/ɪ/ [74 ms]	*/ɪ/ [349 Hz]
☞ /ɪ/			*	*
/i/	*!	*		

The example of tableaus (15) and (16) is a relatively short high vowel. For a Scot, such a token must represent the vowel in *sheep*, because the vowel in *ship* tends to be much more open, and both vowels are short. For a Southern Brit, the same auditory event must represent the vowel in *ship*, because the vowel in *sheep* tends to be much longer, and both vowels are high. These observations are reflected here in the continuous cue constraint families "an auditory F1 of [*x* Hz] should not be perceived as the phonological vowel category /*y*/" and "an auditory duration of [*x* ms] should not be perceived as the phonological vowel category /*y*/". In these tableaus, we again see the language-specificity of perception, as well as the partial arbitrariness of the mapping from auditory to phonological. If the reader does not consider the arbitrariness idea convincing (perhaps because auditory F1 could map to a phonological height feature and auditory duration could map to a phonological length feature), the reader might want to ponder the case of the word-final obstruent voicing contrast in English, which involves a single phonological voice feature but multiple auditory cues such as vowel duration, consonant duration and burst strength.

The simplest case of arbitrary categorization constraints is the case of the categorization of a single auditory continuum, say F1, into a finite number of phonological classes, say /a/, /e/, and /i/. Tableau (17) shows how an F1 of [380 Hz] can be perceived as /e/ in language with three vowel heights (from Boersma 2005b).

(17) *Classifying F1 into vowel height*

[380 Hz]	*/a/ 320 Hz	*/a/ 380 Hz	*/i/ 460 Hz	*/e/ 320 Hz	*/a/ 460 Hz	*/i/ 380 Hz	*/e/ 380 Hz	*/i/ 320 Hz	*/e/ 460 Hz
/a/		*!							
☞ /e/							*		
/i/						*!			

The number of such constraints is very large. Fortunately, the ranking can be learned under the guidance of the lexicon (Boersma 1997; Escudero & Boersma 2003, 2004).

6 Fitting this into a model of bidirectional phonology and phonetics

We have seen that perception can be modelled in OT. But is it also necessary to model it in OT. Why not neural nets or so?

One of the reasons is that perception is restricted by the same structural constraints as production, which every OT phonologist agrees should be modelled in OT. The complete grammar model in Figure 1 makes this explicit.

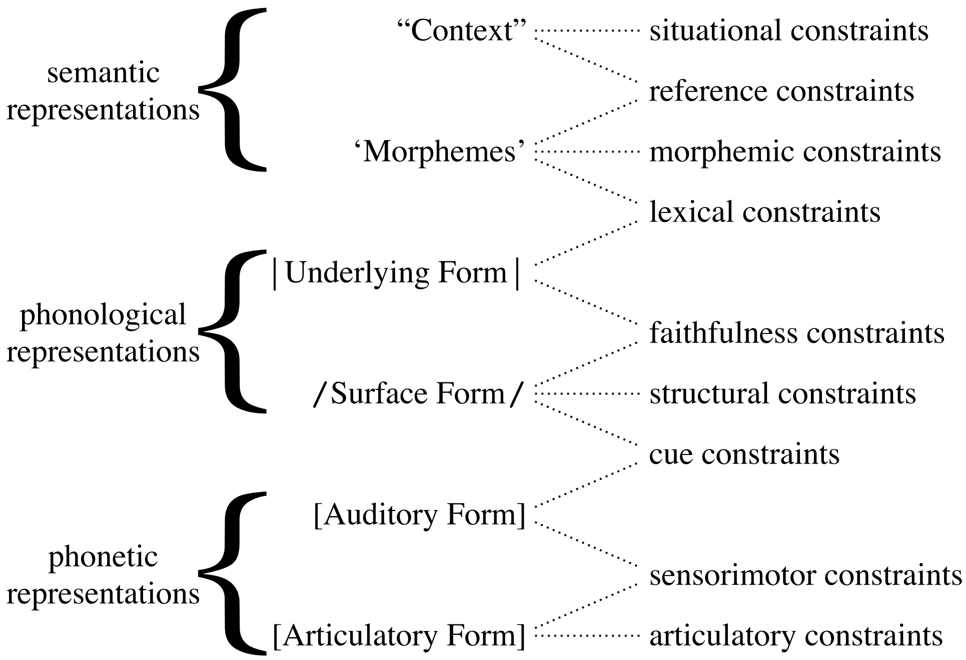

Fig. 1 A complete grammar model for bidirectional phonology and phonetics.

The figure illustrates that the structural constraints evaluate the output of the mapping from Underlying Form to Surface Form (i.e. merely-phonological production), as well as the mapping from Auditory Form to Surface Form (i.e. prelexical perception). If these constraints are ranked in the OT way, they in order to make the most out of them, they should be integrated in our model of perception to the same extent as they are integrated in our model of production (or even more so, as Boersma 1998 used to argue). This argument was valid when Tesar & Smolensky formulated it for overt forms and stress parsing, and it is equally valid for a larger system of representations and constraints, as the one advocated in Figure 1.

References

Boersma, Paul (1998). *Functional phonology: formalizing the interactions between articulatory and perceptual drives*. Doctoral dissertation, University of Amsterdam.

Boersma, Paul (2000). The OCP in the perception grammar. *Rutgers Optimality Archive* **435**.

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

Boersma, Paul (2005a). Some listener-oriented accounts of hache aspiré in French. Ms., University of Amsterdam. [Rutgers Optimality Archive **730**] [Revised version to appear in *Lingua*]

Boersma, Paul (2005b). Prototypicality judgments as inverted perception. Ms., University of Amsterdam. [Rutgers Optimality Archive **742**] [To appear in a book edited by Gisbert Fanselow & Caroline Féry]

Boersma, Paul, and Paola Escudero (2004). Learning to perceive a smaller L2 vowel inventory: an Optimality Theory account. *Rutgers Optimality Archive* **684**.

Dupoux, Emmanuel, Kazuhiko Kakehi, Yuki Hirose, Christophe Pallier, Stanka Fitneva, and Jacques Mehler (1999). Epenthetic vowels in Japanese: a perceptual illusion. *Journal of Experimental Psychology: Human Perception and Performance* **25**: 1568–1578.

Escudero, Paola (2005). *The attainment of optimal perception in second-language acquisition*. Ph.D. thesis, University of Utrecht.

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

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

Goldsmith, John (1976). *Autosegmental Phonology*. Doctoral thesis, MIT, Cambridge. [IULC. Garland Press, New York 1979]

Jusczyk, Peter (1997). *The discovery of spoken language*. Cambridge, Mass.: MIT Press.

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

Leben, William (1973): *Suprasegmental Phonology*. Doctoral thesis, MIT, Cambridge. [Garland Press, New York 1980]

Myers, J. Scott (1997). OCP effects in Optimality Theory. *Natural Language and Linguistic Theory* **15**. 847–892.

Pater, Joe (2004). Bridging the gap between receptive and productive development with minimally violable constraints. In René Kager, Joe Pater & Wim Zonneveld (eds.) *Constraints in phonological acquisition*. Cambridge: Cambridge University Press. 219–244.

Piggott, Glyne (1992). Variability in feature dependency: The case of nasality. *NLLT* **10**: 33–78.

Polivanov, Evgenij Dmitrievič (1931). La perception des sons d'une langue étrangère. *Travaux du Cercle Linguistique de Prague* **4**: 79–96. [English translation: The subjective nature of the perceptions of language sounds. In E.D. Polivanov (1974): *Selected works: articles on general linguistics*. The Hague: Mouton. 223–237]

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

Tesar, Bruce, and Paul Smolensky (1998). Learnability in Optimality Theory. *Linguistic Inquiry* **29**: 229–268.

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

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