Cue constraints and their interactions in phonological perception and production[*]

Paul Boersma, 26 July 2009

This paper shows how one can formalize the phonology-phonetics interface within constraint-based frameworks such as Optimality Theory (OT) or Harmonic Grammar (HG) and why it is necessary and advantageous to do so. I will describe the phonology-phonetics interface in terms of *cue constraints* and illustrate their workings in interaction with each other and in interaction with other classes of constraints. Beside their general usefulness in describing prelexical perception and phonetic implementation, cue constraints help to account for special phenomena such as poverty of the base, the prototype effect, foreign-language perception, and loanword adaptation.

1.   Where is the phonology-phonetics interface?

My first task is to make the phonology-phonetics interface explicit. Figure 1 shows where it resides in an explicit multi-level model of phonology and phonetics (Boersma 1998, 2007a; Apoussidou 2007). In the following two sections I briefly clarify why phonological theory requires at least the five representations shown in Figure 1, and why the phonology-phonetics interface must be where it is in Figure 1.
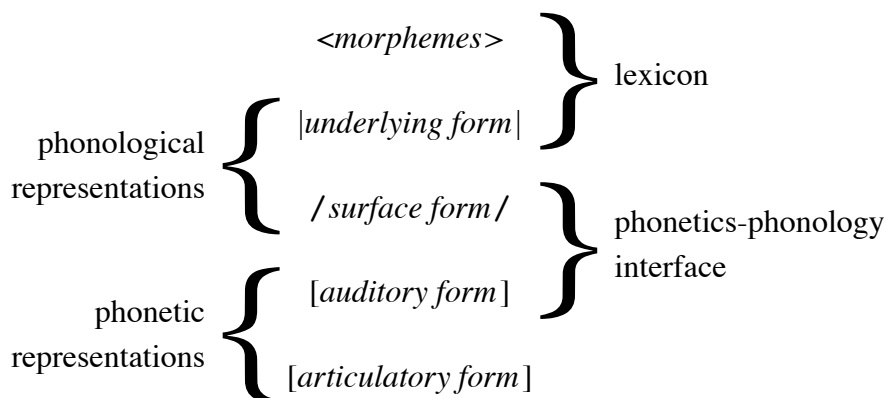
Fig. 1   The BiPhon model (Boersma 2007a, Apoussidou 2007): the five levels of representation minimally required for doing bidirectional phonology and phonetics.

1.1. Phonological theory requires at least five representations

Multiple levels of representation such as those in Figure 1 are common in psycholinguistic models of speech production (e.g. Levelt 1989) and comprehension (e.g. McClelland and Elman 1986, McQueen and Cutler 1997), but are less often seen in phonological theory (I will discuss the exceptions later on). As in my earlier work, I argue in this paper that phonological theory would profit from going beyond its usual two levels, especially by taking the phonetics seriously.

Figure 1, then, contains the minimal number of phonological representations that phonologists can be comfortable working with: the Underlying Form (often called the *input* in the OT literature) and the Surface Form (often called the *output* in the OT literature). The underlying form is the discrete representation of the phonological structure of morphemes in the language user's mental lexicon; the Morpheme (or Lemma) mediates in connecting the phonological underlying form to semantic features in the lexicon (Saussure 1916), which again are probably connected to the meaning of the utterance and from there to the pragmatic context (these are not shown in Figure 1, since they are likely to concern phonological theory to a lesser extent). The surface form is the discrete representation of the phonological surface structure and consists of prosodic elements (feet, syllables, segments) and phonological substance (features, autosegments). For most phonologists, the surface form does not contain any concrete continuous phonetic detail, and this is something I agree with; it means that one can do insightful investigations in many areas of phonology by just considering the two discrete phonological representations.

As for phonetic processing, this is usually regarded as something that comes 'after' the phonology (e.g. Pierrehumbert 1987, Keating 1985, Gussenhoven 2004). Phonologists tend to argue that the phonetics is therefore not really relevant for autonomous phonological processing (e.g. Hale and Reiss 1998), or that it might be relevant but that its modelling is not a priority for phonological theory (e.g. Hayes 1999: fn. 7). On the basis of the abundant existence of seemingly phonetically-inspired processes in segmental phonology, some phonologists have nevertheless tried to include phonetic considerations of articulatory effort and auditory contrast into the usual two-level model of phonology consisting only of underlying and surface form; by doing so, one must either propose that the phonological surface structure somehow includes continuous phonetic detail (Jun 1995, Kirchner 1998; cf. Flemming 1995 for a model without underlying forms but with a phonetically detailed surface form) or that discrete phonological processing is somehow sensitive to extralinguistic information on phonetic detail (Steriade 1995, 2001). Following Boersma (1998), I take the third possible stand, which takes seriously the possible relevance of phonetics for phonological theory without sacrificing the representational modularity of the phonology and the phonetics: in Figure 1, therefore, the phonetic representations are separate from the phonological representations, but are taken just as seriously. The minimum number of phonetic representations that phoneticians can be comfortable working with are two: the Auditory Form and the Articulatory Form. The auditory form is the continuous representation of sound; it consists of noises, pitches, spectra, silences, transitions, and durations. The articulatory form is the continuous representation of the gestures of the human sound-producing mechanism; it consists of the activities of the relevant muscles of the lungs, tongue, throat, larynx, lips and nose and their coordinations.

In the end, whether the phonology and the phonetics are as separated as they appear in Figure 1 is an empirical question. For the moment, however, it seems to suffice to observe that the multi-level model comes with a learning algorithm that has been shown to generate automatically at least three phenomena: (1) the *prototype effect* (Boersma 2006), which leads to the evolutionary emergence of optimal auditory contrast in inventories (Boersma and Hamann 2008); (2) *licensing by cue* (Boersma 2008); and (3) the relation between various properties formerly ascribed to the concept of *markedness* (namely frequency and phonological activity; Boersma 2008). No such explanatory force has been shown to hold for any of the two-level models with which people have tried to explain these phenomena (namely: the MINDIST constraints by Flemming 1995, the P-map by Steriade 2001, and the markedness constraints by Prince and Smolensky 1993 and much following work in OT). The present paper refers to these phenomena and their explanations where they fit naturally in my discussion of the bidirectional use of cue constraints.

To sum up, the present paper assumes (and requires) the five levels of representation mentioned in Figure 1. In real users of language there may well turn out to be more representations than five. For instance, the Underlying Form may turn out to have to be divided up into a Stem Level and a Word Level (Kiparsky 1985, Bermúdez-Otero 1999), and the Auditory Form may turn out to have to be divided into a representation 'before' speaker normalization and a representation 'after' speaker normalization. For the purposes of the present paper, however, the five levels of Figure 1 suffice. The next question is how the two phonological representations connect to the semantic and phonetic ones.

1.2. The phonology-phonetics interface is between Surface and Auditory Form

Given the five levels of representation mentioned in Figure 1, we are left with a couple of possibilities for how and where the phonology interfaces with the morphology (and hence with the semantics) and where it interfaces with the phonetics.

The division of labour between the two phonological representations in this respect seems to be clear: in all published grammar models that make a distinction between underlying and surface form, it is the underlying form that connects to morphemes and meaning (via the lexicon), and it is the phonological surface form that connects to the phonetics (via the phonology-phonetics interface). I see no reason to deviate from this common viewpoint. Figure 1 therefore illustrates my assumption that the connection to the morphology and the semantics is made from the underlying form, and the connection to the phonetics is made from the phonological surface form.

It is a somewhat more controversial matter, however, which of the two phonetic representations (auditory or articulatory) connects to the phonological surface form. The Direct Realist theory of speech perception (Fowler 1986) proposes that auditory speech is directly interpreted in terms of articulatory gestures and that it is these perceived gestures that connect to the phonology. That theory, when confronted with the five representations of Figure 1, would therefore propose that the interface between the phonology and the phonetics resides in a connection between Surface Form and Articulatory Form. While it is thinkable that an explicit model of phonology and phonetics could be based on Direct Realism, such an exercise has yet to be performed. For the present paper I hold the simpler and probably more common assumption that the

auditory-phonetic form connects directly to the phonological surface form. This choice is based partly on theoretical simplicity, since it economizes on one level of representation in the speech comprehension process: a listener who follows Figure 1 just starts out with an Auditory Form and can subsequently process the phonology with the ultimate goal of accessing the semantics, all without ever passing through the articulatory form. Moreover, the observation that children can successfully access meaning from sound, while constructing adultlike phonological representations, well before they can produce any speech (Jusczyk 1997), also points to a direct connection between auditory and surface form (Boersma 1998: ch.14).

To sum up, the interface between phonology and phonetics resides in a connection between the phonological surface form and the auditory-phonetic form. It is now time to make explicit what this connection is about.

1.3.   The phonology-phonetics interface consists of cues

Now that we assume that the phonology-phonetics interface resides in a connection between auditory forms and phonological surface forms, it becomes relevant to ask how the phonetic literature has been talking about this connection. It turns out that phoneticians talk about this connection in terms of *cues*. In English, for instance, auditory vowel duration (in milliseconds) can be a *cue* to the value (plus or minus) of the phonological voicing feature of the following obstruent, both in comprehension and production: English listeners use vowel duration as a cue for reliably perceiving the value of the phonological voicing of the following obstruent (Denes 1955, Hogan and Rozsypal 1980), and English speakers *implement* (or *enhance*) obstruent voicing by using the vowel duration cue (House and Fairbanks 1953; Peterson and Lehiste 1960). This use of cues is a language-specific issue (Zimmerman and Sapon 1958): while most languages lenghten their vowels slightly before voiced consonants, English does it to an especially large extent (Peterson and Lehiste mention a ratio of 2:3 for an unspecified variety of American English).

1.4.   In OT, cues are formalized as cue constraints

Following much earlier work in OT, the present paper assumes that the five levels of representations in Figure 1 are linked by local connections that are implemented as constraints, as in Figure 2.

Figure 2 mentions six types of constraints. The *faithfulness constraints* and the *structural constraints* are the same ones that phonologists have been familiar with since Prince and Smolensky (1993), although the explicit division between Underlying Form and Surface Form, and therefore the formulation of the faithfulness constraints, follows more closely the Correspondence account by McCarthy and Prince (1995); the faithfulness constraints therefore evaluate the similarity between underlying and surface form, and the structural constraints evaluate the surface form alone. The *articulatory constraints* are the ones that were proposed by Kirchner (1998) and Boersma (1998), and measure articulatory effort; following Boersma (1998), these constraints evaluate the articulatory-phonetic form, not the phonological surface form. The *lexical constraints* express the relation between underlying form and morphemes (or meaning) in the lexicon; they were discussed by Boersma (2001) and Escudero (2005: 214–236) and formulated in terms of a connection between two separate levels of representation

(as in Figure 2) by Apoussidou (2007). The *cue constraints* express the language user's knowledge of cues (§1.3), i.e. the relation between auditory form and phonological surface form; these constraints appeared in Boersma (1998, 2000), Escudero and Boersma (2003, 2004), and Pater (2004), although the term 'cue constraint' was not introduced before Boersma (2007a) and Escudero (2005). Finally, the *sensorimotor constraints* (Boersma 2006) express the language user's knowledge of the relation between articulation and sound; with them, the speaker knows how to articulate a given sound and can predict what a certain articulatory gesture will sound like.
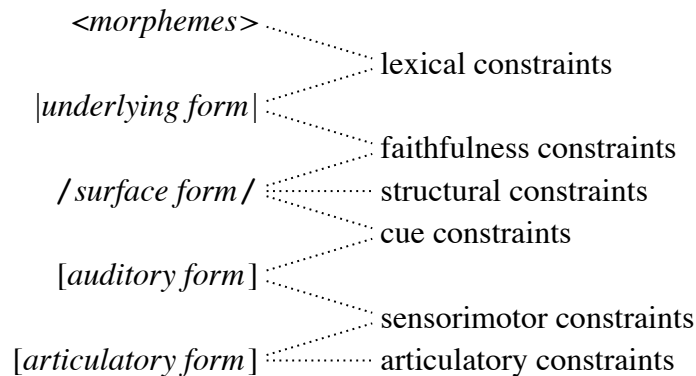

<morphemes> ⋯⋯⋯
                      lexical constraints

|underlying form| ⋯
                      faithfulness constraints

/ surface form / ⋯⋯ structural constraints
                      cue constraints

[auditory form] ⋯
                      sensorimotor constraints

[articulatory form] ⋯⋯ articulatory constraints

Fig. 2   Formalization of bidirectional phonology and phonetics by means of constraints.

The cue constraints, which are the central subject of this paper, will be seen to be able to interact with all of the remaining five types. This ability crucially relies on two properties of the BiPhon model: *bidirectionality of constraints*, which is the use of the same constraints and rankings by both the listener and the speaker, and *cross-level paralellism*, which is the capability of all the constraints of all the levels to interact with each other. I discuss these two properties in the next two sections.

1.5. Bidirectionality of constraints

Nearly all the constraints in Figure 2 are used both by the speaker and by the listener. The task of the speaker is to turn an intended meaning into an articulation, and the task of the listener is to turn an incoming auditory form into a meaning. Several cases of bidirectional use of constraints have been discussed in the literature, and I review them here.

      Bidirectionality of faithfulness constraints. Since Prince and Smolensky (1993) we have known that 'phonological production', i.e. the mapping from underlying to surface form in Figure 2, involves an interaction of structural and faithfulness constraints. Figure 2 makes this interaction explicit by showing that the structural constraints evaluate the output of this mapping, while the faithfulness constraints evaluate the relation between the input and the output of this mapping. Smolensky (1996) has shown that this mapping can be reversed: the mapping from surface to underlying form ('phonological comprehension') is evaluated by the same faithfulness constraints that evaluate phonological production, and with the same rankings. In Figure 2 we see that this mapping is evaluated by faithfulness constraints alone, because there are no constraints that evaluate its output (namely the underlying form). Smolensky makes explicit the point that the structural constraints cannot be involved in this

mapping, because these constraints evaluate its input (namely the surface form), which is identical for all candidates; therefore, structural constraints cannot be used bidirectionally in a grammar model with just two levels of representation (underlying and surface form).

Bidirectionality of structural constraints. In a grammar model with three levels of representation, structural constraints *can* be used bidirectionally, at least if the surface form is the intermediate level. Tesar (1997) proposed a model for metrical phonology with an underlying form, a surface form, and a more 'phonetic' *overt form*; in that model, structural constraints evaluate both the output of the speaker's phonological production as well as the output of the listener's mapping from overt to surface form ('robust interpretive parsing'). Since the latter mapping may involve additional cue constraints, I discuss this subject in detail in §4.

Bidirectionality of lexical constraints. In comprehension, lexical constraints have been shown to help word recognition, for instance by disambiguating phonologically ambiguous utterances (Boersma 2001, Escudero 2005: 214–236), and in production they have been shown to be able to regulate allomorphy (Apoussidou 2007: ch.6). A discussion of their interactions with cue constraints appears in §8.

Bidirectionality of cue constraints. Cue constraints have been shown to be able to handle the listener's 'prelexical perception', i.e. the mapping from auditory to surface form (Boersma 1998; Escudero and Boersma 2004; Escudero 2005), as well as the speaker's phonetic implementation (Boersma 2006, 2007a, 2008; Boersma and Hamann 2008). The present paper illustrates both of these roles of cue constraints, especially as they interact with structural, faithfulness, articulatory, and lexical constraints.

1.6. Cross-level parallelism

Interactions of cue constraints with all the other types of constraints are not automatically allowed by just any model of processing.

Consider, for instance, the serial processing model in Figure 3. On the left side we see the processing task of the listener, namely mapping an incoming auditory form (sound) all the way to morphemes and meaning. In the serial model of Figure 3, this task consists of three subtasks that process the incoming information sequentially: first, the module of prelexical perception maps the auditory form to a surface form; then, the module of word recognition maps this surface form to an underlying form; finally, the module of lexical access connects the underlying form to the morpheme and meaning.[1] On the right side of Figure 3 we see the processing task of the speaker, namely mapping a morphological (and perhaps semantic) representation all the way to an articulation. In the serial model depicted here, this task consists of four subtasks: a module of lexical retrieval, whose output is the underlying form, which is the input to the phonological production module, whose output is the surface form, which is the input to the phonetic implementation module, whose output is an auditory form (i.e. the speaker's view of what she will sound like), which is the input to the final sensorimotor processing module.

---

[1] The dotted curve from auditory to articulatory form in Figure 3 is not part of the comprehension task, but is predicted to occur nevertheless: when a sound comes in, articulatory representations will be automatically activated. The activity of mirror neurons in real human brains may be a sign of this.
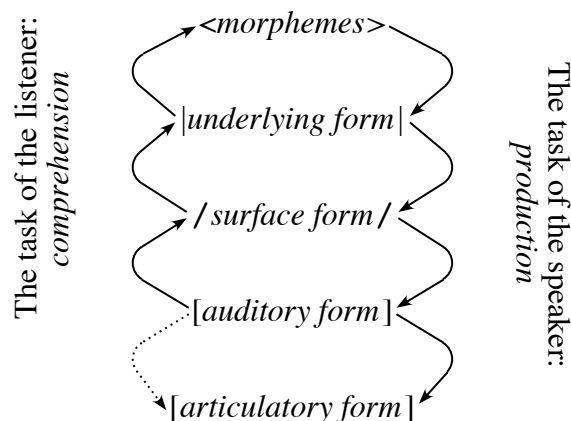
Fig. 3   The processing routes of serial comprehension and serial production.

In the serial ('modular') view of comprehension in Figure 3, the extent to which cue constraints can interact with other constraints is quite limited. When comparing the arrows in Figure 3 with the constraints in Figure 2, we see only three modules in which constraints of various types are allowed to interact: on the comprehension side, prelexical perception is handled by an interaction of structural constraints and cue constraints, and on the production side, phonological production is handled by an interaction of structural and faithfulness constraints while sensorimotor processing is handled by an interaction of articulatory and sensorimotor constraints. The remaining four modules are handled by a single type of constraint: word recognition by faithfulness constraints, lexical access and lexical retrieval by lexical constraints, and phonetic implementation by cue constraints. The only interaction that is allowed for cue constraints is an interaction with structural constraints, and this interaction only occurs in comprehension.

The situation is strikingly different for the parallel (or 'interactive') processing model in Figure 4. On the left side we see that comprehension now involves a parallel handling of prelexical perception, word recognition, and lexical access; in Optimality-Theoretic terms, the listener, given an auditory form as input, has to decide on a simultaneously optimal triplet of surface form, underlying form, and morphemes. On the right side of Figure 4 we see that production now involves a parallel handling of lexical retrieval, phonological production, phonetic implementation, and sensorimotor processing; in Optimality-Theoretic terms, the speaker, given a sequence of morphemes as input, has to decide on a simultaneously optimal quadruplet of underlying, surface, auditory and articulatory forms.
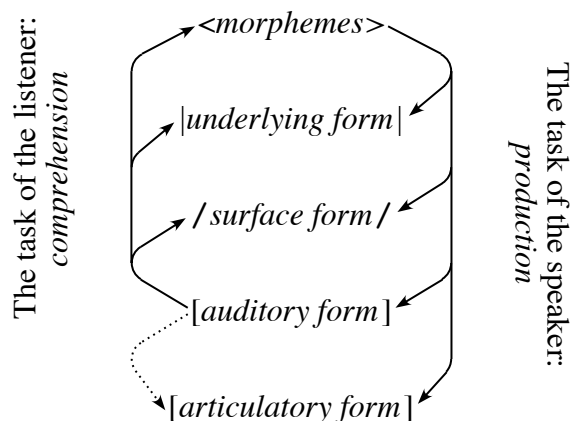
Fig. 4 The processing routes of parallel comprehension and parallel production.

The parallel view of Figure 4 allows many interactions between many more types of constraints than the serial model of Figure 3 does. In the comprehension direction, cue constraints, structural constraints, faithfulness constraints and lexical constraints can all interact with each other, though not with sensorimotor or articulatory constraints. In the production direction, all the six types of constraints that occur in Figure 2 can interact with each other.

The extent to which linguistic processing is serial or parallel is an open question. Especially in production a good case for extensive parallelism can be made: we have witnessed interactions of faithfulness and articulatory constraints (Boersma 1998), interactions of articulatory and cue constraints (Boersma 2006, Boersma and Hamann 2008), and even quadruple interactions of faithfulness, structural, cue, and articulatory constraints (Boersma 2007a). The present paper discusses most of the possible types of interactions between cue constraints and other constraints, i.e. most of the interactions that involve the phonology-phonetics interface. The focus of the paper (§2-§4) is on interactions of cue constraints and structural constraints in comprehension, because such interactions are predicted both by the serial and by the parallel model; a formalization of these interactions explains several old issues in phonological theory and shows that phonetic considerations have to be formalized in the same way as phonological considerations, because they interact in the same process.

2. Perception and its formalization

Of the several linguistic processes that involve cue constraints, the first and main one that I formalize is (prelexical) perception. The primacy of this process lies in the fact that it is the process that least controversially involves the phonology-phonetics interface, and at the same time shows that phonological considerations (structural constraints) are in direct competition with more phonetic considerations (cue constraints); their interaction shows that *perception is phonological*.

2.1. 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 phonological structure'. But 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 language-specific 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 identification tasks that phoneticians conceive in the lab. Other possible terms for the same thing are *prelexical perception* and *phonetic parsing*.

2.2. Modelling robust language-specific perception in OT

To start modelling perception in Optimality Theory, I single out the auditory-to-surface mapping from Figures 2, 3, and 4. This yields Figure 5, which shows both the processing (as a curved arrow) and the grammar (the constraints).
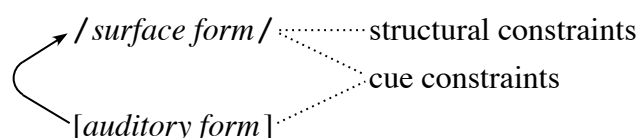

Fig. 5 Prelexical perception.

The structural constraints evaluate the output of the perception process (surface form, SF), and the cue constraints evaluate the mapping between the input (auditory form, AudF) and the output (SF). The structural constraints are the same ones as in production (Prince and Smolensky 1993), where they interact with faithfulness constraints. The cue constraints 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. Just as the word *faithfulness* the term *cue* implies a relation between two representations ("a surface form can be *faithful to* an underlying form"; "an auditory form can be a *cue for* a surface form"). The cue constraints that have been proposed in the OT literature are OCP (Boersma 1998) and the generalized

categorization constraint family "$[x]_{AudF}$ is not perceived as $/y/_{SF}$" (Escudero and Boersma 2003, 2004; Boersma and Escudero 2008; Escudero 2005). Some examples from the pre-OT literature are Polivanov (1931) and Cornulier (1981); I have discussed the latter elsewhere (Boersma 2007a), and discuss the former in the next section.

3.   Polivanov: Japanese learners of Russian[2]

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. This section shows that Polivanov's proposal can be formulated in all details with the decision mechanism of Optimality Theory (Prince and Smolensky 1993) and fits well within the models of Figures 1, 2, 3, 4, and 5.

3.1.   Perception

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.

In a narrower transcription, the auditory form is [ˌˈaᵏ]: as you can see in a spectrogram, this sound consists of (at least) a silence ([ˍ]), followed by a high-frequency brief noise ("burst") ([ˈ]), followed by a loud periodic ("sonorant") sound with formants around 1000 Hz ([a]), followed by another silence, followed by a burst with a peak around 2500 Hz ([ᵏ]). A listener of Russian will have to map this sound 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. an auditorily periodic and 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 such as /a/, /e/ and /i/, periodic (i.e. voiced) non-sonorant consonant phonemes such as /b/, and non-periodic (i.e. voiceless) phonemes such as /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 (= loud periodic) auditory form cannot refer to a plosive. This is the second-highest constraint. 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 is very unlikely that the speaker could have intended to put an underlying |i| into your head. That is the third constraint. It must also be slightly unlikely that the speaker's intention was |e|. That is the fourth constraint. 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 appears in the last column.

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

Now consider the auditory form [ˍᵗaˍᵏ] 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 [ˍᵗaˍᵏ], so the perception /.tak./ is unlikely. So what will a Japanese learner of Russian do when first hearing a Russian say the utterance [ˍᵗaˍᵏ]?

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 [ˍᵏáˍᵏu̥]. 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 [ˍᵏáˍᵏ]. Thus, Japanese listeners are used to interpreting a silence, i.e. the auditory form [ ], as the vowel /u/ (at least after a release burst). They will perceive the Russian [ˍᵗaˍᵏ] 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

[_ˈa_ᵏ]	NOPLOSIVECODA	*/ / [ᵏ]	*/o/ [ ]	*/u/ [ ]
/.tak./	*!			
/.ta./		*!		
☞ /.ta.ku./				*
/.ta.ko./			*!	

In Tableau (2) the Japanese ban on plosive codas has been formalized as the constraint NOPLOSIVECODA; it assigns a violation mark to any plosive that occurs in a coda position in the surface structure. The listener's resistance to ignoring the [ᵏ] release burst is expressed as the cue constraint */ /[ᵏ]. The cue constraints against hallucinating the vowels /o/ and /u/ are written as */o/[ ] and */u/[ ]; the Japanese-specific routine of filling in the vowel /u/ is reflected as a low ranking of the latter constraint.

This Japanese behaviour of hallucinating a vowel 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 and Mehler 1999). It is the cause behind Japanese loanword adaptations, such as /.e.ki.su.to.ɾa./ 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 [_ˈa_ᵏ] as the surface structure /.tak./. In tableaus like (2), such an outcome can be achieved by a much lower ranking of NOPLOSIVECODA. The language-specificity of perception, then, corresponds to the freedom that every language possesses to rank the constraints in its own order.

The second 'European' word that Polivanov discusses is *drama*. Its auditory form in Russian is [˷ᵈrama], 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" (high-frequency) plosive burst.

A Russian listener would perceive this auditory form as the phonological structure /.dɾa.ma./. A Japanese listener will not perceive it as /.dɾa.ma./, because that form contains a syllable onset that consists of two consonants, 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

[◌̃{*hi-freq*, *burst*}rama]	*/.CC/	*/ / [*burst*]	*/du/	*/*dor*/ [*hi-freq*]	*/+*cont*/ [*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 [◌̃ᵈrama] 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 high-frequency (alveolar) burst. As in the case of [◌̃ˈaᵏ], 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 an /u/ analogously to the /.ta.ku./ case. But Japanese happens not to allow the structure /du/ on the surface. This is what the third constraint expresses. It is another structural constraint.

The fourth option is to perceive /.gu.ɾa.ma./. This has the allowed sequence /gu/. But this candidate, with its /*dor*/ (dorsal) value for the phonological /*place*/ feature, ignores the high-frequency cues for alveolar place, as expressed by the fourth constraint.

The fifth option is to perceive /.zu.ɾa.ma./, a phonotactically allowed sequence that would be pronounced as [◌̃ᵈᶻuɾama]. This does honour the spectral place cues but ignores the auditory cue for plosiveness (namely the burst), positing instead a phonological fricative (denoted in the tableau with the feature value /+*cont*/). 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

[˷{*hi-freq*, *burst*}rama]	*/.CC/	*/ / [*burst*]	*/du/	*/*dor*/ [*hi-freq*]	*/o/ [ ]	*/+*cont*/ [*burst*]	*/u/ [ ]
/.dɾa.ma./	*!						
/.ɾa.ma./		*!					
/.du.ɾa.ma./			*!				*
/.gu.ɾa.ma./				*!			*
☞ /.zu.ɾa.ma./						*	*
/.do.ɾa.ma./					*!		

Polivanov says that for 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 [˷ᵈrama]), 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

[˷{*hi-freq*, *burst*}rama]	*/.CC/	*/ / [*burst*]	*/du/	*/*dor*/ [*hi-freq*]	*/+*cont*/ [*burst*]	*/o/ [ ]	*/d/ [ ]
/.dɾa.ma./	*!						
/.ɾa.ma./		*!					
/.du.ɾa.ma./			*!				*
/.gu.ɾa.ma./				*!			*
☞ /.zu.ɾa.ma./					*		*
☞ /.do.ɾa.ma./						*	

This concludes Polivanov's story. It involves three inviolable structural constraints and two competing cue constraints. In the light of possible abstract analyses of Japanese surface forms, however, the story is not yet complete. I defer this issue to §3.4.

Now that Polivanov's Japanese learner of Russian has perceived the sound [˷ᵈrama] as the phonological surface structure /.do.ɾa.ma./, the next question is what the learner does with this: in what form will he store it in his lexicon, and how will he pronounce it? Although the answer may seem obvious (underlying form |dorama|, pronunciation [˷ᵈoɾama]), we have to check that our OT grammar indeed generates those forms. This is what sections 3.2 and 3.3 do.

3.2.   The lexicon: poverty of the base

If Polivanov's Japanese listener wants to learn Russian, he will want to store the *drama* word in his early L2-Russian lexicon. Alternatively, if the European concept of 'drama' is useful and distinctive enough to include into his Japanese vocabulary, he may want to store it as a loanword in his native L1-Japanese lexicon. In either case, the form in

which the word is stored is likely to be influenced by his Japanese foreign-language perception of Russian. There are two straightforward strategies for this.

The first straightforward strategy for including the *drama* word into his lexicon relies on *serial comprehension*: the listener takes the output of the prelexical perception, which is /.do.ra.ma./, and uses this as the input for the next process, that of word recognition. This is summarized in Figure 6.
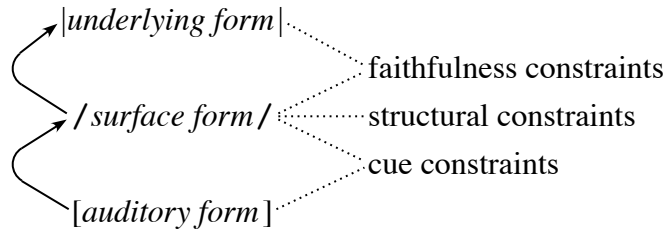
Fig. 6   Serial comprehension.

Once the 'serial' listener, given the sound [˷ᵈrama], has constructed /.do.ra.ma./, he will subsequently map it to the fully faithful underlying form |dorama|:

(6)   'Recognition' of drama

/.do.ra.ma./	*/.CC/	DEP	MAX
☞ |dorama|			
|drama|		*!	
|dorma|		*!	
|doramaribo|			*!***

It is worthwhile to look into the details of (6).

First, the structural constraint */.CC/ evaluates surface forms, not underlying forms; therefore, it can only evaluate the input /.do.ra.ma./, which is the same for all four candidates and which does not violate this constraint. Specifically, the candidate |drama| does *not* violate */.CC/, because |drama|, as an underlying form, is not affected by structural constraints (see also Figure 6). The same point was made explicit by Smolensky (1996) in a discussion of how young children may be hindered by structural constraints in their productions but not in their comprehension.

Second, the candidate |drama| violates DEP (McCarthy and Prince 1995), a constraint against having surface material (the /o/ in /.do.ra.ma./) that is not present underlyingly (in |drama|). Note that although DEP is usually thought of as being an anti-insertion constraint (in a production tableau), in a recognition tableau such as (6) it acts as an anti-deletion constraint: the perceived /o/ is deleted from the recognition output.

Finally, the candidate underlying form |doramaribo| violates fourfold MAX (McCarthy and Prince 1995), a constraint against having underlying material (|ribo|) that does not surface (in /.do.ra.ma./).

In the end, the winner of the recognition process is |dorama|, which is completely faithful to the surface form /.do.ra.ma./. This principle of complete faithfulness is

known in phonological theory from the identical process of *lexicon optimization* (Prince and Smolensky 1993), by which actually existing underlying forms come to reflect the language's high-ranked structural constraints indirectly, simply because the surface forms tend to honour these constraints. Here, this idea extends to forms that have been filtered in the first perception step, such as [˷ᵈrama]. Thus, the mapping from [˷ᵈrama] to |dorama| does not involve a violation of (phonological) faithfulness, because the end result |dorama| is completely faithful to the intermediate form /.do.ɾa.ma./.[3]

The second straightforward strategy to include the word *drama* in the lexicon relies on *parallel comprehension*: the two processes of perception and recognition are handled at the same time and in interaction. Figure 7 summarizes this.
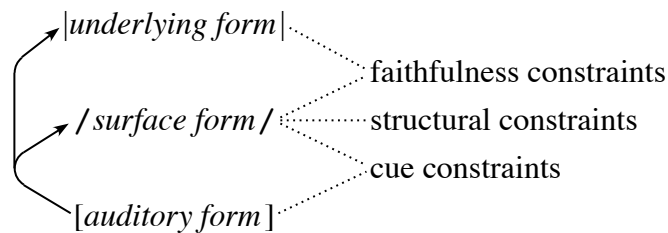

Fig. 7 Parallel (or interactive) comprehension.

The parallel mapping from the auditory form to the surface and underlying forms can be implemented in OT by freely combining pairs of surface and underlying form as candidates in a single *parallel comprehension tableau*, as in (7).

(7)   Perception and recognition in parallel

[˷{*hi-freq, burst*}rama]	*/.CC/	*/du/	DEP	*/+cont/ [burst]	MAX	*/o/ [ ]	*/u/ [ ]
/.dɾa.ma./|dorama|	*!				*		
/.dɾa.ma./|drama|	*!						
/.du.ɾa.ma./|durama|		*!					*
/.zu.ɾa.ma./|zurama|				*!			*
☞ /.do.ɾa.ma./|dorama|						*	
/.do.ɾa.ma./|drama|			*!			*	

In (7) it does not matter how the faithfulness constraints DEP and MAX are interspersed among the structural and cue constraints: since the perceptually optimal candidate pair /.do.ɾa.ma./|dorama| violates neither DEP nor MAX, this pair will be optimal for the whole comprehension process independently of whether DEP and MAX are ranked high or low (or in the middle, as here).

---

[3] Cases where faithfulness is instead violated in the recognition process will occur when there are paradigms with alternations. See Apoussidou (2007: ch. 6) and §8.

The lexicon, then, will be genuinely limited by the structural constraints, as long as these outrank the relevant cue constraints. These limitations on lexical structures have been called *poverty of the base* (Boersma 1998: 395; Broselow 2004; see also §4).

3.3. Production

Now that we know that Japanese listeners perceive an overt [ˍᵈrama] as /.do.ɾa.ma./ and store it in their lexicon as |doɾama|, what will they pronounce it as? Here again, we distinguish between serial and parallel processing.

In a serial view, phonological-phonetic production consists of three steps, as summarized in Figure 8. First, a process of *phonological production* maps a phonological underlying form to a phonological surface form; subsequently, a process of *phonetic implementation* first maps this surface form to an auditory-phonetic form, and finally maps this auditory form to an articulatory-phonetic form.
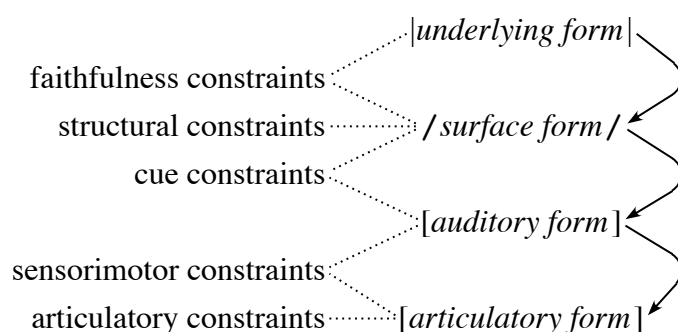

faithfulness constraints

structural constraints

cue constraints

sensorimotor constraints

articulatory constraints

|underlying form|

/ surface form /

[auditory form]

[articulatory form]

Fig. 8 Serial production.

The first of the three subprocesses, phonological production, maps the underlying |doɾama| straightforwardly to /.do.ɾa.ma./, a form that satisfies both the structural constraints and the faithfulness constraints, as shown in (8).

(8)   Phonological production

|doɾama|	*/.CC/	*/du/	DEP	MAX	IDENT
/.dɾa.ma./	*!			*	
/.du.ɾa.ma./		*!			*
/.zu.ɾa.ma./					*!*
☞ /.do.ɾa.ma./					

The second subprocess maps /.do.ɾa.ma./ to [ˍ{*hi-freq,burst*}oɾama], a form that satisfies all high-ranked cue constraints, as shown in (9). To rule out weird pronunciations like [zuɾama], I have added the cue constraint */−cont/[noise], which militates against connecting phonological plosiveness to auditory frication noise, and the cue constraint */o/[u], which militates against having the auditory vocalic material [u] that is inappropriate for the corresponding phonological vowel /o/. Further please note that the notation of phonetic candidates in terms of IPA symbols, such as [ˍᵈo], is just a shorthand for an expression in terms of continuous auditory features, such as [*voicing murmur*; *brief high-frequency noise*; *loud periodicity with mid F1 and low F2*];

17

especially, no representational identity with similar-looking surface structures such as /.do./ is intended; indeed, the latter is a shorthand for a discrete phonological structure such as /(σ C, *–cont*, *–nas*, *cor*, *+voi*; V, *–high*, *–low*, *+back*)σ/ (or any alternative formulation that suits one's particular theory of phonological features).

(9) Phonetic implementation

/.do.ra.ma./	*/+*cont*/* [*burst*]	*/−*cont*/* [*noise*]	*/o/* [u]	*/o/* [ ]	*/u/* [ ]
[zuɾama]		*!	*		
[zoɾama]		*!			
[˔ᵈuɾama]			*!		
☞ [˔ᵈoɾama]					
[˔ᵈrama]				*!	

In the mapping from surface form to auditory form, structural constraints play no role: they would just evaluate the input to this process, which is an invariable /.do.ra.ma./ shared among all candidates, all of which would therefore violate the exact same structural constraints. Therefore, the mapping is entirely determined by cue constraints, and they are in favour of the candidate [˔ᵈoɾama], which violates none of them.

The third subprocess presumably maps an auditory [˔ᵈoɾama] to an articulatory [˔ᵈoɾama]. I have not formalized the sensorimotor constraints here, and for simplicity I assume that they favour the articulatory form [˔ᵈoɾama]. If articulatory constraints are not in the way (see §6), then the speaker will indeed pronounce [˔ᵈoɾama].

In the parallel view, production consists of a simultaneous mapping from the underlying form to an optimal triplet of surface form, auditory form, and articulatory form, as in Figure 9.
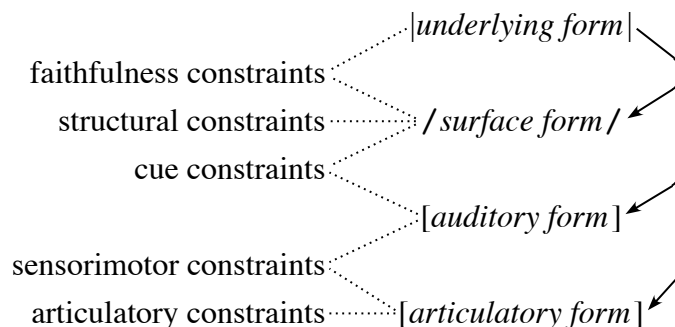

Fig. 9 Parallel (or interactive) production.

Ignoring the articulatory form again, an underlying |doɾama| will be mapped on the surface-auditory pair /.do.ra.ma./[˔ᵈoɾama], just as in the serial view. This is shown in tableau (10).

(10)   Parallel production

|doɾama|	*/.CC/	*/du/	*/+cont/ [*burst*]	MAX	IDENT	*/o/ [u]	*/o/ [ ]
/.dɾa.ma./[̰ᵈɾama]	*!			*			
/.dɾa.ma./[̰ᵈoɾama]	*!			*			
/.du.ɾa.ma./[̰ᵈuɾama]		*!			*		
/.zu.ɾa.ma./[zuɾama]					*!*		
/.zu.ɾa.ma./[̰ᵈɾama]			*!		**		
☞ /.do.ɾa.ma./[̰ᵈoɾama]							
/.do.ɾa.ma./[̰ᵈɾama]							*!
/.do.ɾa.ma./[̰ᵈuɾama]						*!	

Since the winner in tableau (10) violates none of the constraints, it is optimal regardless of the constraint ranking. The result of parallel production is therefore the same as the result of serial production (in this example; for examples where the two are different see §7.2 and Boersma 2007a, 2008).

The end result is that the modelled Japanese learner of Russian hears an incoming auditory [̰ᵈrama] but produces an articulatory [̰ᵈorama].

We have to note that in order for this case to have been so interesting that Polivanov used this as an example, the native Russian perception plays a crucial role as well: Polivanov, with his Russian perception, interpreted the original auditory [̰ᵈrama] as the Russian-compatible surface structure /.dra.ma./, whereas he interpreted the Japanese pronunciation [̰ᵈorama] as the Russian-compatible surface structure /.do.ra.ma./; the discrete difference between the two surface structures is what must have led him to take this as an example for his study on foreign language perception.

3.4.   Abstract surface forms

Something is missing in the story of §3.1 through §3.3: the case of [̰ˡr]. The European word *extra*, for instance, is borrowed into Japanese as /.e.ki.su.to.ɾa./ and not as /.e.ki.su.tu.ɾa./, with the less audible /u/ vowel. In its choice of the hallucinated vowel the case of [̰ˡr] is therefore similar to the case of [̰ᵈr], and a satisfactory account of the perception of Japanese consonant clusters should preferably generalize over the [̰ᵈr] and [̰ˡr] cases. We could therefore posit an inviolable constraint */tu/, analogously to */du/, and probably collapse the two into a formulation such as */{C,–*cont*,–*nas,cor*}u/.

If one maintains an even more abstract view of Japanese surface forms, however, the analysis with */tu/ (or a generalized formulation) will go wrong. We can see this because the traditional abstract view of the Japanese syllable requires that the surface form /.tu./ exists: on the basis of considerations of distribution and alternations, the Japanese syllable that is pronounced as [̰ᵗsu] is regarded as having the surface form /.tu./. For instance, a word that means 'connection' has the surface form /.tu.gi./ although it is pronounced [̰ᵗsu ̰ᵍi]. If the abstract view is correct, there cannot therefore exist a high-ranked constraint */.tu./. Thus, what rules out /.tu.ɾa./ as the perceptual

result of [ˌˈra] is not the structure /.tu./ in itself, but its associated pronunciation [ˌᵗˢu]. The solution then lies in the fact that the naked auditory release burst [ˈ], without affrication noise (as in [eˌᵏsˌˈra]), cannot be a good representative of the structure /tu/, which must be pronounced with a full affricate [ˌᵗˢ]. The way to handle this is with the cue constraint */tu/[*no noise*], i.e. "the structure /tu/ cannot go without auditory frication noise".

The cue constraint */tu/[*no noise*] handles both perception and production. In (prelexical) perception we get tableau (11).

(11)   Japanese foreign-language perception of Russian -tra

[{*hi-freq,burst, no noise*}ra]	*/.CC/	*/ / [*burst*]	*/tu/ [*no noise*]	*/dor/ [*hi-freq*]	*/+cont/ [*burst*]	*/o/ [ ]	*/u/ [ ]
/.tra./	*!						
/.ra./		*!					
/.tu.ra./			*!				*
/.ku.ra./				*!			*
/.su.ra./					*!		*
☞ /.to.ra./						*	

In (11) we see the /tu/-specific cue constraint in the same position as */du/ in previous tableaus; it accomplishes the required elimination of the candidate /.tu.ra./. In production (phonetic implementation), the same cue constraint is again crucial; without it, the surface form /.tu.gi./ would be pronounced [ˌˈu ˌˠi] instead of [ˌᵗˢu ˌˠi]:

(12)   Phonetic implementation

/.tu.gi./	*/u/ [o]	*/tu/ [*no noise*]	*/−cont/ [*noise*]
[ˌˈu ˌˠi]		*!	
☞ [ˌᵗˢu ˌˠi]			*
[ˌˈo ˌˠi]	*!	*	

Perhaps Japanese has a general cue constraint */−cont/[*noise*] that says that phonological plosives should not be phonetically affricated; such a constraint would account for the affricationless pronunciations of /ta/, /te/, /to/, /pa/, /pe/, /pi/, /po/, /pu/, /ka/, /ke/, /ki/, /ko/, and /ku/, and their voiced counterparts. For the affricated exceptions this general constraint has to be overridden by more specific constraints such as */tu/[*no noise*]. Perhaps these exceptions could be handled by a more general */{C,−cont,−nas,cor}{V, +high}/[*no noise*], which would also turn /ti/ into [ˌᵗᶜi] and /du/ into [ˌᵈᶻu]. Whether such a more general constraint is viable depends on the extent to which [ˌˈi] is allowed to contrast with [ˌᵗᶜi] in Japanese and on the analysis of the merger of underlying |du| and |zu| into [ˌᵈᶻu]. In-depth investigations that could shed light on this matter are outside the scope of the present paper, but I mention the issues here in order to illustrate that one can do 'real' phonology around cue constraints.

The structural constraints discussed in §3 have restricted perception alone; I did not discuss any effect they might have on production. Their truly bidirectional effects are illustrated in §4, and for a Japanese-like case in §5.4.

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 and 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 and 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 and Smolensky 1998, 2000), according to which the listener succeeds in making sense of any *overt form* (in Tesar and Smolensky's example this is a sequence of syllables marked for stress) by converting it to a sensible surface structure (in Tesar and 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 and 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*)[4], such a tableau can be drawn easily, as here in tableaus (13) and (14) which use a subset of Tesar's constraints.

---

[4] Tesar's (1997) Table 4, for instance, contains the optimal form in comprehension (called 'winner'), but also contains an even more harmonic form, namely the optimal form in *production* (called 'loser'). This makes it clear that the goal of such tableaus is not to model the comprehension process, but to compare forms on behalf of a learning algorithm. In later work, Tesar (1999: Tableau 8) does provide a tableau like (13) or (14).

(13)   Metrical perception in a left-aligning iambic language

[σ 'σ σ]	FEETLEFT	IAMBIC	TROCHAIC	FEETRIGHT
☞ /(σ 'σ) σ/			*	*
/σ ('σ σ)/	*!	*		

(14)   Metrical perception in a right-aligning trochaic language

[σ 'σ σ]	FEETRIGHT	TROCHAIC	IAMBIC	FEETLEFT
/(σ 'σ) σ/	*!	*		
☞ /σ ('σ σ)/			*	*

While Tesar and 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 (13) and (14). Are they really auditory forms? After all, the input form [σ 'σ 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 (13) and (14) are even more abstract and high-level than this overt input form. What we see in (13) and (14), then, is a step on the way from the universal auditory form to the language-specific phonological surface structure. Thus, tableaus (13) and (14) 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 (13) and (14) as perception tableaus, thus identifying Tesar and Smolensky's interpretive parsing with the perception process.

The grammatical framework by Tesar and 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 and Smolensky follow the usual Optimality-Theoretical standpoint that structural constraints are violable, i.e. reside in CON. This violability is crucial in tableaus (13) and (14) 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 [_'a_ᵏ]. Tesar and 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 (13), for instance, may have at a certain point during her acquisition period the grammar FEETLEFT ≫ TROCHAIC ≫ IAMBIC ≫

FEETRIGHT. This left-aligning trochaic grammar is incorrect, since it causes an underlying |σ σ σ| to be produced as the surface form /('σ σ) σ/. When such a child hears the correct overt form [σ 'σ σ], however, she will interpret it as /(σ 'σ) σ/, which can easily be seen by reversing the two foot-form constraints in (13). 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 /('σ σ) σ/ and /(σ 'σ) σ/, and can take action, perhaps by reversing the ranking of TROCHAIC >> IAMBIC in her grammar. Thus, Tesar and 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.

All of Tesar and Smolensky's examples of robust interpretive parsing are handled with structural constraints alone; in none of their examples do they address the issue of cue constraints. In a full account of stress perception one would have to include constraints for the mapping of stress cues. For instance, language-specific auditory events (intensity, pitch, duration) are cues to phonological stress (i.e. phonological foot headedness for all stressed syllables, and phonological-word headedness for primary-stressed syllables). I will not pursue this any further here. An example of structural constraints that are crucially bidirectional (i.e. that restrict perception as well as production in non-trivial and non-identical ways) and crucially interact with cue constraints is provided in §5.4.

5. More examples of perception in OT

This section reviews some more examples of how perception has been formalized in Optimality Theory. I investigate none of these examples in full detail; rather, I provide them here in order to familiarize the reader with the directions that full phonological investigations into cue constraints may take.

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. Myers (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 other pairs of opposites such as ALIGNFEETLEFT and ALIGNFEETRIGHT, or IAMBIC and TROCHAIC). The explicit definitions of the two constraints are given in (15).

(15)   Autosegmental constraints

    a. OCP (*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 (*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 (16) and (17) 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.

(16)   Shona perception of a suprasyllabic high tone

[ɓáŋgá]	$[\acute{\sigma}] \rightarrow /\overset{H}{\underset{\sigma}{|}}/$	Ocp (H, }σ{σ)	Lcc (H, }σ{σ)
☞ / ɓ a ŋ g a / (H spread)			*
/ ɓ a ŋ g a / (H H)		*!	
/ ɓ a ŋ g a / (single H)	*!		

(17)   Mandarin perception of a sequence of syllabic high tones

[ʂáfá]	$[\acute{\sigma}] \rightarrow /\overset{H}{\underset{\sigma}{|}}/$	Lcc (H, }σ{σ)	Ocp (H, }σ{σ)
/ ʂ a f a / (H spread)		*!	
☞ / ʂ a f a / (H H)			*
/ ʂ a f a / (single H)	*!		

In Shona, phonological processes of spreading and deletion indicate that disyllabic words with two high-toned syllables, such as [ɓáŋgá] 'knife', have only one underlying H tone (Myers 1997). If prelexical perception is aimed at maximally facilitating lexical access,[5] a sequence of two high-toned syllables should therefore preferably be interpreted on the phonological surface level as having a single high tone (H). Tableau (16) 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 (16) is ruled out because

---

[5] This aim of prelexical perception has been formulated by psycholinguists (e.g. McQueen and Cutler 1997), and has been formulated for OT by Boersma (2000). In multi-level OT, the similarity of surface and underlying forms is generally advocated by faithfulness constraints. Whether faithfulness constraints indeed help the emergence of the OCP and LCC rankings in (16) and (17) during acquisition, or whether they would instead simply override the preferences of OCP and LCC in a parallel comprehension model such as that in Figure 6, could be determined by computer simulations of the concurrent acquisition of structural, cue, and faithfulness constraints, perhaps along the lines of Boersma (2008).

there is a cue constraint that says that any high-toned "syllable" in the auditory form[6] has to correspond to 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.

For the Shona case, the result in (16) reflects the common autosegmental analysis. For Mandarin Chinese, I here try out the slightly more controversial non-autosegmental position, which maintains that in contour-tone systems such as Mandarin the contour tones are "phonologically unitary [...] and not structurally related to a system of level tones" (Pike 1948: 8), a description that extends to any single level tone that a contour-tone language might have (Pike 1948: 12), as is the case in Mandarin. Phonologically speaking, Mandarin Chinese is different from Shona in the sense that every syllable has a separate underlying specification for one of the four possible tones of this language, as a result of which the alternations could perhaps best be described in terms of contour features rather than just H and L (Wang 1967). Even within autosegmental phonology, structural differences between the cases in (16) and (17) have been proposed before (Yip 1989: 166; 2002: 50–56), although they remain controversial (Duanmu 1994). More relevant within the present framework is the observation that if prelexical perception is aimed at maximally facilitating lexical access, a sequence of two high-toned syllables, such as [ʂáfá] 'sofa', should preferably be perceived with two separate H tones, as it is in (17). Indeed the assumption in the phonetic literature is that Mandarin listeners interpret auditory pitch in terms of their four tones rather than in terms of H and L (e.g. Gandour 1978: 45–47).

I have included the difference between the two tone language types here in order to illustrate the possible language-specific perception of phonetic tone stretches, another example of the idea that the phonology-phonetics interface should be handled by linguistic means.

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 (18) and (19), again from Boersma (2000), show examples for nasality.

---

[6] Of course the auditory form does not really contain syllables, which are phonological structures. I make the same simplification here as Tesar and Smolensky (see §4).

(18) Guaraní perception of suprasyllabic nasality

[tũpã]	$[\tilde{V}] \rightarrow /\overset{N}{\underset{V}{\mid}}/$	OCP (*nas*, }σ{σ)	LCC (*nas*, }σ{σ)
☞ / N / / ⟋＼ / / t u p a /			*
/ N N / / ∣ ∣ / / t u p a /		*!	
/ N / / ∣ / / t u p a /	*!		

(19) French perception of segmental nasality

[ʃãsɔ̃]	$[\tilde{V}] \rightarrow /\overset{N}{\underset{V}{\mid}}/$	LCC (*nas*, }σ{σ)	OCP (*nas*, }σ{σ)
/ 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 at the surface level. Tableau (18) 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 [ʃãsɔ̃] 'song', [lapɛ̃] 'rabbit', [ʃapo] 'hat', and [pɔso] '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, French perception must map the two nasalized vowels in [ʃãsɔ̃] to two different /N/ feature values in the phonological surface structure, as in tableau (19).

Many phonological issues remain that I cannot fully address here. The domain of the single nasal specification in (18), as well as of the single H tone in (16), is the word. But on the prelexical level listeners do not hear word boundaries. The question then is: are phonetic high-tone stretches and phonetic nasality stretches interrupted by word boundaries or not, e.g., would listeners interpret [tũpã] as having two nasals if there were a word boundary between [tũ] and [pã]? They could indeed do this if the lexicon is allowed to pass on information about word boundaries to the lower prelexical level, as in (7). I defer an account of such an interaction to §7.1.

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 (20) and (21) give the example (from Boersma 2000: 21–22, 2003) 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*/[*nonnasal*], */V,−*nas*/[*nasal*], */C,+*nas*/[*nonnasal*] and */C,−*nas*/[*nasal*] express the favoured interpretation of nasality cues for vowels and consonants, respectively.

(20) Desano adaptation of Portuguese

[ʒwẽw̃]	*/{C,−*nas*}{V,+*nas*}/	*/{σ,−*nas*}{σ,+*nas*}/	*/V,+*nas*/[*nonnasal*]	*/V,−*nas*/[*nasal*]	*/C,+*nas*/[*nonnasal*]
N / | / ʒ u	*!				
☞ N / /\ / ɲ u					*
/ ʒ u /				*!	

(21) Desano adaptation of Portuguese

[sɐbẽw̃]	*/{C,−*nas*}{V,+*nas*}/	*/{σ,−*nas*}{σ,+*nas*}/	*/V,+*nas*/[*nonnasal*]	*/V,−*nas*/[*nasal*]	*/C,+*nas*/[*nonnasal*]
N / | / s a b o	*!				
N / /\ / s a m o		*!			*
N / /\\ / n a m o			*!		**
☞ / s a b o /				*	

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. Korean

Sometimes a phonological process seems to be different in perception than in production. Kabak and Idsardi (2007) mention that Korean avoids [km] (and other)

sequences in different ways depending on the direction of processing: speakers turn an underlying |hak+mun| 'learning' into the sound [haŋmun], with assimilation of manner, but often perceive nonnative [km]-containing sounds in the same manner as the Japanese of §3 with epenthesis, e.g. [hakmun] as /hakɯmun/. Kabak and Idsardi interpret this as evidence against phonology in perception (p.33): "if Korean listeners hear epenthetic vowels in consonant clusters, they are likely to interpret pairs such as [pʰa**km**a] versus [pʰa**kʊm**a] to be the same. If, on the other hand, native phonological processes apply to perception, they should hear pairs such as [pʰa**km**a] versus [pʰa**ŋm**a] to be the same."

I will now show that within a three-level account, both the perception and the production are phonological. The idea (also shown in Boersma and Hamann this volume) is that the comprehension process involves a mapping [hakmun] → /.ha.kɯ.mun./ → |hakɯmun|, whereas the production process involves a mapping |hak+mun| → /.haŋ.mun./ → [haŋmun]. We see here that in both directions the /km/ sequence is avoided, but in different ways. Apparently, a single phonological constraint like */km/ is at work, and it interacts with different types of constraints in perception than in production.

In perception, the structural constraint interacts with cue constraints much in the same way as in the Japanese examples of (2) and (3):

(22)   Korean foreign-language perception of English

[hakmun]	*/km/	*/+*nas*/ [*burst*]	*/ɯ/ [ ]
/.hak.mun./	*!		
/.haŋ.mun./		*!	
☞ /.ha.kɯ.mun./			*

Tableau (22) expresses the idea that */km/ is an inviolable constraint of Korean, and that throwing away the positive nonnasality cue of a plosive burst is worse than hallucinating a vowel for which there are no auditory cues.

In production, the structural constraint interacts instead with faithfulness constraints:

(23)   Korean production

|hak+mun|	*/km/	DEP	IDENT(NAS)
/.hak.mun./	*!		
☞ /.haŋ.mun./			*
/.ha.kɯ.mun./		*!	

Tableau (23) expresses the idea that */km/ is an inviolable constraint of Korean, and that inserting a non-underlying vowel in production (violating DEP; McCarthy and Prince 1995) is worse than changing the value of the nasality feature.

I have thus given an account of an apparent perception-production difference fully in terms of the three levels and the native processing model of Boersma (1998 et seq.), without having to take recourse to any devices specific to foreign-language perception or loanword phonology. This is strikingly different from later accounts of phonological perception in loanword phonology (Kenstowicz 2001, Broselow 2004, Yip 2006), all of which work within a two-level model of phonology and therefore have to posit different faithfulness constraints (or different rankings) in production than in perception. Here as well we *could* say (with Kenstowicz, Broselow and Yip) that in perception the identity constraint outranks the anti-insertion constraint whereas in production the anti-insertion constraint outranks the identity constraint. However, in the three-level model the former anti-insertion constraint is a cue constraint whereas the latter anti-insertion constraint is a faithfulness constraint, and both types of constraints are independently needed to account for native phonological processing. The two types are fundamentally different, and each type comes with its own causes for being ranked high or low. Specifically, the low ranking of */ɯ/[ ] can be explained by the fact that in a noisy environment not all possible auditory cues will always be present, so that listeners have learned to freely hypothesize features and segments for which there is no positive auditory evidence, whereas no such causation mechanism is available for the ranking of DEP (Boersma 2007a: 2021).

We have seen that perception and production do not undergo the same phonological *process* (as Kabak and Idsardi indeed found), but they do undergo the influence of the same phonological *constraint*. By using the decision mechanism of OT, in which a single *constraint* can cause different *processes*, I have thus reconciled one of the phonology-versus-perception debates. For related examples of the interaction of structural and cue constraints in loanword adaptation, see Boersma and Hamann (to appear).

5.5. Arbitrary relations between auditory and surface forms

The cue constraints in (16) to (21) 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 (24) and (25), from Escudero and 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/.

(24)   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/		*		*

29

(25) 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 (24) and (25) 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 yet (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 (26) shows how an F1 of [380 Hz] can be perceived as /e/ in language with three vowel heights (from Boersma 2006).

(26) 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 and Boersma 2003, 2004).

6. The interaction of cue constraints with articulatory constraints

In sections 3 to 5 we saw interactions of cue constraints with structural constraints. The present section focuses on their interaction with articulatory constraints. I use the example of final voiced obstruents in English.

In English, there are at least two auditory cues to the voicing or voicelessness of a final obstruent: the presence or absence of periodicity (as in most languages), and the lengthening or shortening of the preceding vowel (the size of the effect is specific to English: Zimmerman and Sapon 1958). We can translate this into four cue constraints:

(27) English cue constraints

*/+*voi*/[*nonperiodic*]

*/−*voi*/[*periodic*]

*/−*son, +voi*/[*nonlengthened vowel*]

*/−*son, −voi*/[*lengthened vowel*]

The first use of the cue constraints is in prelexical perception, as before. Most often, the relevant cues agree, so that perception works well. I illustrate this in (28) and (29).

(28) A perception tableau where the two cues agree

[niːd]	*/−*son, −voi*/ [*lengthened vowel*]	*/−*voi*/ [*periodic*]
/.nit./	*!	*
☞ /.nid./		

(29) Another perception tableau where the two cues agree

[niːt]	*/−*son, +voi*/ [*nonlengthened vowel*]	*/+*voi*/ [*nonperiodic*]
☞ /.nit./		
/.nid./	*!	*

But sometimes the cues disagree. Perception experiments in the lab find that periodicity is the main cue (e.g. Hogan and Rozsypal 1980), but since vowels are much louder than consonant closures, the vowel lengthening constraint must outrank the direct periodicity cue in more natural noisy settings, as in (30).

(30) A perception tableau with a conflict

[niːt]	*/−*son, −voi*/ [*lengthened vowel*]	*/+*voi*/ [*nonperiodic*]
/.nit./	*!	
☞ /.nid./		*

The same cue constraints that are used in comprehension are also used in the production process, namely in phonetic implementation. It is tempting to regard phonetic implementation as being the inverse of prelexical perception, as in Figure 10.
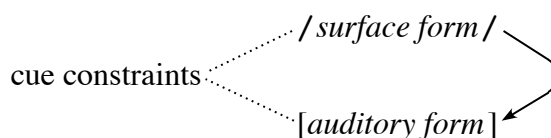
Fig. 10   Phonetic implementation (preliminary version).

Let us see what phonetic implementation would look like if Figure 10 were correct, i.e. if it were handled by cue constraints alone. The most economical assumption to make about the ranking of the cue constraints in phonetic implementation is that this ranking is identical to the ranking of the cue constraints that is optimal for comprehension, i.e. the ranking in (28) to (30). If a speaker of English reuses this ranking in production, she will try to have both cues right, as tableau (31) shows.

(31)   Phonetic implementation with cue constraints only

/.nid./	*/−son, +voi/ [*nonlengthened vowel*]	*/+voi/ [*nonperiodic*]
[niːt]	*!	*
[niːːt]		*!
[niːd]	*!	
☞ [niːːd]		

But phonetic implementation is not just about rendering cues. It is also about doing so efficiently, i.e. with the minimum expenditure of articulatory effort. Therefore, phonetic implementation is a parallel process that maps from a phonological surface form to a pair of auditory and articulatory form, as in Figure 11.
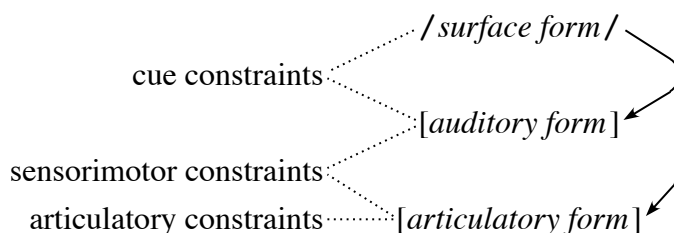
Fig. 11   Parallel phonetic implementation (full version).

The articulatory form has to be linked to the auditory form in some way. In this paper I simplifyingly assume that this sensorimotor knowledge is perfect, i.e. sensorimotor constraints are either ranked very high or very low. The articulatory-phonetic form itself is evaluated by articulatory constraints (Kirchner 1998, Boersma 1998). In the case at hand, we observe that it is especially difficult to pronounce periodicity in a final plosive. I express this simply with the constraint *[*periodic, final plosive*]. In a complete phonetic implementation tableau, this constraint must interact with cue constraints. If the articulatory constraint outranks the lower-ranked cue constraint, speakers will implement only the most important cue:

(32) Interaction of articulatory and cue constraints

/.nid./	*/−son, +voi/ [nonlengthened vowel]	*[periodic, final plosive]	*/+voi/ [nonperiodic]
[niːt]	*!		*
☞ [niːːt]			*
[niːd]	*!		
[niːːd]		*!	

As we saw in the "conflicting perception" tableau, listeners will still perceive this [niːːt] as the intended /.nid./. This means that speakers will easily get away with saying [niːːt].

I have simplified away from a large amount of possible detail. A full modelling of the case probably would require making the sensorimotor constraints explicit, and it would require arbitrary cue constraints like those in §5.5, i.e. the families */±voi/[x percent periodicity] and */±voi/[x milliseconds vowel duration].

The 'superspeaker' of (31), i.e. a speaker who always implements the best cues, probably corresponds to what real humans do when confronted with a prototype task in the lab, where they have to select the best auditory realization of a phonological category by selecting it from among a large number of auditorily presented tokens. That has been modelled within the present framework by Boersma (2006). In real humans, the cue constraints will be counteracted by articulatory constraints, so that an equilibrium emerges (Boersma 2006). This automatic balancing mechanism may lead to the achievement of stable degrees of auditory dispersion in language change, as has been shown in computer simulations of multiple generations of learners (Boersma and Hamann 2008, Van Leussen 2008).

7. The interaction of cue constraints with faithfulness constraints

While the interaction of cue and articulatory constraints discussed in §6 could be said to take place entirely in the phonetics, there are also cases where cue constraints interact with constraints at the other side of the phonology-phonetics interface. This section discusses their interactions with faithfulness, both in comprehension and in production.

7.1. Interaction of cues and faithfulness in comprehension

If we want to account more fully for the tone and nasality perceptions discussed in §5.1 and §5.2, we have to be able to model explicitly the interactions between cue and faithfulness constraints in parallel comprehension. The present section gives a basic account of a case where it looks as if the lexicon influences prelexical perception.

The example I will discuss is that of a shift of the boundary between two categories on a single auditory continuum. It has been shown that the existence of a form in the lexicon can bias the listener's reported category towards the one that occurs in an existing word, especially if the auditory form is ambiguous between the two categories (Ganong 1980).

I will discuss an example. Suppose the auditory form is a sound that sounds like a typical Spanish *barte* (which is a nonsense word) or *parte* (which means 'part'), or

something in between. The following tableaus ignore every auditory aspect of this sound except the voice onset time (VOT) of the initial plosive. I assume that the perceptual boundary between /b/ and /p/ in Spanish lies at –10 milliseconds.

In a serial view of comprehension, prelexical perception is followed by word recognition, as in Figure 6. Step one is prelexical perception, i.e. the mapping from the given Auditory Form to a phonological surface structure (Surface Form). The cue constraints are ranked by distance to the boundary. The worst token of /p/ is one with a very negative VOT such as –100 ms, so the cue constraint that says that */p/[−100] is high-ranked. Likewise, constraints that connect large positive VOT values to /b/ are also high-ranked. An appropriate ranking for perceiving Spanish voicing must be similar to that in tableaus (33) to (35).

(33)   Spanish classification of voicing

[−100 ms]	*/p/ [−100]	*/b/ [+30]	*/p/ [−20]	*/b/ [−20]	*/p/ [+30]	*/b/ [−100]
☞ /.baɾ.te./						*
/.paɾ.te./	*!					

(34)   Spanish classification of voicing

[−20 ms]	*/p/ [−100]	*/b/ [+30]	*/p/ [−20]	*/b/ [−20]	*/p/ [+30]	*/b/ [−100]
☞ /.baɾ.te./				*		
/.paɾ.te./			*!			

(35)   Spanish classification of voicing

[+30 ms]	*/p/ [−100]	*/b/ [+30]	*/p/ [−20]	*/b/ [−20]	*/p/ [+30]	*/b/ [−100]
/.baɾ.te./		*!				
☞ /.paɾ.te./					*	

Step two is word recognition. I will include both the underlying form and the morpheme in the candidates. The lexical entry |paɾte| <part> exists, the underlying form |baɾte| does not (i.e., it is not connected to any morpheme). The perceived form /.paɾ.te./ will easily be recognized with the help of faithfulness constraints:

(36)   Word recognition

/.paɾ.te./	*< > |X|	*|m| /p/	*|m| /b/	*|b| /p/	*|p| /b/
☞ |paɾte| <part>					
|baɾte| < >	*!			*	
|maɾte| <Mars>		*!			

The constraint *<>|X| militates against throwing away phonological material in lexical access (i.e. against linking the material X to no morpheme). Since the winning candidate in (36) violates no constraints at all, a more interesting input form is /.baɾ.te./. What the resulting accessed underlying form is depends on the ranking of the faithfulness constraints. If nasality faithfulness outranks voicing faithfulness, the listener again recognizes |paɾte| <part>:

(37)   Word recognition

/.baɾ.te./		*<>|X|	*|m|/p/	*|m|/b/	*|b|/p/	*|p|/b/
☞	|paɾte| <part>					*
	|baɾte| < >	*!				
	|maɾte| <Mars>			*!		

If, by contrast, voicing faithfulness outranks nasality faithfulness, the listener will recognize |maɾte| <Mars> instead:

(38)   Word recognition

/.baɾ.te./		*<>|X|	*|m|/p/	*|p|/b/	*|m|/b/	*|b|/p/
	|paɾte| <part>			*!		
	|baɾte| < >	*!				
☞	|maɾte| <Mars>				*	

We cannot predict which of the two options, (37) or (38), people will choose. In any case, the choice between these two tableaus does not depend on the degree of ambiguity of the auditory VOT: once prelexical perception has chosen the category, without the help of the lexicon, the probability of subverting the category in the word recognition phase no longer depends on the auditory form.

The situation is different in the parallel model of Figure 7. We first provide a ranking that makes the listener perceive a VOT of −100 ms as /.baɾ.te./, never mind that the faithful lexical item |baɾte| does not exist. If the lexicon is still capable of telling the listener that the word the speaker intended was |paɾte|, the ranking must be similar to that in tableau (39).

(39)   Perception possibly but not really influenced by lexical access

[−100 ms]	*/p/ [−100]	*/b/ [+30]	*<>|X|	*|b|/p/	*|p|/b/	*/p/ [−20]	*/b/ [−20]	*/p/ [+30]	*/b/ [−100]
/.baɾ.te./ |baɾte| < >			*!						*
/.paɾ.te./ |baɾte| < >	*!		*	*					
☞ /.baɾ.te./ |paɾte| <part>					*				*
/.paɾ.te./ |paɾte| <part>	*!								

In the case of a VOT of –20 ms, which was perceived as /b/ in the sequential model, the perception now becomes /p/, as shown in tableau (40):

(40)   Perception possibly and really influenced by lexical access

[–20 ms]	*/p/ [−100]	*/b/ [+30]	*<> |X|	*|b| /p/	*|p| /b/	*/p/ [−20]	*/b/ [−20]	*/p/ [+30]	*/b/ [−100]
/.baɾ.te./ |baɾte| <>			*!				*		
/.paɾ.te./ |baɾte| <>			*!	*		*			
/.baɾ.te./ |paɾte| <part>					*!		*		
☞ /.paɾ.te./ |paɾte| <part>						*			

In this tableau we see that the cue constraints prefer /b/, but the faithfulness constraint, forced by *<>|X|, prefers /p/. What we see in (39) and (40) is that the perceptual shift occurs only in the vicinity of the auditory boundary between the two categories. The parallel comprehension model therefore seems to be more consistent with the Ganong effect than the serial model. This distinction between bottom-up (serial) models and interactive (parallel) models of speech processing has been known for some time. For instance, the TRACE model of speech perception (McClelland and Elman 1986) was designed to be able to produce interactive effects, and one of the simulations performed in the original paper was indeed the Ganong effect. The bottom-up model of speech perception is not dead yet, however: McQueen and Cutler (1997) and Norris, McQueen and Cutler (2000) argue that listeners in the lab base their reported perceptions partly on the phonological surface form and partly on the underlying form, and that this mix can explain the observed boundary shift. This issue seems not to have been settled.

A remaining question is whether the constraint *<>|X| can ever be violated in a winning form. It can, if it is outranked by faithfulness. In such a case, tableau (39) would become tableau (41).

(41)   *Recognizing a nonsense word*

[–100 ms]	*/p/ [−100]	*/b/ [+30]	*|b| /p/	*|p| /b/	*<> |X|	*/p/ [−20]	*/b/ [−20]	*/p/ [+30]	*/b/ [−100]
☞ /.baɾ.te./ |baɾte| <>					*				*
/.paɾ.te./ |baɾte| <>	*!		*		*				
/.baɾ.te./ |paɾte| <part>				*!					*
/.paɾ.te./ |paɾte| <part>	*!								

If both the cue constraints and the faithfulness constraints are ranked high enough, the auditory form is apparently capable of creating an underlying form not yet connected to a morpheme; perhaps this is the moment for the creation of a new morpheme (Boersma 2001).

A more detailed account of these effects would require computer simulations of the acquisition of all levels of comprehension, building on the simulations in Boersma

(2006). Such simulations would also be needed to account for the rankings in §5.1 and §5.2.

7.2. Interaction of cues and faithfulness in production

Computer simulations of the acquisition of cue constraints and faithfulness constraints were performed by Boersma (2008). That paper studied a relatively universal case in which plosive consonants have better place cues than nasal consonants and coronal consonants are more frequent (in word-final position) than labial consonants.

The simulations revealed that both faithfulness constraints and ("identity-preferring") cue constraints ended up being ranked higher for plosives than for nasals and higher for labials than for coronals. The results for the faithfulness constraints explain the automatic mechanism behind Steriade's (1995, 2001) *licensing by cue* as well as the automatic mechanism behind the concept of *markedness*, i.e. the relation between a feature value's frequency and its degree of phonological activity. For the cue constraints, the results show rankings by distance to the boundary, such as those in (41), and the preference of reliable cues over unreliable cues.

8. The interaction of cue constraints with lexical constraints

Computer simulations of the whole trajectory from the auditory form to the morpheme were performed by Apoussidou (2007). The simulated learners, given pairs of sound and morpheme, had to construct both intermediate forms (surface and underlying), as well as the ranking of all the constraints involved. At the lexical level, the relation between morpheme and underlying form had to be determined by a ranking of lexical constraints, i.e. there were multiple candidate underlying forms for each morpheme. Learners typically came up with rankings that favoured single underlying forms for each morpheme (rather than with allomorph selection), together with a phonology that changed these underlying forms to potentially rather diffferent surface forms.

As for a direct interaction between cue and lexical constraints, Boersma (2007b) considers the case of lexical selection in production: if there is a single morpheme <water>, and it has the two possible underlying forms |#watr#| and |#aː#|,[7] then the choice between the two could partly be based on cue constraints, i.e. on how well the sounds [watr] and [aː] connect to the phonological structures /.watr./ and /.aː./. One can then observe that especially in postconsonantal position the auditory cues for a phonological syllable boundary (and hence the cues for the underlying word boundary) are poorer in [aː] than in [watr]. On this basis, the speaker might select the |#watr#| form even though the lexicon (by means of the ranking of the lexical constraints) may prefer the |#aː#| form (e.g. by means of its slightly better semantic features).

We see here a case of near-maximum interactivity of cue constraints, as they compete directly with constraints that guide connections in the lexicon. This is therefore an interaction that completely bypasses the whole phonology.

---

[7] The "#" sign is the word boundary. This case is loosely based on what must have happened with the 'water' words in Old Germanic. I make here the simplification that the two underlying forms share the same morpheme. It is probably more likely that each is connected to its own morpheme, and that the intended lexical semantic features are instead the same. If so, the cue constraints interact not with the lexical phonological constraints of Figure 1, but with lexical semantic constraints, which can connect the morpheme to a representation above those of Figure 1.

9. Conclusion

Every OT phonologist agrees that structural constraints, when they appear in the production process, are of a phonological nature. If the same structural constraints dictate the perception process, then the conclusion that I like to draw is that perception is phonological as well. In other words, the perception process is restricted by the same phonological constraints as the production process is. The structural constraints evaluate the output of the mapping from Underlying Form to Surface Form (i.e. 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, then 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. This argument was valid when Tesar and 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. This means that if the structural constraints that restrict perception are ranked in the OT way (or weighted in the HG way), the cue constraints that compete with them must also be ranked in the OT way (or weighted in the HG way). The present paper has shown how these constraints can interact with other phonetic constraints, with phonological constraints, and with constraints in the lexicon. The resulting grammar model is representationally modular, but entirely interactive when it comes to processing.

References

Apoussidou, Diana   2007   *The Learnability of Metrical Phonology*. Ph.D. dissertation, University of Amsterdam.

Bermúdez-Otero, Ricardo 1999   *Constraint Interaction in Language Change: Quantity in English and Germanic*. Ph.D. dissertation, University of Manchester.

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

Boersma, Paul   1998   *Functional Phonology: Formalizing the Interactions between Articulatory and Perceptual Drives*. Ph.D. 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 and Joe Pater (eds.), *Papers in Experimental and Theoretical Linguistics*. Volume 6, 24–35. Edmonton: University of Alberta.

Boersma, Paul   2003   Nasal harmony in functional phonology. In: Jeroen van de Weijer, Vincent van Heuven and Harry van der Hulst (eds.), *The Phonological Spectrum*, Volume 1: *Segmental Structure*. 3–35. Amsterdam: John Benjamins.

Boersma, Paul   2006   Prototypicality judgments as inverted perception. In: Gisbert Fanselow, Caroline Féry, Matthias Schlesewsky and Ralf Vogel (eds.), *Gradedness in Grammar*. 167–184. Oxford: Oxford University Press.

Boersma, Paul   2007a   Some listener-oriented accounts of *h*-aspiré in French. *Lingua* 117: 1989–2054.

Boersma, Paul   2007b   The evolution of phonotactic distributions in the lexicon. Talk presented at the Workshop on Variation, Gradience and Frequency in Phonology, Stanford, July 8, 2007.

Boersma, Paul   2008   Emergent ranking of faithfulness explains markedness and licensing by cue. *Rutgers Optimality Archive* 954.

Boersma, Paul, and Paola Escudero 2008   Learning to perceive a smaller L2 vowel inventory: an Optimality Theory account. In: Peter Avery, Elan Dresher and Keren Rice (eds.), *Contrast in Phonology: Theory, Perception, Acquisition.* 271–301. Berlin: Mouton de Gruyter.

Boersma, Paul, and Silke Hamann 2008   The evolution of auditory dispersion in bidirectional constraint grammars. *Phonology* 25: 217–270.

Boersma, Paul, and Silke Hamann   this volume   Introduction: models of phonology in perception.

Boersma, Paul, and Silke Hamann   to appear   Loanword adaptation as first-language phonological perception. To appear in: Andrea Calabrese and W. Leo Wetzels (eds.), *Loan Phonology.* Amsterdam: John Benjamins. [*Rutgers Optimality Archive* 975]

Broselow, Ellen 2004   Language contact phonology: richness of the stimulus, poverty of the base. *North-Eastern Linguistic Society* 34: 1–22.

Cornulier, Benoit de   1981   H-aspirée et la syllabation: expressions disjonctives. In: Didier L. Goyvaerts (ed.), *Phonology in the 1980's.* 183–230. Ghent: Story-Scientia.

Denes, Peter   1955   Effect of duration on the perception of voicing. *Journal of the Acoustical Society of America* 27: 761–764.

Duanmu, San   1994   Against contour tone units. *Linguistic Inquiry* 25: 555–608.

Dupoux, Emmanuel, Kazuhiko Kakehi, Yuki Hirose, Christophe Pallier 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   *Linguistic Perception and Second-Language Acquisition.* Ph.D. dissertation, Utrecht University.

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. [non-misprinted version: *Rutgers Optimality Archive* 439, 2001]

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.

Flemming, Edward   1995   *Auditory Representations in Phonology.* Ph.D. dissertation, UCLA. [published in 2002 by Routledge, London]

Fowler, Carol A. 1986   An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics* 14. 3–28.

Gandour, Jackson 1978   The perception of tone. In: Victoria Fromkin (ed.), *Tone,* 41–76. New York: Academic Press.

Ganong, William F. III   1980   Phonetic categorization in auditory word perception. *Journal of Experimental Psychology*: *Human Perception and Performance* 6: 110–125.

Goldsmith, John 1976   *Autosegmental Phonology*. Ph.D. dissertation, MIT, Cambridge. [published in 1979 by Garland Press, New York]

Gussenhoven, Carlos   2004   *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press.

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

Hayes, Bruce   1999   Phonetically-driven phonology: the role of Optimality Theory and inductive grounding. In: Michael Darnell, Edith Moravcsik, Michael Noonan, Frederick Newmeyer and Kathleen Wheatley (eds.), *Functionalism and Formalism in Linguistics*, Volume I: *General Papers*. 243–285. Amsterdam: John Benjamins.

Hogan, John T., and Anton J. Rozsypal   1980   Evaluation of vowel duration as a cue for the voicing distinction in the following word-final consonant. *Journal of the Acoustical Society of America* 67: 1764–1771.

House, Arthur S., and Grant Fairbanks   1953   The influence of consonant environment upon the secondary acoustical characteristics of vowels. *Journal of the Acoustical Society of America* 25: 105–113.

Jun, Jongho   1995   Place assimilation as the result of conflicting perceptual and articulatory constraints. In: José Camacho, Lina Choueiri and Maki Watanabe (eds.), *Proceedings of the 14th West Coast Conference on Formal Linguistics*. 221–237. Stanford, Calif.: CSLI.

Jusczyk, Peter   1997   *The Discovery of Spoken Language*. Cambridge, Mass.: MIT Press.

Kabak, Barış, and William Idsardi 2007   Perceptual distortions in the adaptation of English consonant clusters: syllable structure or consonantal contact constraints? *Language and Speech* 50: 23–52.

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

Keating, Patricia 1985   Universal phonetics and the organization of grammars. In: Victoria Fromkin (ed.), *Phonetic Linguistics: Essays in Honor of Peter Ladefoged*. 115–132. Orlando: Academic Press.

Kenstowicz, Michael   2001   The role of perception in loanword phonology. Ms. to appear in *Linguistique africaine*.

Kiparsky, Paul   1985   Some consequences of Lexical Phonology. *Phonology Yearbook* 2: 85–138.

Kirchner, Robert 1998   *Lenition in Phonetically-based Optimality Theory*. Ph.D. dissertation, UCLA.

Leben, William 1973   *Suprasegmental Phonology*. Ph.D. dissertation, MIT, Cambridge. [published in 1980 by Garland Press, New York]

Levelt, Willem   1989   *Speaking: From Intention to Articulation*. Cambridge, Mass.: MIT Press.

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

McClelland, James L., and Jeffrey L. Elman 1986   The TRACE model of speech perception. *Cognitive Psychology* 18: 1–86.

McQueen, James M., and Anne Cutler   1997   Cognitive processes in speech perception. In: William J. Hardcastle and John Laver (eds.), *The Handbook of Phonetic Sciences*. 566–585. Oxford: Blackwell.

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

Norris, Dennis, James M. McQueen and Anne Cutler 2000   Merging information in speech recognition: feedback is never necessary. *Behavioral and Brain Sciences* 23: 299–370.

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

Peterson, Gordon, and Ilse Lehiste 1960   Duration of syllable nuclei in English. *Journal of the Acoustical Society of America* 32: 693–703.

Pierrehumbert, Janet   1987   *The Phonology and Phonetics of English Intonation*. Bloomington: Indiana University Linguistics Club.

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

Pike, Kenneth   1948   *Tone Languages*. Ann Arbor: University of Michigan Press.

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]

Prince, Alan, and Paul Smolensky 1993   *Optimality Theory: Constraint Interaction in Generative Grammar*. Technical Report TR-**2**, Rutgers University Center for Cognitive Science. [published in 2004 by Blackwell, Malden]

Saussure, Ferdinand de   1916   *Cours de linguistique générale*. Edited by Charles Bally and Albert Sechehaye in collaboration with Albert Riedlinger. Paris: Payot & Cie.

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

Steriade, Donca 1995   Positional neutralization. Two chapters of an unfinished manuscript, Department of Linguistics, UCLA.

Steriade, Donca 2001   The phonology of perceptibility effects: the P-map and its consequences for constraint organization. Unpublished manuscript, Department of Linguistics, UCLA.

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

Tesar, Bruce   1999   Robust interpretive parsing in metrical stress theory. In: Kimary Shahin, Susan Blake and Eun-Sook Kim (eds.), *Proceedings of the 17th West Coast Conference on Formal Linguistics*. 625–639. Stanford, Calif.: CSLI.

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.

Van Leussen, Jan-Willem 2008   *Emergent optimal vowel systems*. M.A. thesis, University of Amsterdam. *Rutgers Optimality Archive* 1006.

Walker, Rachel 1998   *Nasalization, Neutral Segments, and Opacity Effects*. Ph.D. dissertation, University of California, Santa Cruz.

Wang, William 1967   The phonological features of tone. *International Journal of American Linguistics* 33: 93–105.

Yip, Moira   1989   Contour tones. *Phonology* 6: 149–174.

Yip, Moira   2002   *Tone*. Cambridge: Cambridge University Press.

Yip, Moira   2006   The symbiosis between perception and grammar in loanword phonology. *Lingua* 116: 950–975.

Zimmerman, Samuel A., and Stanley M. Sapon   1958   Note on vowel duration seen cross-linguistically. *Journal of the Acoustical Society of America* 30: 152–153.