



## UvA-DARE (Digital Academic Repository)

### A programme for bidirectional phonology and phonetics and their acquisition and evolution

Boersma, P.

**DOI**

[10.1075/la.180.02boe](https://doi.org/10.1075/la.180.02boe)

**Publication date**

2011

**Document Version**

Submitted manuscript

**Published in**

Bidirectional Optimality Theory

[Link to publication](#)

**Citation for published version (APA):**

Boersma, P. (2011). A programme for bidirectional phonology and phonetics and their acquisition and evolution. In A. Benz, & J. Mattausch (Eds.), *Bidirectional Optimality Theory* (pp. 33-72). (Linguistik Aktuell/Linguistics Today; Vol. 180). John Benjamins Publishing Company. <https://doi.org/10.1075/la.180.02boe>

**General rights**

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

**Disclaimer/Complaints regulations**

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

*UvA-DARE is a service provided by the library of the University of Amsterdam (<https://dare.uva.nl>)*

# A programme for bidirectional phonology and phonetics and their acquisition and evolution

Paul Boersma, 10 April 2011, paul.boersma@uva.nl

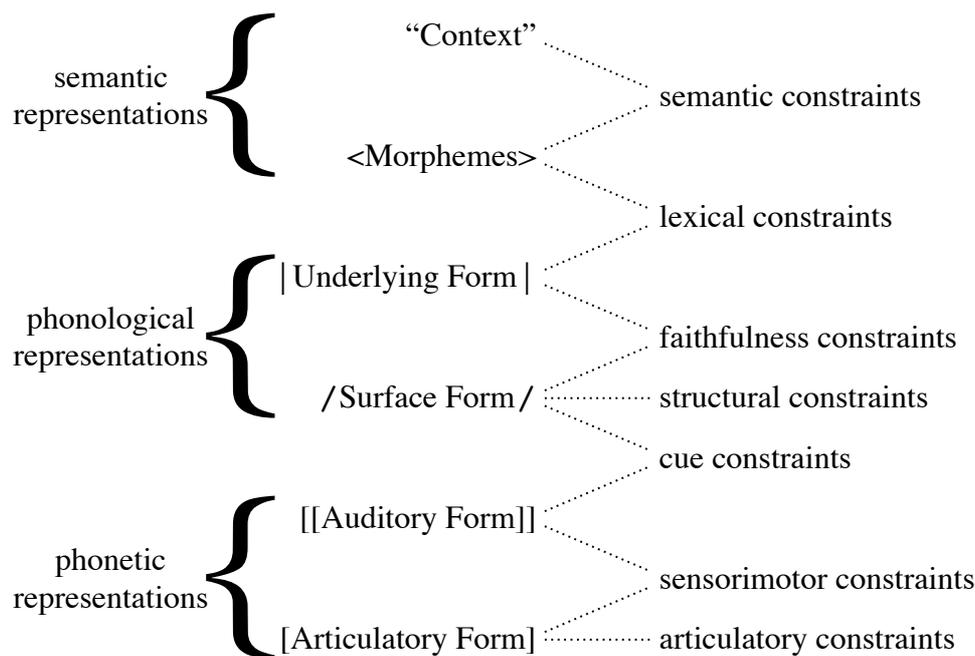
## Abstract

This paper summarizes an existing bidirectional six-level model of phonology and phonetics (and a bit of morphology). Bidirectionality in this case refers to the modelling of both the speaking process (production) and the listening process (comprehension). The elements of the grammar (the constraints) are bidirectional in the sense that the speaker and listener use the same sets of constraints, with the same rankings. In contrast with Blutner's and Mattasusch's bidirectional OT models, the evaluation is the simplest possible, i.e. it is performed unidirectionally in both directions of processing; still, listener-oriented effects tend to emerge from having learning algorithms for the comprehension direction alone. This paper describes a great number of learning algorithms in both directions of processing, and their typical results across one or multiple generations.

This paper presents an Optimality-Theoretic (OT) grammar model that is intended to be capable of handling 'all' of phonology: its representations with their relations, its processes with their relations, its connection to the semantics, its acquisition by the child, its evolution over the generations, and its typology across languages. The goal of the model is to achieve explanatory adequacy by doing *whole-language simulations* of the acquisition and evolution of a language.

I start by giving the whole grammar model, then zoom in on smaller parts of it, starting with the very small.

Figure 1, then, shows the proposed minimal but comprehensive model of phonological grammar (based on Boersma, 1998, 2007; Apoussidou, 2007). It is comprehensive in the sense that it is meant to be able to handle 'all' phonological and related phenomena. And it is 'minimal' in the sense that it contains what I think is the minimum number of representations that we need to do interesting phonology, namely two phonological representations that are connected to each other and to two semantic and two phonetic representations.



**Fig. 1** The grammar model.

Figure 1 is not just meant to be a model of grammar, but a model of processing as well. In fact, it is meant to be a *bidirectional* model of processing: the task of the listener is to travel up the figure, starting from the Auditory Form (the sound) and ending up with a change in the Context; the task of the speaker is to travel down the figure, starting from an intended change in the Context and ending up with an Articulatory Form (the pronunciation as implemented by the speech organs). When travelling up or down the figure, the speaker or listener will visit a number of intermediate representations (Surface Form, Underlying Form, Morphemes). During this processing, Optimality-Theoretic constraints evaluate either a single level of representation (structural and articulatory constraints) or a relation between two levels of representation (sensorimotor, cue, faithfulness, lexical and semantic constraints). Following Smolensky (1996), the constraints are used *bidirectionally*, i.e., a language user uses the same constraints when she speaks as when she listens, with the same rankings. This will be seen to lead to apparent effects of bidirectional processing (i.e. the speaker appears to take the listener into account, and/or the listener appears to take the speaker into account), although no listener-orientedness is explicitly modelled in speakers, nor speaker-orientedness in listeners.

Sections 2 through 6 introduce phonological and phonetic representations, constraint families, processes and learning algorithms to an increasing degree of comprehensiveness. Sections 7 through 9 link these to the semantics. Finally, section 10 discusses the assumptions and wider issues associated with the model.

## 1. Phonological representations: Underlying and Surface Form

The minimum number of phonological representations capable of handling any interesting phonological phenomena seems to be two: at the very least we seem to require the traditional distinction between Underlying Form and Surface Form. The

**Underlying Form** is usually regarded as a sequence of pieces of phonological material copied from the lexicon, with discernible morpheme structure, for example |an+pa|, where ‘+’ is a morpheme boundary. The **Surface Form** is typically a treelike structure of abstract phonological elements such as features, segments, syllables, and feet, for instance (in linearized form) /.am.pa./, where ‘.’ is a syllable boundary. For these two representations, the following subsections describe their relations, their roles in merely-phonological processes, and their roles in merely-phonological acquisition and evolution.

### 1.1. The relation between Underlying Form and Surface Form

From Prince & Smolensky (1993) on, the relation between these two representations has been modelled in terms of the **faithfulness constraints** that are shown in Figure 1. For instance, the combination of the underlying form |an+pa| and the surface form /.am.pa./ constitutes a *faithfulness violation* because the underlying |n| corresponds to a surface /m/ that is not identical to it (McCarthy and Prince, 1995).

### 1.2. The process of merely-phonological production

Prince & Smolensky, and most OT-ists following them, have regarded phonology as being primarily concerned with the unidirectional process of *phonological production*, i.e. the mapping *from* Underlying Form *to* Surface Form, as in Figure 2.

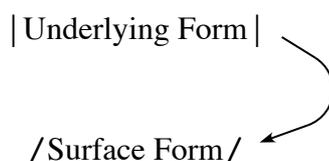


Fig. 2 The merely-phonological production process.

In this mapping, the faithfulness constraints can interact with the **structural constraints** that are also shown in Figure 1. For instance, the mapping from underlying |an+pa| to surface /.am.pa./ could be due to a constraint against codas that do not share their place with a following onset. In order to force the surfacing of /.am.pa./, this constraint has to outrank the faithfulness constraint mentioned in the previous paragraph. Tableau (1) shows how this works in OT. The constraint names contain subscripts for the representations that they evaluate (S for Surface, US for Underlying & Surface).

(1) *Phonological production without semantics or phonetics*

an+pa	*CODAWITHSEPARATEPLACES <sub>S</sub>	IDENTPLACE <sub>US</sub>
an+pa  /.an.pa./	*!	
☞  an+pa  /.am.pa./		*
an+pa  /.aŋ.pa./	*!	*

The notation in the **production tableau** (1) is slightly different from the usual notation in that the candidate cells contain paired representations, i.e., the ‘input’ (the

underlying form |an + pa|) has been included in each cell. This manner of writing candidates will become especially relevant when we consider cases with more than two representations, and cases with bidirectional learning tableaux.

The interpretation of production tableaux like (1) is that the candidates listed are all those that share the input representation(s) from the top left cell. Thus, GEN (the OT candidate generator) generates three UF–SF pairs that contain the representation |an + pa|, and these three are the candidates in (1). This entails that for the evaluation of the constraints, we have to look only at the candidates themselves. For instance, the violation of IDENTPLACE<sub>US</sub> by the second candidate can be detected by sole inspection of the pair |an + pa| / .an.pa./, without comparing the candidate to the input |an + pa| separately.

### 1.3. The process of merely-phonological comprehension

The OT model with two representations (Underlying and Surface Form) has been used to a limited extent bidirectionally, i.e., it has been used to model not only phonological production, as above, but also *phonological comprehension*, as in Figure 3.

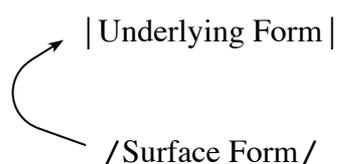


Fig. 3 The merely-phonological comprehension process.

Smolensky (1996) mentions the case of a learner with a high-ranked structural constraint, who cannot produce all forms that she can comprehend. Smolensky’s argument can be expressed with the same example as above: the ranking of tableau (1) can be regarded as the grammar of a child who cannot yet produce / .an.pa./ . If we suppose that adults of the same language have the reverse ranking, these adults must be able to produce / .an.pa./, and this form will occur in the learner’s environment. Tableau (2) shows that the learner successfully comprehends this form. The finger points backwards (“☞”) to mark the candidate that wins in the comprehension direction.

#### (2) *Phonological comprehension without semantics or phonetics*

	/ .an.pa./	*CODAWITHSEPARATEPLACES	IDENTPLACE <sub>US</sub>
☞	an + pa  / .an.pa./	*	
	am + pa  / .an.pa./	*	*!
	aŋ + pa  / .an.pa./	*	*!

The candidates in the **comprehension tableau** (2) are now all the thinkable paired representations that share the Surface Form / .an.pa./ . Crucially in (2), all of these doublets violate the structural constraint with the long name, so that the decision falls to the faithfulness constraint. This is how Smolensky solved what he called the

‘production-comprehension dilemma’, the case where a learner’s comprehension skills precede her production skills.

#### 1.4. Merely-phonological acquisition

A **bidirectional learning tableau** is a tableau with two input representations rather than just one. It helps in improving the relation between the two representations. For the case of mere phonology, it helps to improve both of the mappings in Figure 4.

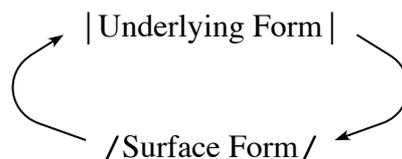


Fig. 4 Bidirectional acquisition of mere phonology.

In tableau (3), the assimilating learner of tableau (2) has obtained knowledge that the underlying form |an + pa| and the surface form /.an.pa./ can go together (i.e. the parents do not assimilate). This pair therefore appears in the top left cell of the tableau. The candidates in the tableau are all the underlying-surface pairs that share one or two of the representations, i.e. all the possible pairs that contain either |an + pa| or /.an.pa./, or both, i.e. the candidates from tableaux (1) and (2) combined.

#### (3) Phonological acquisition without semantics or phonetics

an + pa  /.an.pa./	*CODAWITHSEPARATEPLACE <sub>S</sub>	IDENTPLACE <sub>US</sub>
√   an + pa  /.an.pa./	*→	
  an + pa  /.am.pa./		←*
an + pa  /.aŋ.pa./	*	*
am + pa  /.an.pa./	*	*
aŋ + pa  /.an.pa./	*	*

The interpretation of the pair |an + pa| /.an.pa./ in the top left cell is as follows. The learner has just heard the surface form /.an.pa./, and successfully ‘comprehended’ it, which means that something (for instance the discourse context) has told her that the correct underlying form that corresponds to this instance of the surface form /.an.pa./ was |an + pa|. On the basis of this information the learner regards |an + pa| /.an.pa./ as a correct combination of underlying and surface form.

The interpretation of the three marks in the candidate cells in (3) is as follows. The backward-pointing finger (“”) marks the candidate that would win in the comprehension direction, if nothing more than the ‘correct’ surface form /.an.pa./ were given. This form is shared by candidates 1, 4, and 5, and of these three candidates candidate 1 is the most harmonic given the constraint ranking. The forward-pointing finger (“”) marks the candidate that would win in the production direction if nothing more than the ‘correct’ underlying form |an + pa| were given. This form is shared by candidates 1, 2, and 3, and candidate 2 is the most harmonic of

these. The check mark (“√”), finally, marks the most harmonic of all the candidates that share both the underlying form |an + pa| and the surface form /.an.pa./. There is obviously only one candidate that shares both forms, namely candidate 1, so candidate 1 is immediately the most harmonic of all such candidates.

The interpretation of the arrows in (3) is as follows. The candidate with the check mark is regarded by the learner as the *correct candidate*. The *forward winner* (“☞”) differs from this correct candidate, so the learner has evidence that the forward winner is an incorrect pair. As a result, the learner will take action by taking a *learning step*: she raises the rankings of all the constraints that are violated in the incorrect forward winner (in this case, only IDENTPLACE<sub>US</sub>), and lowers the rankings of all the constraints that are violated in the correct candidate (in this case, only \*CODAWITHSEPARATEPLACE<sub>S</sub>). These raisings and lowerings are depicted with arrows in the tableau. With ‘raising’ and ‘lowering’ I mean that these constraints move a small distance along the ranking scale of Stochastic OT (Boersma, 1997; Boersma and Hayes, 2001). These movements make it thereby more likely that a future |an + pa| will be produced as /.an.pa./. This half of the bidirectional learning procedure is identical to the application of the Gradual Learning Algorithm (Boersma, 1997) to merely-phonological cases (Boersma and Hayes, 2001).

The other half of the bidirectional learning procedure happens if the *backward winner* (“☜”) differs from the ‘correct’ candidate. This does not happen in tableau (3), but tableau (4) shows a case. This tableau shows a learner who performs no place assimilation, although place assimilation does occur in her environment, as indicated by the learning pair |an + pa| /.am.pa./.

(4) *Phonological acquisition without semantics or phonetics*

an + pa  /.am.pa./	IDENTPLACE <sub>US</sub>	*CODAWITHSEPARATEPLACE <sub>S</sub>	X
☞  an + pa  /.an.pa./		←*	
√  an + pa  /.am.pa./	*→→		
an + pa  /.aŋ.pa./	*	*	
☜  am + pa  /.am.pa./			←*
aŋ + pa  /.am.pa./	*		

In tableau (4), the mismatch between the ‘correct’ candidate and the backward winner leads to a learning step analogous to the one described above: the constraints violated in the ‘correct’ candidate are lowered, as indicated in the tableau by the second right-pointing arrow, and the constraints violated by the incorrect backward winner are raised; in order to be able to make this visible, I added a mysterious constraint X to the tableau, which is violated by the fourth candidate (for instance, this constraint could punish the underlying form |am| somehow; see §1.6 for what this constraint could really look like). The changes in the constraint rankings raise the likelihood that a future occurrence of /.am.pa./ will be comprehended as |an + pa|.

### 1.5. Merely-phonological evolution

In the examples of §1.4, and more generally in merely-phonological learning, the learner will usually end up in exactly the same language as her parents. That is, given a certain underlying form she and her parents will produce the same surface form, and given a certain surface form, she and her parents will usually comprehend the same underlying form. Boersma and Pater (2008) showed that if learners have the same constraint set as their parents, and both the parents and the children can only entertain languages generated by that constraint set, the algorithm in (4) usually causes the children to end up in a language identical to that of their parents, where the term ‘usually’ is based on the observation that the algorithm correctly converged for 99.6% of 100,000 randomly generated languages. If the parents have one of those 99.6% of possible languages, the language is predicted not to change over the generations; if they have one of the remaining 0.4%, the children will create a different language, but their children seem unlikely to change it any further.

### 1.6. What is wrong with merely-phonological grammars?

Sections §1.2 through §1.5 assumed that phonology lives on an island. In reality, it is connected upwards to the semantics (and the syntax, and the pragmatics) and downwards to the phonetics, and these connections are felt throughout the phonology. This section points out problems with the production model of §1.2, the comprehension model of §1.3, the acquisition model of §1.4, and the evolution model of §1.5.

The merely-phonological production model of §1.2 has problems accounting for many observed phonological typologies, such as universal hierarchies of frequency of occurrence (an aspect of ‘markedness’), universal hierarchies of the degree of phonological activity (another aspect of ‘markedness’), and universal hierarchies of faithfulness rankings (yet another aspect of ‘markedness’). The OT literature on merely-phonological grammars has proposed innate rankings of structural constraints (Prince and Smolensky, 1993), innate rankings of faithfulness constraints (Beckman, 1998), and extralinguistic knowledge of auditory contrast (Steriade, 1995, 2001). These proposals come with their own problems: supposedly universal rankings turn out to have exceptions wherever such exceptions would be functionally advantageous (see Steriade 1995 against positional faithfulness, and Boersma 1998 against the sonority hierarchy), and rankings involving perceptual contrast turn out to do so in a language-specific way rather than with reference to universal auditory contrast (e.g. Boersma & Escudero, 2008).

The solution is to allow the phonology to interact with the phonetics. Many observed phenomena like *auditory enhancement* (Flemming, 1995), *licensing by cue* (Steriade, 1995), and things often attributed to innate *markedness* will fall out automatically as side effects of learning. Merely-phonological production tableaux will often no longer be valid: the choice for  $|\text{an} + \text{pa}| / .\text{am}.\text{pa} /$  in (1), for instance, does not have to be determined by a structural constraint at all: constraints further down Figure 1, most notably the articulatory constraints, could take care of that, as shown in §6.1.

The merely-phonological comprehension model of §1.3 also has its problems. Suppose that the lexicon of the listener contains the items  $|\text{an}|$ ,  $|\text{am}|$ , and  $|\text{a}\eta|$ , and

that each of them can be concatenated with |pa|. In an assimilating language like the one in (3), all three cases will end up as the surface structure /.am.pa./. However, when confronted with the surface form /.am.pa./, the listener has no option but to comprehend the underlying form |am + pa|, as tableau (5) shows.

(5) *A failure of phonological comprehension without semantics*

/.am.pa./	*CODAWITHSEPARATEPLACES	IDENTPLACEUS
an + pa  /.am.pa./		*!
☞  am + pa  /.am.pa./		
aŋ + pa  /.am.pa./		*!

This problem with Smolensky's (1996) proposal was first noted by Hale and Reiss (1998). The solution they proposed was that comprehension is not handled by tableaux like (5) but instead follows a procedure that yields a list of underlying forms that produce the same surface form. In the present case, all three candidate underlying forms (|an + pa|, |am + pa|, and |aŋ + pa|) yield the requested surface form /.am.pa./, so that all three remain as comprehension candidates, to be disambiguated higher up by syntactic, semantic, and pragmatic processing. Thus, although Hale & Reiss criticize OT for not handling comprehension well, their own proposal does have to working with lists of candidates, as OT does.

Not surprisingly, then, the problem with Smolensky's proposal turns out not to be a problem with bidirectionality or OT, but a problem with the number of levels considered. Within bidirectional OT, the solution is to allow the phonology to interact with the independently needed higher levels, such as the semantics. The choice for |am + pa| /.am.pa./ in (5) is then not entirely determined by faithfulness constraints: constraints further up Figure 1 could play a role. For instance, a semantic-pragmatic constraint could say that |am + pa| is not an appropriate sequence of morphemes in the present discourse context. This is also the constraint X in tableau (5). More details are in §9.1.

Next, the merely-phonological acquisition model in §1.4 has its problems. Tableau (3) only works if the learner has knowledge both of the underlying form and of the surface form. But in reality, that information is not directly available to the learner. Both forms must be based on something the learner has heard, perhaps an auditory-phonetic form such as [ampa]. From this, the learner first has to construct the abstract discrete phonological surface form /.am.pa./. This perceptual construction process has to rely on language-specific knowledge of the relation between phonetic detail and discrete phonological elements, and the process itself is language-specific and interacts with the phonology (see §3.2). The second thing the learner has to construct is the underlying form |an + pa|. Something must have told her that this form is correct, rather than a competing underlying form |am + pa| that could also be in the lexicon. The recognition process has to rely on language-specific knowledge of the relation between phonological structures and meaning, and the process itself is language-specific and interacts with the phonology (as seen above and in §9.1).

The solution is to model both the perceptual construction process and the recognition process in OT, because both are language-specific and interact with the phonology.

Finally, the merely-phonological evolution model in §1.5 has its problems. Phonological change is predicted to occur very rarely if at all, although in reality it happens all the time. Also, the model cannot handle the existence of **transmission noise**, which is the phenomenon that what the learner hears has been distorted by background noise. Of course, it is possible to think that this transmission noise is precisely what causes phonological change (e.g. Ohala 1981), but that can be shown to be incorrect (§4.3).

The solution is to model the phonetics in OT and have it interact with the phonology. It will turn out that automatic biases arising during acquisition will counteract the transmission noise, so that equilibria are allowed to emerge within a few generations (§4.3).

The following four sections describe the phonetic and semantic representations and their relations, and how they can interact with the phonology.

## 2. Phonetic representations: Auditory and Articulatory Form

The minimum number of phonetic representations seems to be two.

The **Auditory Form** is a sequence of events on auditory continua such as pitch, noise, spectral peaks and valleys, and silences, their durations, and their relations such as simultaneity and order. For instance, the *microscopic auditory transcription* [[aãm\_pa]] (Boersma 1998) is a shorthand notation for vocalic material with a high first formant, followed by the same but with a nasal spectral peak and valley added, followed by the spectral resonance that reflects the nasal cavity and a long oral sidebranch, followed by a silence, followed by a burst with low spectral features, followed by vocalic material with a high first formant (it is the cross-linguistically most common sound associated with the phonological structure /.am.pa./).

The **Articulatory Form** is a sequence of gestures by the multiple articulatory muscles that move, hold, tense, or relax the glottis, the larynx, the epiglottis, the pharynx walls, the tongue tip, the tongue body, the velum, the lips, the cheeks, the jaw, and the lungs. For instance, the phonetic transcription [aãmpa] is a very rough approximating shorthand for an articulation with constantly applied lung pressure and glottal adduction, starting with a lowered jaw, a low tongue, open lips, and a closed velum, followed by a lowering of the velum, followed by a closure of the lips (with jaw raising), followed by a raising of the velum, followed by an opening of the lips (with jaw lowering). This description of the Articulatory Form in terms of movements is still rather sketchy, because a description in terms of muscle activities (Boersma, 1998) is much more precise.

For these two representations, the following subsections describe their relations and their roles in merely-phonetic comprehension, production, acquisition, and evolution.

## 2.1. The relation between Auditory Form and Articulatory form

As a speaker/listener, you have knowledge of what your articulations will sound like, and conversely of how to implement articulatorily a sound that you want to produce. This is expressed by **sensorimotor constraints**, which say such things as “an auditory high F1 (first formant) does not correspond to an articulatory raised jaw.”

The ranking of these constraints is less language-specific than that of other constraints, because the shapes of our vocal tracts do not depend on the language we are learning. So the rankings of these constraints, *if and when* they have been learned (i.e. *if* the relevant sounds and articulations are used in the language at all, and *when* the learner has finished acquiring their relations), are universal (or perhaps they depend on the speaker). Areas in auditory or articulatory space that your language does not use at all will probably lead to poor sensorimotor knowledge (variable constraint ranking) in those areas, though.

## 2.2. The process of merely-phonetic articulation

Merely-phonetic articulation is the mapping from a target Auditory Form to an Articulatory Form, as in Figure 5.

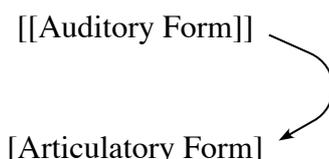


Fig. 5 The merely-phonetic articulation process.

As can be seen in Figure 1, the process is analogous to merely-phonological production in the sense that the result is due to an interaction between constraints that connect the input to the output (the sensorimotor constraints) and constraints that evaluate the output, the **articulatory constraints**. Observed cases of place assimilation, for instance, do not necessarily have to be due to a phonological process such as the one in (1), but could also have a purely articulatory source: the pronunciation [anpa] requires both a tongue-tip movement and a lip movement, whereas [ampa] requires only the lip movement. Tableau (6) shows how the target sound [[aã\_n\_pa]], which is specified for nasal coronality (say, high F2), could be pronounced if articulatory constraints outrank sensorimotor constraints.

### (6) Phonetic production without phonological input

[[aã_n_pa]]	*LIPS <sub>ART</sub>	*TONGUE TIP <sub>ART</sub>	*[[high F2]] [lips]	*[[high F2]] [tip]
[[aã_n_pa]] [anpa]	*	*!		*
☞ [[aã_n_pa]] [ampa]	*		*	

The ranking of the sensorimotor constraints \*[[high F2]] [lips] >> \*[[high F2]] [tip] reflects the idea that the spectral auditory cue [[high F2]] is more compatible with a tongue tip articulation than with a lip articulation. While these two constraints

together favour the first ‘faithful’ candidate, the higher-ranked articulatory constraint against tongue-tip gestures forces an observable assimilation.

### 2.3. The processes of merely-phonetic audition

The process of merely-phonetic articulation can be reversed, as in Figure 6. This process answers the question: given the Articulatory Form [ampa], what Auditory Form is associated with it?

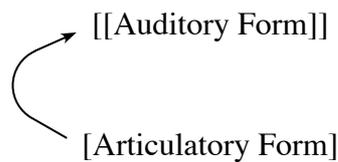


Fig. 6 Reversed merely-phonetic articulation.

This process cannot really be called ‘audition’, because audition is something that starts in your ears. The process can be regarded as the speaker’s internalized view of what her articulations will sound like. Tableau (7) shows that the articulation [ampa] will in this way be interpreted as producing the sound [[aãm\_pa]].

#### (7) *Sensorimotor expectation*

[ampa]	*LIPS <sub>ART</sub>	*TONGUE TIP <sub>ART</sub>	*[[high F2]] [lips]	*[[low F2]] [lips]
[[aãn_pa]] [ampa]	*		*!	
☞ [[aãm_pa]] [ampa]	*			*

The articulatory constraints are now ineffective, because they evaluate the input articulatory form, which is identical for all candidates. The choice is then made by the sensorimotor constraints, whose ranking  $*[[high F2]] [lips] \gg *[[low F2]] [lips]$  expresses the fact that lip-closing gestures tend to generate low rather than high F2 values.

### 2.4. Merely-phonetic acquisition

Every time you speak, you can improve your knowledge of the relation between articulation and sound, as in Figure 7. For an infant, such learning occurs every time she tries out her speech apparatus in vocal play.

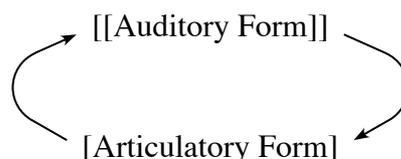


Fig. 7 Bidirectional acquisition of mere sensorimotor knowledge.

Analogously to tableau (3), we can combine tableaus like (6) and (7) into a learning tableau. Presumably, if the learner produces the articulation [anpa], she will hear it as

[[aã̃n\_p̃a]]. The learning pair for tableau (8) is therefore [[aã̃n\_p̃a]] [anpa]. The tableau combines all other candidates that share the sound [[aã̃n\_p̃a]] or the articulation [anpa].

(8) *Sensorimotor acquisition*

[[aã̃n_p̃a]] [anpa]	*TONGUE TIP <sub>ART</sub>	*[[low F2]] [lips]	*[[high F2]] [tip]	*[[low F2]] [tip]	*[[high F2]] [lips]
√ [[aã̃n_p̃a]] [anpa]	*→		*→→		
☞ [[aã̃n_p̃a]] [ampa]					←*
[[aã̃m_p̃a]] [ampa]		*			
☞ [[aã̃m_p̃a]] [anpa]	*			←*	

Tableau (8) represents a learner with non-optimal sensorimotor knowledge: she thinks that lip movements are associated with high F2 and tongue-tip movements with low F2. Under this immature view, candidate 2 is the most harmonic of all the candidates that share the auditory form [[aã̃n\_p̃a]], and candidate 4 is the most harmonic of the candidates that share the articulation [anpa]. Since the most harmonic candidate that has both [[aã̃n\_p̃a]] and [anpa] (“√”) is different from both the forward winner (“☞”) and the backward winner (“☞”), the learner can profit from two “mistakes”. From the forward “mistake” she will raise the ranking of \*[[high F2]] [lips] and lower the ranking of \*[[high F2]] [tip], so that her knowledge about articulations that produce a high F2 improves; from the backward “mistake” she will raise \*[[low F2]] [tip] and again lower \*[[high F2]] [tip], so that her knowledge about the sound produced by tongue-tip movements improves. Another facet of tableau (8) is that the learner’s forward “mistake” (a comparison between candidates 1 and 2) causes a lowering of the articulatory constraint \*TONGUE<sub>TIP</sub><sub>ART</sub>.<sup>1</sup> In general, sensorimotor learning tableaux like (8) cause all articulatory constraints to become ranked so low that they stop determining the articulatory output.

Articulatory constraints may be different from other kinds of constraints to the extent that they are connected to the articulatory periphery. If constraint reranking is a process that takes place entirely in the cerebral cortex, it is not sure that articulatory constraints can be reranked in tableaux like (8): perhaps the ranking of these constraints is directly determined by articulatory effort (Kirchner 1998 considers this a possibility).

Independently of whether articulatory constraints can move or not, forward sensorimotor learning with tableaux such as (8) will always render the articulatory constraints jobless: in (8), the sensorimotor constraint \*[[high F2]] [lips] will inevitably rise above \*TONGUE<sub>TIP</sub><sub>ART</sub> after some time; the other relevant constraint in (8), namely \*[[low F2]] [tip], will analogously rise above \*LIP<sub>ART</sub> as a result of tableaux involving learning pairs such as [[aã̃m\_p̃a]] [ampa]. The end result is perfect sensorimotor knowledge, i.e., the learners typically end up with a firmly fixed relation between Auditory Form and Articulatory Form, where the sensorimotor constraints

<sup>1</sup> In *backward* learning, \*TONGUE<sub>TIP</sub><sub>ART</sub> does not move, because this constraint is violated equally often in the two relevant candidates (1 and 4).

against incorrect auditory-articulatory relations are high-ranked, and those against correct auditory-articulatory relations are low-ranked. This means that the adult speakers-listeners that we model elsewhere in this paper can be assumed to have perfect sensorimotor knowledge, so that we can usually collapse the Auditory Form and the Articulatory Form into one single “Phonetic Form”.

## **2.5. Merely-phonetic evolution**

The very simple acquisition model of §2.4 does not in itself seem to lead to changes over the generations: there is no way I can put my very personal sensorimotor knowledge into my child’s head, so my child has to learn the articulation-audition relation from scratch, with her own physiology. Any changes between generations necessarily involve at least one higher level of representation, and are therefore discussed in §3.6 and §4.3.

## **3. The phonology-phonetics interface**

Having established that there are two phonological representations connected to each other (§1) and two phonetic representations connected to each other (§2), it remains to be established which phonological representations are connected to which phonetic representations. There exist several theories about this. From the phonological side, there is usually a single connecting representation, namely the Surface Form, and this is illustrated in Figure 1. From the phonetic side, the situation is less clear: Figure 1 proposes that the connecting representation is the Auditory Form, but the theory of *Direct Realism* (Fowler, 1986) proposes that the connection is made via the Articulatory Form instead. I will assume that Figure 1 is correct, because it can be shown to work quite well. Direct Realists are invited to interchange the two phonetic representations in Figure 1 and to show that that alternative grammar model works equally well or better; more about this, and about a third model (connecting both Auditory and Articulatory Form to the Surface Form, depending on the direction of processing) is discussed in §4.4.

### **3.1. The relation between Surface Form and Auditory Form**

Following Figure 1, the phonetics-phonology interface is a relation expressed in terms of *cues*: auditory events in the Auditory Form can be *cues to* phonological elements in the Surface Form.

Cross-linguistically speaking, auditory cues are arbitrarily related to phonological elements. In English, a major cue to the phonological feature /voiced/ at the end of a word is the duration of the preceding vowel (House and Fairbanks, 1953): the vowel of /li:də/ ‘leader’ is produced longer than that of /li:tə/ ‘litre’, and the difference is even larger in monosyllables such as /ɪoud/ ‘road’ and /ɪout/ ‘wrote’. No such gigantic differences are found in most other languages, like e.g. between German productions of /li:də/ ‘songs’ and /li:tə/ ‘litre’. In production, therefore, auditory cues for voicing are used differently in German than in English.

The cross-linguistic differences in cue use are bidirectional. The differences in cue use in production are reflected in how the cues are weighted in perception. Thus, English listeners but not Arab listeners rely strongly on the duration of the preceding

vowel when having to decide whether they heard a voiced or voiceless consonant (Crowther and Mann, 1994). The same correlation between production and perception holds between closely related varieties of the same language. For instance, the differences between an /i/ and /ɪ/ produced by Scottish English lies mainly in spectral differences, namely the F1 (first formant), whereas the same phonological difference in Southern British English is implemented to a large degree in duration as well, both in production and in comprehension (Escudero and Boersma, 2003).

Clearly, the interface between phonology and phonetics is bidirectional, so that the **cue constraints** have to be formulated bidirectionally. For the case of English voiced consonants, one of the relevant cue constraints can be written as follows (the index *i* denotes correspondence; the symbol “·” means that the two adjacent elements are in the same morpheme):

(9) *A vocalic cue constraint for voicing*

$*/V_i \cdot \{C, +\text{voi}\} / [[-\text{lengthened } v_i]]$

“non-lengthened auditory vocalic material does not correspond to a phonological vowel before a tautomorphemic voiced consonant”

The formulation in (9) is not entirely correct yet, because the notation of the auditory value  $[[-\text{lengthened}]]$  suggests discreteness (it is either plus or minus). In general, there will be a continuous range of possible values along every auditory continuum. This is why Escudero and Boersma (2003) proposed continuous ranges of cue constraints such as those in (10).

(10) *Arbitrary cue constrains for vowel classification*

$*/\text{ɪ} / [[F1 = x \text{ Hz}]]$

“a first formant of *x* Hz does not correspond to the phonological vowel /ɪ/”

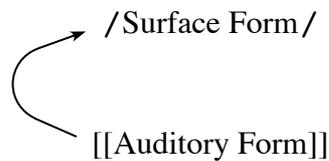
Constraints as in (10) are then thought to exist for every possible value of *x* between, say, 200 and 1200 Hz (more correctly, the frequency scale should be in auditory units such as Bark or ERB rather than Hz, which are acoustic units). Thus, a conspicuous property of the constraints in (10) is that they are *arbitrary*, i.e., they exist even for first formant values that are typical of the vowel /ɪ/. With arbitrary cue constraints it is the task of the constraint ranking, not the task of the constraint set, to make sure that /ɪ/ connects to plausible auditory events.

The formulations in (9) and (10) are bidirectional, e.g., (9) can be read equally well as “non-lengthened auditory vocalic material should not be perceived as a phonological vowel before a tautomorphemic voiced consonant” and as “a phonological vowel before a tautomorphemic voiced consonant should not be produced as non-lengthened auditory vocalic material.” This bidirectionality has the advantage that the acquisition of comprehension (§3.2, §3.3) helps in achieving appropriate production skills (§3.4, §4).

### 3.2. The process of prelexical perception

If the mapping from Auditory Form to Surface Form is regarded in isolation, it does not involve any processing at higher levels, most notably the lexicon. For this reason,

psycholinguists call this process *prelexical perception*. Phoneticians, who tend to be less involved with matters lexical, usually stay with the term *perception*.



**Fig. 8** The prelexical perception process.

Since the relation between Auditory and Surface Form is expressed in terms of cues, an OT modelling of the mapping from Auditory to Surface Form is expected to involve cue constraints. An account of the perception of English /i/ and /ɪ/ in terms of cue constraints alone was given by Escudero and Boersma (2003, 2004). As an example, they considered the auditory event [[vocalic material, F1 = 349 Hz, duration = 74 ms]], to be abbreviated as [[349 Hz, 74 ms]]. In a Scottish English environment this event will be perceived as the vowel /i/, because 349 Hz would be a too low first formant for /ɪ/ in that variety of English:

(11) *Vowel perception in Scottish English*

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

In a Southern British English environment, the same auditory event will be perceived as /ɪ/, perhaps because it is too short to be a plausible Southern /i/:

(12) *Vowel perception in Southern British English*

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

But cue constraints are not the only constraints that pose restrictions on the outcome of prelexical perception. We can see in Figure 1 that structural constraints directly evaluate Surface Forms, so they ought to interact with cue constraints in perception. The first account of what can (with hindsight) be called perception with structural constraints in OT is that by Tesar (1997, 1998, 1999) and Tesar & Smolensky (1998, 2000). In their examples of *robust interpretive parsing*, an *overt form*, which is a string of syllables marked for stress but not for phonological foot structure, is interpreted as a *full structural description*. The overt form [[σ'σσ]], for instance, is a sequence of an unstressed, a stressed, and an unstressed syllable. In the left-aligning iambic language of tableau (13) this overt form is interpreted as a left-aligned phonological iamb: / (σ'σ) σ /.

(13) *Perception of metrical structure in a left-aligning iambic language*

[[σ 'σ σ]]	FOOT BIN <sub>S</sub>	FOOT LEFT <sub>S</sub>	IAMBIC <sub>S</sub>	TROCHAIC <sub>S</sub>	FOOT RIGHT <sub>S</sub>
☞ / (σ 'σ) σ / [[σ 'σ σ]]				*	*
/σ ('σ σ) / [[σ 'σ σ]]		*!	*		
/σ ('σ) σ / [[σ 'σ σ]]	*!	*			*

In the right-aligning trochaic language of tableau (14) the same overt form is interpreted as a right-aligned phonological trochee: /σ ('σ σ)/.

(14) *Perception of metrical structure in a right-aligning trochaic language*

[[σ 'σ σ]]	FOOT BIN <sub>S</sub>	FOOT RIGHT <sub>S</sub>	TROCHAIC <sub>S</sub>	IAMBIC <sub>S</sub>	FOOT LEFT <sub>S</sub>
/ (σ 'σ) σ / [[σ 'σ σ]]		*!	*		
☞ /σ ('σ σ) / [[σ 'σ σ]]				*	*
/σ ('σ) σ / [[σ 'σ σ]]	*!	*			*

Because of the striking parallels, ‘robust interpretive parsing’ can be equated with prelexical perception, the ‘overt form’ can be regarded as a somewhat abstract variety of Auditory Form, and the ‘full structural description’ can be equated with the Surface Form (Boersma, 2003); this is the interpretation assumed by Apoussidou and Boersma (2003, 2004) and the reason for the notations in (13) and (14).

The big point that Tesar and Smolensky made was that structural constraints are needed both in the production direction (§1.2) and in the comprehension direction (the metrical examples above). Hence, if these structural constraints are ranked in an OT manner and if they influence production in an OT way, then they also influence comprehension in an OT way. Hence, interpretive parsing (which is prelexical perception) should be handled in OT if phonological production is.

The example by Tesar and Smolensky does not seem to involve cue constraints, but that is only because they did not consider candidates with different numbers of syllables or different stress patterns. One can imagine cue constraints for the relation between auditory intensity, pitch and duration on the one hand and phonological stress (headship of a foot) on the other, and one can imagine that these cue constraints interact in a parallel manner with the structural constraints in (13) and (14). For instance, in an iambic language a higher pitch on the first syllable may not turn that syllable into a foot head, whereas in a trochaic language it may.

The most general cases of perception involve an interaction of structural and cue constraints. For these, see Boersma (2006b) on the McGurk effect, Boersma (2007) on *h-aspiré* in French, and Boersma (2009a) on Polivanov’s idea of phonological perception.

### 3.3. Unidirectional acquisition of prelexical perception

I now explain the perceptual learning algorithm proposed by Boersma (1997), with an example from Escudero and Boersma (2004).

Suppose that a Scottish English child, at some point during her acquisition period, has a grammar that would be appropriate for listening to Southern British English, i.e. with the ranking of tableau (12). Now suppose that a Scottish English adult pronounces the auditory form [[349 Hz, 74 ms]]. As tableau (15) shows with the backward pointing finger, the child will perceive this as /ɪ/.

(15) *Vowel perception in Southern British English*

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

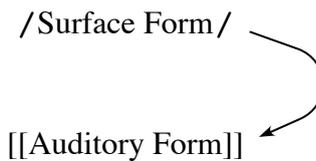
However, it is likely that the Scottish adult speaker intended the vowel /i/ instead. It is quite possible that the child will detect this. Perhaps the speaker said the morpheme <please>, so that the child's lexicon, which contains an underlying form |plɪz| but not |plɪz|, can already tell her that she should have perceived an /i/. Or the speaker intended the morpheme <sheep> and the child's lexicon was satisfied with recognizing the morpheme-underlying form pair <ship> |ʃɪp|, but subsequent semantic and conceptual processing in the given situational context made her decide that the speaker had actually intended <sheep> |ʃɪp|. Either way, the child will know that she has made a **perception error**, and mark the second candidate in (15) as 'correct'. As a result, the child will move the relevant constraints according to the arrows in (15), analogously to tableaux (3) and (4). After many of such demotions and promotions, the constraints will become ranked in a way appropriate for the Scottish language environment; in fact, the child will become a *probability matcher*, i.e., she will come to rank the constraints in such a way that an F1 of  $x$  Hz that is intended as /i/  $y$  percent of the time, will be perceived by her as /i/  $y$  percent of the time (Escudero and Boersma, 2003).

The error-driven procedure in (15) is called *lexicon-driven learning of perception*: the ultimately recognized Underlying Form *supervises* the learner's perception, i.e. determines what she should have perceived. The existence of this form of learning has been confirmed in the lab (Eisner, 2006).

A conspicuous property of tableau (15) is that it does not consider forward learning, i.e. it does not include candidates with the surface form /i/ but different auditory forms. So this is a case of unidirectional (only backward) learning. Section §3.4 has more to say about this.

### 3.4. The process of prototype selection

The process of prelexical perception can be reversed, as in Figure 9. This process answers the question: given the phonological Surface Form /i/, what is the best Auditory Form associated with it?



**Fig. 9** The prototype selection process.

Boersma (2006a) argued that this process cannot really be called ‘production’, because articulatory considerations are not involved. Not bound by articulatory effort, the resulting winning auditory form may well be much more ‘peripheral’ (lower F1, higher duration) than the average auditory realization. In fact, the learning algorithm described in §3.3 leads to such a situation. The idea is that the average token of /i/ may have an F1 of 330 Hz, but that 290 Hz is an even better token (a ‘prototype’) because it has less chance of being perceived as anything but /i/ (e.g. the typical /ɪ/ token has an F1 around 500 Hz).

(16) *Prototype selection in Scottish English*

/i/	*/i/ [[74 ms]]	*/i/ [[349 Hz]]	*/i/ [[200 ms]]	*/i/ [[290 Hz]]
/i/ [[349 Hz, 74 ms]]	*!	*		
/i/ [[290 Hz, 74 ms]]	*!			*
/i/ [[349 Hz, 200 ms]]		*!	*	
☞ /i/ [[290 Hz, 200 ms]]			*	*

For a more detailed gigantic tableau see Boersma (2006a).

**3.5. Acquisition of prototype selection?**

I have no separate learning algorithm for prototype selection, nor does there need to be one. This is because prototype selection is just a **paralinguistic task**, i.e. you can find the effect in the laboratory (Johnson, Flemming and Wright, 1993), but it is not a **linguistic task** like production and comprehension, which human evolution has optimized. Prototype selection is acquired automatically as a side effect of the perception learning algorithm described in §3.3.

**3.6. The evolution of the phonology-phonetics interface**

The learning algorithm described in §3.3 is not necessarily stable over the generations. This is because the child learns to mimic the auditory frequency distributions that she hears in her environment. When she becomes a parent, she will produce these same auditory distributions. But her child will not exactly hear these distributions: there will be an additional **transmission noise** caused by wind and speaker variation. For instance, the fact that Dutch listeners use duration instead of F2 as the main cue to distinguish /a/ from /ɑ/ (Gerrits, 2001: 89) is because the F2, being highly regionally dependent, is less reliable in the environment. So the auditory environment is different from the cue use of any single speaker. It would seem, then, that if the child mimics this variation, she will end up having a much broader

distribution of auditory values than her parents. Fortunately, the prototype effect described in §3.4 counteracts this drift, as we will see in §4.3.

Some straightforward biases can be imagined. Bone conduction makes the sound of our own speech slightly different from that of others, so that if cue constraints that we have optimized for comprehending others are reused by us in our own productions, the result may be slightly different. Likewise, speaker normalization will be an issue: if the same cue constraints for F1 and F2 that have optimized a child's vowel perception are reused by her in her own productions, she will try to mimic the adult formant values and thereby produce articulatorily less open vowels; this will lead to a general raising of vowel heights, especially for those vowels whose formants are relatively reliable, i.e. for long vowels. A mechanism like this may well be behind the Middle English vowel shifts (Boersma, 1998: 413).

### 3.7. Is this how the phonology-phonetics interface works?

As for which of the two phonetic representations is (or are) connected to the Surface Form, two other hypotheses are thinkable, and I discuss these in §4.4. As for whether the mapping from Auditory to Surface Form can always be regarded in isolation, as here in §3, at least McQueen and Cutler (1997) argue that perception indeed works this way, namely that prelexical perception is modular and receives no feedback from higher processing, such as from the lexicon or from the conceptual systems. Nevertheless, we will see an apparent case of parallel comprehension in §5.2. In production, a good case for parallelism can be made, as I show in the next section.

## 4. The three 'low' representations:

### Articulatory Form – Auditory Form – Surface Form

In §3.4, we saw that Auditory Form and Surface Form are not sufficient for modelling production. We therefore now include Articulatory Form as well.

#### 4.1. The process of phonetic production

Once we have three representations, there are two ways to get from one end to the other: serial and parallel. Figure 10 shows the serial edition of getting from Surface Form to Articulatory Form in phonetic production:

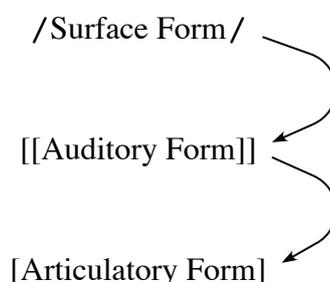
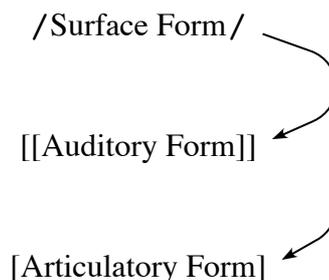


Fig. 10 The phonetic production process, serial edition.

What this means is that the speaker, given the phoneme /i/, first computes an auditory prototype (by means of **cue constraints**), say [[F1=280 Hz]], then turns this

prototype into an articulation, which because of articulatory constraints only produces an F1 of 330 Hz. It is possible that this works. The **sensorimotor constraints** would prefer to generate an articulation that produces an F1 of 280 Hz, but the **articulatory constraints** would prevent this. However, the learning algorithm in §2.4 predicts that learners have no route to a situation in which articulatory constraints stably outrank sensorimotor constraints. This problem does not occur in the parallel model, which I describe next.

The parallel edition of phonetic implementation looks as Figure 11.



**Fig. 11** The phonetic production process, parallel edition.

What this means is that the Auditory and Articulatory Forms are computed at the same time, and that the cue constraints can interact with the articulatory constraints. Tableau (17) shows how the phonological surface form /*.an.pa./* can be pronounced as [ampa].

(17) *Phonetic production with phonological input*

<i>.an.pa./</i>	*LIPS <sub>ART</sub>	*TONGUE TIP <sub>ART</sub>	*/n/ [[low F2]]	*/n/ [[high F2]]
<i>.an.pa./</i> [[aãn <sup>p</sup> a]] [anpa]	*	*!		*
☞ <i>.an.pa./</i> [[aãm <sup>p</sup> a]] [ampa]	*		*	

In tableau (17) I have ignored the sensorimotor constraints by assuming a perfect relationship between Auditory Form and Articulatory Form, as explained in §2.4; this is why in every candidate cell the Auditory Form, e.g. [[n]], corresponds perfectly to the Articulatory Form, e.g. [n]. The cue constraint ranking \*/n/[[low F2]] >> \*/n/[[high F2]] prefers that the phonological surface element /n/ is pronounced with a high F2 (i.e. as coronal) rather than with a low F2 (i.e. as labial). Nevertheless, the articulatory constraint \*TONGUETIP<sub>ART</sub> overrides this preference because of its high ranking. The result is a crucial interaction between different levels of the grammar: a ‘later’ constraint (at Articulatory Form) overrides ‘earlier’ constraints (between Surface and Auditory Form), leading to a choice for [[aãm<sup>p</sup>a]] that the serial model of Figure 10 can never produce.

#### 4.2. The acquisition of phonetic knowledge

Boersma and Hamann (2008) use the unidirectional perceptual learning algorithm of §3.3 to show how a child learns to produce sibilants like /s/ and /ʃ/, under the

assumptions that (1) phonetic production is parallel, (2) sensorimotor knowledge is perfect, and (3) articulatory constraints have a fixed ranking. Figure 12 shows the processes involved (prelexical perception and parallel phonetic production).

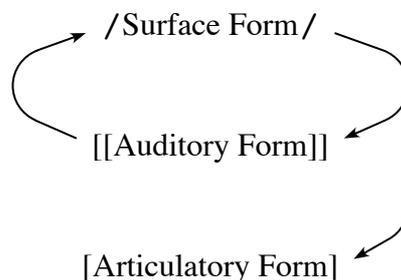


Fig. 12 The acquisition of parallel phonetic production.

Boersma and Hamann show that acquisition comes with an automatic bias towards balancing distinctivity and articulatory effort, without the assumption that the learner has any knowledge of auditory distances (hence no need for Flemming’s 1995 MINDIST constraints). This effect relies, then, on the idea that the ranking of the cue constraints has been optimized in perceptual learning (§3.3) and that the resulting ranking is reused by the speaker in production, not just for mapping the Surface Form to the Auditory Form (as in §3.4) but for mapping it to the two phonetic forms in parallel.

A unidirectional production learning procedure is also imaginable. Imagine, for instance, that a child has the grammar in (18), which does not allow her to produce place assimilation, as the forward point finger indicates. Imagine at the same time that adults in the child’s environment do assimilate. In that case, the child will hear auditory forms such as  $[[a\tilde{a}m\_pa]]$ , even if the phonological surface form is  $/.an.pa./$  (I ignore higher-level processes such as faithfulness violations for the moment).

(18) *The acquisition of phonetic implementation*

$/.an.pa./$ $[[a\tilde{a}m\_pa]]$	$*/n/$ $[[low\ F2]]$	$*TONGUE$ $TIP_{ART}$	$*/n/$ $[[high\ F2]]$
$☞$ $/.an.pa./$ $[[a\tilde{a}n\_pa]]$ $[anpa]$		$\leftarrow*$	$\leftarrow*$
$\checkmark$ $/.an.pa./$ $[[a\tilde{a}m\_pa]]$ $[ampa]$	$*\rightarrow$		

The result is that the articulatory constraint rises, a situation that was impossible in the case of sensorimotor learning in tableau (8). After many surface-auditory pairs such as the one supplied in (18), the articulatory constraint will emerge above the cue constraint  $*/n/[[low\ F2]]$ , so that the child will assimilate her  $/n/$ , just as the adults around her.

**4.3. The evolution of phonetic implementation**

Boersma and Hamann (2008) showed that the perceptual acquisition model in §4.2 leads to a stable evolution of the language over the generations. Even if a language starts with a skewed and rare set of sibilants, say  $[ç]$  and  $[ʂ]$ , it will achieve a stable

equilibrium of [s] and [ʃ] within a few generations. The equilibrium is achieved when the transmission noise is exactly counterbalanced by the articulatory effort associated with extreme auditory values. An optimal balance between articulatory ease and perceptual distinctivity is thus obtained without the assumption that the learner has any knowledge of auditory distances. This result is achieved solely by the assumption that cue constraints whose ranking has been optimized in perception are reused with the same ranking in production, and by the assumption that the mapping from Surface Form to Auditory Form runs in parallel with the mapping from Auditory to Articulatory Form, so that articulatory and cue constraints can interact. Whether and how this result should be modified if the unidirectional acquisition model of (18) is included, has not been investigated yet.

#### **4.4. Is this how the phonetic representations are connected to the phonology?**

Throughout §3 and §4 I have been assuming that the only phonetic representation that connects to the phonological Surface Form is the Auditory Form, and that the Articulatory Form connects only to the Auditory Form. As mentioned at the beginning of §3, there are two other possibilities.

The first other possibility is *Direct Realism* (Fowler, 1986), which claims that listeners directly perceive articulatory gestures. This view of perception can be summarized by a modification of Figure 10 in which the Auditory and Articulatory Forms are reversed; that is, listeners receive an Auditory Form, map this via their sensorimotor knowledge to an Articulatory Form, and use this Articulatory Form to get at a phonological Surface Form. This view of speech comprehension is fully compatible with the model described in this paper, with the difference that the cue constraints would have to involve articulatory rather than auditory cues. In speech *production* (an analogously modified Figure 11), however, such a model may become problematic: it would predict that phonetic targets are articulatory rather than auditory, which is incompatible with the results of bite-block experiments (Lindblom, Lubker and Gay, 1979), in which speakers, when confronted with artificial articulatory restrictions, apparently adapt their articulations to reach fixed auditory goals.

The second other possibility (noted by an anonymous reviewer) is that both the Auditory and Articulatory Form are connected to the Surface Form, so that listeners could directly map incoming sounds to the phonology without activating any articulations, and speakers could directly map the phonology to articulatory gestures without computing any sounds. Such a model would need both the auditory cue constraints of Figure 1 and the articulatory cue constraints that Direct Realism would need. Beside sharing with Direct Realism the bite-block problem in production, such a model would be especially difficult to reconcile with the learning procedures of this paper: whereas the model of this paper optimizes the cue constraint ranking in perception (§3.3) and is able to reuse this ranking in production (§4.2), the model under discussion here would be able to optimize the rankings of its *auditory* cue constraints in perception, but be at a loss when confronted with the task of learning appropriate rankings of *articulatory* cue constraints in production: because the two sets of cue constraints are separate, acquiring a ranking for one set does not help in acquiring the ranking of the other set; and the learning algorithm of (18) cannot work

either, because it relies on the simultaneous availability of auditory and articulatory representations.

## 5. The three ‘middle’ representations: Auditory Form – Surface Form – Underlying Form

In this section we go one level up from the triplet of representations discussed in §4. Since the triplet Auditory-Surface-Underlying does not include the Articulatory Form, the only process we can handle is comprehension. This can be done serially or in parallel.

### 5.1. The serial edition of the process of phonetic-phonological comprehension

In a serial view of comprehension, prelexical perception is followed by word recognition, as in Figure 13.

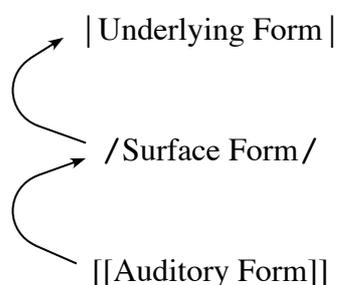


Fig. 13 The phonetic-phonological comprehension process, serial edition.

I will discuss an example (after an analogous example from Boersma 2009a, which involves voice onset time). Suppose the auditory form is a sound that sounds like a typical Scottish Standard English *clin* (which is a nonsense word) or *clean*, or something in between. In the following tableaux I assume, quite simplifyingly, that the only auditory aspect in which these sounds differ is F1.

Step one in the serial model is prelexical perception, i.e. the mapping from the given Auditory Form to a phonological surface structure (Surface Form). The cue constraints have become ranked (by the acquisition procedure) according to their distance to the category boundary, which is at, say, 400 Hz (Escudero and Boersma, 2004). The worst token of /ɪ/ is one with a very low F1 such as 300 Hz, so the cue constraint that says that [[300]] should not be perceived as /ɪ/ is high-ranked. Likewise, constraints that connect large F1 values to /ɪ/ are also high-ranked. An appropriate ranking for perceiving the Scottish contrast must be similar to that in tableaux (19) to (21).

#### (19) Scottish English classification of vowel height

[[300 Hz]]	*/ɪ/ [[520]]	*/ɪ/ [[300]]	*/ɪ/ [[450]]	*/ɪ/ [[450]]	*/ɪ/ [[300]]	*/ɪ/ [[520]]
☞ /.klin./ [[300 ms]]					*	
/.kln./ [[300 ms]]		*!				

(20) *Scottish English classification of vowel height*

[[450 Hz]]	*/i/ [[520]]	*/ɪ/ [[300]]	*/i/ [[450]]	*/ɪ/ [[450]]	*/i/ [[300]]	*/ɪ/ [[520]]
/klin./ [[450 Hz]]			*!			
☞ /kɪn./ [[450 Hz]]				*		

(21) *Scottish English classification of vowel height*

[[520 Hz]]	*/i/ [[520]]	*/ɪ/ [[300]]	*/i/ [[450]]	*/ɪ/ [[450]]	*/i/ [[300]]	*/ɪ/ [[520]]
/klin./ [[520 Hz]]	*!					
☞ /kɪn./ [[520 Hz]]						*

We see that, as expected, F1 values below the boundary of 400 Hz are perceived as /i/, and that those above 400 Hz are perceived as /ɪ/.

Step 2 in the serial model is word recognition. The underlying form |klin| exists (it means ‘clean’, i.e. it is connected to the morpheme <clean>), the underlying form |kɪn| does not. The perceived form /klin./ will easily be recognized with the help of faithfulness constraints such as \*|æ|/i/ (‘an underlying |æ| does not connect to a surface /i/’):

(22) *Word recognition*

/klin./	*<>  x	* æ  /i/	* æ  /ɪ/	* ɪ  /i/	* i  /ɪ/
☞ <clean>  klin  /klin./					
<>  kɪn  /klin./	*!			*	
<clan>  klæn  /klin./		*!			

Here it is necessary to include some minimal information from the morpheme level, namely about whether the underlying form corresponds to a morpheme or not. The lexical constraint \*<>|x|, then, militates against having any underlying form |x| that does not correspond to any morpheme. The winning candidate violates no constraints at all. A more interesting surface form is /kɪn./:

(23) *Word recognition*

/kɪn./	*<>  x	* æ  /i/	* æ  /ɪ/	* ɪ  /i/	* i  /ɪ/
☞ <clean>  klin  /kɪn./					*
<>  kɪn  /kɪn./	*!				
<clan>  klæn  /kɪn./			*!		

In this case one still recognizes <clean>|klin|, although a different ranking of some faithfulness constraints would have led one to recognize <clan>|klæn| instead:

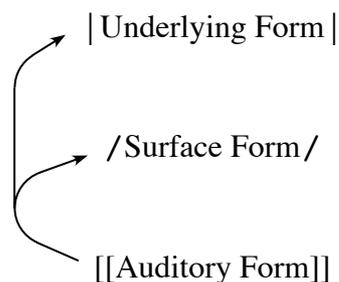
(24) *Word recognition*

/kɫɪn./	*<>  x	* æ  /i/	* i  /ɪ/	* æ  /ɪ/	* ɪ  /i/
<clean>  klin  /kɫɪn./			*!		
<>  kɫɪn  /kɫɪn./	*!				
☞ <clan>  klæn  /kɫɪn./				*	

We cannot predict which of the two options people will choose. In any case, the choice between tableaux (23) and (24) does not depend on the degree of ambiguity of F1; that is, once the listener has perceived /kɫɪn./, the chances of recognizing |klæn| do not increase when F1 rises. This may be a disadvantage of the serial model, as the next section argues.

**5.2. The parallel edition of the process of phonetic-phonological comprehension**

The situation is different in the parallel model of Figure 14.



**Fig. 14** The phonetic-phonological comprehension process, parallel edition.

We first provide a ranking that makes the listener perceive an F1 of 520 Hz as /kɫɪn./, never mind that the faithful underlying form |kɫɪn| does not exist in the lexicon. If the lexicon is still capable of telling the listener that the word the speaker intended was <clean>|klin|, the ranking can be the one in tableau (25).

(25) *Perception possibly but not really influenced by lexical access*

[[520 Hz]]	*/i/ [[520]]	*/ɪ/ [[300]]	*<>  x	* ɪ  /i/	* i  /ɪ/	*/i/ [[450]]	*/ɪ/ [[450]]	*/i/ [[300]]	*/ɪ/ [[520]]
<>  kɫɪn  /kɫɪn./			*!						*
<>  kɫɪn  /kɫɪn./	*!		*	*					
☞ <clean>  klin  /kɫɪn./					*				*
<clean>  klin  /kɫɪn./	*!								

In the case of an F1 of 450 ms, which was perceived as /ɪ/ in the sequential model, the perception now becomes /i/, as shown in tableau (26):

(26) Perception possibly and really influenced by lexical access

[[450 Hz]]	*/i/ [[520]]	*/ɪ/ [[300]]	*<>  x	* ɪ/ /i/	* i/ /i/	*/i/ [[450]]	*/ɪ/ [[450]]	*/i/ [[300]]	*/ɪ/ [[520]]
<>  kɪn  /.kɪn./			*!				*		
<>  kɪn  /.klin./			*!	*		*			
<clean>  klin  /.kɪn./					*!		*		
☞ <clean>  klin  /.klin./						*			

In this tableau we see that the cue constraints prefer /ɪ/, but the faithfulness constraint, forced top-down by \*<>|x|, prefers /i/. If we compare this to tableau (20), we see that the availability of the lexical item <clean>|klin| has shifted the auditory boundary between the categories /i/ and /ɪ/ towards the /ɪ/ side. This is an effect that has been found with human listeners in the lab (for a different auditory continuum) by Ganong (1980); it is predicted within McClelland and Elman's (1986) interactive TRACE model of speech comprehension, but not in McQueen and Cutler's (1997) serial models (see Norris, McQueen & Cutler 2000 for a defence of one of these models).

A remaining question is whether \*<>|x| can ever be violated in a winning form. The answer is that it can, if it is outranked by both faithfulness and cue constraints. In that case, tableau (25) would become tableau (27).

(27) Recognizing a nonsense word

[[520 ms]]	*/i/ [[520]]	*/ɪ/ [[300]]	* ɪ/ /i/	* i/ /i/	*<>  x	*/i/ [[450]]	*/ɪ/ [[450]]	*/i/ [[300]]	*/ɪ/ [[520]]
☞ <>  kɪn  /.kɪn./					*				*
<>  kɪn  /.klin./	*!		*		*				
<clean>  klin  /.kɪn./				*!					*
<clean>  klin  /.klin./	*!								

If both the cue constraints and the faithfulness constraints are ranked high enough, the auditory form is apparently capable of creating a new underlying form. This is explicit in tableau (27), but one can also see it from Figure 1 by regarding the cue and faithfulness constraints in that figure as strong connections, and the lexical constraints against non-existing morphemes as weak.

## 6. The quadruplet Underlying – Surface – Auditory – Articulatory

### 6.1. The process of phonological-phonetic production

The typical process in this quadruplet is *phonological-phonetic production*. It is hard to model phonetic influences on phonological decisions if one does not assume that

this process is parallel, as in Figure 15. For instance, in Boersma (2008) and Boersma (2009a), faithfulness constraints must crucially interact with both articulatory and cue constraints.

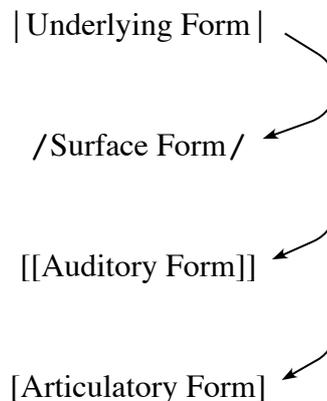


Fig. 15 The phonological-phonetic production process, fully parallel edition.

The example I want to address here is again the example of nasal place assimilation. After the ‘merely-phonological’ assimilation of tableau (1) and the ‘merely-phonetic’ assimilation of tableau (6), tableau (28) shows our third way to obtain assimilation, namely phonetically-based phonological assimilation.

(28) *Phonological-phonetic production*

an + pa	*/n/ [[low F2]]	*TONGUE TIP <sub>ART</sub>	IDENT PLACE <sub>US</sub>
an + pa  / .an.pa./ [[aã <sub>n</sub> <sup>p</sup> a]] [anpa]		*!	
an + pa  / .an.pa./ [[aã <sub>m</sub> <sup>p</sup> a]] [ampa]	*!		
☞  an + pa  / .am.pa./ [[aã <sub>m</sub> <sup>p</sup> a]] [ampa]			*

In (28), the high-ranked low-level articulatory constraint \*TONGUE<sub>TIP<sub>ART</sub></sub> and the high-ranked low-level cue constraint \*/n/[[low F2]] together force the violation of the low-ranked high-level faithfulness constraint IDENT<sub>PLACE<sub>US</sub></sub>. Such ‘feedback’ from lower to higher levels in production is only possible in a parallel (or ‘interactive’) model such as the one in Figure 15: in a serial model, IDENT<sub>PLACE<sub>US</sub></sub> would invariably (in the absence of structural constraints) turn |an + pa| into / .an.pa./.

**6.2. The acquisition of phonological-phonetic production**

One way to learn phonological-phonetic production is to interpret an incoming Auditory Form in terms of a Surface Form and an Underlying Form, then given this Underlying Form, to compute the Surface Form, Auditory Form and Articulatory Form that the learner herself would have produced. This is shown in Figure 16.

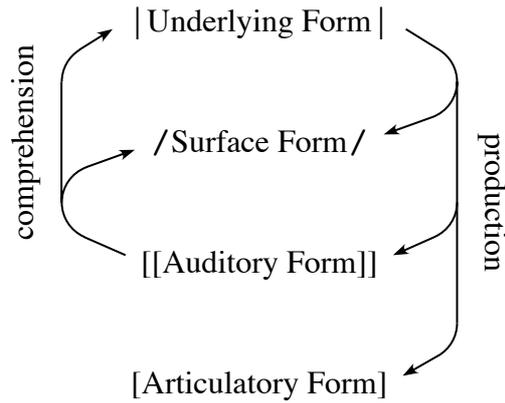


Fig. 16 Bidirectional acquisition of the phonological-phonetic production process.

Tableau (29) gives an example of a child who does not assimilate, in a language environment where adults do assimilate. Presumably, the child will hear auditory forms like  $[[a\tilde{a}m_p a]]$  from which she can deduce an underlying form like  $|an + pa|$ .

(29) *Phonological-phonetic acquisition*

$ an + pa $ $[[a\tilde{a}m_p a]]$	$*/n/$ $[[low F2]]$	IDENT PLACE <sub>US</sub>	$*TONGUE$ TIP <sub>ART</sub>
☞ $ an + pa $ /.an.pa./ $[[a\tilde{a}n_p a]]$ [anpa]			←*
$ an + pa $ /.an.pa./ $[[a\tilde{a}m_p a]]$ [ampa]	*		
√ $ an + pa $ /.am.pa./ $[[a\tilde{a}m_p a]]$ [ampa]		*→	

Given the child’s ranking of  $*/n/[[low F2]] \gg IDENTPLACE_{US}$ , the third candidate is the most harmonic of the two candidates that include both  $|an + pa|$  and  $[[a\tilde{a}m_p a]]$ . As a result, the child will regard this as the correct quadruplet. As a result of that, the child will lower her faithfulness constraint and raise her articulatory constraint, leading ultimately (i.e. after more learning from similar data) to the situation in (28), where she mimics her environment in assimilating her nasals.

Boersma (2008) shows that some observed universal rankings of faithfulness constraints (between Underlying Form and Surface Form) are predicted to be automatic results of parallel phonological-phonetic production. The learning algorithm in (29) predicts rankings of faithfulness constraints by frequency and auditory cue quality, without the need for innately ranked positional faithfulness constraints (Beckman, 1998), rankability by extralinguistic knowledge of auditory distances (Steriade’s 2001 P-map), or rankability by linguistically computed confusability (Boersma, 1998).

## 7. Semantic representations

Figure 1 includes only the two semantic representations that are of most interest to phonologists: the morpheme (for establishing morphemic identity), and the context (which influences expectations in comprehension). Semanticists would probably want to include more, such as the semantic underlying form (semantic features associated

with morphemes in the lexicon) and the literal meaning of an utterance (Henk Zeevat, p.c.). They would also probably not regard the Morpheme as exclusively semantic, because e.g. the morpheme ‘Nominative Singular’ expresses a syntactic function rather than a semantic role. And semanticists have equivalent (or nearly equivalent) names for the Context (or Context Change), such as ‘message meaning’, ‘situation’, ‘discourse representation structure’, ‘pragmatic context’, or even ‘pragmatic form’, some of which suggest that this representation is not exclusively semantic either.

OT semanticists tend to be interested in the relation between Semantic Form and Context Change, or just between Form and Meaning, where ‘Form’ is the Morpheme (e.g. *him* or *himself*) and Meaning can be a part of the Context (e.g. the person referred to be the pronoun or anaphor). All this is far away from the interests of phonologists, but it is important to note that OT semanticists have invented *Bidirectional Optimality Theory* for the solution of their problems, especially for the problem of how to explain the difference between *to kill* and *to cause to die*, or the division of labour between *him* and *himself* (Blutner 2000, Mattausch 2004). Boersma & Hamann (2008) noted that any distinctivity emerging from the bidirectional unidirectionality described above in Section 4.3 could well explain the partial blocking effects that Blutner and Mattausch ascribe to much more complex types of evaluation; this has been confirmed in simulations by Boersma (2009b).

## 8. The phonology-semantics interface: the lexicon

### 8.1. Relations

Since Saussure (1916), lexical entries have been regarded as ‘form-meaning’ pairs, whose ‘form’ part is the Underlying Form and whose ‘meaning’ part we can identify with the Morpheme. Saussure’s own terms were *signifiant* (‘signifier’ = form) and *signifié* (‘signified’ = meaning), and he insisted that their relation is *arbitrary*, i.e. there are no cross-linguistic universals on what form goes with what meaning (except for some cases of onomatopoeia).

The relation between form and meaning in the lexicon is usually regarded as fixed. This happens even in OT. Boersma (1999/2001), for instance, used ‘lexical constraints’ such as \*<wheel>|ɾad|, but this example just expressed the listener’s reluctance to access the single lexical item <wheel>|ɾad|, as opposed to, say, the item <rat>|ɾat|, with which it could be in competition during word recognition (in Dutch). Escudero (2005:214–236) went a bit further, proposing a competition between multiple lexical items with the same meaning, with ‘lexical constraints’ such as \*<girl>|tʃika| and \*<girl>|tʃika| for Dutch learners of Spanish. But only Apoussidou (2007: ch.6) investigated the relation between Underlying Form and Morpheme as a violable **lexical constraint**, in an application to the interaction between lexical and grammatical stress in Greek. For instance, if the lexical constraint \*<sea>|θalas| outranks \*<sea>|θálas|, then the morpheme <sea> likes to be |θálas| (with lexically specified stress) in the Underlying Form, whereas if the ranking is the reverse, the morpheme <sea> likes to be |θalas| (without any lexical stress specification) in the Underlying Form.

## 8.2. The process of lexical retrieval in production

If you ignore everything outside the lexicon (or if you have a serial modular view of production), then you will believe in the existence of a local process that retrieves the Underlying Form |dɔg|, given the Morpheme <dog>, as in Figure 17.

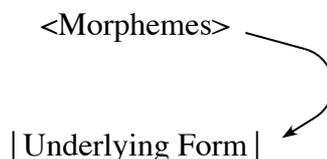


Fig. 17 The isolated lexical retrieval process in production.

## 8.3. The process of the access of meaning in comprehension

Analogously, if the listener has an Underlying Form at her disposal, she can access its corresponding Morpheme (and hence its lexical meaning), as in Figure 18.

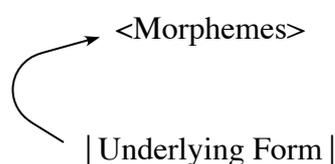


Fig. 18 The isolated lexical access process in comprehension.

## 8.4. The acquisition of lexical relations

An isolated acquisition of word-meaning pairs, as in Figure 19, would occur if a learner is presented with a given set of underlying forms and a given set of morphemes. It is difficult to see how such a learning situation could work without help from higher levels of representation (the Context) or lower levels of representation (the Surface Form). In §9, therefore, I include the Surface Form, thus enabling us to look at interesting interactions of the lexicon with the phonology.

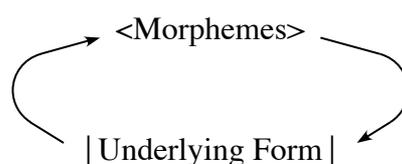


Fig. 19 Isolated bidirectional lexical acquisition.

# 9. The triplet Morphemes – Underlying Form – Surface Form

## 9.1. The influence of Morphemes (and Context) on word recognition

Boersma (1999/2001) modelled the connection from Surface Form to Underlying Form as an interaction of faithfulness constraints and lexical constraints, which militated against certain combinations of form and meaning. For instance, the mapping from the Dutch surface form /.rat./ to either the lexical form-meaning pair <wheel>|rad| or the lexical form-meaning pair <rat>|rat| would be decided by 'lexical constraints' that were conditioned by the Context, e.g. if the Context is "turn", then \*<rat>|rat|/"turn" will probably outrank \*<wheel>|rad|/"turn". In the present

model, it would be the Morpheme-Context relation that decides this, i.e. the ranking \**“turn”*<rat> >> \**“turn”*<wheel>, but the idea is the same. As pointed out by Boersma (1999/2001), this solves Smolensky’s problem in (5) without having to invoke Hale & Reiss’ analysis-by-synthesis model. If we assume, just as an example, that |an| means ‘wheel’, |am| means ‘rat’, |aŋ| means ‘guinea pig’, and the context is “turn” (perhaps because |pa| means something like ‘turn’), tableau (5) can be corrected as tableau (30).

(30) *The success of phonological comprehension with semantics*

“turn” / .am.pa./	*CODA WITH SEPARATE PLACES	*“turn” <guinea pig>	*“turn” <rat>	*“turn” <wheel>	IDENT PLACE <sub>US</sub>
☞ ☞ “turn” <wheel>  an + pa  / .am.pa./				*	*
“turn” <rat>  am + pa  / .am.pa./			*!		
“turn” <guinea pig>  aŋ + pa  / .am.pa./		*!			*

As long as two of the three semantic constraints outrank the faithfulness constraint for place, the listener will succeed in recognizing / .am.pa./ correctly, i.e. in finding the correct Underlying Form. For this to work, the higher-level semantic constraints have to be able to override the lower-level faithfulness constraints in the comprehension direction, i.e. word recognition and the access of meaning have to run in parallel, as in Figure 20.

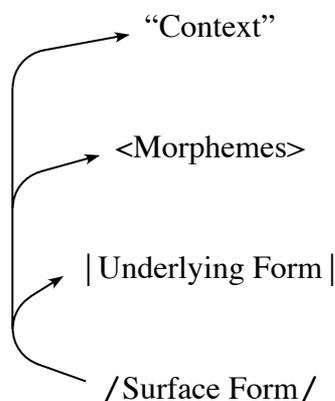
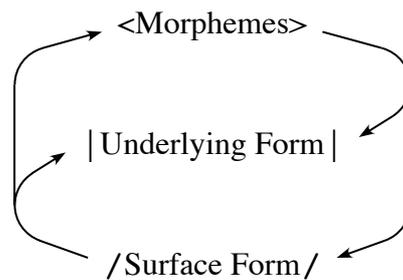


Fig. 20 Parallel access of lexical form and meaning.

## 9.2. Acquisition

Analogously to Figure 16, the mapping from Morphemes to Surface Form can be learned by first interpreting an incoming Surface Form in terms of a pair of Underlying Form and Morpheme, then computing the Underlying and Surface Form that the learner herself would have produced given this interpreted Morpheme, as in Figure 21.



**Fig. 21** The acquisition of underlying forms.

Apoussidou (2007: ch.6) investigates this procedure for the case of lexical versus grammatical stress in Greek (see §8.1 above). She shows that there is an automatic acquisition bias towards creating one Underlying Form for each morpheme, without the need for intelligent repair mechanisms like ‘surgery’ (Tesar, Alderete, Horwood, Merchant, Nishitani and Prince, 2003).

## 10. Discussion

### 10.1. The larger picture: whole-language simulations

In the above comprehension acquisition models I have been simplifying severely by considering no more than three representations at a time. A more realistic model of phonological-phonetic comprehension will include at least the quadruplet

Auditory Form – Surface Form – Underlying Form – Morphemes

where the Auditory Form and the Morphemes are known, but the Surface Form and the Underlying Form have to be *constructed* (in both senses of the word, i.e. gradually by the learner and on the fly by the listener). The only example in the present paper are tableaux (25), (26), and (27), where the general lexical constraint \*<>|x| appears. Because these tableaux have the Auditory Form as their input, they involve all four representations; however, in a full model of the whole language the substantive lexical constraints of §9.1 have to be included.

In production, the fifth representation, namely Articulatory Form, has to come in as well. With these five we may start to be capable of doing **whole-language simulations**, i.e. computer simulations of the acquisition process that use as much realistic data from the language as possible (as well as transmission noise) and thereby derive the complete phonological-phonetic system of that language. Repeating the process for several generations should generate predictions about the stability and evolution of the sound system.

### 10.2. The assumptions: naïve bidirectionality and multi-level parallelism

As stated in the introduction, the kind of bidirectionality defended in this paper is the ‘naïve’ kind in which both listening and speaking are performed by unidirectional evaluation, following Smolensky (1996). Several later proposals in the literature propose instead that the speaker’s production process explicitly takes into account the listener’s comprehension process (Boersma, 1998; Jäger, 2003), or the listener’s

comprehension process explicitly takes into account the speaker's production process (Lieberman and Mattingly, 1985; Hale and Reiss, 1998), or both (Blutner, 2000). In OT, unidirectional evaluation seems to be less complex than bidirectional evaluation, because unidirectional evaluation involves just one long list of candidates whereas bidirectional evaluation of e.g. production would, if written out, involve a long list of candidates each of which in itself contains a long list of candidates for the reverse direction of processing (e.g. comprehension); it is no wonder that every published example of bidirectional evaluation works with a rather short candidate list for at least one of the two directions.

I must admit here that the difference between the two approaches may be smaller than I just described: the parallel multi-level evaluations described in this paper come, if written out, with a number of candidates that is typically exponential in the number of levels of representation. In the end it is an empirical question which of the bidirectional models, if any, reflects human language processing realistically. As for the bidirectional use of parallel multi-level unidirectional evaluation (with the same constraints and rankings) described in this paper, we can only say that it correctly predicts the prototype effect (§3.4–5), auditory dispersion (§4.2), the Ganong effect (§5.2), the frequency-dependence of phonological activity (§6.2), and licensing by cue (§6.2); it remains to be seen how or whether other bidirectional methods could account for these observed phenomena.

## References

- Apoussidou, D. (2007). *The Learnability of Metrical Phonology*. PhD thesis, University of Amsterdam.
- Apoussidou, D. and Boersma, P. (2003). The learnability of Latin stress. *Proceedings of the Institute of Phonetic Sciences Amsterdam*, 25:101–148.
- Apoussidou, D. and Boersma, P. (2004). Comparing two Optimality-Theoretic learning algorithms for Latin stress. *WCCFL* 23:29–42.
- Beckman, J. N. (1998). *Positional Faithfulness*. PhD thesis, University of Massachusetts, Amherst.
- Blutner, R. (2000). Some aspects of optimality in natural language interpretation. *Journal of Semantics*, 17:189–216.
- Boersma, P. (1997). How we learn variation, optionality, and probability. *Proceedings of the Institute of Phonetic Sciences Amsterdam*, 21:43–58.
- Boersma, P. (1998). *Functional Phonology: Formalizing the interactions between articulatory and perceptual drives*. PhD thesis, University of Amsterdam.
- Boersma, P. (2000). The OCP in the perception grammar. *Rutgers Optimality Archive* 435.
- Boersma, P. (2001). Phonology-semantics interaction in OT, and its acquisition. In Kirchner, R., Wikeley, W., and Pater, J., editors, *Papers in Experimental and Theoretical Linguistics*. Vol. 6, pages 24–35. University of Alberta, Edmonton.
- Boersma, P. (2003). Review of Tesar & Smolensky (2000): *Learnability in Optimality Theory*. *Phonology* 20:436–446.
- Boersma, P. (2006a). Prototypicality judgments as inverted perception. In Fanselow, G., Féry, C., Schlewsky, M., and Vogel, R., editors, *Gradedness in Grammar*, pages 167–184. Oxford University Press, Oxford.
- Boersma, P. (2006b). A constraint-based explanation of the McGurk effect. *Rutgers Optimality Archive* 869.
- Boersma, P. (2007). Some listener-oriented accounts of *h-aspiré* in French. *Lingua* 117: 1989–2054.
- Boersma, P. (2008). Emergent ranking of faithfulness explains markedness and licensing by cue. *Rutgers Optimality Archive* 954.
- Boersma, P. (2009a). Cue constraints and their interactions in phonological perception and production. In Boersma, P. and Hamann, S., editors, *Phonology in Perception*, pages 55–110. Mouton De Gruyter, Berlin.
- Boersma, P. (2009b). Unidirectional optimization of comprehension can achieve bidirectional optimality. Talk presented at 10th Szklarska Poręba Workshop on the Roots of Pragmasemantics, Szklarska Poręba, March 13, 2009.

- Boersma, P. and Escudero, P. (2008). Learning to perceive a smaller L2 vowel inventory: an Optimality Theory account. In Avery, P., Dresher, E., and Rice, K., editors, *Contrast in Phonology: Theory, Perception, Acquisition*, pages 271–301. Mouton De Gruyter, Berlin.
- Boersma, P. and Hamann, S. (2008). The evolution of auditory dispersion in bidirectional constraint grammars. *Phonology*, 25:217–270.
- Boersma, P. and Hayes, B. (2001). Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry*, 32:45–86.
- Boersma, P. and Pater, J. (2008). Convergence properties of a gradual learner in Harmonic Grammar. *Rutgers Optimality Archive* 970.
- Crowther, C. S. and Mann, V. (1994). Use of vocalic cues to consonant voicing and native language background: the influence of experimental design. *Perception and Psychophysics*, 55:513–525.
- Eisner, F. (2006). *Lexically-Guided Perceptual Learning in Speech Processing*. PhD thesis, Nijmegen University.
- Escudero, P. (2005). *The Attainment of Optimal Perception in Second-Language Acquisition*. PhD thesis, Utrecht University.
- Escudero, P. and Boersma, P. (2003). Modelling the perceptual development of phonological contrasts with Optimality Theory and the Gradual Learning Algorithm. In Arunachalam, S., Kaiser, E., and Williams, A., editors, *Proceedings of the 25th Annual Penn Linguistics Colloquium. Penn Working Papers in Linguistics* 8.1:71–85.
- Escudero, P. and Boersma, P. (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition*, 26:551–585.
- Flemming, E. (1995). *Auditory Representations in Phonology*. PhD thesis, UCLA.
- Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics*, 14:3–28.
- Ganong, W. F. III (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, 6:110–125.
- Gerrits, E. (2001). *The Categorisation of Speech Sounds by Adults and Children*. PhD thesis, Utrecht University.
- Hale, M. and Reiss, C. (1998). Formal and empirical arguments concerning phonological acquisition. *Linguistic Inquiry*, 29:656–683.
- House, A.S. and Fairbanks, G. (1953). The influence of consonant environment upon the secondary acoustical characteristics of vowels. *Journal of the Acoustical Society of America*, 25:105–113.
- Jäger, G. (2003). Learning constraint sub-hierarchies: the Bidirectional Gradual Learning Algorithm. In Zeevat, H. and Blutner, R., editors, *Optimality Theory and Pragmatics*, pages 251–287. Palgrave Macmillan, Basingstoke.
- Johnson, K., Flemming, E., and Wright, R. (1993). The hyperspace effect: Phonetic targets are hyperarticulated. *Language*, 69:505–528.
- Kirchner, R. (1998). *Lenition in Phonetically-Based Optimality Theory*. PhD thesis, UCLA.
- Liberman, A. and Mattingly, I. (1985). The motor theory of speech perception revised. *Cognition*, 21:1–36.
- Lindblom, B., Lubker, J., and Gay, T. (1979). Formant frequencies of some fixed-mandible vowels and a model of speech motor programming by predictive simulation. *Journal of Phonetics*, 7:147–161.
- Mattausch, J. (2004). *On the Optimization and Grammaticalization of Anaphora*. PhD thesis, Humboldt Universität, Berlin.
- McCarthy, J. J. and Prince, A. (1995). Faithfulness and reduplicative identity. In Beckman, J., Walsh Dickey, L., and Urbanczyk, S., editors, *Papers in Optimality Theory*. University of Massachusetts Occasional Papers 18:249–384. Graduate Linguistic Student Association, Amherst, MA.
- McClelland, J. L., and Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18:1–86.
- McGurk, H. and MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264:746–748.
- McQueen, J. M. and Cutler, A. (1997). Cognitive processes in speech perception. In Hardcastle, W. J. and Laver, J., editors, *The Handbook of Phonetic Sciences*, pages 566–585. Blackwell, Oxford.
- Norris, D., McQueen, J. M., and Cutler, A. (2000). Merging information in speech recognition: feedback is never necessary. *Behavioral and Brain Sciences*, 23:299–370.
- Ohala, J. J. (1981). The listener as a source of sound change. *CLS*, 17:178–203.
- Pater, J. (2004). Bridging the gap between receptive and productive development with minimally violable constraints. In Kager, R., Pater, J., and Zonneveld, W., editors, *Constraints in Phonological Acquisition*, pages 219–244. Cambridge University Press, Cambridge.

- Polivanov, E. D. (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*, pages 223–237. Mouton, The Hague: Mouton]
- Prince, A. and Smolensky, P. (1993). Optimality Theory: Constraint interaction in generative grammar. Technical Report 2, Rutgers University Center for Cognitive Science.
- Saussure, F. de (1916). *Cours de linguistique générale*. Edited by Charles Bally & Albert Sechehaye in collaboration with Albert Riedlinger. Paris: Payot & C<sup>ie</sup>.
- Smolensky, P. (1996). On the comprehension/production dilemma in child language. *Linguistic Inquiry*, 27:720–731.
- Steriade, D. (1995). Positional neutralization. Two chapters of an unfinished manuscript, Department of Linguistics, UCLA.
- Steriade, D. (2001). Directional asymmetries in place assimilation. In Hume, E. and Johnson, K., editors, *The Role of Speech Perception in Phonology*, pages 219–250. Academic Press, San Diego.
- Tesar, B. (1997). An iterative strategy for learning metrical stress in Optimality Theory. In Hughes, E., Hughes, M., and Greenhill, A., editors, *Proceedings of the 21st Annual Boston University Conference on Language Development*, pages 615–626. Cascadilla, Somerville, MA.
- Tesar, B. (1998). An iterative strategy for language learning. *Lingua*, 104:131–145.
- Tesar, B. (1999). Robust interpretive parsing in metrical stress theory. *WCCFL* 17:625–639.
- Tesar, B., Alderete, J., Horwood, G., Merchant, N., Nishitani, K., and Prince, A. (2003). Surgery in language learning. *WCCFL* 22:477–490.
- Tesar, B. and Smolensky, P. (1998). Learnability in Optimality Theory. *Linguistic Inquiry*, 29:229–268.
- Tesar, B. and Smolensky, P. (2000). *Learnability in Optimality Theory*. The MIT Press, Cambridge, MA.