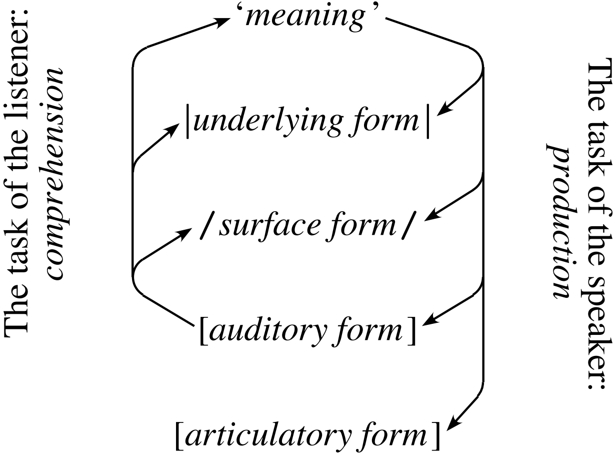
# The emergence of ranking by cue

Paul Boersma
Workshop on Experimental Approaches to OT
Ann Arbor, May 19, 2007

# Phonology and phonetics separate but connected



The task of the listener: *comprehension*

‘*meaning*’

|*underlying form*|

/*surface form*/

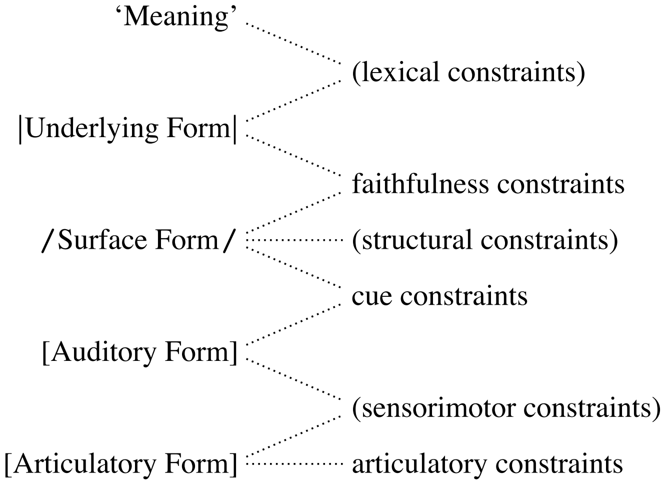
[*auditory form*]

[*articulatory form*]

The task of the speaker: *production*

(Boersma 1998, 2005; Apoussidou 2006)

# Global evaluation, but local connections



‘Meaning’

(lexical constraints)

|Underlying Form|

faithfulness constraints

/Surface Form/

(structural constraints)

cue constraints

[Auditory Form]

(sensorimotor constraints)

[Articulatory Form]

articulatory constraints

# Bidirectionality of constraint rankings

**Bidirectionality of constraints and their rankings** means the use of the same constraints, with the same rankings, by the listener as well as the speaker.

**Faithfulness constraints** were used bidirectionally by Smolensky (1996).

**Structural constraints** were used bidirectionally by Tesar (1996), Tesar & Smolensky (2000), Boersma (2000), Pater (2004), Moreton (today).

**Cue constraints** were used bidirectionally by Boersma (1998, 2005) and Boersma & Hamann (2007). An example now follows of their use in (1) comprehension and (2) production...

# Case: bidirectional cues in English

In English, there are at least two auditory cues to the voicing or voicelessness of a final obstruent: the presence or absence of periodicity (as in most languages), and the lengthening or shortening of the preceding vowel (this is specific to English). We can translate this into four cue constraints (“*” stands for “don’t have this”):

*/+voi/[–periodicity]

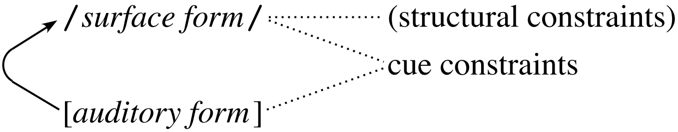
*/–voi/[+periodicity]

*/obs, +voi/[–lengthened vowel]

*/obs, –voi/[+lengthened vowel]

# Cue constraints in perception

With ‘perception’ I mean the first step in comprehension, i.e. the mapping from an auditory form to the phonological surface form:



/*surface form*/ (structural constraints)

cue constraints

[*auditory form*]

(some psycholinguists call this ‘prelexical perception’ because it does not involve the lexicon)

# English cue constraints in perception: agreement

Most often, the relevant cues agree, so that perception works well:

| [niːːd] | */obs, –voi/ [+lengthened vowel] | */–voi/ [+periodicity] |
|---|---|---|
| /.nit./ | *! | * |
| ☞ /.nid./ | | |

| [niːt] | */obs, +voi/ [–lengthened vowel] | */+voi/ [–periodicity] |
|---|---|---|
| ☞ /.nit./ | | |
| /.nid./ | *! | * |

## English cue constraints in perception: conflict

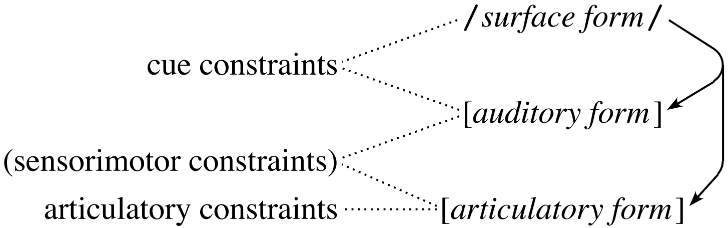
But sometimes the cues disagree. Perception experiments have shown that in that case, the vowel lengthening constraint outranks the direct periodicity cue:

| [niːːt] | */obs,–voi/<br>[+lengthened vowel] | */+voi/<br>[–periodicity] |
|---|---|---|
| /.nit./ | *! | |
| ☞ /.nid./ | | * |

# Cue constraints in production

The listener is also a speaker.

As a speaker, she uses the same cue constraints, with the same rankings, in phonetic implementation:



cue constraints

(sensorimotor constraints)

articulatory constraints

/*surface form*/

[*auditory form*]

[*articulatory form*]

# English cue constraints in production: agreement

The reuse of comprehension-based cue constraints in production is why speakers of English will try to have both cues right.

Here is an (incomplete) phonetic implementation tableau:

| /.nid./ | */obs,+voi/ [−lengthened vowel] | */+voi/ [−periodicity] |
|---|---|---|
| [niːt] | *! | * |
| [niːːt] | | *! |
| [niːd] | *! | |
| ☞ [niːːd] | | |

## Articulatory constraints in production

But phonetic implementation is not just about rendering cues. It is also about doing so efficiently, i.e. with the minimum expenditure of articulatory effort. So we need articulatory constraints to evaluate the articulatory-phonetic form.

In the case at hand, we observe that it is especially difficult to pronounce periodicity in a final plosive. I express this simply as:

*[+periodicity, final plosive]

In a complete phonetic implementation tableau, this constraint must interact with cue constraints.

# English interaction of articulatory and cue constraints

If the articulatory constraint outranks the lower-ranked cue constraint, speakers will implement only the most important cue:

| /.nid./ | */obs,+voi/ [–lengthened vowel] | *[+periodicity, final plosive] | */+voi/ [–periodicity] |
|---|---|---|---|
| [niːt] | *! | | * |
| ☞ [niːːt] | | | * |
| [niːd] | *! | * | |
| [niːːd] | | *! | |

As we saw on slide 8, listeners still perceive this [niːːt] as the intended /.nid./. This means speakers easily get away with saying [niːːt]. And so they often say [niːːt].

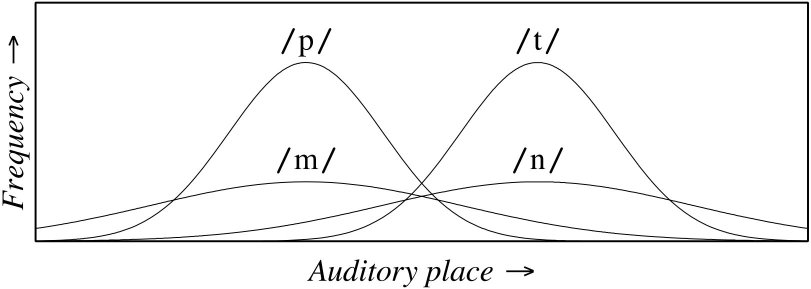
# Case: nasal place assimilation

I will illustrate my point with the example of place assimilation of nasals, i.e. the case that an underlying |n| can be realized as /m/ but an underlying |t| cannot be realized as /p/. This happens in Malayalam, for instance (Mohanan 1981).

Ohala (1990) ascribes this difference to a difference in the availability of place cues between nasals and plosives. Auditory place is influenced by *transmission noise* (the speaker’s muscle noise, the non-normalizable part of between-speaker variation, acoustic background noise, blood running through the listener’s ear...). The transmission noise is more likely to obscure the weaker place cues of the nasals than the stronger place cues of the plosives.

## The auditory language environment

The listener hears both nasals and plosives with continuous place values along an auditory continuum, perhaps according to Gaussian distributions. The distributions of intended surface nasals will be wider than those of intended surface plosives:



# What are our cue constraints?

In reality there may be 30,000 possible auditory place values. In my computer simulations, however, I simplifyingly assume that there are only four possible auditory place values. For nasals they are [m], [M], [N], and [n], where [M] and [N] are auditorily intermediate between [m] and [n].

So we have 16 thinkable combinations of the eight possible auditory forms and the two possible place feature values:

| | | | |
|---|---|---|---|
| */n/[n] | */n/[N] | */n/[M] | */n/[m] |
| */m/[n] | */m/[N] | */m/[M] | */m/[m] |
| */t/[t] | */t/[T] | */t/[P] | */t/[p] |
| */p/[t] | */p/[T] | */p/[P] | */p/[p] |

# What are our faithfulness constraints?

The underlying form and the phonological surface form are made up of the same kind of material: discrete phonological elements such as segments, features, syllables, and feet. Faithfulness constraints aim at making the two forms identical.

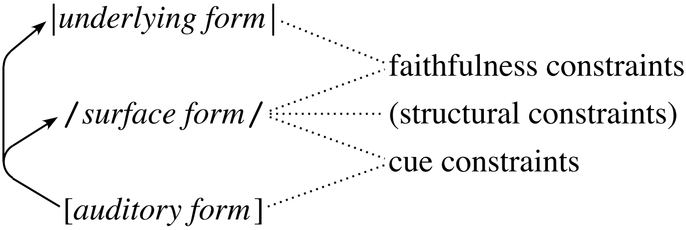
For the case of place assimilation, the relevant ones are
IDENTPLACE: IDPL(|t|) IDPL(|p|) IDPL(|n|) IDPL(|m|)

# What are our articulatory constraints?

Following Kirchner (1998), I use today an oversimplified single constraint LAZY which is violated maximally in an articulatory [np] (or [mt] or [tm] or [pn]), somewhat less in [Np] (or [Mt]...), somewhat less again in [Mp], and not at all in [mp].

# Learning model: comprehension first

The first task of the learner is to acquire a good mapping from auditory form to underlying form. I assume for simplicity that the learner is provided with pairs of auditory and underlying forms (i.e. the lexicon has been built up correctly), and that she has to *invent* the hidden phonological surface form.



|*underlying form*|

faithfulness constraints

/*surface form*/

(structural constraints)

cue constraints

[*auditory form*]

# Learning ingredient 1: real parallel comprehension

As said before, I assume that the child is given pairs of underlying form and auditory form. A possible input pair is therefore |an+pa|[aMpa]. For instance, the following tableau may occur at some point during acquisition:

<table>
<tr><th>|an+pa|<br>[aMpa]</th><th>IDPL (|n|)</th><th>*/n/ [M]</th><th>(LAZY)</th><th>*/m/ [M]</th></tr>
<tr><td>√ |an+pa|/.an.pa./[aMpa]</td><td></td><td>*</td><td>(*)</td><td></td></tr>
<tr><td>|an+pa|/.am.pa./[aMpa]</td><td>*!</td><td></td><td>(*)</td><td>*</td></tr>
</table>

The form marked “√” is the one that is the most harmonic of all triplets that have both |an+pa| and [aMpa]. The child will consider this the “correct” form.

## Learning ingredient 2: virtual parallel production

After having established a “correct” triplet, the child computes the triplet that she herself would have produced, given the same underlying form |an+pa|.

| \|an+pa\| | IDPL (\|n\|) | */n/ [M] | LAZY | */m/ [M] |
|---|---|---|---|---|
| √ \|an+pa\|/.an.pa./[aMpa] | | *! | * | |
| \|an+pa\|/.am.pa./[aMpa] | *! | | * | * |
| ☞ \|an+pa\|/.an.pa./[anpa] | | | *** | |

The “correct” triplet is also included in this tableau. The form marked “☞” (the winner) is different from the form marked “√”, so that *learning* will occur.

# Learning ingredient 3: parallel error-driven reranking

| \|an+pa\|<br>[aMpa] | IDPL<br>(\|n\|) | */n/<br>[M] | LAZY | */m/<br>[M] |
|---|---|---|---|---|
| √ \|an+pa\|/.an.pa./[aMpa] |  | *→ | * |  |
| \|an+pa\|/.am.pa./[aMpa] | * |  | * | * |
| ☞ \|an+pa\|/.an.pa./[anpa] |  |  | ←*** |  |

All the constraints that prefer the form marked “☞” will move by a small step down the constraint hierarchy (“→”), and all constraints that prefer the “correct” form (“√”) will move up (“←”). This will make it more likely that in a future occurrence of |an+pa|[aMpa] the “☞” and “√” forms will agree.

# Computer simulation of learning

A virtual child grows up in an environment where people do not assimilate their nasals at all (*articulatorily*). The transmission noise leads to the following underlying/*auditory* distribution:

<table>
<tr><th><i>form</i></th><th><i>frequency</i></th><th><i>form</i></th><th><i>frequency</i></th></tr>
<tr><td>|an+pa|[anpa]</td><td>36</td><td>|at+ma|[atma]</td><td>57</td></tr>
<tr><td>|an+pa|[aNpa]</td><td>31</td><td>|at+ma|[aTma]</td><td>34</td></tr>
<tr><td>|an+pa|[aMpa]</td><td>22</td><td>|at+ma|[aPma]</td><td>8</td></tr>
<tr><td>|an+pa|[ampa]</td><td>11</td><td>|at+ma|[apma]</td><td>1</td></tr>
<tr><td>|am+ta|[amta]</td><td>36</td><td>|ap+na|[apna]</td><td>57</td></tr>
<tr><td>|am+ta|[aMta]</td><td>31</td><td>|ap+na|[aPna]</td><td>34</td></tr>
<tr><td>|am+ta|[aNta]</td><td>22</td><td>|ap+na|[aTna]</td><td>8</td></tr>
<tr><td>|am+ta|[anta]</td><td>11</td><td>|ap+na|[atna]</td><td>1</td></tr>
</table>

Thus, the transmission noise leads to more auditory assimilation for nasals than for plosives.

In order for the learning algorithm to succeed, it requires some sanity data:

| *form* | *frequency* | *form* | *frequency* |
|---|---|---|---|
| \|in+ti\|[inti] | 100 | \|it+ni\|[itni] | 100 |
| \|im+pi\|[impi] | 100 | \|ip+mi\|[ipmi] | 100 |

**Initial stage:** the simulated learner (virtual baby) starts with all 21 constraints ranked at the same height.

**Acquisition:** the learner is subsequently fed a million input pairs randomly drawn from the above input distribution. She learns with the learning algorithm of slide 20.

# Result: a cue-based ranking

I taught this to not 1 but 1,000 virtual babys.

Nearly all virtual children end up ranking their **cue constraints** by cue reliability, e.g. 993 out of 1000 have:

*/t/[P] above */n/[M]

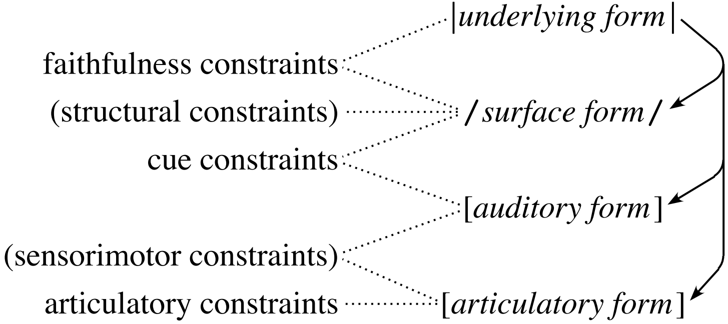
Nearly all virtual children end up ranking their **faithfulness constraints** by cue reliability, e.g. 998 out of 1000 have:

IDPL(|t|) above IDPL(|n|)

The page opposite shows a tableau with the median final rankings of all 1,000 learners, for every constraint.

# Bidirectional consequence

The child will use in production the same faithfulness constraints, with the ranking that she has learned in comprehension. They can interact with articulatory constraints, because phonological and phonetic production are handled in parallel:



faithfulness constraints
(structural constraints)
cue constraints
(sensorimotor constraints)
articulatory constraints

|*underlying form*|
/*surface form*/
[*auditory form*]
[*articulatory form*]

# Bidirectional consequence

Given the cue-based ranking IDPL(|t|) >> IDPL(|n|), the learner has become more likely to assimilate a nasal than to assimilate a plosive:

| \|an+pa\| | IDPL(\|t\|) | LAZY | IDPL(\|n\|) |
|---|---|---|---|
| /.an.pa./[anpa] |  | *! |  |
| ☞ /.am.pa./[ampa] |  |  | * |

| \|at+ma\| | IDPL(\|t\|) | LAZY | IDPL(\|n\|) |
|---|---|---|---|
| ☞ /.at.ma./[atma] |  | * |  |
| /.ap.ma./[apma] | *! |  |  |

# Summary

Even if her parents do not have a preference for assimilating nasals over plosives, the child will still develop such a preference as an automatic result of the differential transmission noise and the learning algorithm.

The cause of this is simply that the child uses the same constraint ranking in comprehension and production.

The concept of extralinguistic devices such as Steriade's (2001) "P-map" is superfluous for explaining the greater phonological strength of |t| as compared to |n|.

# Discussion

The learning algorithm does not always work in the same way: twenty-five of the 1,000 children end up with very low nasal faithfulness constraints; they have full neutralization.

The convergence of the learner is not robust: e.g., replacing IdPl(|t|) and IdPl(|p|) with IdPl(|*plosive*|) causes the learner to move all faithfulness constraints to the bottom of the hierarchy.

Harmonic Grammar (additive constraint weights) with additive evaluation noise and a gradient ascent learning algorithm works as well as Stochastic Optimality Theory with the Gradual Learning Algorithm.

# Conclusion

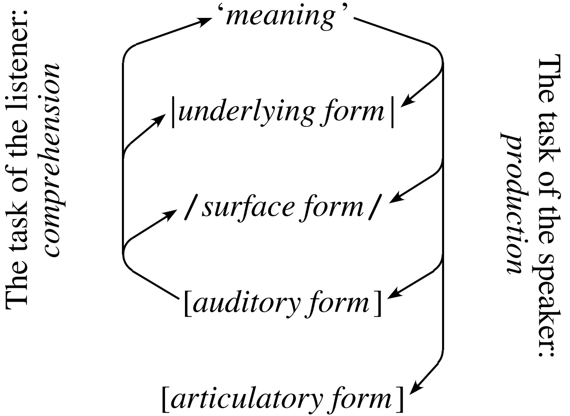
Some phenomena in phonology **emerge** as the result of an automatic acquisition bias.

If we can identify **emergent** phenomena, phonological theory will need **less innatism** and **less synchronic functionalism**:

(1) less innatism, because we need no innate ranking of IDPL(|t|) over IDPL(|n|) or of *|t|/P/ over *|n|/M/.

(2) less synchronic functionalism, because we need not assume that the speaker explicitly takes into account the biases of the listener, neither for cue reliability (Steriade 1995/2001, Boersma 1998), nor for frequency (Boersma 1998, Hume 2004).

# The processing model



The task of the listener:
*comprehension*

‘*meaning*’

|*underlying form*|

/*surface form*/

[*auditory form*]

[*articulatory form*]

The task of the speaker:
*production*