Inducing Functionally Grounded Constraints*

Kathryn Flack

University of Massachusetts Amherst

1. Introduction

Phonologists have been long concerned with finding phonetic properties which allow phonological patterns to be seen as 'natural' or 'grounded' (e.g. Stampe (1973), Hooper [Bybee] (1976), Ohala (1990), Archangeli and Pulleyblank (1994)). With the advent of Optimality Theory(Prince and Smolensky, 1993/2004), this often takes the form of identifying functional grounding for specific OT constraints (e.g. Hayes (1999), Smith (2002), Steriade (1999; 2001), and papers in Hayes et al. (eds.) (2004)).

There is general agreement that 'functionally grounded' constraints prefer forms which are more perceptually or psycholinguistically salient, or less articulatorily challenging, to less prominent or more difficult forms. Beyond this, however, there is very little agreement about the nature of the connection between phonetic facts and constraints. Most work is agnostic on this matter, enthusiastically finding phonetic facts which correlate with constraint activity while remaining uncommitted to a particular relationship between the two. Prince and Smolensky (2004) originally proposed that all constraints in the universal constraint inventory are innate. If this is true, any functional factors which shape the constraint inventory must have acted at an earlier stage of evolution. Alternatively, Hayes (1999), Smith (2002), and Steriade (1999; 2001), among others, discuss various means by which individual learners could induce functionally grounded constraints directly from their linguistic experience.

This paper takes the position that functional grounding shapes individual learners' constraint inventories: each learner induces functionally grounded constraints based on their immediate linguistic experience. An argument for this view comes from cognitive economy. Assume for the moment that learners have consistent access to phonetic data

* Thanks to Michael Becker, Tim Beechey, Gaja Jarosz, Shigeto Kawahara, John Kingston, Dan Mash, Andrew McCallum, John McCarthy, Marianne McKenzie, Joe Pater, Jason Riggle, Nathan Sanders, Matt Wolf, and the UMass Phonology Group for lots of helpful discussions and suggestions. This paper is a revised version of chapter four of Flack (2007).

demonstrating the relative perceptual salience, articulatory difficulty, and so on of segments (or features) in particular phonetic and phonological contexts. Further assume that learners have some reliable mechanism for evaluating this experience, allowing them to induce constraints which are grounded in these functional factors.

The consistent availability of this information to learners would make innate specifications of these constraints redundant. Assuming that innate mechanisms for language acquisition should be maximally simple, learners should use as much external information as possible. Innate specifications should therefore only be posited when the information in learners' experience is insufficient to the learning task.

Ultimately, the question of whether (and how) constraints are induced from phonetic data is an empirical one. As described above, induction is only possible if learners can observe phonetic facts which could motivate particular constraints, and if there is some mechanism which could induce the attested set of constraints from this data. Much of the current work on functional grounding addresses the first point. Smith (2002) and Steriade (1999; 2001) propose mechanisms for inducing constraints from this sort of data, and Hayes' (1999) Inductive Grounding model integrates articulatory facts with such a constraint inducer. Of course, demonstrating that constraint induction is logically possible is not the same as showing that learners actually induce constraints. Any proposed induction mechanism is at best a hypothesis about learners' behavior, and must be tested against actual speakers.

The goal of this paper is to propose a mechanism for the induction of perceptually grounded constraints, and to show how a computational model can be used to test proposals about constraint induction against real phonetic data. The empirical focus of this paper is the constraint *#P, 'No word-initial unaspirated *p*', which is argued to be functionally grounded in learners' perceptual experience. This constraint is active in languages including Cajonos Zapotec (Nellis and Hollenbach, 1980) and Ibibio (Akinlabi and Urua, 2002; Connell, 1994; Essien, 1990), as described in section 2. Experimental data reported in section 3 show that initial *p* is uniquely perceptually difficult as a result of its acoustic similarity to initial *b*, suggesting that *#P is phonetically natural. The latter portion of this paper describes a computational model in which virtual learners induce this constraint from precisely these phonetic facts. This model achieves realistic perception of acoustically realistic segments; the relative perceptibility of these segments then forms the basis for perceptually grounded constraint induction. The 'production' and 'perception' components of the model are described in section 5, and these form the basis for a mechanism which allows learners to consistently induce *#P from these functional factors, described in section 6.

2. Phonological restrictions on initial *p*

In Cajonos Zapotec (Nellis and Hollenbach, 1980) and Ibibio (Akinlabi and Urua, 2002; Connell, 1994; Essien, 1990), unaspirated *p* contrasts with *b* in non-initial positions, but only *b* may surface initially. These languages allow other voicing contrasts in initial position (e.g. *t* and *d*, *k* and *g*; *t* and *k* are also unaspirated).

In Cajonos Zapotec, coronal and velar stops contrast for voicing initially, medially, and finally. Labials contrast for voicing only medially and finally. All labial-initial native words begin with *b*, rather than *p*. This restriction was productive until recently. Older Spanish loans borrowed initial /p/ as [b], as in *béj* 'sash' (Sp. *paño*) and *bêd* (Sp. *Pedro*). Newer loans faithfully retain initial /p/, as in *pát* 'duck' (Sp. *pato*).

(1) Cajonos Zapotec

*pèn		tò	'one'	kóc	'pig'
bèn	'do!'	dò	'string'	góc	'gunny sack'
gòpée	'fog'	yítàʔ	'the squash'	wáké	'it can'
dòbée	'feather'	yîdàʔ	'the leather'	wágé	'firewood'
jáp	'will care for'	yèt	'tortilla'	wák	'it can'
jáb	'will weave'	zèd	'disturbance'	wág	'firewood'

A similar restriction against word-initial *p* is found in Ibibio (data below from Essien (1990)), though the distribution of voicing and length contrasts in Ibibio stops is more complex. Intervocalic stops are typically geminates, as intervocalic singletons are typically lenited. Ibibio has no voiced velar stop, and coronals are devoiced syllable-finally and in geminates. See Essien (1990) and especially Akinlabi and Urua (2002) for further discussion of Ibibio morphophonology.

Most interestingly for the present discussion, Ibibio licenses *b* but not *p* word-initially. Medial *p* and *b* contrast in *dɨ̂ppé* 'lift up' versus *dɨ̂bbé* 'hide oneself', and finally in *bɔ́p* 'build (something)' versus *bɔ́ɔb* 'build many things'. While there are *b*-initial words like *bàt* 'count', there are no *p*-initial words (**pàt*). Unlike labials, coronals contrast initially as in *tàppa* 'call someone's attention' and *dàppa* 'remove something from a fire'.

(2) Ibibio

*pàt		tàppa	'call someone's attention'	kárá	'govern'
bàt	'count'	dàppa	'remove s.t. from a fire'	*gárá	
dɨ̂ppé	'lift up'	sɨtté	'uncork'	dàkká	'move away'
dɨ̂bbé	'hide oneself'	*sɨddé		*dàggá	
bɔ́p	'build (something)'	wèt	'write'	sák	'laugh'
bɔ́ɔb	'build many things'	*wèd		*ság	

The remainder of this paper will investigate the functional grounding of this restriction on word-initial *p* (*#P). Section 3 will show that in French, where initial unaspirated *p* is attested, this segment is particularly difficult for listeners to identify. Sections **Error! Reference source not found.**–6 will then propose an induction mechanism which allows learners of both French-type languages (with initial *p*) and also Cajonos Zapotec-type languages (without initial *p*) to induce *#P from the acoustic and perceptual properties of initial stops.

3. Phonetic properties of initial *p*

Initial *p* is uniquely difficult to identify, as it is more similar to initial *b* than other voiceless stops are to their voiced counterparts. These facts demonstrate that *#P is phonetically natural. This section describes perceptual and acoustic experiments which identify these properties of initial and intervocalic *p*, *b*, *t*, and *d*.

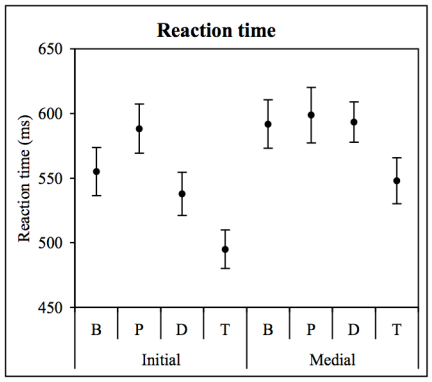
3.1 Perceptual experiment

First, participants in the perceptual experiment heard stimuli containing a target stop (*p*, *b*, *t*, or *d*) and flanking vowels; these stimuli were extracted from French words produced by a native speaker of Parisian French. Stimuli contained stops in either word-initial or intervocalic position, e.g. [#pa] from *paragraphe*, [#bɔ] from *bordeaux*; [edi] from *comedie*, [iti] from *itineraire*. 15 native speakers of French participated in the experiment. Each heard a total of 45 unique stimuli in each of 8 conditions: 4 stops (*p*, *b*, *t*, *d*) × 2 positions (initial, medial). The task was to identify stops by pressing a button as quickly as possible after each stimulus. See Flack (2007: ch. 3) for more about the experimental methods.

The hypothesis investigated in this experiment has three parts: (1) initial *p* is more difficult to accurately identify than initial *b*, (2) no similar asymmetry emerges word-medially, and (3) this is a particular property of initial *p* rather than a general property of initial voiceless stops (that is, initial *t* is not similarly more difficult than *d*). This perceptual difficulty could reveal itself either in inaccurate identification of initial *p* or slow reaction times to initial *p* (Pisoni and Tash, 1974). Thus recognition of initial *p* should be slower and/or less accurate than initial *b*, and neither initial *t* and *d* nor medial *p* and *b* should differ in the same way. To test this hypothesis, preplanned two-sample one-tailed t-tests were performed on reaction times and accuracy scores.

Participants identified initial *p* significantly more slowly than initial *b*, as shown in Figure 1a. The average reaction time for accurate initial *p* responses was 588 ms. This is marginally significantly greater than the average response time for initial *b* responses (555 ms; $t(88) = 2.445$, $p = 0.016$).[1] Reaction times for medial *p* (599 ms) are not significantly different from reaction times for medial *b* (592 ms; $t(82) = 0.485$, $p = 0.629$), so *p* is more slowly recognized than *b* only in initial position. Turning to the coronal stops, response times for initial *t* (495 ms) were quicker than those for initial *d* (538 ms; $t(85) = 3.742$, $p < 0.001$), indicating that slow responses are not a general property of initial voiceless stops but rather a unique property of initial *p*. Participants' slow reaction times do not correlate with any difference in the accuracy of responses to initial *p* versus *b*. 93% of initial *p* stimuli were identified correctly, and 94% of initial *b* stimuli were identified correctly, as shown in Figure 1b. This difference is not significant ($t(88) = 0.314$, $p = 0.754$).

[1] Because these comparisons cover four conditions (*p*, *b*, *t*, and *d* in a given context), a Bonferroni correction is applied to $\alpha = 0.05$ such that α here is equal to 0.05/4, 0.0125, for all t-tests. While significance is only marginal here, no other pairs of stops shows a comparable pattern with even marginal significance. Therefore, initial *p*'s perceptual difficulty is unique and reliable.

a. 

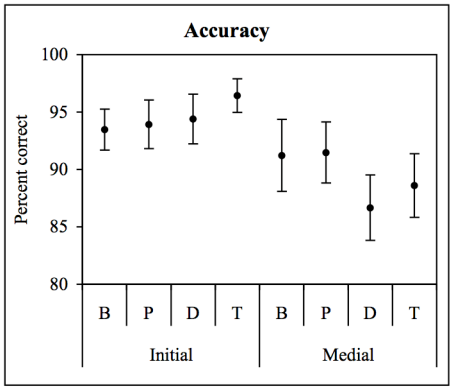
b. 

Figure 1. Average reaction times (a) and accuracy (b) in each condition, with 95% confidence intervals (from items analysis).

These results indicate that word-initial *p* is, as predicted, uniquely perceptually difficult. While *p* and *b* are recognized with equal accuracy in both initial and medial position, initial *p* is significantly slower than initial *b* while medial *p* and *b* are identified with equal speed. Responses to *t* and *d* follow a different pattern: both initial and medial *t* are recognized more quickly than *d*, and *t* responses tend to be more accurate overall than *d* responses as well. Additional analysis in Flack (2007: ch. 3) shows that these results are not artifacts of flanking vowels or segmental frequency. The typological observation that only *p* may be banned in word-initial position correlates with these perceptual facts, supporting the hypothesis that the constraint *#P is functionally grounded.

3.2 Acoustic experiment

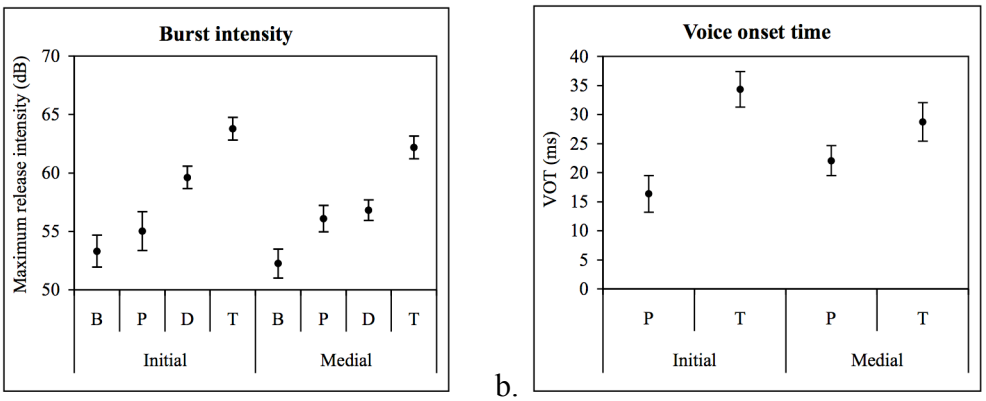
The perceptual difficulty of initial *p* correlates with its similarity to *b*. Initial *p* is acoustically more similar to *b* than initial *t* is to *d*, or medial *p* to *b* or *t* to *d*. Acoustic similarity was measured in terms of the stops' maximum release burst intensities and voice onset times (VOT), both of which are major cues to voicing (Lisker and Abramson, 1964; Repp, 1979). Burst intensity and VOT were calculated for the French stops used in the perceptual experiment. See Flack (2007: ch. 3) for more details about these analyses.

Looking first at the intensity measures, the maximum burst intensities of all homorganic stop pairs differ significantly, with the important exception of initial *p* and *b*, whose bursts are not significantly different. These results are shown in Figure 2a and Table 1. Initial *p–b* are therefore have more similar bursts than medial *p–b*, or initial *t–d*. The similarity between initial *p* and *b* is thus a specific fact about initial labials, rather than a general property of initial or voiceless stops.

Turning to the VOT measures, initial *p–b* are again more similar than medial *p–b* or initial or medial *t–d*. This follows from the fact that initial *p* has the shortest VOT of all four voiceless stops; initial *p* thus has the weakest VOT cue to voicelessness. In the weakness of this cue, initial *p* is the most like (or, least unlike) its voiced counterpart.[2]

[2] VOT was measured only for voiceless stops followed by non-high vowels. High vowels were often partially or fully devoiced after voiceless stops, giving these stops artifically long VOTs.

Initial *p*'s 16 ms VOT is significantly shorter than that of medial *p*, so initial *p–b* are more similar than medial *p–b*. Initial *p*'s VOT is also significantly shorter than initial *t*'s, so initial *p–b* are also more similar than initial *t–d*. Further, while initial labial stops are more similar than medial labial stops, the pattern is reversed for coronals: medial *t* has a significantly shorter VOT than initial *t*. The tendency for initial *p–b* to have similar VOTs compared to medial *p–b* is thus a specific property of labials, rather than a general property of all stops. These results are summarized in Figure 2b and Table 2.



a. b.

Figure 2. Average maximum release burst intensity (a) in each condition, within 5 ms of release, (a); VOT of initial and medial voiceless labial and coronal stops followed by non-high vowels (b); with 95% confidence intervals.

Maximum burst intensity							
Initial				Medial			
	Mean (dB)	Difference (dB)	*p* value		Mean (dB)	Difference (dB)	*p* value
b	53	2	0.129 $t(93) = 1.530$	*b*	52	4	<0.001 $t(93) = 4.495$
p	55			*p*	56		
d	60	4	<0.001 $t(91) = 5.937$	*d*	57	5	<0.001 $t(93) = 7.969$
t	64			*t*	62		

Table 1. Maximum release burst intensity measures for initial and medial *p*, *b*, *t*, and *d*, with differences and *p* values (from preplanned two-sample t-tests) for pairs of stops differing in voicing.

VOT							
Initial				Medial			
	Mean (ms)	Difference (ms)	*p* value		Mean (ms)	Difference (ms)	*p* value
p	16	18	<0.001 $t(64) = 7.995$	*p*	22	7	0.018 $t(57) = 2.432$
t	34			*t*	29		
Initial vs. medial *p*		6	0.008 $t(63) = 2.719$	Initial vs. medial *t*		5	0.03 $t(56) = 2.432$

Table 2. VOT measures for initial and medial voiceless stops, with differences and *p* values (from preplanned two-sample t-tests).

These acoustic results have shown that initial *p* and *b* are similar. But the perceptual study indicated that these segments are not symmetrically confusable; instead, initial *p* is uniquely perceptually difficult.[3] The source of this asymmetry must be some property of *p* itself, rather than simply the similarities between *p* and *b* that have been discussed so far. A likely acoustic source of this asymmetrical difficulty lies in *p*'s greater acoustic variability. Initial *p*'s burst values have a greater standard deviation than initial *b*'s: 6.0 and 4.7 dB, respectively.[4] Similarly, while VOT was not measured for voiced stops, initial *p*'s VOT is the most variable of all voiceless stops: its standard deviation is 9.4 ms, while that of initial *t* is 8.7 ms, medial *p* 7.1 ms, and medial *t* 8.9 ms.

Because initial *p*'s burst is more variable than initial *b*'s, more initial *p* tokens have burst intensities equal to, or even less than, the mean burst intensity for initial *b* than there are initial *b* tokens with burst intensities equal to or greater than the average for initial *p*. Put more simply, there are more *b*-like initial *p*s than there are *p*-like initial *b*s, in terms of burst intensity. Initial *p*'s VOT is also more variable than, and shorter than, that of any other voiceless stop. This suggests that there are also more tokens of initial *p* with extremely short VOTs similar to that of a voiced stop than there are of other voiceless segments: there are again more *b*-like initial *p*s than *b*-like medial *p*s, or *d*-like *t*s in any position. It could thus be the case that while the general similarity between initial *p* and *b* makes them difficult to distinguish, the large variability of initial *p* makes it more *b*-like than initial *b* is *p*-like, accounting for initial *p*'s unique perceptual difficulty. Section 5.2.2 shows how a perceptual model can explore this hypothesis.

4. Interim summary: *#P, functional grounding, and constraint induction

So far, this paper has shown that word-initial *p* is phonologically marked. Evidence of this is found in languages like Cajonos Zapotec (Nellis and Hollenbach, 1980) and Ibibio (Essien, 1990), where *p* and *b* contrast in all non-initial positions, but only *b* is licensed initially. This phonological markedness correlates with initial *p*'s acoustic and perceptual properties. French speakers find word-initial *p* uniquely perceptually difficult. Word-initial *p* and *b* are also uniquely acoustically similar. The cross-linguistic dispreference for initial *p* is likely the result of its perceptual difficulty in this position, which in turn follows from its acoustic properties. The constraint *#P therefore appears to be functionally grounded in initial *p*'s perceptual difficulty.

For *#P to be induced by learners, they must have consistent experience of this perceptual difficulty, and must also have some mechanism for reliably deriving the attested constraint from this perceptual experience. A computational model of a learner's perceptual experience can be used to evaluate whether *#P could be induced in this way.

[3] Nearly all misidentifications in the perceptual experiment were voicing mistakes. This suggests that slow reaction times were also caused by difficulty identifying stops' voicing. Initial *p* was identified more slowly than initial *b*, so *p* tokens apparently sounded more *b*-like than vice versa.

[4] The standard deviations of other consonants' bursts: initial *d* = 3.4; initial *t* = 3.4; medial *b* = 4.3; medial *p* = 4.0; medial *d* = 3.1; medial *t* = 3.5 (all dB).

The remainder of this paper describes such a model, in which realistic patterns of perception are based on realistic acoustic representations of initial and medial *p*, *b*, *t*, and *d*. When a constraint induction algorithm evaluates this perceptual experience, the constraint *#P can be consistently induced in languages where initial *p* is attested and also in languages where there is no initial *p*.

The model has three components. First, in the production component, a virtual adult speaker pronounces words with initial and medial stops whose acoustic properties are those described above. This is the input to the perception component, where a virtual learner develops acoustic criteria for identifying these stops in initial and medial position using a simple prototype model. At the end of phonetic learning, the learner's perceptual behavior is equivalent to that of participants in the perceptual experiment: initial *p* is uniquely difficult. Finally, in the induction component, the learner uses its perceptual experience to induce constraints against segments which meet 'innate' criteria for perceptual difficulty. The learner reliably induces the attested constraint *#P without inducing other, unattested constraints; this occurs whether the learner is exposed to pseudo-French, where initial *p* occurs, or pseudo-Cajonos Zapotec ('pseudo-CZ'), where it is unattested.

5. Modelling production and perception

The production component of the model represents an adult speaker's acoustically realistic productions of initial and medial stops. The output of production is the input to perception, which represents a learner who listens to adult speech, develops acoustic prototypes of segments, and learns to identify stops based on their acoustic properties. The perception component is also realistic, as it finds word-initial *p* uniquely difficult. These properties of the perception model make it a reliable foundation for the model of constraint induction discussed in section 6.

5.1 How the model works

This section will describe the structure of first the production model and then the perception model. The following section will demonstrate that the perception model faithfully represents human perception.

5.1.1 Production and acoustic representations

In each cycle of the model, the virtual speaker produces an utterance of the form *CaCa*. Each *C* is randomly chosen from the inventory *p*, *b*, *t*, *d*, and these stops are produced with realistic acoustic properties. The virtual learner's task is to learn to identify stops based on their acoustics.

Stops in the model have four acoustic properties: maximum burst intensity (*burst*), VOT, closure voicing (*voicing*), and place. The speaker randomly chooses an appropriate numeric value for each property each time a stop is produced. These sets of acoustic values constitute the spoken utterance, as shown in (3). The learner 'hears' these

sets of acoustic values and uses its developing knowledge of prototypical acoustic values for each stop to identify new stops.

(3)
```
SPOKEN: "tada!"
   Initial t:  Place = 100   Voicing = 0     VOT = 47  Burst = 61
   Medial d:   Place = 100   Voicing = 100   VOT = 18  Burst = 54
```
[5]

Each acoustic value of each stop in an utterance is randomly chosen from a normal distribution with a specified mean and variance. The parameters of these normal distributions represent stops' actual acoustic properties. Each acoustic value is an integer between 0 and 100. If a normal distribution with a specified mean and variance would allow some chance for values below 0 or above 100, those values were replaced by additional 0 or 100 values, respectively.

Burst intensity and VOT values were taken from the experimental data in section 3.2. Maximum burst intensity was measured for each stop in each position, and the means and variances of the normally distributed burst values for each stop are shown in (4).

(4)

	p		***b***		***t***		***d***	
BURST	mean	*variance*	mean	*variance*	mean	*variance*	mean	*variance*
Initial	55	*36*	53	*22*	64	*11*	60	*11*
Medial	56	*16*	52	*19*	62	*12*	57	*10*

The model's 'VOT' property reflects only the positive portion of each stop's voice onset time. This is distinct from the model's 'closure voicing' property described below. This distinction reflects speakers' tendency to process the presence vs. absence of prevoicing categorically, while fine-grained distinctions among positive VOTs are interpreted gradiently (Hay, 2005). For the voiceless stops *p* and *t*, VOT means and variances were taken directly from the experimental measures. As voiced stops do not have positive voice onset times, the VOT means for *b* and *d* were set to 0, and these stops' variances were set to the averaged variance of all voiceless stops' VOTs.[6] Stops' possible VOT values in the model are summarized in (5).

(5)

	p		***b***		***t***		***d***	
VOT	mean	*variance*	mean	*variance*	mean	*variance*	mean	*variance*
Initial	16	*89*	0	*74*	34	*76*	0	*74*
Medial	22	*51*	0	*74*	29	*80*	0	*74*

In order to model the categorical nature of closure voicing perception, the possible closure voicing values of each stop are distributed such that there is a robust binary distinction between voiced and voiceless stops. Voiceless stops have closure

[5] Throughout this chapter, `this font` will be used to show data from the model.

[6] These variances for voiced stops' VOTs are almost certainly too large. However, section 5.3 suggests that smaller, more realistic variances for these stops' VOTs would make the model behave increasingly realistically, as this change would make initial *p* even more uniquely perceptually difficult.

voicing values of essentially zero (their mean is zero, and the variance is extremely small), while voiced stops have closure voicing values of essentially 100.

(6)

	Voiceless *p,t*		**Voiced *b,d***	
	mean	*variance*	mean	*variance*
CLOSURE VOICING	0	*2*	100	*2*

Finally, the 'place' cue also produces a binary distinction between labial stops (with values of essentially 0) and coronal stops (with values of essentially 100). As this model is concerned with voicing distinctions within a single place rather than the perception of place distinctions themselves, these extremely simple values are placeholders for more realistic sets of detailed acoustic cues to place.

(7)

	Labial *p,b*		**Coronal *t,d***	
	mean	*variance*	mean	*variance*
PLACE	0	*2*	100	*2*

The output of each round of production is a single word of the form *CaCa*, where initial and medial stops are randomly chosen. Each stop's acoustic values for place, closure voicing, VOT, and burst intensity are chosen from normal distributions with the means and variances specified above. Because the acoustic properties of these 'spoken' stops are taken from acoustic measurements of naturally produced stops, the production component of the model accurately represents a learner's acoustic experience. The result of a round of production is repeated in (8).

(8)
```
SPOKEN: "tada!"
   Initial t:  Place = 100   Voicing = 0     VOT = 47  Burst = 61
   Medial d:   Place = 100   Voicing = 100   VOT = 18  Burst = 54
```

5.1.2 Perception: Hearing, phoneme identification, and category learning

The perception component of the model consists of three subcomponents. First, during hearing, the set of acoustic values produced by the virtual speaker is heard somewhat imperfectly. During identification, the learner guesses which stops were produced by comparing the heard values to prototypical values for each stop. Finally, during learning, the learner adjusts the stop prototypes to reflect the new acoustic information. The perception component of the model takes experimentally-determined acoustic properties as its input, and produces a pattern of perceptual accuracy consistent with experimental results. Crucially, the model finds word-initial *p* more perceptually difficult than initial *b*, without similarly finding medial *p* more difficult than medial *b*, or initial *t* more difficult than initial *d*; these results are presented in section 5.2.

Hearing: Not all spoken acoustic values are perceived accurately

In order to model imperfect perception (as in a noisy environment), transmission of the spoken acoustic values is imperfect in two ways: some acoustic values are not heard at

all, and those which are heard may be perturbed slightly. The spoken properties given above can thus be heard as a subset of imperfectly-transmitted values as in (9), where the learner fails to hear the medial *d*'s closure voicing and burst cues at all and its VOT property is inaccurately transmitted. This imperfect cue transmission introduces realistic randomness into the model. Its specific details are not crucial to the model, so they are not discussed here; see Flack (2007: ch. 4) for further details.

(9)
```
SPOKEN: "tada!"
   Initial t:  Place = 100   Voicing = 0     VOT = 47  Burst = 61
   Medial d:   Place = 100   Voicing = 100   VOT = 18  Burst = 54

HEARD:
   Initial:    Place = 100   Voicing = 0     VOT = 47  Burst = 61
   Medial:     Place = 100   Voicing =       VOT = 17  Burst =
```

Identification: Comparing heard values to prototypes

In order to identify the spoken consonant from its acoustic properties, the learner compares the set of heard acoustic properties to prototypes of each consonant. The learner guesses that the prototype most similar to the set of heard properties is the consonant produced by the speaker. A prototype is a four-dimensional vector whose coordinates represent the average value for each of a consonant's four acoustic properties, based on the tokens of that consonant heard by the learner thus far. Examples of these prototype coordinates are given in (10).

(10) Prototypes:

Initial	Place	Voicing	VOT	Burst
p:	3.9	4.3	16.5	54.1
b:	3.5	96.1	4.5	53.4
t:	94.8	7.4	30.2	65.3
d:	96.8	96.2	3.7	60.1

Medial	Place	Voicing	VOT	Burst
p:	4.1	5.6	20.9	55.4
b:	4.3	94.0	5.6	52.7
t:	95.5	5.6	22.9	63.0
d:	96.9	95.7	5.2	57.3

In order to guess which stops were heard in a particular *CaCa* word, the listener calculates the distance between the points represented by the heard acoustic values and those of each prototype. When all four cues are heard, as for the initial stop in (9), distance is calculated by the formula in (11). As the model is concerned with the details of voicing identification but not with place identification, there are three cues to voicing but only one for place. To compensate for this simplification, the single place cue is more heavily weighted in the distance calculation.

(11) *distance* =

$$\text{sqrt}\left(3*(\text{place}_{\text{Heard}}-\text{place}_{\text{C}})^2 + (\text{voi}_{\text{Heard}}-\text{voi}_{\text{C}})^2 + (\text{vot}_{\text{Heard}}-\text{vot}_{\text{C}})^2 + (\text{burst}_{\text{Heard}}-\text{burst}_{\text{C}})^2 \right)$$

The distance between each prototype and the set of heard values is calculated, producing a set of distances between the heard stop and each prototype as in (12). The

learner guesses that the heard stop is in the category of the nearest prototype. In this example, the learner guesses correctly that the initial stop was *t*.

(12) HEARD: Initial: Place = 100 Voicing = 0 VOT = 47 Burst = 61

Initial	DISTANCE (X~prototype)		Place	Voicing	VOT	Burst
p:	169		3.9	4.3	16.5	54.1
b:	198	Prototype	3.5	96.1	4.5	53.4
t:	**21**	coordinates:	94.8	7.4	30.2	65.3
d:	106		96.8	96.2	3.7	60.1

GUESS: Initial t (correct)

When the listener fails to hear all of the acoustic properties of some stop, as is the case for the medial stop in example (9) above, distance is calculated based on only those properties heard. For example, when the learner hears only place and VOT values, the distance between this stop and each prototype is calculated using only the prototypes' place and VOT values. The reduced distance equation for this case is given in (13).

(13) When only Place and VOT cues are heard:

$$distance = \text{sqrt}(3*(place_{Heard} - place_{C})^2 + (vot_{Heard} - vot_{C})^2)$$

Learning: Prototypes are adjusted based on new acoustic information

Finally, the learner must learn what the stops sound like. This is accomplished by updating prototype coordinates to reflect the new information acquired in each round. In this way, over time, the prototypes come to represent stops' canonical acoustic properties.

In a round where an initial *t* is produced, the learner updates the coordinates of the initial *t* prototype – even if the learner misidentifies the initial stop. Accomplishing this requires the learner to know which stops were actually produced. In assuming that the learner has this information, the model is similar to learning algorithms for OT constraint rankings where the learner compares observed surface forms to underlying representations from which they were derived (see e.g. Tesar and Smolensky (1994 et seq.), Boersma and Hayes (2001)). In giving this phonetic learner access to the 'underlying representation' of the speaker's utterance, in addition to surface acoustic values, the model focuses on learning the relationship between a given set of categories and their possible acoustic realizations. The assumption that the learner knows segments' true identities is not critical. Just as learners discover underlying representations as well as constraint rankings in elaborated models of phonological learning (Jarosz, 2006; Merchant and Tesar, to appear), this model could be enriched, allowing the learner to discover phonetic categories itself as in de Boer's model of vowel inventories (2001).

It is important to note that this knowledge about segments' identities is available only to the learning component of the model. During identification, the learner tries to identify segments based only on their acoustic properties; the learner effectively 'finds out' which segments were truly produced only after it guesses which were heard.

Learning via updating prototypes proceeds as follows. The value of a prototype coordinate for, say, initial *t*'s VOT property is the average of all VOT values for initial *t* heard by the learner. At the beginning of a simulation, each prototype's coordinates are the default values given in (14). A default value is the average of the four consonants' mean values for a particular acoustic property. For example, each word-initial stop prototype has an initial 'place' value of 50 because this is the average of the mean place values of initial *p* (0), *b* (0), *t* (100), and *d* (100). These simulation-initial default values give the model no inherent bias towards any of the four stops.

(14)

Simulation-initial defaults	Place	Voicing	VOT	Burst
Initial pbtd:	50	50	12.5	58
Medial pbtd:	50	50	12.8	56.8

After each round of identification, the learner adjusts the prototype coordinates for the actual initial and medial segments produced. For example, the initial *t* discussed above is the twelfth initial *t* heard by the model. The most recent set of acoustic values for *t* is averaged with the previous 11 sets of values to get a new set of coordinates for the initial *t* prototype. This new prototype, given in (15), represents the learner's entire experience with initial *t*, and is used in the next round of identification.[7]

(15)

Adjusted prototypes	Place	Voicing	VOT	Burst
Initial t -->	95.2	6.6	31.8	64.8
Medial d -->	97.0	95.7	5.9	57.3

The learner also collects information about stops' relative perceptual difficulty. This information is the basis for perceptually grounded constraint induction. The learner calculates two aspects of perceptual difficulty: accuracy and false alarms. Accuracy measures the learner's ability to correctly identify tokens of a particular consonant: 'Of all the tokens of initial *p* the learner has heard, how many have been correctly identified?' False alarms measure the learner's ability to guess that some particular consonant was heard only when this is true: 'Of all the times the learner guessed that it heard initial *p*, how many of those guesses were wrong?' The formulas for calculating these two scores are given in (16), and sample accuracy and false alarm rates are in (17).

(16) For some segment *x*:

$$\text{Accuracy}(x) = 100 * [\ \#\ x \text{ tokens correctly identified}\ ] \div [\ \#\ x \text{ tokens heard}\ ]$$

$$\text{FalseAlarm}(x) = 100 * [\ \#\ x \text{ tokens incorrectly identified}\ ] \div [\ \#\ x \text{ responses}\ ]$$

[7] As this model alternates between token identification and prototype learning, it has a similar structure to an Expectation Maximization model (Dempster et al., 1977). However, the present model doesn't implement true EM: here, learning is supervised and incremental; further, there is no explicit effort to identify prototypes which allow maximally accurate identification of all tokens heard to date. In the future, the model could be relatively straightforwardly revised to incorporate EM.

(17)

Initial	Accuracy		False alarm	
p:	80%	(12 of 15)	8%	(1 of 13)
b:	88%	(15 of 17)	12%	(2 of 17)
t:	92%	(11 of 12)	21%	(3 of 14)
d:	94%	(17 of 18)	6%	(1 of 18)

5.2 Results and discussion

The production and perception components of the model described above provide a simple yet accurate representation of human production and perception. Section 5.2.1 shows that the model, like human listeners, finds word-initial *p* uniquely perceptually difficult. Section 5.2.2 then demonstrates that the source of this difficulty is the variability of initial *p*'s VOT values. In this way, the model can be used to generate and refine hypotheses for future perceptual experiments.

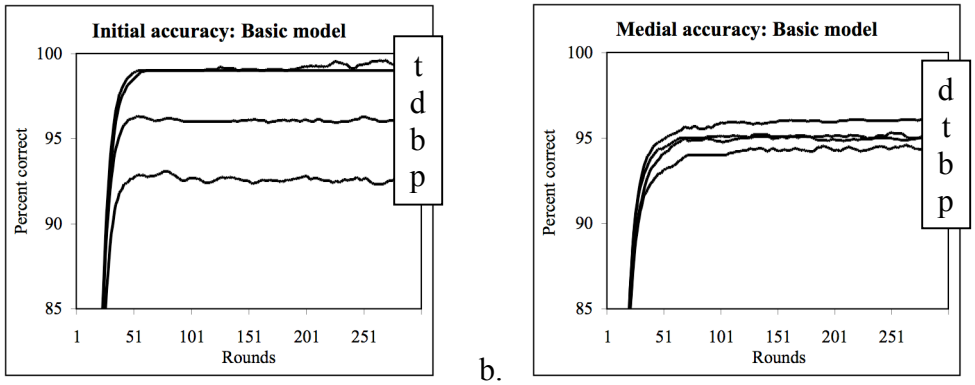
5.2.1 Initial *p* is perceptually difficult

The perceptual model behaves very much like participants in the perceptual experiment, finding word-initial *p* uniquely difficult. Initial *p* is more frequently misidentified than initial *b*, while no similar relationship holds between medial *p* and *b* or initial *t* and *d*. A difference between the model and human listeners lies in the indicator of perceptual difficulty, which is measured in the model only by accuracy scores. This, like many aspects of the model, is a simplification of real behavior, where perceptual difficulty can be indicated by either accuracy or reaction times (Ashby and Maddox, 1994; Pisoni and Lazarus, 1973). The perceptual experiment found indications of initial *p*'s perceptual difficulty in participants' reaction times, but not in their overall accuracy; this is likely a consequence of the specific task. Similarly, it is a consequence of the structure of this simple model that perceptual difficulty is measured here only in terms of accuracy; these two indicators of initial *p*'s difficulty are taken to be comparable for the discussion here.

The model's overall rates of accurately identifying various segments are determined by averaging the results of many simulations, much like human listeners' overall ability to perceive segments is typically evaluated by averaging experimental results from many subjects. Each simulation represents the progress of a single learner towards stable phonetic categories which allow that learner to identify segments at stable rates of accuracy. Figure 3 represents the model's developing ability to accurately identify each initial and medial consonant, averaged over 20,000 simulations. The accuracy measure for each consonant begins at chance. As the learner acquires phonetic experience, its ability to accurately identify each segment first increases and then stabilizes over the course of a 300-round simulation.

The model's percent correct for some segment in some round is measured as follows: out of 20,000 simulations, given some segment (e.g. initial *p*) and some point in time (e.g. round 200), how many accurate identifications of that segment in that round are there, vs. the total number of times that segment was produced in that round? Here, out of 20,000 simulations, the model 'heard' initial *p* in round 200 approximately 5,000 times; initial *p* was accurately identified in 91% of those 5,000 rounds. Lines are labeled by the

segments in boxes at the right of each graph, which show segments' order from most to least accurate.



a. b.

Figure 3. Model accuracy for each initial (a) and medial (b) consonant, averaged across 20,000 simulations of 300 rounds each. The lines represent moving averages across 15-round windows.

Figure 3 shows that the perceptual model gives the same primary result as the perceptual experiment: initial *p* is consistently more difficult for the model to identify than initial *b*. This difference is unique to labials in initial position, as no comparable accuracy difference is seen in medial *p* and *b*. There is also no comparable difference between the accuracies of initial *t* and *d*, so initial *p* is uniquely perceptually difficult.

5.2.2 The source of initial *p*'s perceptual difficulty: VOT variances

Initial *p*'s perceptual difficulty correlates with its close acoustic similarity to initial *b*. Section 3.2 speculated that initial *p* is more difficult than initial *b* because initial *p* is more acoustically variable than initial *b*. Because the perceptual model accurately represents both the acoustics and the relative perceptibility of these segments, it can be used to investigate the acoustic source of initial *p*'s perceptual difficulty. By removing individual acoustic features (e.g. VOT and burst intensity) from the model, and by changing segments' parameters for these properties, the acoustic properties which make initial *p* difficult to perceive can be identified. These results can suggest the direction for further perceptual experiments.

First, the model can be simplified in order to examine the effects of burst and VOT on perception. The binary closure voicing cue, which provides the learner with a perfect cue to voicing, can be removed. The learner then hears all other available cues in all utterances. Under these conditions, any perceptual difficulty in the model comes from the inherent properties of VOT and burst cues. Figure 4a shows the performance of this 'place-VOT-burst' model, where the closure voicing cue is never heard and place, VOT, and burst are always heard. This is compared to the basic model in Figure 4b. The two are fundamentally similar – initial *p* is always identified much less accurately than initial *b*; both are less accurately identified than *t* and *d*, whose accuracies are fairly similar.

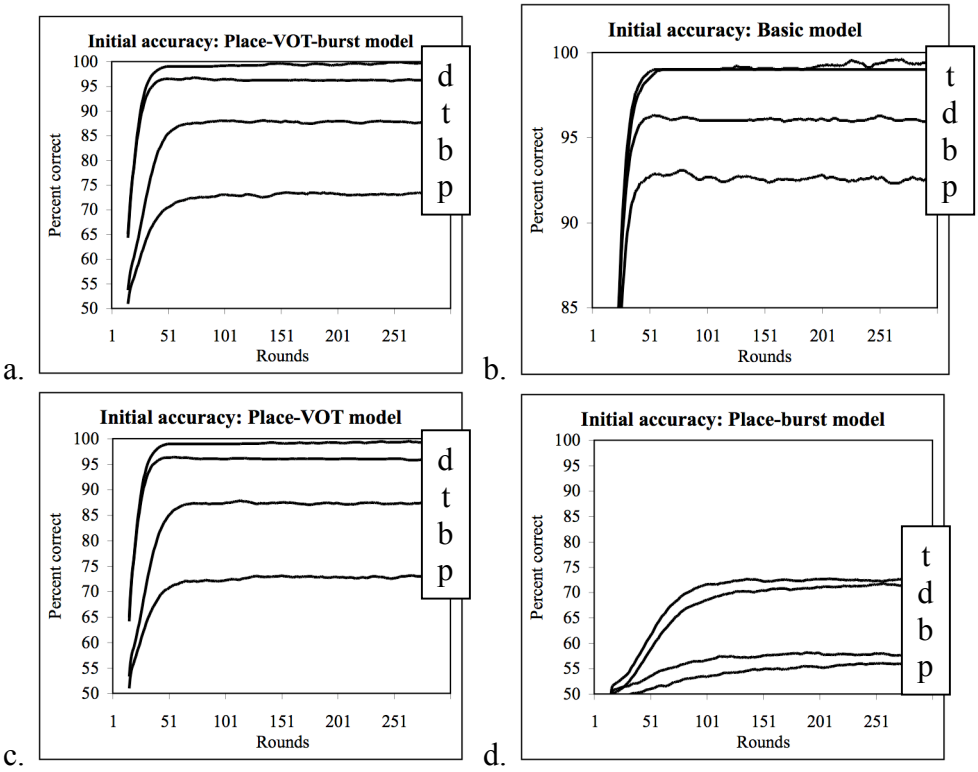


Figure 4. Model accuracy for each initial consonant when place, VOT, and burst cues are always heard, closure voicing is never heard (a); for the basic version of the model (b); when place and VOT cues are always heard, closure voicing and burst cues are never heard (c); when place and burst cues are always heard, closure voicing and VOT cues are never heard (d). Values are averaged over 20,000 300-round simulations.

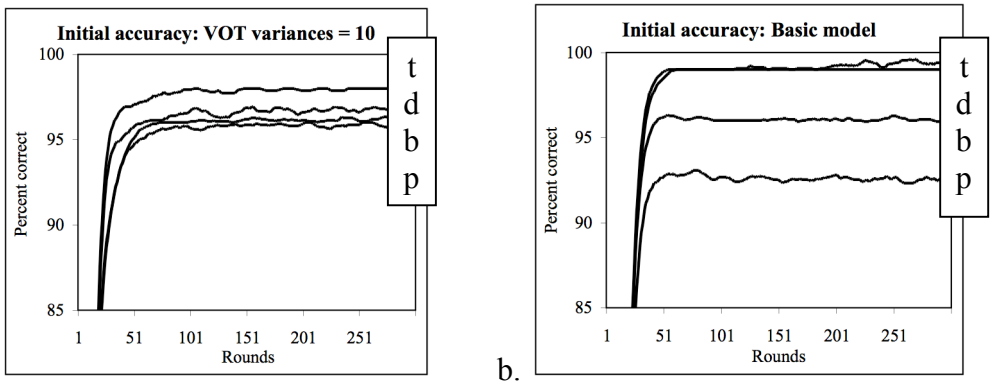
Using the place-VOT-burst model as a baseline, the perceptual results of models which make voicing decisions based on only VOT or only burst cues can be explored. If one of these models similarly finds initial p uniquely perceptually difficult, the single voicing cue in that model contributes significantly to this perceptual result. If, instead, the presence of only one cue to voicing changes the perceptual results dramatically, then that cue is not responsible for the basic pattern.

Figure 4 also shows the performance of models in which only (c) place and VOT cues vs. (d) place and burst cues are heard. The place-VOT model in Figure 4c is extremely similar to the place-VOT-burst model in Figure 4a. The place-burst model in Figure 4d, though, produces a pattern of perception quite unlike the others. While initial *p* is still somewhat less accurately identified than initial *b*, this difference is comparable in size to the difference between initial *d* and *t*. Unlike the place-VOT-burst model, the place-VOT model, the basic model, and experimental results, the perceptual difficulty of initial *p* is not particularly unique. Taken together, these modified models indicate that VOT cues, rather than burst cues, are responsible for initial *p*'s perceptual difficulty.

Finally, the model can be further adjusted to show that the variances of segments' VOTs largely determine these perceptual results, (partially) confirming the hypotheses raised in section 3.2. Segments' VOT variances in the basic model are quite large, as shown in (18). When the basic model is modified such that each segment's VOT variance is 10, the resulting pattern of perception is quite different. As shown in Figure 5a, all four initial segments are now identified with very nearly equal accuracy. Crucially, the asymmetry between *p* and *b* disappears.

(18)

	p		***b***		***t***		***d***	
VOT	mean	*variance*	mean	*variance*	mean	*variance*	mean	*variance*
Initial	16	*89*	0	*74*	34	*76*	0	*74*



a. b.

Figure 5. Model accuracy for each initial consonant where the VOT variance of each initial segment is 10 (a) and in the basic model (b). Values are averaged over 20,000 300-round simulations.

This suggests that it is in fact segments' large VOT variances – and initial *p*'s particularly large variance – which give rise to initial *p*'s perceptual difficulty. As the model typically makes voicing mistakes, this makes sense as follows. The VOT means of initial *p* and *b* are relatively similar (compared to those of initial *t*–*d*), so these segments have a relatively large chance of being inaccurately misidentified as each other. Initial *p* is even less accurate than initial *b* because its VOT variance is larger than that of *b*. There are more *p*s with *b*-like VOTs, so more *p*s are misidentified as *b* than vice versa. These results suggest that future perceptual experiments could manipulate stop VOTs in order to determine whether VOT variances are also the source of human listeners' perceptual difficulty with initial *p*.

5.3 Discussion and conclusion

This model is based on realistic acoustic representations of voicing cues for the initial and medial stops *p*, *b*, *t*, and *d*. From this data, the virtual learner develops criteria for identifying each stop, ultimately presenting a pattern of perception very similar to that found in the perceptual experiment. Word-initial *p* is more perceptually difficult (as indicated by the model's less accurate identification of initial *p*) than initial *b*. This is not a general property of labials: the model, like humans, finds medial *p* no more difficult

than medial *b*. Neither is it a general property of initial voiceless stops: initial *t* is no more difficult than initial *d*. Within the model, initial *p*'s unique perceptual difficulty is due to both the similarity between the VOT means for initial *p* and *b* and especially to initial *p*'s greater VOT variance.

As explained in section 5.1.1, the VOT variances for voiced stops were arbitrarily set to the average of the VOT variances of initial and medial *p* and *t*. This gives the voiced stops quite large VOT variances, effectively allowing some of the model's voiced stops to have voiceless intervals following their release. This sort of brief post-release voicelessness in voiceless stops is reported by Mikuteit (2006), justifying this basic representation. However, the assumption that the variance of voiced stops' (positive) VOTs is as large as that of voiceless stops is entirely arbitrary. If anything, these post-release voiceless periods are likely shorter and rarer than assumed in this model. The variance of voiced stops' positive VOTs is thus likely smaller than assumed here. As initial *p*'s perceptual difficulty follows primarily from its large VOT variance relative to initial *b*, any revised the model in which voiced stops' VOT variance were smaller would also produce – and, in fact, enhance – this perceptual asymmetry between initial *p* and *b*.

These results about the source of initial *p*'s perceptual difficulty are, of course, necessarily true of only the model. Whether humans process these acoustic features in the same way, and thus whether humans' perceptual difficulty with initial *p* also follows primarily from its short, variable VOT, is a matter for further experimental study. The model is useful in that it allows preliminary exploration of this sort of acoustically motivated hypothesis much more rapidly than can be done with actual experimentation.

6. Modelling constraint induction

This model of learners' acoustic and perceptual experience can be used to explore the induction of functionally grounded constraints. Section 1 argued that learners induce constraints based on data from their immediate linguistic experience; section 6.1 considers the varieties of linguistic experience which play a role in constraint induction. Crucially, induction cannot refer to certain kinds of perceptual experience, e.g. infants' perception of their own speech. Further, raw linguistic experience must be evaluated in a structured way in order to give rise to phonological constraints, as Hayes (1999) argues. Towards this end, section 6.2 suggests a general schema which structures the induction of perceptually grounded constraints, and shows how a specific instantiation of this schema can consistently lead learners to induce *#P from realistic acoustic and perceptual data.

6.1 Constraint induction and linguistic experience

The general goal of constraint induction is for all learners to induce a consistent set of functionally grounded constraints from their immediate linguistic experience. A realistic model of constraint induction depends on a precise, realistic characterization of this experience. Learners' experience can vary along two dimensions. Different learners can be exposed to languages with different phonological properties, and so can have differential exposure to individual segments or structures. Each learner also has various kinds of experience with language: learners perceive adult speech, and also articulate and

perceive their own babbling and early speech. This section argues that each functionally grounded constraint is consistently induced from any learner's experience with any language. However, perceptually grounded constraints can only be induced from learners' experience perceiving adult speech.

This paper has discussed two phonotactic possibilities for word-initial *p*. Initial *p* can be present in a language, in which case it is perceptually difficult as in French and the model of pseudo-French in section 5. Initial *p* can also be absent, as in Cajonos Zapotec and Ibibio. Learners of these two types of languages will have fundamentally different experiences of initial *p*; the constraint inducer discussed here is designed to consistently identify initial *p* as perceptually difficult in either situation, and so consistently induce *#P in either type of language.

In a language like French where initial *p* is licensed, its perceptual properties are readily available to learners as they induce constraints. This process can be straightforwardly implemented in a model of pseudo-French. If the inducer has a way of tracking the relative perceptual difficulty of segments in particular phonotactic positions, it will observe that *p* is more difficult to accurately identify than other word-initial segments. This information can trigger the induction of *#P. In a language like Cajonos Zapotec, however, this experience of initial *p*'s acoustic properties and their perceptual consequences is unavailable to learners. If a learner of (pseudo-)Cajonos Zapotec is to consistently induce *#P from its perceptual experience like the (pseudo-)French learner, it must refer to a different aspect of perceptual experience.

It is frequently assumed that constraints like *#P, which can prevent learners of some languages from ever being exposed to perceptually difficult segments, are nonetheless universally grounded in the segment's perceptual difficulty. After all, we assume that languages like Cajonos Zapotec ultimately ban initial *p* is because it is perceptually difficult. This suggests that the constraint responsible for this restriction should represent each speaker's knowledge of initial *p*'s perceptual difficulty.

This perspective is difficult to reconcile with the claim that functionally grounded constraints are induced rather than innate, as some learners lack perceptual experience of adult word-initial *p*. To resolve this conflict, it is often tacitly assumed that learners inducing constraints – typically through a proposed mechanism like Inductive Grounding (Hayes, 1999) or the Schema/Filter model of CON (Smith, 2002) – refer to something other than perceptual experience of adult forms of the ambient language, or their own articulations of these same forms in the case of articulatorily grounded constraints.

One possible way in which learners could experience unattested segments' articulatory and perceptual properties is through their own early productions. Hayes' discussion of Inductive Grounding focuses on learners' induction of articulatorily grounded constraints. If learners could refer to articulatory experience from babbling, they could learn about segments' articulatory properties in positions where they are not attested in an adult language (like initial *p*). This information could then perhaps be used to induce a complete set of typologically attested, articulatorily grounded constraints.

In the Schema/Filter model of CON, perceptually grounded constraints emerge from a mechanism similar to Inductive Grounding. Smith discusses the induction of perceptual augmentation constraints within this model; these constraints prefer perceptually salient forms to minimally different, less perceptually salient forms.[8] For example, the augmentation constraint HEAVYσ prefers more salient long vowels to less salient short vowels. In order to determine the relative perceptual salience of segments or structures unattested in learners' target languages, learners could again examine the psycholinguistic consequences of their own early productions of unattested structures.

With respect to the perceptually grounded constraint *#P, however, it is unlikely that infants' own early productions of *p*-initial forms could provide them with the same perceptual data as French-speaking adults' pronunciations. First of all, evidence of such forms would be relatively rare, and highly inconsistent across learners. While children's early babbling occasionally includes unattested segments and phonotactic structures, later stages of babbling quickly come to reflect the segmental frequency and phonotactics of the target language (Jusczyk, 1997: 177-9). Various child phonology processes such as truncation, consonant harmony, and other unfaithful mappings can also give rise to phonotactic structures unattested in adult language (Vihman, 1996: 218-21), but children vary widely in their use of these processes (as well as in the phonetic inventories and structures used in babbling). So while it is likely that many children learning languages without initial *p* could occasionally produce initial *p*, it is unlikely that this experience would be frequent enough, or consistent enough across learners, for universal induction of *#P.

A further reason why infants' early productions provide perceptual data unlike that garnered from adult speech is that infants' speech is much more articulatorily variable than adult speech (Jusczyk, 1997: 181). In fact, while young children's articulations may be impressionistically similar to various adult segments, children only very rarely produce adult-like segments before approximately 6 months, at which point the segmental content of babbling very quickly comes to resemble that of early child speech (Oller, 2000). The acoustic experiments discussed above, as well as the production component of this model, suggest that the perceptual difficulty of initial *p* follows from its relatively fine-grained acoustic properties. As infants' articulations are much more variable than those of adults, it is unlikely that an infant's own rare productions of unattested segments would be articulatorily and acoustically similar enough to those of adult speakers to trigger the same patterns of perception as adult productions. For these reasons, learners should refer only to their perceptual experience with adult speech in inducing perceptually grounded constraints.[9]

A Cajonos Zapotec learner therefore cannot induce *#P from the same knowledge of initial *p*'s perceptibility that a learner of French uses. Cajonos Zapotec and French

[8] Perceptual salience is a psychoacoustic measure, perhaps of neural response magnitude.

[9] Learners' articulatory experience of their own productions poses similar difficulties for constraint induction. In addition to children's articulatory inaccuracy and the scarcity of unattested segments and phonotactic structures, the size and shape of an infant's mouth (along with the initial absence of teeth) may give infants substantially different experience of articulatory difficulty than that of adult speech, which is typically assumed to shape adult phonology.

learners know fundamentally different things about initial *p*: a French learner knows that it is dispreferred – and so induces *#P – based on the difficulty of accurately identifying initial *p*. A Cajonos Zapotec learner instead knows that initial *p* is dispreferred simply because it is unattested in adult speech.

Reflecting this diverse knowledge about initial *p*, the induction mechanism proposed here refers to correspondingly diverse aspects of perceptual difficulty. In general, the constraint inducer tracks segments' perceptual properties, identifies segments which are relatively perceptually difficult in particular phonotactic positions, and generates constraints against these segments in these positions. In order to induce constraints against segments with which learners have actual perceptual experience as well as those which are absent from a particular phonotactic position, the inducer refers to two measures of perceptual difficulty: accuracy (which reflects correct identification of a segment) and false alarms (which reflect incorrect guesses that a segment was heard).

The next section describes how the mapping from perceptual data to induced constraints is governed by schemata for perceptually grounded constraints. These constraint schemata provide accuracy and false alarm criteria for labeling segments 'perceptually difficult', and for inducing constraints against these segments.

6.2 How the model works

The input to the constraint induction component of the model comes from the perception component, which hears acoustically realistic representations of initial and medial *p*, *b*, *t*, and *d* and perceives them realistically: word-initial *p* is uniquely perceptually difficult. The induction component induces the constraint *#P from this perceptual experience. A functionally grounded constraint schema defines the phonotactic positions which can be targeted by the induced constraints, as well as what exactly is meant by 'perceptually difficult.' Schemata are thus sets of phonotactic and perceptual (as well as articulatory, psycholinguistic, etc.) criteria for constraint induction.

Section 6.2.1 first describes the general structure of schemata for functionally grounded constraints, as well as the particular schema which governs the assessment and comparison of accuracy and false alarm scores in this model. Section 6.2.2 presents the specific criteria comparing segments' accuracy scores, allowing consistent induction of *#P in a French-type language where learners hear initial *p*. Induction of the same constraint from false alarm scores, as in a Cajonos Zapotec-type language where learners never hear initial *p*, is discussed in section 6.2.3.

6.2.1 General schemata for perceptually grounded constraint induction

The goal of the induction mechanism is to consistently induce the constraint *#P from word-initial *p*'s unique perceptual difficulty. Hayes (1999) demonstrates that constraints cannot simply emerge from raw phonetic data; rather, learners must come know how to map phonetic information to phonological constraints. This paper proposes that four basic elements of the induction of perceptually grounded constraints must be specified. These

are summarized, along with the particular parameters of the constraint induction model described here, in (19).

(19) Schemata specify four basic features of perceptual constraint induction:

(i) What kind of phonological element could be perceptually difficult.

Here: Individual segments.

(ii) Phonotactic positions where perceptual difficulty is considered.

Here: Word-initial position.

(iii) What makes a segment 'perceptually difficult'.
(A procedure for comparing measures of perceptual difficulty.)

How many recent tokens' accuracy/false alarm scores are considered.

Here: 400 tokens of each segment.

Properties of segments' relative accuracy and false alarm scores that trigger induction.

Here: See sections 6.2.2 and 6.2.3.

(iv) Definition of the induced constraints.

Here: If a segment *x* is relatively perceptually difficult in $Context_Z$:

*x/Context$_Z$ Assign one violation mark for each instance of *x* in $Context_Z$.

First, a constraint schema defines the type of phonological element over which perceptual difficulty is calculated. In the present model, individual segments are judged perceptually difficult. In other models, features, or sets of segments all sharing a feature or features, could be judged perceptually difficult as well.

An induction schema must also specify the phonotactic positions in which segments' perceptual difficulty will be evaluated. With no such specifications, learners would track perceptual difficulty in all logically possible phonotactic positions. This is undesirable, as some positions have no known phonological relevance: for example, no constraint targets third-syllable onsets. In this model, the inducer is further simplified in that it looks only at word-initial position, ignoring intervocalic stops. This is because stops are not typically banned intervocalically, as that is where their cues are most salient.

The third element specified by a constraint schema is the set of criteria for labelling a segment 'perceptually difficult'. In the present model, the inducer tracks two perceptibility measures: accuracy and false alarms (definitions repeated in (20)). The accuracy and false alarm scores of individual segments are compared using the criteria described in sections 6.2.2 and 6.2.3.

(20) For some segment *x*:

Accuracy(*x*) = 100 * [# *x* tokens correctly identified] ÷ [# *x* tokens heard]

FalseAlarm(*x*) = 100 * [# *x* tokens incorrectly identified] ÷ [# *x* responses]

Before these scores can be compared, the schema must specify both how and when they are calculated. In the early part of a simulation, before robust criteria for identifying segments have developed, a learner has low accuracy scores and high false alarm scores for all segments. For this reason, the inducer does not begin tracking accuracy or false alarm scores until phonetic categories and accuracy rates have stabilized. In the simulations reported here, induction begins after 150 rounds of production and perception. A time where prototype coordinates have stabilized (and so when induction should begin) could also be dynamically identified in each simulation in a more complex model.

A learner must also know how much of its experience to take into consideration in calculating these scores. For the sake of efficiency, learners in this model do not consider every token of every segment in their entire experience. In order to be resilient in the face of noisy data, however, learners also should not consider too little experience. Learners here consider accuracy and false alarm scores of the most recent 400 tokens of each segment, and so induce constraints only from persistent patterns of perceptual difficulty.

These scores are tracked as follows. In a given round (after round 150), each segment heard by the learner gets an accuracy score of 1 if it is correctly identified and 0 otherwise, as shown in (21). Similarly, each segment which the learner guesses it heard gets a false alarm score of 0 if the guess was correct and 1 if the guess was incorrect. Once the learner has collected 400 such accuracy or false alarm scores for a segment, they can be averaged to obtain a representative score for that segment, as in (22). Segments which are typically accurately identified have average accuracy scores close to 1, and segments for which the learner typically makes accurate guesses have false alarm scores close to 0.

(21)
```
HEARD:   Initial p      -->      Initial p accuracy    = 0
GUESS:   Initial b      -->      Initial b false alarm = 1
```

(22)
```
ACCURACY:   Initial  p: 0.913    b: 0.920    t: 0.957    d: 0.970
FALSE ALARMS:           0.080       0.093       0.027       0.040
```

Finally, after defining the elements that can be judged perceptually difficult, the phonotactic positions in which these elements' perceptibility is evaluated, and the criteria for finding particular segments perceptually difficult, a functionally grounded constraint schema must also define the constraints that are induced against perceptually difficult segments. Here, a positional markedness constraint of the form defined in (23) is induced. (*#P is an abbreviation for **p*/#__.)

(23) *x/Context$_Z$ Assign one violation mark for each instance of *x* in *Context$_Z$*.

The model described here has one additional property which is a significant simplification of any actual learners' induction processes.[10] This model is only concerned with the relative perceptibility of voiced and voiceless homorganic stop pairs. For this reason, the acoustic and perceptual differences between *p* and *b*, and *t* and *d*, are accurately modeled. Differences between other pairs of segments, however, are not. Therefore while the model can accurately assess the relative perceptual difficulty of *p* and *b* or *t* and *d*, any judgment it would make about the relative perceptibility of *b* and *d*, *p* and *t*, or other heterorganic pairs does not accurately reflect speakers' judgments about these segments' perceptibility. For this reason, the model never compares the perceptual difficulty of a segment to anything other than its homorganic counterpart.

The virtual speaker in the production component can speak either pseudo-French or pseudo-Cajonos Zapotec (pseudo-CZ); the only difference between these two languages is whether or not they allow word-initial *p*, as shown in (24). A learner of either pseudo-language compares accuracy and false alarm scores using the two methods described below. Section 6.2.2 describes the comparison of accuracy scores that allows pseudo-French learners to induce *#P, and section 6.2.3 describes the comparison of accuracy and false alarm scores that allows pseudo-CZ learners to also induce *#P.

(24)

	Initial Cs	Medial Cs
pseudo-French:	p b t d	p b t d
pseudo-Cajonos Zapotec:	b t d	p b t d

6.2.2 Induction from accuracy scores: Pseudo-French

The virtual learner induces constraints against segments which it finds perceptually difficult. In intuitive terms, a pseudo-French learner knows that initial *p* is more perceptually difficult than initial *b* simply because initial *p* is recognized less accurately than initial *b*. In order for the learner to judge perceptual difficulty in a consistent way, it must have an explicit procedure for identifying difficulty which is persistent and significant enough to merit encoding in a constraint. This section describes the procedure which allows pseudo-French learners to consistently induce *#P.

The constraint inducer measures segments' accuracy scores against both absolute and relative criteria. A segment is perceptually difficult if its accuracy score (over the most recent 400 tokens of the segment) is lower than the absolute threshold of 0.9, and also significantly lower than that of its homorganic counterpart.

(25) Some segment *x* is perceptually difficult in $Context_Z$ if:

$$\text{Accuracy}(x/Context_Z) < 0.9$$

and

$$\text{Accuracy}(x/Context_Z) < \text{Accuracy}(y/Context_Z)$$

The two accuracy scores must be significantly different. ($\alpha = 0.01$)

[10] Restricting the model to *CaCa* words is another such simplification.

According to these criteria, initial *p* is perceptually difficult only if it is accurately identified less than 90% of the time, and if its accuracy score is significantly lower than that of initial *b* (as determined by a t-test, where $\alpha = 0.01$). The absolute difficulty measure ensures that only significant, persistent perceptual problems will be penalized by induced constraints. The relative measure further captures the inherently comparative character of markedness constraints: constraints are induced only against segments which are more difficult than others.

The results of pseudo-French simulations where constraints are induced through these accuracy criteria are summarized in Figure 6. The graph shows the sets of constraints induced in 250 pseudo-French simulations of 40,000 rounds each. Initial *p*'s accuracy score is consistently both sufficiently low and also significantly lower than that of initial *b*. Therefore the inducer consistently observes that initial *p* is perceptually difficult, and so induces *#P in nearly every simulation; the very small number of simulations in which *#P is not induced would disappear if simulations were slightly longer. Because initial *p* is so much more perceptually difficult than initial *b*, the inducer has evidence for the opposite constraint *#B only extremely rarely.

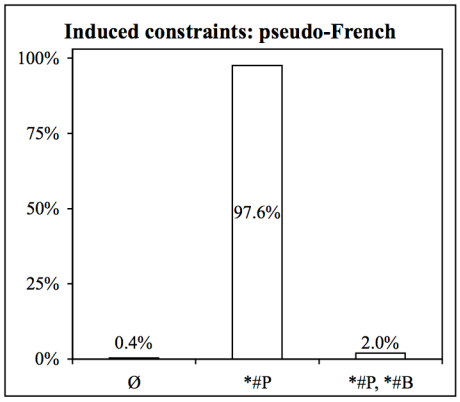


Figure 6. Constraints induced in 250 pseudo-French simulations of 40,000 rounds each.

Unlike initial *p* and *b*, initial *t* and *d* are essentially equally perceptually difficult. While either may occasionally be significantly less accurately identified than the other, and either may very occasionally have an accuracy score below 0.9, these aspects of perceptual difficulty consistently fail to coincide. Learners therefore have no evidence for the induction of either *#T or *#D.

6.2.3 Induction from false alarm scores: Pseudo-Cajonos Zapotec

In this model of constraint induction, two segments' accuracy scores cannot be compared until a learner has collected enough accuracy scores for each segment to provide a reliable picture of the segments' overall perceptibility. This is enforced through the requirement that 400 tokens of each segment must be heard before the segments' accuracy scores can be compared. Learners who never hear any tokens of initial *p* never develop comparable accuracy scores for *p* and *b*, so a pseudo-CZ learner can never induce *#P via comparison of accuracy scores. It was argued above that all learners of all

languages must induce the functionally grounded constraint *#P from their immediate experience with adult speech; for this reason, learners must be able to use perceptual measures in addition to accuracy score comparisons in order to identify initial *p* as perceptually difficult and induce *#P. The perceptual measure which allows pseudo-CZ learners to induce this constraint capitalizes on basic properties of the perceptual model implemented here, as follows.

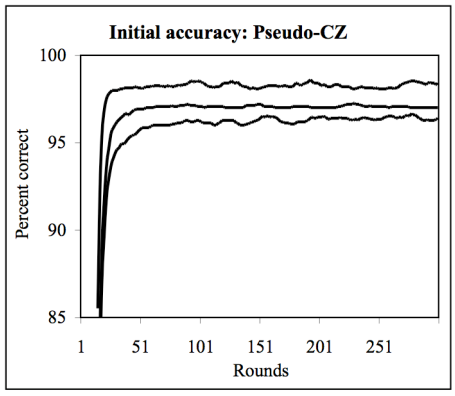
The perception component of the model assumes that learners know which segments occur in the ambient language before they begin to learn segments' acoustic properties in each phonotactic position. That is, before learners undertake the perceptual learning procedure described above, they first identify an initial inventory of sounds present in their language. During perceptual learning, learners then expect to hear each of these in each phonotactic position.[11] In the case of a pseudo-CZ learner, *p* is present in pseudo-CZ, though it is never heard word-initially. But because the learner expect to hear each segment in the pseudo-CZ inventory – including initial *p* – in initial position, it occasionally misidentifies another initial stop as *p*. In this way, pseudo-CZ learners acquire false alarms for unattested initial *p*.

Learners therefore have a unique kind of perceptual experience with phonotactic gaps: segments which are missing in a particular position (e.g. word-initially) incur more false alarms than accurate identifications in that position. These false alarms are relatively rare but do consistently occur, as illustrated in Figure 7. Figure 7a shows the pseudo-CZ learner's accuracy with the three attested initial stops *b*, *t*, and *d*. This learner, like the pseudo-French learner described in section 5.1.2, identifies initial *t* and *d* with roughly comparable accuracy.

The pseudo-CZ learner is overall more accurate in its identification of initial *b* than the pseudo-French learner. This is because the pseudo-CZ learner has no knowledge of the detailed acoustic properties of initial *p* – crucially, this learner has no knowledge of the actual degree of similarity between initial *p* and *b*, and so this learner is less likely than the pseudo-French learner to misidentify initial *b* as *p*. Even without this knowledge, however, there is a small but consistent chance that the pseudo-CZ learner will make exactly this mistake, guessing that an atypical initial *b* is the expected but thus far unattested initial *p*. As the confusion matrix in Figure 7b shows, this misidentification occurs for 0.2% of initial *b* tokens.

[11] This initial inventory is stipulated in the present model; it could also be learned from the statistical properties of the segments that it hears, as proposed by Maye (2000) and as modelled by de Boer (2000). This initial inventory does not necessarily correspond to the language's actual phoneme inventory but instead is simply the learner's initial hypothesis space for early categorization.

a.



b.

		Response			
		p	**b**	**t**	**d**
Segment	**p**				
	b	0.2%	98.4%	0.0%	1.4%
	t	0.1%	0.0%	97.1%	2.8%
	d	0.2%	1.2%	2.2%	96.4%

Figure 7. Model accuracy for each initial pseudo-CZ consonant, averaged across 20,000 simulations of 300 rounds each; lines represent moving averages across 15-round windows (a). Confusion matrix for initial pseudo-CZ consonants, from the last 15 rounds of each of 20,000 simulations (b).

This property of 'gapped' segments like pseudo-CZ initial *p* can motivate the induction of *#P as follows. In addition to comparing different segments' accuracy scores, the model also compares each segment's false alarm score to its accuracy score. If some segment's false alarm score is not lower than its accuracy score – that is, if the false alarm score is higher than the accuracy score, or if there is a false alarm score but no accuracy score – the false-alarm-prone segment qualifies as perceptually difficult. Using this measure of perceptual difficulty, every simulated pseudo-CZ learner observes that initial *p* is prone to false alarms, and so induces the constraint *#P as shown in Figure 8.

(26) Some segment *x* is perceptually difficult in $Context_Z$ if:

$$\text{Accuracy}(x/Context_Z) \not> \text{FalseAlarm}(x/Context_Z)$$

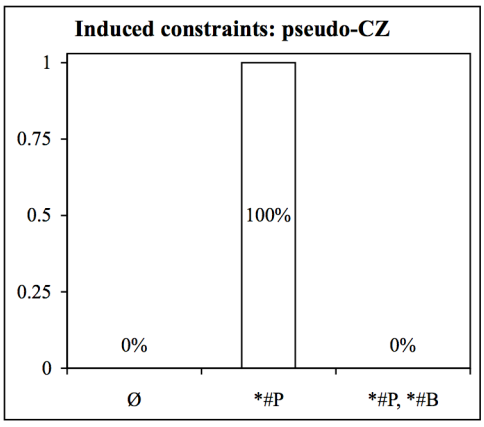


Figure 8. Constraints induced in 250 pseudo-CZ simulations of 40,000 rounds each.

6.3 Summary of the constraint induction model

In this model, a learner of any (pseudo-)language considers accuracy and false alarm scores for recent tokens of each segment. Learners examine individual segments'

accuracy scores, testing those which are below 0.9 to see whether they are significantly lower than those of their homorganic counterparts. At the same time, learners also compare individual segments' accuracy and false alarm scores. A segment is labeled perceptually difficult if either its accuracy score is below 0.9 and is significantly lower than that of some other segment, or if its false alarm score is greater than its accuracy score. These criteria are summarized in (27).

(27) Some segment *x* is perceptually difficult in $Context_Z$ if either:

Accuracy(*x*/$Context_Z$) < 0.9

and

Accuracy(*x*/$Context_Z$) < Accuracy(*y*/$Context_Z$) ➔ Constraint **x*/$Context_Z$

This difference must be significant ($\alpha = 0.01$).

…or…

Accuracy(*x*/$Context_Z$) ≯ FalseAlarm(*x*/$Context_Z$) ➔ Constraint **x*/$Context_Z$

Whenever a segment *x* is found to be perceptually difficult in some phonotactic position $Context_Z$, a positional markedness constraint of the form **x*/$Context_Z$ is induced. Learners of languages like pseudo-French, where initial *p* is present but less perceptible than initial *b*, consistently induce the constraint *#P through comparison of accuracy scores. Learners of languages like pseudo-CZ, where initial *p* is absent, consistently induce the same constraint *#P through comparison of accuracy and false alarm scores.

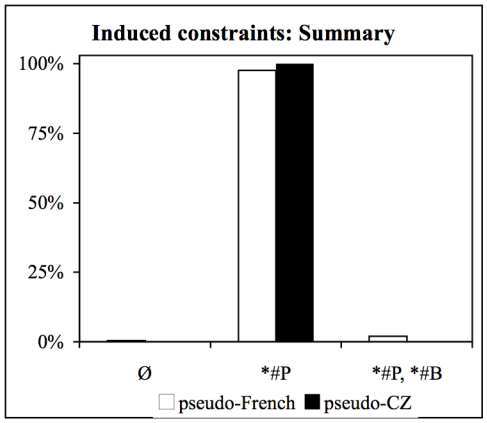


Figure 9. Constraints induced in each of 250 simulations of 40,000 rounds each.

7. General discussion and conclusion

The constraint induction component of this model has demonstrated that a perceptually grounded positional markedness constraint against word-initial *p* can be consistently induced from the diverse perceptual experiences of learners who hear this perceptually difficult segment, as well as those learning languages where *p* is banned word-initially.

This is possible because the inducer makes use of two measures of perceptual difficulty. The relative accuracy of initial *p* and *b* demonstrates *p*'s perceptual difficulty in languages like French. In languages without initial *p* like Cajonos Zapotec, learners mistakenly expect to hear initial *p* and so occasionally misidentify initial *b* as *p*. This makes *p*'s false alarm score higher than its accuracy score, which also indicates that initial *p* is difficult to accurately identify. The constraint *#P can be induced from perceptual difficulty in either case.

This case study of constraint induction illuminates the structure and role of constraint schemata in the induction of functionally grounded constraints. These schemata tell learners which positions and marked elements can be targeted by constraints. The functionally grounded schema governing the perceptual induction process described here tells the learner what how induced constraints should be defined, which kinds of phonological elements may be considered perceptually difficult, which phonotactic positions segments' perceptual difficulty should be assessed in, and how to compute and compare perceptual difficulty measures (i.e. accuracy and false alarm scores). Schemata for constraints grounded in facts of articulation, psycholinguistics, or other aspects of perception presumably make similar specifications.

While this paper has shown that the constraint *#P can be consistently induced from learners' immediate linguistic experience, it has not undertaken to show that learners actually do induce *#P in this way. The induction model demonstrated that realistic acoustic and perceptual properties of initial stops can be consistently mapped to attested constraints on initial segments; however, the question of whether real learners actually use this mechanism, or one like it, to induce *#P from the perceptual data considered here is left for future investigation. In determining that induction is possible in principle, a computational model can be used to develop hypotheses about constraint induction for further experimentation. Similarly, section 5.2.2 showed that the perceptual component of the model can also generate hypotheses about the acoustic source of initial *p*'s perceptual difficulty, which could be tested in further perceptual experiments. In general, computational models of the sort developed here are valuable tools for showing whether constraint induction and other aspects of phonological learning are possible in principle, and also for developing further experimental investigations of these matters.

References

Akinlabi, Akinbiyi, and Eno E. Urua. 2002. Foot structure in the Ibibio verb. *Journal of African Languages and Linguistics* 23:119-160.

Archangeli, Diana, and Douglas Pulleyblank. 1994. *Grounded Phonology*. Cambridge, MA: MIT Press.

Ashby, F. Gregory, and W. Todd Maddox. 1994. A response time theory of separability and integrality in speeded classification. *Journal of Mathematical Psychology* 38:423-466.

Boersma, Paul, and Bruce Hayes. 2001. Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry* 32:45-86.

Connell, Bruce. 1994. The Lower Cross languages: A prolegomena to the classification of the Cross River languages. *Journal of West African Linguistics* 24:3-46.

de Boer, Bart. 2001. *The Origins of Vowel Systems*. Oxford: Oxford University Press.

de Boer, Bart G. 2000. Self-organization in vowel systems. *Journal of Phonetics* 28:441-465.

Dempster, Arthur P., Nan M. Laird, and Donald B. Rubin. 1977. Maximum Likelihood from incomplete data via the EM algorithm. *Journal of Royal Statistics Society* 39:1-38.

Essien, Okon E. 1990. *A Grammar of the Ibibio Language*. Ibadan: University Press Limited.

Flack, Kathryn. 2007. The Sources of Phonological Markedness, University of Massachusetts Amherst: Doctoral dissertation.

Hay, Jessica. 2005. How Auditory Discontinuities and Linguistic Experience Affect the Perception of Speech and Non-Speech in English- and Spanish-Speaking Listeners, University of Texas Austin: Doctoral dissertation.

Hayes, Bruce, Robert Kirchner, and Donca Steriade eds. 2004. *Phonetically Based Phonology*. Cambridge: Cambridge University Press.

Hayes, Bruce P. 1999. Phonetically driven phonology: The role of Optimality Theory and inductive grounding. In *Formalism and Functionalism in Linguistics, vol. 1*, eds. M. Darness, E. A. Moravcsik, F. Newmeyer, M. Noonan and K. M. Wheatley, 243-285. Amsterdam: Benjamins.

Hooper [Bybee], Joan. 1976. *An Introduction to Natural Generative Phonology*. New York: Academic Press.

Jarosz, Gaja. 2006. Rich Lexicons and Restrictive Grammars - Maximum Likelihood Learning in Optimality Theory, Johns Hopkins University: Doctoral dissertation.

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

Lisker, Leigh, and Arthur Abramson. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. *Word* 20:384-422.

Maye, Jessica. 2000. Learning Speech Sound Categories from Statistical Information, University of Arizona: Doctoral dissertation.

Merchant, Nazzaré, and Bruce Tesar. to appear. Learning underlying forms by searching restricted subspaces. In *The Proceedings of CLS 41*. Chicago: Chicago Linguistics Society.

Mikuteit, Simone. 2006. A Cross Linguistic Inquiry on Voice, Quantity and Aspiration, Universität Konstanz: Doctoral dissertation.

Nellis, Donald G., and Barbara E. Hollenbach. 1980. Fortis versus lenis in Cajonos Zapotec phonology. *International Journal of American Linguistics* 46:92-105.

Ohala, John J. 1990. There is no interface between phonology and phonetics: A personal view. *Journal of Phonetics* 18:153-171.

Oller, D. Kimbrough. 2000. *The Emergence of the Speech Capacity*. Mahwah, N.J.: Lawrence Erlbaum Associates.

Pisoni, David B., and Joan House Lazarus. 1973. Categorical and noncategorical modes of speech perception along the voicing continuum. *Journal of the Acoustical Society of America* 55:328-333.

Pisoni, David B., and J. Tash. 1974. Reaction times to comparisons within and across phonetic categories. *Perception and Psychophysics* 15:285-290.
Prince, Alan, and Paul Smolensky. 1993/2004. *Optimality Theory: Constraint Interaction in Generative Grammar*. Malden, MA & Oxford: Blackwell.
Prince, Alan, and Paul Smolensky. 2004. *Optimality Theory: Constraint Interaction in Generative Grammar*. Malden, MA & Oxford: Blackwell.
Repp, Bruno. 1979. Relative amplitude of aspiration noise as a voicing cue for syllable-initial stop consonants. *Language and Speech* 22:173-189.
Smith, Jennifer L. 2002. Phonological Augmentation in Prominent Positions, University of Massachusetts Amherst: Doctoral dissertation.
Stampe, David. 1973. A Dissertation on Natural Phonology, University of Chicago: Doctoral dissertation.
Steriade, Donca. 1999. Alternatives to the syllabic interpretation of consonantal phonotactics. In *Proceedings of the 1998 Linguistics and Phonetics Conference*, eds. O. Fujimura, B. Joseph and B. Palek, 205-242. Prague: The Karolinum Press.
Steriade, Donca. 2001. The phonology of perceptibility effects: The P-map and its consequences for constraint organization. Ms. Los Angeles.
Tesar, Bruce, and Paul Smolensky. 1994. The learnability of Optimality Theory. In *Proceedings of the Thirteenth West Coast Conference on Formal Linguistics*, eds. Raul Aranovich, William Byrne, Susanne Preuss and Martha Senturia, 122-137. Stanford, CA: CSLI Publications.
Vihman, Marilyn May. 1996. *Phonological Development: The Origins of Language in the Child*. Oxford: Blackwell.

Department of Linguistics
South College
University of Massachusetts
Amherst, MA 01003

flack@linguist.umass.edu