To appear in *Proceedings of NELS 39*. (See published version for correct page numbers.)

A Bidirectional Stochastic OT Analysis of Production and Perception in Yucatec Maya*

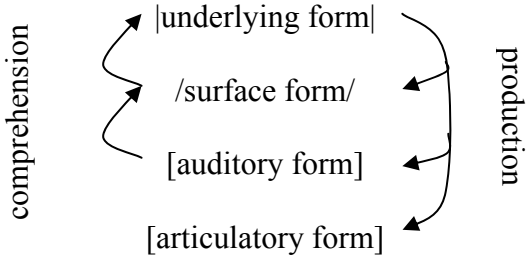
Melissa Frazier

University of North Carolina, Chapel Hill

1. Introduction

Bidirectional Stochastic OT (Boersma 1997, 2006, 2007a,b) provides a compelling model for the analysis of the phonetics-phonology interface. This paper applies this model to the production and perception of pitch and glottalization in Yucatec Maya. In this language, HIGH TONE vowels are produced with initial high pitch, long duration, and modal voice throughout production, while GLOTTALIZED vowels are produced with initial high pitch, long duration, and creaky voice during the middle or last portion of the vowel.[1] Importantly, GLOTTALIZED vowels are produced with higher pitch during the initial portion of the vowel than HIGH TONE vowels (Frazier 2009). Thus, HIGH TONE and GLOTTALIZED vowels differ in the production of both pitch and voice quality.

(1) Boersma's (2007a: 2031) "bidirectional model of phonology and phonetics with parallel production":



In this paper, I test the success of the Bidirectional model in accounting for the phonetic production of pitch and glottalization in the HIGH TONE and GLOTTALIZED

* I would like to thank the participants in both the production and perception experiments, Santiago Domínguez for invaluable assistance in Santa Elena, Jennifer Smith, Elliott Moreton, David Mora-Marin, and Ian Clayton for helpful feedback at various stages of this project, Chris Wiesen for statistical consulting, and the NELS reviewers and audiences. This work was supported by the Luis Quirós Varela Graduate Student Travel Fund and the Jacobs Research Fund (Whatcom Museum, Bellingham, WA).

[1] In this paper, I use small caps to denote underlying categories. Hence, "HIGH TONE" refers to specific underlying category in Yucatec Maya, whereas "high tone" refers to a property of a phonetic output.

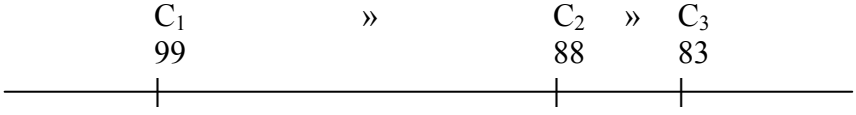
vowels as well as the use of these cues by the listener. As illustrated in (1), the Bidirectional model elucidates how the speaker uses an underlying form to generate an abstract phonological surface form (what is normally considered the output of the phonology) as well as two distinct phonetic forms, which encode both auditory (acoustic) and articulatory information. These same forms (except the articulatory form) are then used by the listener, who first maps an auditory form onto the abstract surface form and then maps this form onto a stored underlying form in order to complete the task of comprehension.

The Bidirectional model predicts that the tasks of production and perception make use of the same *cue constraints* with the same ranking. In this way, the production grammar directly encodes the speaker's desire to be correctly perceived. For example, if the language-user's grammar says that an auditory form of [Y] will likely be heard as /X/ in the task of perception, then the same constraints controlling this preference will also be present in the production grammar, telling the speaker to produce /X/ as [Y]. Now, it may be the case that [Y] is highly marked and will not actually be produced even if its production would guarantee the perception of /X/, because the markedness constraints that penalize certain articulatory forms are not active in perception. This approach reflects the idea, present in the literature on contrast (e.g. Padgett 1997, Flemming 2001), that the production grammar must balance the competing goals of producing distinct phonemes and minimizing effort. I believe the Bidirectional model to be advantageous because it assumes that *comprehension* is the motivation for the *production* of distinct phonemes by working with constraints that function in both tasks, thus getting more mileage with fewer components of the grammar.

Another benefit of Boersma's model is that, because it uses stochastic evaluation, it is able to account for variation. This is especially important when doing phonetic analysis, as variation is always present in continuous phonetic productions. For example, a surface phonological form may be consistently marked for high tone, but individual phonetic forms will be produced with different pitch values. Stochastic OT (StOT) is able to account for both the categorical presence of high tone in the surface form and the variable pitch values present in phonetic forms.

In general, a constraint ranking can be defined by the *ranking value* of each constraint. If C_1 has a larger ranking value than C_2, then C_1 » C_2. In classic OT, constraints have the same ranking value at every point of evaluation, as shown in (2). For this reason, OT analyses usually reference dominance relations instead of ranking values.

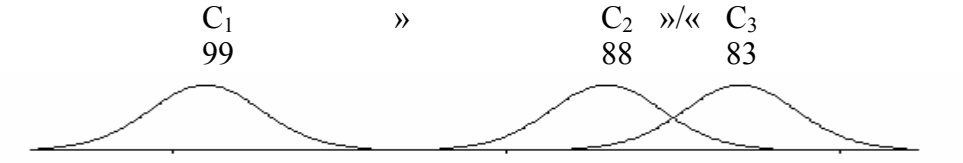
(2)



In StOT, a constraint's rank is defined by a *mean ranking value*. At each point of evaluation, statistical noise is added to each constraint's mean ranking value in order to determine the ranking value of that constraint at that time. This means that the ranking values of a given constraint follow a normal distribution, as shown in (3). Here, we see

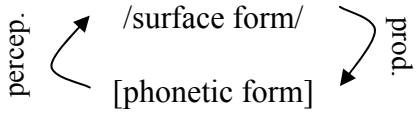
that, because the mean ranking value of C_1 is significantly larger than the mean ranking value of C_2, there is almost zero probability that C_2 will dominate C_1 at any point of evaluation. C_2 and C_3, on the other hand, have mean ranking values that are close enough that, at some points of evaluation, C_2 will dominate C_3, and, at other points of evaluation, C_3 will dominate C_2. At any point of evaluation in StOT, the constraint ranking, and hence the winning candidate, may differ from another point of evaluation.

(3)



Returning to the model in (1), Boersma (2007b: 4) says that "the interface between phonology and phonetics resides in a connection between the phonological surface form and the auditory-phonetic form". Thus, in this paper, I will use the simplified model in (4), where I have relabeled the auditory form as the phonetic form for expository convenience.

(4) phonetics-phonology interface in Bidirectional StOT:



In §4 we will develop a grammar with the Bidirectional StOT model that can account for the distribution of phonetic forms occurring as productions of a specific surface form in Yucatec Maya. We will then turn the production grammar into a perception grammar in order to predict how often certain phonetic forms will be mapped onto certain surface forms. The production and correlated perception grammar show that pitch and glottalization are important cues in the contrast between HIGH TONE and GLOTTALIZED vowels. However, the results of a perception experiment show that participants are only attending to glottalization while performing a discrimination task. It is possible that participants were focusing on glottalization in response to the fact that the stimuli had been manipulated. If so, this result is in line with the evidence that categorical perception is more likely to occur with more natural stimuli (van Hessen and Schouten 1999) and motivates the need for a model of how the perception grammar adjusts to the quality of the language input.

This paper proceeds as follows. I present a brief description of the phonology and phonetics of Yucatec Maya in §2 and then provide further details about the Bidirectional StOT model and the use of the Gradual Learning Algorithm to develop ranking values for constraints in §3. The production grammar of Yucatec Maya is discussed in §4, followed by the perception grammar in §5. I address the implications of this analysis and present conclusions in §6.

2. Phonological and Phonetic Description of Yucatec Maya

Yucatec Maya is a member of the Yucatecan branch of the Mayan language family, spoken by about 700,000 in Yucatan, Campeche, and Quintana Roo, Mexico (1990 census, Gordon 2005). It is the vowel system of this language that is important for our purposes. In Yucatec Maya, the canonical five vowel qualities – [i e a o u] – are contrastive. Additionally, four different combinations of suprasegmental features (length, pitch, and glottalization) can be combined with each vowel quality, resulting in 20 unique possible syllable heads. I refer to each contrastive set of suprasegmental features as a *vowel shape*. Examples and descriptions of each vowel shape are given in (5). Only the HIGH TONE and GLOTTALIZED vowels will be discussed further in this paper.

(5) vowel shapes of Yucatec Maya (Bricker et al. 1998):

SHORT	/v/	*chak*	'red'
LOW TONE	/v̀v/	*chaak*	'boil'
HIGH TONE	/v́v/	*cháak*	'rain'
GLOTTALIZED	/v́v̰/	*cha'ak*	'starch'

According to the abstract surface representations given in (5), we see that both HIGH TONE vowels and GLOTTALIZED vowels are long and are marked by high tone on the initial portion of the vowel, and that GLOTTALIZED vowels are marked by creaky voice on the final portion of the vowel. Phonetic analysis of these vowel shapes (as spoken in isolated target words) shows, though, that the two significantly differ in both pitch and glottalization. For full details of the production experiment that provides these results, see Frazier (2009); in the rest of this section, I present details from a subset of the data collected during that study.[2]

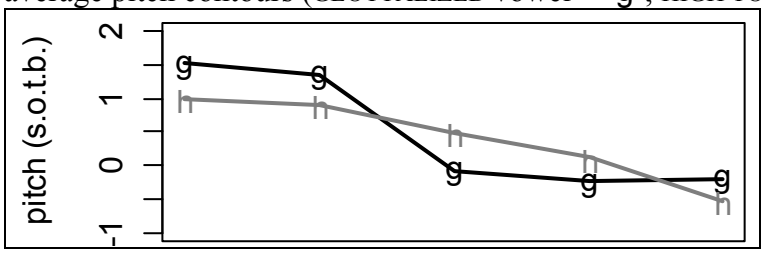
As should be expected, HIGH TONE vowels and GLOTTALIZED vowels differ in the production of glottalization (either creaky voice or a full glottal stop): 3% of HIGH TONE vowels are produced with some form of glottalization, while 50% of GLOTTALIZED vowels are produced with some form of glottalization. When glottalization is produced, it is usually in the form of creaky voice, though 2% of GLOTTALIZED vowels are produced with a glottal stop that interrupts a long modal voiced vowel (i.e. [v́ʔv]).

In addition to glottalization differences, there are also significant pitch differences between the HIGH TONE and GLOTTALIZED vowels such that, during the initial portion of the vowel, pitch is higher in GLOTTALIZED vowels than in HIGH TONE vowels. This is shown in (6), where we see average pitch contours (as defined by five normalized time points) for all participants combined. Here we see that both HIGH TONE and GLOTTALIZED vowels display a falling pitch contour. However, initial pitch is significantly higher in GLOTTALIZED vowels than in HIGH TONE vowels (p=.03,

[2] Only data from participants from Santa Elena, Yucatan is used here, and nonce forms are ignored. Hence, data in this paper comes from 12 participants (5 males, 7 females) who each produced 19 words with a GLOTTALIZED vowel and 20 words with a HIGH TONE vowel.

t(427)=2.21, using a linear mixed regression model to account for multiple observations within subjects).

(6) average pitch contours (GLOTTALIZED vowel = 'g'; HIGH TONE vowel = 'h')



Pitch is measured as semitones over a given speaker's baseline. In order to obtain these measurements, PRAAT (Boersma and Weenink 2006) is used to extract pitch values in Hz, and then Hz is converted to semitones over each speaker's baseline, where the baseline is the average pitch value produced by a given speaker at the mid point of low tone vowels (see the formula in (7)). This conversion allows for cross-speaker comparison regardless of gender (see Nolan 2003 for experimental support of semitones to measure pitch spans and Pierrehumbert 1980 for similar methods in using a baseline that is relative to an individual speaker).

(7) conversion of Hz to semitones over the baseline (s.o.t.b.):
$12*\log_2$(produced Hz/baseline Hz)

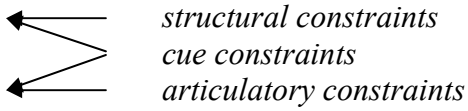
To summarize, we have seen that glottalization is more likely to occur with the GLOTTALIZED vowel (though it occurs only half the time) and that initial pitch is generally higher in GLOTTALIZED vowels than in HIGH TONE vowels. We will see in §4 that the production grammar is able to account for these differences with a high degree of accuracy.

3. Bidirectional Stochastic OT and the Gradual Learning Algorithm

With regard to the phonetics-phonology interface, the production grammar accounts for how the speaker uses a discrete surface form to generate a continuous phonetic form that the speaker actually says, while the perception grammar accounts for how the listener uses the continuous phonetic form that is heard to identify a discrete surface form. The constraints used for the analysis of production and perception are shown in (8). *Cue constraints* act like faithfulness constraints in that they compare two forms (surface and phonetic) and assign violation marks accordingly. For example, the cue constraint "*/a/, [F_1=500 Hz]" says that an /a/ is not paired with an F_1 of 500 Hz and assigns a violation mark any time a surface form with /a/ is paired with a phonetic form with an F_1 of 500 Hz. Importantly, cue constraints assign violation marks whether a surface form is the input (production grammar) or a phonetic form is the input (perception grammar). Thus, cue constraints are used in the analysis of both production and perception. The *structural constraints* and *articulatory constraints* act like markedness constraints in that they look at only one form (surface and phonetic,

respectively) and assign violation marks as applicable. Because markedness constraints can only assign violation marks to output candidates (and not to inputs), articulatory constraints are only applicable in the production grammar, while structural constraints are only applicable in the perception grammar.[3, 4]

(8) constraints for production and perception (Boersma 2007b: 5)



In order to illustrate how these constraints are used in the production and perception grammar, example tableaux are given below. In (9) we see an example tableau for production. Here the input /a/ is mapped onto the winning candidate with an F_1 of 600 Hz. The losing candidates crucially violate cue constraints that penalize the pairing of /a/ with their specific F_1 values. The last candidate also violates the articulatory constraint *[F_1 = 700 Hz]. Note that no candidate can violate the structural constraint */back/ because this constraint penalizes surface forms which are not the candidates in a production tableau.

(9) example production tableau

/a/	*/a/,[F_1=500 Hz]	*/a/,[F_1=700 Hz]	*[F_1=700 Hz]	*/back/	*/a/,[F_1=600 Hz]
[F_1=500 Hz]	*!				
☞ [F_1=600 Hz]					*
[F_1=700 Hz]		*!	*		

The perception tableau in (10) shows how the cue constraints are used in the perception grammar. Additionally, structural constraints assign violation marks to offending surface forms. In this grammar it is the articulatory constraints that have no effect because they assign violations to phonetic forms, which are not the candidates in a perception tableau.

(10) example perception tableau

[F_1=600 Hz]	*/u/,[F_1=600 Hz]	*/o/,[F_1=600 Hz]	*[F_1= 600 Hz]	*/back/	*/a/,[F_1= 600 Hz]
☞ /a/				*	*
/o/		*!		*	
/u/	*!			*	

We have now seen how the production and perception grammars work in the Bidirectional model. Before proceeding to an analysis, we have yet to determine how to develop a StOT constraint ranking. This is a nontrivial issue as StOT rankings are more complex than classic OT rankings. It is not enough to simply derive a dominance relation; we must identify mean ranking values that can be used to define the probability

[3] Structural constraints are of course applicable in the production grammar as a whole, when underlying forms are inputs and surface forms compete as candidates.

[4] Because articulatory constraints assign violation marks to the [articulatory form] and cue constraints compare a /surface form/ to an [auditory form], the [phonetic form] in the simplified model is actually a conflation of both the [articulatory form] and [auditory form].

distributions of various input-output pairings. For this reason, it is useful to employ an algorithm for determining the mean ranking value of each constraint that most accurately accounts for the data.

The Gradual Learning Algorithm (GLA, Boersma and Hayes 2001) can be used to develop a StOT ranking. This algorithm is a model of language acquisition insofar as it models how the learner adjusts an interim constraint ranking when faced with data that contradicts that ranking. As illustrated in (11), the mean ranking value of certain constraints is adjusted when the learning datum (✋) contradicts the winning candidate (☞). In this case the learner's current grammar predicts an incorrect winner. In order to modify the grammar in favor of the learning datum, all constraints that favor the incorrect winner over the learning datum are demoted (mean ranking values are decreased by a small amount) and all constraints that favor the learning datum over the incorrect winner are promoted (mean ranking values are increased by a small amount).

(11) adjustment of mean ranking values with the GLA

	→	→		←	→	←
/surface form/	C_1	C_2	C_3	C_4	C_5	C_6
✋ [phonetic form$_1$]	*	*			*	
☞ [phonetic form$_2$]				*		*

4. Production Grammar

In order to generate a production grammar that accounts for the Yucatec Maya data, learning simulations with the GLA are run with PRAAT (all default settings, see Boersma 1999 or PRAAT manual). I assume that the surface forms of the HIGH TONE and GLOTTALIZED vowels are identical to their underlying forms and hence that no phonology happens that might interfere with the analysis of the phonetics-phonology interface. PRAAT requires two things to run learning simulations: the known (or desired) distribution of paired inputs and outputs and the constraints that compose the grammar (including information about how violation marks are assigned).

In these simulations, the distributions of paired inputs and outputs came directly from the production study. Only GLOTTALIZED vowels and HIGH TONE vowels were possible inputs (henceforth abbreviated /gl/ and /hi/). Outputs consisted of initial pitch values (s.o.t.b.) and glottalization types. For example, a possible output might be [2, glottal stop], which would be a production with an initial pitch of 2 s.o.t.b. and a full glottal stop. The number of times each possible input (/gl/ or /hi/) is paired with this particular output is determined by the number of times participants actually produced such an output when producing a GLOTTALIZED vowel or a HIGH TONE vowel.

Though phonetic forms are ideally continuous, in the sense that a speaker may produce *any* pitch value (in a certain range), whether that value is 1.483 or 1.481 s.o.t.b., it is the case that, for practical purposes, phonetic forms have to be categorized in some way (or there would be an infinite number of candidates and an infinite number of constraints). It would preferable to use categories that match the limits of what humans

distinguish perceptually. However, for simplicity, I will work with rather broad categories. In some ways, these broad categories are against the spirit of the phonetics-phonology interface, but they are necessary in order to develop an understandable grammar in a short space.[5] In the production grammar presented below, I use four possible categories each for initial pitch and glottalization type. Because each type of initial pitch can be combined with each type of glottalization, this yields 16 possible phonetic forms. The four types of glottalization are [modal] (for modal voice throughout the production of the vowel), [short creak] (for creaky voice that is less than or equal to 40 ms long), [long creak] (for creaky voice that is greater than 40 ms long), and [glottal stop] (for the production of a full glottal stop). The four types of initial pitch are abbreviated as [L], [ML], [MH], and [H], which stand for values of ≤ 0, ≤ 1, ≤ 3 and > 3 s.o.t.b., respectively. I use these abbreviations because they are readily recognizable as standing for low, mid-low, mid-high, and high pitch. However, because we are dealing with average values that are at the high end of the pitch scale, the abbreviation [L], for example, does not refer to overall low pitch, but to low pitch relative to the average productions of both HIGH TONE and GLOTTALIZED vowel shapes. In (12) we see the output distributions, given these phonetic categories, for each input vowel shape.

(12) output distributions (from production study)

	L		ML		MH		H		
modal	39	18%	29	13%	22	10%	21	10%	/gl/ n=215
short creak	13	7%	7	3%	9	4%	6	3%	
long creak	22	10%	9	4%	22	10%	11	5%	
glottal stop	3	1%	1	0%	1	0%	0	0%	
modal	103	46%	51	23%	50	22%	15	7%	/hi/ n=225
short creak	0	0%	0	0%	1	0%	0	0%	
long creak	2	1%	1	0%	2	1%	0	0%	
glottal stop	0	0%	0	0%	0	0%	0	0%	

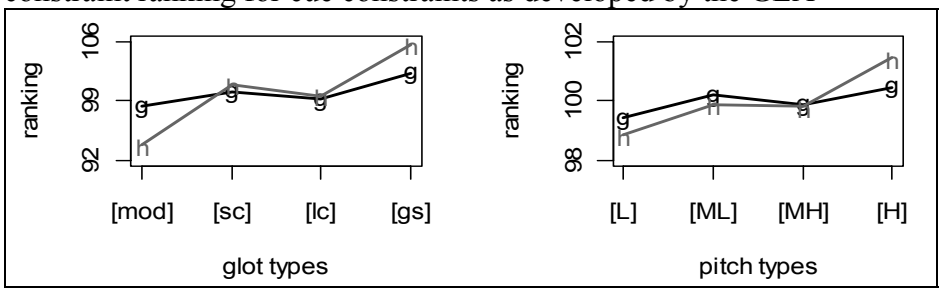
Sixteen cue constraints are used that penalize the pairing of each possible input with each possible output value for pitch or glottalization. Hence, */gl/,[L] is a relevant cue constraint, as are */hi/,[L]; */gl/,[modal]; */hi/,[modal]; etc. Again for simplicity, articulatory and structural constraints are ignored.

When we use this data to run a learning simulation, we get the constraint ranking presented in (13). These graphs show the mean ranking value for each of the cue constraints, e.g. the leftmost 'g' in the first graph tells us that the cue constraint */gl/,[modal] has a mean ranking value of just under 99 (and the mean ranking value of this constraint is higher than the constraint */hi/,[modal]). Because the cue constraints are negatively formulated, low mean ranking values denote preferred input-output pairings. For example, the graph on the left tells us that the pairing of /hi/ with [modal] is more preferable than the pairing of /hi/ with any other glottalization type. This is what we expect, given the known distribution of outputs for HIGH TONE vowels shown in (12).

[5] I have run simulations with more finely-grained constraints that have yielded similar results, but the resulting grammars are too cumbersome to work with for the purposes of this paper.

A closer look at the graphs does not yield any surprises – the types of glottalization and initial pitch that are the most commonly produced for a given input are correlated with those cue constraints that have the lowest mean ranking values.

(13) constraint ranking for cue constraints as developed by the GLA



The graphs above show the mean ranking values for these constraints, and so it is important to remember that statistical noise is added to these means at every point of evaluation. Just because */hi/,[modal] has the lowest mean ranking value does not mean that a production with [modal] will win every time the input is /hi/. In order to illustrate this point, and to further illustrate how the production grammar works, an example tableau for a single point of evaluation is given in (14). At this point of evaluation, for the input /gl/, the winning candidate is [L, modal]. Note that, for example, if the constraint */gl/,[L] had a ranking value of only half a point higher, [H, modal] would be the winner. At some other point of evaluation, this would be the case.

(14) example production tableau for single point of evaluation

	103.2	103.0	101.8	100.3	100.2	99.5	99.2	98.8
/gl/	*/gl/, [ML]	*/gl/, [gs]	*/gl/, [lc]	*/gl/, [MH]	*/gl/, [sc]	*/gl/, [H]	*/gl/, [L]	*/gl/, [mod]
☞[L, mod]							*	*
[L, sc]					*!		*	
[L, lc]			*!				*	
[L, gs]		*!					*	
[ML, mod]	*!							*
[ML, sc]	*!				*			
[ML, lc]	*!		*					
[ML, gs]	*!	*						
[MH, mod]				*!				*
[MH, sc]				*!	*			
[MH, lc]			*!	*				
[MH, gs]		*!		*				
[H, mod]						*!		*
[H, sc]					*!	*		
[H, lc]			*!			*		
[H, gs]		*!				*		

We have now defined a StOT grammar that can be used to predict the output distribution for a given input. In (15) we see a side-by-side comparison of the known

output distributions as obtained through the production study and the predicted output distributions as defined by the grammar developed through the GLA. What this table shows us is that the StOT grammar generated by the GLA predicts output distributions that closely mimic those obtained through experimentation. Therefore, the GLA and StOT can handle real linguistic data to develop a constraint ranking that accurately predicts the phonetic output distributions for a given surface form.

(15) percentage of times a specific input is paired with a specific phonetic form: empirical percentage (bold) compared to predicted percentage (italics)

	L		ML		MH		H		
modal	**18**	*17*	**13**	*10*	**10**	*13*	**10**	*9*	/gl/
short creak	**6**	*6*	**3**	*4*	**3**	*4*	**3**	*3*	
long creak	**10**	*10*	**4**	*6*	**10**	*8*	**5**	*5*	
glottal stop	**1**	*1*	**0**	*1*	**0**	*1*	**0**	*0*	
modal	**46**	*44*	**23**	*23*	**22**	*23*	**7**	*7*	/hi/
short creak	**0**	*0*	**0**	*0*	**0**	*0*	**0**	*0*	
long creak	**1**	*1*	**0**	*0*	**1**	*0*	**0**	*0*	
glottal stop	**0**	*0*	**0**	*0*	**0**	*0*	**0**	*0*	

5. Perception Grammar

Because the cue constraints are used in both the production and perception grammars, the ranking developed above also defines a perception grammar for Yucatec Maya. In the perception grammar, the listener takes the phonetic form that is heard and maps it onto a surface form. Specifically, a phonetic form (e.g. [L, modal]) is the input and the surface forms (/gl/, /hi/) compete as output candidates.

Note that the cue constraints that conflict in perception are not the same as the constraints that conflict in production. In production, all constraints that penalize the pairing of a specific surface form with any phonetic form conflict, but in perception it is all the constraints that penalize the pairing of a specific phonetic form with any surface form that compete. Referring back to (13), constraints on a given line (e.g. all constraints denoted by a '*g*') conflict in production, while all constraints in a given column (e.g. constraints in the [modal] column) conflict in perception.

Using the mean ranking values for the cue constraints as determined in §4, we can predict the percentage of times a given phonetic form will be mapped onto one of the possible surface forms in perception. In (16), we see how often each possible phonetic form is predicted to be perceived as a GLOTTALIZED vowel. Because a HIGH TONE vowel is the only other option, any time a phonetic form is not perceived as a GLOTTALIZED vowel, it is perceived as a HIGH TONE vowel, e.g. the phonetic form [L, modal] is perceived as a GLOTTALIZED vowel 32% of the time and, thus, as a HIGH TONE vowel 68% of the time.

(16) percentage of times the perception grammar predicts that a given input (phonetic form) will be heard as a GLOTTALIZED vowel /gl/:

	L	ML	MH	H
modal	32	39	41	59
short creak	55	55	57	65
long creak	49	49	52	63
glottal stop	88	87	88	88

According to the predictions made in (16), both a glottal stop and creaky voice increase the probability of the stimulus being perceived as a GLOTTALIZED vowel (the glottal stop more so). Furthermore, as pitch increases, the probability of perceiving a GLOTTALIZED vowel increases (except in conjunction with a glottal stop).

This prediction is tested with a perception experiment involving 14 native speakers of Yucatec Maya living in Santa Elena, Yucatan, Mexico (5 males (ages 23, 43, 44, 64, 69); 9 females (ages 21, 21, 21, 23, 26, 31, 34, 38, 65)). Most participants had only lived in Santa Elena, and all were fluent in Spanish, while two were also fluent in English. The perception experiment occurred about one year after the production experiment, and some subjects participated in both experiments.

Participants performed a forced choice task, where they heard a stimulus and were asked to choose between a word with a HIGH TONE vowel and a word with a GLOTTALIZED vowel. Two minimal pairs were used: *k'a'an* 'strong' vs. *k'áan* 'hammock' and *cha'ak* 'starch' vs. *cháak* 'rain'. Stimuli were manipulated from one production of *k'an* 'ripe' and of *chak* 'red' (SHORT vowel with mid pitch and modal voice, produced by a male from Mérida, Yucatan). For each original production, 16 stimuli were manipulated that had each combination of four types of glottalization and four values for initial pitch. Pitch periods were added to the original productions until the vowels were about 200 ms long. Four different pitch contours were created with PRAAT such that the vowel began with 125, 140, 155, or 170 Hz (-0.7, 1.2, 3.0, 4.6 semitones over this speaker's baseline). This pitch value continued for 75 ms, and then fell over the rest of the vowel until reaching 110 Hz at the end of the vowel. Because the original productions were not glottalized, no manipulation was needed for the modal voice category of glottalization. In order to create a short and long duration of creaky voice, the pitch contours were altered to contain a shorter and longer duration of pitch at 35 Hz. In order to mimic a production with a glottal stop, the middle portion of the vowel was removed and replaced with 75 ms of silence.

Participants heard each manipulated stimulus embedded in the frame sentence *Tin wa'alaj* __. 'I said __.' (as spoken naturally by the same male). A frame sentence was used to give listeners some familiarity with the speaker's natural pitch range.[6] Each stimulus was heard three times, for a total of 96 trials (2 minimal pairs x 16 stimuli x 3

[6] Even though the production data comes from words spoken in isolation, Frazier (in preparation) shows that the pitch contours and distribution of glottalization of HIGH TONE and GLOTTALIZED vowels as spoken in a similar frame sentence (*Tu ya'alaj* __. 'S/he said __.") are highly similar to those as spoken in isolation.

repetitions). The results presented below exclude half the data obtained from three different participants. These participants always selected *cháak* (HIGH TONE vowel) when given the choice of *cháak* vs. *cha'ak*, and so the 48 trials involving this minimal pair for these three participants were rejected. In (17) we see the percentage of times that participants selected a word with a GLOTTALIZED vowel for each type of stimulus. Again, if a GLOTTALIZED vowel was not selected, a HIGH TONE vowel was.

(17) percentage of time participants selected a word with a GLOTTALIZED vowel

	L (-0.7)	ML (1.2)	MH (3.0)	H (4.6)
modal	27%	25%	23%	29%
short creak (≈20 ms)	44%	44%	41%	37%
long creak (≈70 ms)	61%	63%	63%	69%
glottal stop	79%	85%	77%	83%

The results show a significant effect of glottalization ($p < .0001$, Wald $\chi^2(3)=189.2$), a nonsignificant effect of pitch ($p = .64$, Wald $\chi^2(3)=1.67$), and a nonsignificant interaction ($p = .92$, Wald $\chi^2(9)=3.81$). This means that even though productions of GLOTTALIZED and HIGH TONE vowels differ by both pitch and glottalization and the perception grammar that correlates with an accurate production grammar predicts pitch to be a cue, listeners are choosing by glottalization alone and not pitch. This is made clear in (18), where we see a side-by-side comparison of the empirical and predicted results.

(18) percentage of times a given stimulus is heard as a GLOTTALIZED vowel: empirical results (bold) and predicted results (italics)

	L		ML		MH		H	
modal	**27**	*32*	**25**	*39*	**23**	*41*	**29**	*59*
short creak	**44**	*55*	**44**	*55*	**41**	*57*	**37**	*65*
long creak	**61**	*49*	**63**	*49*	**63**	*52*	**69**	*63*
glottal stop	**79**	*88*	**85**	*87*	**77**	*88*	**83**	*88*

The perception grammar, as developed by the GLA, seems to make an incorrect prediction. It predicts that listeners will use pitch as a cue to the distinction between GLOTTALIZED and HIGH TONE vowels (with the former having higher pitch), but native listeners are not using this acoustic parameter when making their decision.

6. Implications and Conclusions

We learned in §5 that the perception grammar correlated with a highly accurate production grammar makes an incorrect prediction. Pitch is predicted to be a cue used by listeners of Yucatec Maya in distinguishing the contrast between HIGH TONE and GLOTTALIZED vowels. However, listeners are not using this cue. To generalize, it seems that speakers of Yucatec Maya are actively controlling a phonetic parameter that is not used by the listener.

It is possible that the mismatch between the production and perception grammars

could be resolved by including structural and articulatory constraints in the model. Specifically, there might be articulatory motivation for the production of certain acoustic values, and as this would be regulated by articulatory constraints, it would not influence perception. This explanation seems unlikely in the present case however, because, if we expect very high pitch to be marked, we have a situation where speakers prefer *marked* phonetic forms, not unmarked ones.

One avenue worth exploring is the relation between the manipulated stimuli used in the perception experiment and the natural stimuli obtained through the production experiment. Perhaps the perception grammar developed through the GLA is not wrong, it is just inappropriate for use with manipulated data. One reason to suspect this comes from the literature on categorical perception. Van Hessen and Schouten (1999) show that, as stimulus quality increases, so does categorical perception. They conclude that stimuli of higher complexity distract listeners from focusing on specific acoustic parameters.

One way to explain the noted discrepancy, then, would be to say that the manipulated stimuli somehow caused listeners to focus on glottalization, whereas with 'real-world' stimuli, their attention would be pulled between glottalization and pitch. The question then becomes how the grammar might account for these different perceptual strategies. It is clear from the data that participants were not randomly guessing nor were some using the cue of glottalization while others were using the cue of pitch: all participants made their decisions on the basis of glottalization and not pitch. This means that the grammar of Yucatec Maya must explain this behavior. It will thus be important for future work to investigate how the grammar decides which cues to use for the perception of not just natural stimuli but also less-natural stimuli. This is important not just for language use in the laboratory, but also for language use in less-than-ideal settings. Listeners are not always faced with perfectly produced utterances and must be able to comprehend speech in settings with loud background noise, or as shouted over distances, etc. We wish, then, to continue to explore how the grammar accounts for perceptual differences as a result of stimulus quality.

This paper has applied the GLA to real language data to develop a StOT ranking that accounts for production with a high degree of accuracy. A perception experiment was conducted to test the prediction of the Bidirectional StOT model that perception and production can be accounted for with the same constraints. Though the results of the perception experiment were not predicted by the perception grammar (with the same constraints and rankings as the production grammar), it is possible that this mismatch occurred because the stimuli were manipulated. It is thus important for future work to model how the perception grammar reacts to stimulus quality.

References

Barnes, Jonathan. 2006. Strength and Weakness at the Interface: Positional Neutralization in Phonetics and Phonology. New York: Mouton de Gruyter.

Boersma, Paul. 1997. How We Learn Variation, Optionality, and Probability. In

Proceedings of the Institute of Phonetic Sciences 21, 43–58. Amsterdam: University of Amsterdam.

Boersma, Paul. 1999. Optimality-Theoretic Learning in the PRAAT Program. In Proceedings of the Institute of Phonetic Sciences 23, 17-35. Amsterdam: University of Amsterdam.

Boersma, Paul. 2006. Prototypicality Judgments as Inverted Perception. In Gradience in Grammar, ed. Gisbert Fanselow, Caroline Féry, Matthais Schlesewsky and Ralf Vogel, 167-184. Oxford: Oxford University Press.

Boersma, Paul. 2007a. Some Listener-Oriented Accounts of *h*-Aspiré in French. Lingua 117, 1989-2054.

Boersma, Paul. 2007b. Cue Constraints and their Interaction in Phonological Perception and Production. Ms. University of Amsterdam. [ROA 944.]

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

Boersma, Paul and David Weenink. 2006. PRAAT: Doing Phonetics by Computer (Version 4.4.05) [Computer program]. Retrieved from http://www.praat.org/.

Bricker, Victoria, Eleuterio Poʔot Yah, and Ofelia Dzul Poʔot. 1998. A Dictionary of the Maya Language As Spoken in Hocabá, Yucatán. Salt Lake City: University of Utah Press.

Flemming, Edward. 2001. Scalar and Categorical Phenomena in a Unified Model of Phonetics and Phonology. Phonology 18, 7-44.

Frazier, Melissa. 2009. Tonal Dialects and Consonant-Pitch Interaction in Yucatec Maya. In New Perspectives in Mayan Linguistics, WPLMIT 59, ed. Heriberto Avelino, Jessica Coon, and Elisabeth Norcliffe, 59-82. Cambridge, Mass.: MIT.

Gordon, Raymond G., Jr. (ed.), 2005. Ethnologue: Languages of the World, Fifteenth edition. Dallas, Tex.: SIL International. Online version: http://www.ethnologue.com/.

Nolan, Francis. 2003. Intonational Equivalence: an Experimental Evaluation of Pitch Scales. In Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona, 771-774.

Padgett, Jaye. 1997. Perceptual Distance of Contrast: Vowel Height and Nasality. Phonology at Santa Cruz 5, 63-78.

Pierrehumbert, Janet. 1980. The Phonology and Phonetics of English Intonation. Doctoral dissertation, MIT, Cambridge, Mass.

van Hessen, A.J. and S.E.H. Schouten. 1999. Categorical Perception as a Function of Stimulus Quality. Phonetica 56, 56-72.

Department of Linguistics
128 Smith, CB #3155
University of North Carolina
Chapel Hill, NC 27599-3155

melfraz@email.unc.edu