PRODUCING, PERCEIVING, AND LEARNING VARIATION: PITCH AND GLOTTALIZATION IN YUCATEC MAYA

Melissa Frazier
University of North Carolina at Chapel Hill
melfraz@email.unc.edu

Linguistic Society of America Annual Meeting
January 10, 2009

1. INTRODUCTION

(1) work on the phonetics-phonology interface accounts for the relation between abstract phonological forms and continuous phonetic forms
   a. the language-learner develops phonological categories from variable acoustic input
   b. in the adult production grammar, a single phonological input is mapped onto variable phonetic outputs
   c. in the adult perception grammar, variable acoustic inputs are mapped onto some phonological category

(2) in this paper, we analyze the interaction of pitch and glottalization in Yucatec Maya (a Mayan language of Mexico) and how these cues contribute to contrast in this language's vowel system
   a. primary concern: modeling an accurate production and perception grammar
   b. secondary concern: implications of necessary components of the grammars for language acquisition

(3) vowel contrasts in Yucatec Maya (Bricker et al. 1998)
   a. quality: [i e a o u]
   b. suprasegmental contrasts (*vowel shape*):

• short	/v/
• long, low tone	/v̀v/
• long, high tone	/v́v/
• long, glottalized	/v́v̰/

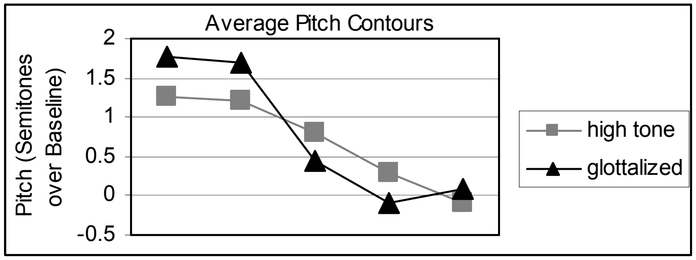
   c. focus of this paper: high tone and glottalized vowels – the production and perception of pitch and glottalization

(4) research questions and preview of answers:
   a. In production, how is the contrast between high tone and glottalized vowels realized?
      ➢ both pitch and glottalization are significantly different in productions of high tone and glottalized vowels
   b. In perception, what cues do listeners use to distinguish between the high tone and glottalized vowels?
      ➢ when faced with natural stimuli, listeners use both pitch and glottalization
      ➢ when faced with manipulated stimuli, listeners use only glottalization
   c. How does the grammar account for the perception of both natural and less-natural stimuli?
      ➢ the language-user has knowledge of which cues are better predictors of underlying form and uses only these cues when faced with less natural stimuli
   d. How does the language learner develop this knowledge?

(5) outline of talk
   a. data from production and perception experiments with native Yucatec Maya speakers
   b. modeling the production and perception of pitch and glottalization
   c. discussion and conclusions

2. THE PRODUCTION AND PERCEPTION OF PITCH AND GLOTTALIZATION IN YUCATEC MAYA

(6) production of high tone and glottalized vowels (Frazier to appear)
   a. due to dialect variation in the production of pitch, only participants from Santa Elena, Yucatán, México are reported on in this paper (applies to both production and perception studies)
   b. participants read target words in isolation
   c. glottalization: glottalized vowels are produced with either creaky voice or a full glottal stop 50% of the time; 3% of high tone vowels are produced with creaky voice
   d. pitch: both vowel shapes start with high pitch, but the glottalized vowels start with significantly higher pitch than the high tone vowels (p=.03, t(427)=2.2, using a linear mixed regression model to account for multiple observations within subjects)



e. high tone and glottalized vowel shapes differ by both pitch and glottalization: glottalized vowels are produced with higher pitch initially and with more glottalization

(7) a note on the measurement of pitch

a. PRAAT (Boersma and Weenink 2006) used to extract pitch values in Hz, then Hz converted to semitones (Nolan 2003), relative to each speaker’s *baseline* (similar to methods of Pierrehumbert 1980), where the baseline is the average pitch value produced by a given speaker at the mid point of low tone vowels
b. equation: $12*\log_2$(produced Hz/baseline Hz)
c. e.g.: a pitch measurement of 2.7 indicates that pitch is 2.7 semitones higher than that speaker’s baseline
d. allows for direct comparison of measurements from males and females

(8) implications of production results

a. the Yucatec Maya language-learner hears input with high tone and glottalized vowels that (1) is variable, (2) contains acoustic values that are permissible in both categories, and (3) uses multiple cues to signal contrast
b. the production grammar must be able to account for the variable nature of the phonetic forms
c. Can the listener successfully distinguish high tone from glottalized vowels? (see Yu 2007 for a case where productions differ significantly but listeners cannot correctly identify which production they heard); If so, what cues does the listener use?

(9) two perception experiments were conducted in Yucatán, México

a. participants: 14 native speakers of Yucatec Maya living in Santa Elena, Yucatán, México
b. 5 males (ages 23, 43, 44, 64, 69); 9 females (ages 21, 21, 21, 23, 26, 31, 34, 38, 65)
c. most had only lived in Santa Elena
d. all fluent in Spanish; two also fluent in English
e. perception experiments occurred one year after production experiment; some subjects participated in both experiments

(10) perception experiment 1: methodology

a. forced choice task: participants heard an unaltered token of either *k’a’an* [k’áa̰n] ‘strong’ or *k’áan* [k’áan] ‘hammock’ and were asked to choose which word they heard
b. the 48 stimuli came from the productions of these words as spoken by all 24 participants of the production study
c. participants heard each stimulus once

(11) perception experiment 1: results

a. participants performed better than chance at selecting the word that matches what the speaker intended to say (*k’a’an* heard as *k’a’an* and *k’áan* heard as *k’áan* 63% of the time, Rao-Scott χ^2 =17.56, p<.0001)
b. participants used both pitch and glottalization to make their decision
   - participants were more likely to select a word with a glottalized vowel if the stimulus had creaky voice and even more so if the stimulus had a glottal stop

glottalization type of stimulus	participant’s selection	
	k’a’an	k’áan
modal	164 (39%)	256 (61%)
creaky	146 (65%)	78 (35%)
glottal stop	26 (93%)	2 (7%)
	Rao-Scott χ^2 =34.5, p < .0001	

   - participants were more likely to select a word with a glottalized vowel if pitch is higher during the first quarter of the vowel, or if pitch is lower during the last quarter of the vowel (for normalized time points where 1 is the start of the vowel and 5 is the end of the vowel; time 1: z=3.36, p < .01; time 2: z=3.42, p < .01; time 4: z= -2.11, p = .03; effect of pitch is nonsignificant at time 3 and 5)

(12) perception experiment 2: methodology

a. forced choice task: participants heard a stimulus and were asked to choose between a word with a high tone vowel and a word with a glottalized vowel (either *k'a'an* 'strong' vs. *k'áan* 'hammock' or *cha'ak* 'starch' vs. *cháak* 'rain')

b. 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, Yucatán)

c. 16 manipulated stimuli for each minimal pair (32 stimuli total)

- four types of glottalization (all modal, ≈30 ms of creaky voice, ≈70 ms of creaky voice, full glottal stop)
- four values for initial pitch (125, 140, 155, 170 Hz; -0.7, 1.2, 3.0, 4.6 semitones over baseline)

d. vowels of all manipulated stimuli ≈200 ms long

e. each stimulus was embedded in the frame sentence *Tin wa'alaj* __. 'I said __.', which was used to give the listener familiarity with the speaker's voice/normal pitch range[1]

f. participants heard each stimulus three times, for 96 trials

g. rejected data: three participants always selected *cháak* when given the choice of *cháak* vs. *cha'ak*, and so these responses were not included in this analysis (48 rejected trials for 3 participants)

(13) perception experiment 2: results

a. percentage of time participants selected a word with a glottalized vowel (all other times a high tone vowel was chosen)

	L	ML	MH	H
modal	27%	25%	23%	29%
short creak	44%	44%	41%	37%
long creak	61%	63%	63%	69%
glottal stop	79%	85%	77%	83%

b. significant effect of glottalization ($p < .0001$, Wald $\chi^2(3)=189.2$), nonsignificant effect of pitch ($p = .64$, Wald $\chi^2(3)=1.67$), nonsignificant interaction ($p = .92$, Wald $\chi^2(9)=3.81$)

(14) by categorizing pitch and glottalization of the stimuli from perception experiment 1 in the same was as they are categorized for perception experiment 2, we can directly compare the results:

percentage of times a glottalized vowel was chosen in perception experiment 1 (bold) and perception experiment 2 (italics)

	L		ML		MH		H	
modal	**41**	*27*	**28**	*25*	**41**	*23*	**60**	*29*
short creak	**57**	*44*	**n/a**	*44*	**54**	*41*	**n/a**	*37*
long creak	**75**	*61*	**43**	*63*	**65**	*63*	**86**	*69*
glottal stop	**86**	*79*	**n/a**	*85*	**100**	*77*	**n/a**	*83*

3. Modeling Production and Perception

(15) review of results

a. in production, pitch and glottalization distinguish high tone from glottalized vowels

b. when listeners hear natural stimuli, perception is influenced by both pitch and glottalization

c. when listeners hear manipulated stimuli, perception is only influenced by glottalization

(16) the phonetics-phonology interface models how the production grammar uses a stored abstract form to generate a continuous phonetic output and how the perception grammar uses a continuous phonetic input to identify a stored abstract form

a. How does the production grammar of Yucatec Maya account for the distribution of glottalization types and initial pitch values associated with high tone and glottalized vowels?

b. How does the perception grammar use the cues of pitch and glottalization to discriminate between high tone and glottalized vowels? How does the grammar account for the results of both perception experiments? Does it need to?

(17) the different results could be related to stimulus quality: van Hessen and Schouten (1999) show that there is an increase in categorical perception as stimulus quality increases:

> "Because considerably more information was available [in tokens of natural speech] … listeners just could not focus their attention on one aspect of the stimuli…; they had to listen to the full spectrum and all its subtle, interacting cues, which is what they normally do." p. 58

a. the perception grammar must be able to account for the use of multiple cues (as demonstrated by perception experiment 1)

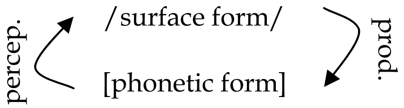
---

[1] Even though production data comes from words as spoken in isolation, Frazier (in preparation) shows that the pitch contours and glottalization types of words spoken in a similar frame sentence (*Tu ya'alaj* __. 'S/he said __.') closely match those of words spoken in isolation.

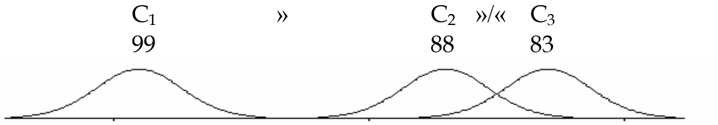
b. Are the results of perception experiment 2 also a function of the (language-specific) grammar – or can they be explained by universal laws (of perception, etc.)?
  - more experimentation (with non-Yucatec Maya speakers) would be needed to determine if this process is language-specific
  - each participant behaved in the same way (using pitch in experiment 1 but not in experiment 2), so if the process is not universal, it is certainly part of the grammar

(18) model of analysis: Bidirectional Stochastic OT (Boersma 1997, 2006, 2007a-b)

a. simplified model (to account for only the phonetics-phonology interface and not for any phonology that might happen between the underlying form and phonological surface form):

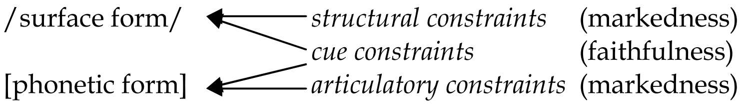


b. Stochastic OT can account for variation: in StOT, a constraint's rank is defined by a *mean ranking value*; at a single point of evaluation, statistical noise is added to each constraint's mean ranking value:



c. at any point of evaluation, the constraint ranking, and hence the winning candidate, may differ from another point of evaluation

d. the Bidirectional model makes use of the same constraints (with the same mean ranking values) in production and perception:



e. in production, a surface form is the input and phonetic forms compete as output candidates; articulatory constraints assign violation marks to outputs and cue constraints assign violation marks to offending input-output pairs:

example production tableau

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

f. in perception, a phonetic form is the input and surface forms compete as output candidates; structural constraints assign violation marks to offending outputs and cue constraints (again) assign violation marks to offending input-output pairs:

example perception tableau

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

g. cue constraints do work in both the production and perception grammars: predicts that a surface form will be correlated with certain acoustic values in both production and perception

(19) learning a StOT ranking: the Gradual Learning Algorithm (Boersma and Hayes 2001)

a. the GLA models how the learner adjusts an interim constraint ranking when faced with data that contradicts that ranking

b. initial state: all constraints have same ranking value

c. the mean ranking value of certain constraints is adjusted when the learning datum (✋) contradicts the winning candidate (☞) (the learner's current grammar predicts an incorrect winner)

d. all constraints that favor the "incorrect winner" (☞) over the learning datum (✋) are demoted and all constraints that favor the learning datum over the incorrect winner are promoted.

	→	→		←	→	←
/s. f./	C_1	C_2	C_3	C_4	C_5	C_6
✋ [$cand_1$]	*	*			*	
☞ [$cand_2$]				*		*

(20) in order to develop a StOT ranking of cue constraints, PRAAT is used to run learning simulations with the GLA (all default settings, see Boersma 1999 or the PRAAT manual)

a. only "glottalized vowels" and "high tone vowels" are possible inputs: /gl/, /hi/

b. outputs consist of pairs of initial pitch values and glottalization types, which are classified with the same four categories as used in perception experiment 2

c. the distributions for specific input and output pairings come from the production study

d. 16 cue constraints are used, with one cue constraint penalizing the pairing of each possible input with each possible output (as defined for a specific phonetic dimension) e.g., because [modal] is a glottalization type, we need the cue constraints: */gl/,[modal] and */hi/,[modal], etc. for each acoustic category

e. after using the GLA to determine mean ranking values for the cue constraints, this grammar can be use to predict *output distributions*

- In the production grammar, how often is a given surface form mapped onto each phonetic output?
- In the perception grammar, how often is a given phonetic form mapped onto each surface form?

(21) the production grammar defined by the cue constraints (with mean ranking values as determined by the GLA) predicts output distributions that closely mimic those obtained empirically:

empirical output distributions (bold) compared to predicted output distributions (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*	

(22) when the same cue constraints (with the same mean ranking values) are used to define a perception grammar, we get the following predicted output distributions:

percentage of times the perception grammar predicts 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

a. resembles the results of perception experiment 1: an increase in pitch and an increase in glottalization are correlated with a higher likelihood of being perceived as a glottalized vowel

b. the Bidirectional model can accurately account for the production and perception of pitch and glottalization

(23) How do we account for the results of perception experiment 2?

a. if we "turn off" the cue constraints that penalize certain pitch values (i.e. if they assign no violation marks), we get the following prediction:

percentage of times a given input will be heard as a glottalized vowel /gl/:

modal	5
short creak	47
long creak	64
glottal stop	91

b. the predicted values are more extreme at the edges (lower predicted percentage for [modal] and higher predicted percentage for [glottal stop]), but the general pattern matches the results from perception experiment 2

(24) Why is glottalization used instead of pitch?

a. glottalization is more closely associated with underlying form than pitch is

	pitch values			
	L	ML	MH	H
%/gl/	44	45	48	74

	glottalization values			
	modal	short creak	long creak	glottal stop
%/gl/	33	91	97	100

b. this association is learned during the language acquisition process and is used to alter the grammar for non-ideal language situations

c. this ‘altered grammar’ is not just relevant for laboratory settings with manipulated/synthesized stimuli but for ‘real-world’ scenarios with non-optimal conditions for language perception

4. Discussion and Conclusions

(25) predictions and theoretical implications of this analysis

a. production: if the perception grammar that accounts for less-than-ideal stimuli is also ‘bidirectional’, then we expect Yucatec Maya speakers to emphasize glottalization when worried the listener will misinterpret the signal

b. diachronic: because sound change is often considered to result from various types of misperception, it is important for theories of sound change to take into account all aspects of the perception grammar

(26) testing Bidirectional Stochastic OT and the Gradual Learning Algorithm

- the GLA is capable of using real language data to develop StOT rankings that accurately account for the use of pitch and glottalization by both the speaker and the listener in Yucatec Maya according to the Bidirectional model

(27) the grammar of a language must be able to account for how the listener adjusts to less-than-ideal stimuli

- as learners develop a grammar that uses multiple cues for the perception of contrast, they also learn which cues are the most reliable and use only these when faced with less-than-ideal stimuli

References

Boersma, Paul. 1997. How We Learn Variation, Optionality, and Probability. *Proceedings of the Institute of Phonetic Sciences*, University of Amsterdam 21: 43–58.

Boersma, Paul. 1999. Optimality-Theoretic Learning in the PRAAT Program. *IFA Proceedings* 23: 17-35.

Boersma, Paul. 2006. Prototypicality Judgments as Inverted Perception. In G. Fanselow, C. Féry, M. Schlesewsky & R. Vogel (eds.), *Gradience in Grammar*, 167-184. Oxford: Oxford University Press.

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

Boersma, Paul. 2007b. Cue Constraints are 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.

Frazier, Melissa. To appear. “Tonal Dialects and Consonant-Pitch Interaction in Yucatec Maya.” In H. Avelino, J. Coon, and E. Norcliffe (eds.), *New Perspectives in Mayan Linguistics. WPL-MIT*.

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

Pierrehumbert, Janet. 1980. The Phonology and Phonetics of English Intonation. Ph.D. dissertation, MIT.

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

Yu, Alan C.L. 2007. Understanding Near Mergers: the Case of Morphological Tone in Cantonese. *Phonology* 24: 187-214.

Acknowledgements

I thank the participants in the production and perception studies, Santiago Domínguez for all of his assistance in Yucatán, and Jennifer L. Smith, Elliott Moreton, David Mora-Marin, and the audience and reviewers at NELS 39 for helpful feedback.

This work was supported by the Jacobs Research Fund (Whatcom Museum, Bellingham, WA) and the Luis Quirós Varela Graduate Student Travel Fund (supplemented by the ISA Mellon Dissertation Fellowship at UNC-CH).