Pitch and Glottalization as Cues to Contrast in Yucatec Maya[*]

Melissa Frazier

Abstract

Production experiments with Yucatec Maya show that two phonemic categories – HIGH TONE and GLOTTALIZED vowels – are systematically distinguished by pitch, glottalization, and (to a lesser extent) duration. In performing a discrimination task with natural stimuli, listeners make use of all of these cues in determining which vowel sound they heard. This result supports grammars that use bidirectional constraints with the same ranking values in both production and perception. Listeners reacted differently to manipulated stimuli by attending to glottalization alone. As glottalization is the phonetic dimension that is most correlated with this contrast, this result suggests that the perception grammar is adapted in degraded linguistic situations to attend to the cue that is most likely to lead to success.

1. Introduction

Yucatec Maya (a Mayan language of Mexico, henceforth abbreviated as YM) is one of the few Mayan languages to use a tonal contrast. Long vowels are produced with either low tone (e.g. /mìis/ *cat*) or high tone (e.g. /míis/ *sweep*). There is also a third type of long vowel used in YM; GLOTTALIZED vowels are produced with initial high pitch and with creaky voice. The results of production experiments show that HIGH TONE and GLOTTALIZED vowels differ in several phonetic dimensions: GLOTTALIZED vowels tend to have higher initial pitch, to have a larger pitch span, to be produced with more creaky voice, and to be longer than HIGH TONE vowels. The fact that multiple phonetic dimensions are at play in this phonological contrast has implications for what strategies listeners will use in differentiating the contrast as well as for how to model the phonological grammar and its relation to phonetics. These implications are explored in this paper via two perception experiments conducted in Yucatan, Mexico.

It has been demonstrated for various phonemic contrasts that the same phonetic properties that distinguish different phonemes in production are used by the listener as cues to perceiving the contrast. For example, both F1 and duration systematically differ in the productions of tense and lax vowels in English (Peterson & Barney 1952; Peterson & Lehiste 1960), though the reliability of such cues varies by dialect. Escudero & Boersma (2004) show that speakers of different dialect groups attend to those cues that are the most correlated with the contrast in that dialect: Scottish English speakers distinguish /i/ from /ɪ/ primarily by F1 and only minimally by duration, whereas duration is predominantly used by Southern British English speakers. As another example, the average VOT (for both voiced and voiceless stops) is greater in English than in Spanish, and when performing a discrimination task, the crossover point (the point on the VOT continuum that begins to elicit more "voiceless" responses than "voiced" responses) has a higher VOT value for English listeners than for Spanish listeners (Liberman et al. 1958; Lisker & Abramson 1964, 1970; Cho & Ladefoged 1999).

[*] I would like to thank Grant McGuire, David Mora-Marin, Elliott Moreton, and Jennifer L. Smith for providing discussion and comments on various aspects of this paper and Christopher Wiesen for providing statistical consulting. All mistakes are of course my own. The data presented here previously appeared in my doctoral dissertation (Frazier 2009). This work was supported by the Luis Quirós Varela Graduate Student Travel Fund (Institute for the Study of the Americas at the University of North Carolina, Chapel Hill) and the Jacobs Research Fund (Whatcom Museum, Bellingham, WA).

Given the above facts, one would expect YM listeners to make use of all available phonetic dimensions – pitch, glottalization, and vowel duration – when distinguishing between HIGH TONE and GLOTTALIZED vowels. This expected result is confirmed when listeners respond to natural stimuli produced by native speakers. However, listeners reacted differently to manipulated stimuli by attending to glottalization alone and ignoring pitch. I argue that the manipulated stimuli forced participants to focus on only one phonetic dimension, and so listeners made use of the phonetic cue that is the most reliable at distinguishing the contrast between HIGH TONE and GLOTTALIZED vowels. This means that the grammar of a language must not only account for how listeners use multiple cues in "ideal" language situations but also how they focus on the most useful cue in "degraded" language situations.

The fact that perception strategies are dependent on the specific phonetic productions of a language has led to the development of language-specific perception grammars and bidirectional grammars, in which the production and perception grammars mirror each other. In Boersma's (2007) Bidirectional Stochastic OT, phonetic cues are related to phonological forms via constraints that have the same ranking values in both production and perception grammars. The results of the YM production and perception experiments provide support for Boersma's model.

This paper proceeds as follows. In §2, I present the phonological properties of the vowel system of Yucatec Maya and the results of production experiments that show how different phonetic properties indicate a contrast between HIGH TONE and GLOTTALIZED vowels. Two perception experiments are discussed in §§3-4.[1] The first perception experiment used natural stimuli to test how accurate listeners are at discriminating between HIGH TONE and GLOTTALIZED vowels and to determine which phonetic dimensions listeners use to make their choice. The second perception experiment used manipulated stimuli in order to test how listeners are influenced by the interaction of pitch and glottalization. Conclusions are presented in §5.

2. Production of Vowel Contrasts in Yucatec Maya

Yucatec Maya is spoken by about 700,000 in Yucatan, Campeche, and Quintana Roo, Mexico and Belize (Gordon 2005). In addition to five contrasting vowel qualities ([i e a o u]), there is a four-way contrast of suprasegmental features, henceforth referred to as *vowel shape*. The four vowel shapes, described in (1) along with a minimal quadruplet, can appear with any of the vowel qualities. Each vowel in YM must occur with one of these shapes. Thus, SHORT vowels are the only vowels not marked for tone; all long vowels bear some tone. Throughout this paper, vowel shapes are denoted by small capital letters so that they are not confusable with generic phonological/phonetic properties (i.e. "GLOTTALIZED" is a phonemic category of YM, whereas "glottalization" is a phonetic property).

[1] The methodology and results of all production and perception experiments are presented in my dissertation (Frazier 2009); the results of production experiment 1 are also reported in Frazier (In Press). In this paper, I focus on the data most relevant to the discussion at hand, and, as such, I present data pooled from different speakers and different tokens than those reported in previous work.

(1) vowel shapes of YM (Bricker et al. 1998)[2]

SHORT	/v/	short duration, mid pitch, modal voice	*chak* 'red'
LOW TONE	/v̀v/	long duration, low pitch, modal voice	*chaak* 'boil'
HIGH TONE	/v́v/	long duration, initial high pitch, modal voice	*cháak* 'rain'
GLOTTALIZED	/v́v̰/	long duration, initial high pitch, creaky voice (during the medial or final portion of the vowel)	*cha'ak* 'starch'

Two production experiments were designed to document the realization of the suprasegmental properties of the vowel system of YM. We will see in this section that the phonetic dimensions of vowel length, pitch, and glottalization all contribute to this contrast, and so the perception experiments of §§3-4 are designed to measure the effect of these properties on listeners' responses in discrimination tasks.

2.1 Methodology

For both production experiments, speakers were recorded from multiple towns in Yucatan, Mexico, and the results of these experiments indicate a dialectal split in the production of pitch and vowel length. Because these dialect differences are not the focus of this paper, I only present data from speakers from Mérida (the capital of Yucatan) and Santa Elena (approximately 65 km south of Mérida).[3] These towns are both located in the western half of Yucatan, and these speakers produce the vowel shapes in ways that are most congruent with previous literature on YM vowels.

The two production experiments mainly differ in the context of the target word. In production experiment 1 the target words were produced in isolation; in production experiment 2 the target words were produced in frame sentences (and data reported on here comes from target words in phrase-final position).

Nineteen speakers (1 female and 6 males from Mérida; 7 females and 5 males from Santa Elena) who participated in production experiment 1 are reported on here. These participants read 100 words (25 with each vowel shape) in isolation. Not all the data collected in production experiment 1 is reported on here. Only words with HIGH TONE or GLOTTALIZED vowels are of interest to this paper. The wordlist contained both nonce forms and existing forms; the results from nonce forms are not included below. Furthermore, the wordlist for speakers from Santa Elena contained some polysyllabic words, where measurements were taken from a vowel in a non-final syllable. Because there is variation in the production of suprasegmental features that is conditioned by the position of the syllable within the word,

[2] Bricker et al. (1998), along with other major phonological descriptions of YM (e.g. Blair & Vermont Salas 1965), represent the GLOTTALIZED vowel as /vʔv/ or /v́ʔv/. These vowels are hence often called "rearticulated vowels" in the literature. The representation /v́v̰/ is supported by phonetic research which shows that this vowel shape is canonically produced with creaky voice (and not a full glottal stop) and with initial high pitch (Frazier 2009, In Press; Avelino et al. 2007; see §2.2.2). Because these vowels are more often creaky than rearticulated, I follow their other naming convention and refer to them as GLOTTALIZED vowels here.

[3] The other towns represented in the production experiments were Sisbicchén, Xocén, and Yax Che, which are all clustered around Valladolid, the largest city in the eastern part of Yucatan. The reader is referred to Frazier (2009) for more information about these dialect differences.

these polysyllabic words are also excluded here. Thus, the results presented below represent the production of 20 words with a HIGH TONE vowel and 19 words with a GLOTTALIZED vowel by each speaker from Mérida and 15 words with a HIGH TONE vowel and 16 words with a GLOTTALIZED vowel by each speaker from Santa Elena. All target words are of the form CVC.

Seventeen participants are reported on here for production experiment 2 (3 males from Mérida; 10 females and 4 males from Santa Elena).[4] These participants read a total of 144 sentences (36 target words (9 with each vowel shape) of the form CVC embedded in four different frame sentences each). The target word was positioned phase-initially, phrase-medially, phrase-finally (after a HIGH TONE vowel), and phrase-finally (not after a HIGH TONE vowel). Only the last of these contexts is reported on here, and the relevant frame sentence is *Tu ya'alaj __*. "S/he said __."[5] Again, only HIGH TONE and GLOTTALIZED vowels are relevant, and so the following data for production experiment 2 comes from 9 HIGH TONE and 9 GLOTTALIZED vowels as spoken by each participant.

Measurements were taken from target words only (as spoken in isolation in production exp. 1 and in phrase-final position in production exp. 2). All measurements were extracted using PRAAT (Boersma & Weenink, 2006). For each target word, the boundaries of the vowel were demarcated (as determined by the onset and offset of F2) so that vowel length could be calculated, pitch values in Hz were extracted at 10 ms intervals for the duration of each vowel, and each vowel was coded for glottalization type.

2.1.1 Coding Glottalization Type

Gordon & Ladefoged (2001) summarize a body of literature that documents the acoustic and aerodynamic differences among modal, creaky, and breathy voice and conclude that, while the reliability of these properties varies from language to language, seven main characteristics can be used to differentiate these phonation types: periodicity, intensity, spectral tilt, fundamental frequency, formant frequencies, duration, and airflow. Though all of these properties are quantifiable, there is no way to determine a cut off point that divides the values of a given property into those that equate with each of the phonation types. For example, spectral tilt (e.g. the amplitude of the second harmonic minus the amplitude of f0) tends to be negative for creaky voice, but this does not mean that one can simply measure spectral tilt and categorize those tokens with negative spectral tilt as being produced with creaky voice and those with positive spectral tilt as being produced with modal voice.[6] Similar problems arise for any of the aforementioned phonetic properties.

In order to determine the glottalization type of each token obtained in the production experiments, I observed the waveform and spectrogram for periodicity, intensity, and f0 (similar to the methods of Redi & Shattuck-Hufnagel 2001). I found that, in YM, a departure from modal voice was most consistently indicated by a weakening of intensity, as seen in Fig. 1. In this token, it is clear that the vowel is not produced with sustained modal voice, but the only visual indicator of a change in phonation type is lowered intensity and a slight lowering

[4] An additional speaker from Santa Elena was recorded, but his data was rejected because he did not produce the requested sentences.

[5] A more detailed gloss for this frame sentence is as follows:

T-uy	a'al-aj	__.
COMP.ASP-3ERG	say-COMP.ASP/TRANS.STATUS	__.

[6] It has also been documented that there can be an interaction between tone and spectral tilt (Blankenship 1997), which could be especially problematic for YM, where tone and creaky voice are used phonemically.

of f0. Such tokens were classified as being produced with creaky voice. Irregularly spaced glottal pulses (i.e. aperiodicity) and irregularity in the intensity of consecutive glottal pulses (possibly due to diplophonia) were less consistently present in productions of creaky voice in YM. See Fig. 2 for a token with creaky voice that exhibits all of these properties. Thus, the visual cues of lowered intensity and aperiodicity were used to indicate a token that was produced with creaky voice during some portion of vowel production.[7] While lowered f0 was commonly present, it was never the sole indicator of a departure from modal voice.

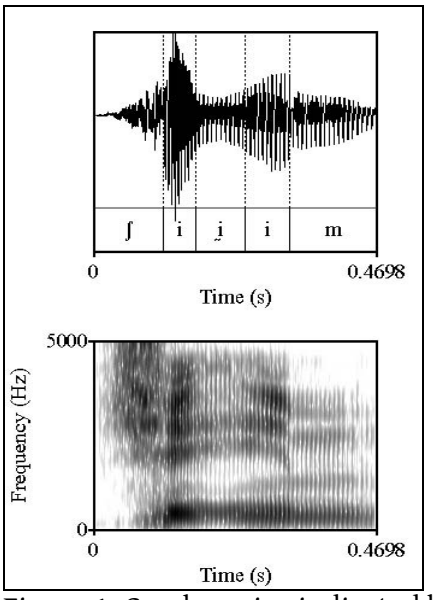

Figure 1: Creaky voice indicated by low intensity
This token of *xi'im* /ʃiḭm/ 'corn' is produced by a male from Mérida.

A small percentage of GLOTTALIZED vowels were produced with a full glottal stop interrupting vowel production. Any vowel that was interrupted by a gap between glottal pulses of 20 ms or more was coded as having a glottal stop. Additionally, I discovered that many tokens in YM did not display all of the canonical properties of creaky voice but rather showed only a momentary but extreme drop in intensity that indicated a departure from modal voice. This pattern occurred often and such tokens gave an auditory impression of creaky voice. Because these tokens were clearly not produced with only modal voice but were also clearly not produced with a lengthy stretch of creaky voice, they are given the label of "weak glottalization".

Four glottalization types are thus referred to in this paper: modal voice (no glottalization at any point in vowel production), weak glottalization (indicated by a brief dip in intensity), creaky voice (at some point during vowel production), and a glottal stop (interrupting vowel production). These glottalization types are henceforth abbreviated as mod., w.g., cr., and g.s. (respectively), and examples are presented in Fig. 2.

[7] Creaky voice most commonly occurred during the medial portion of the vowel, such that the vowel began and ended with modal voice. A smaller set of vowels were produced with creaky voice that began in the middle of the vowel and continued to the end. Only a few tokens showed creaky voice either initially or throughout vowel production.

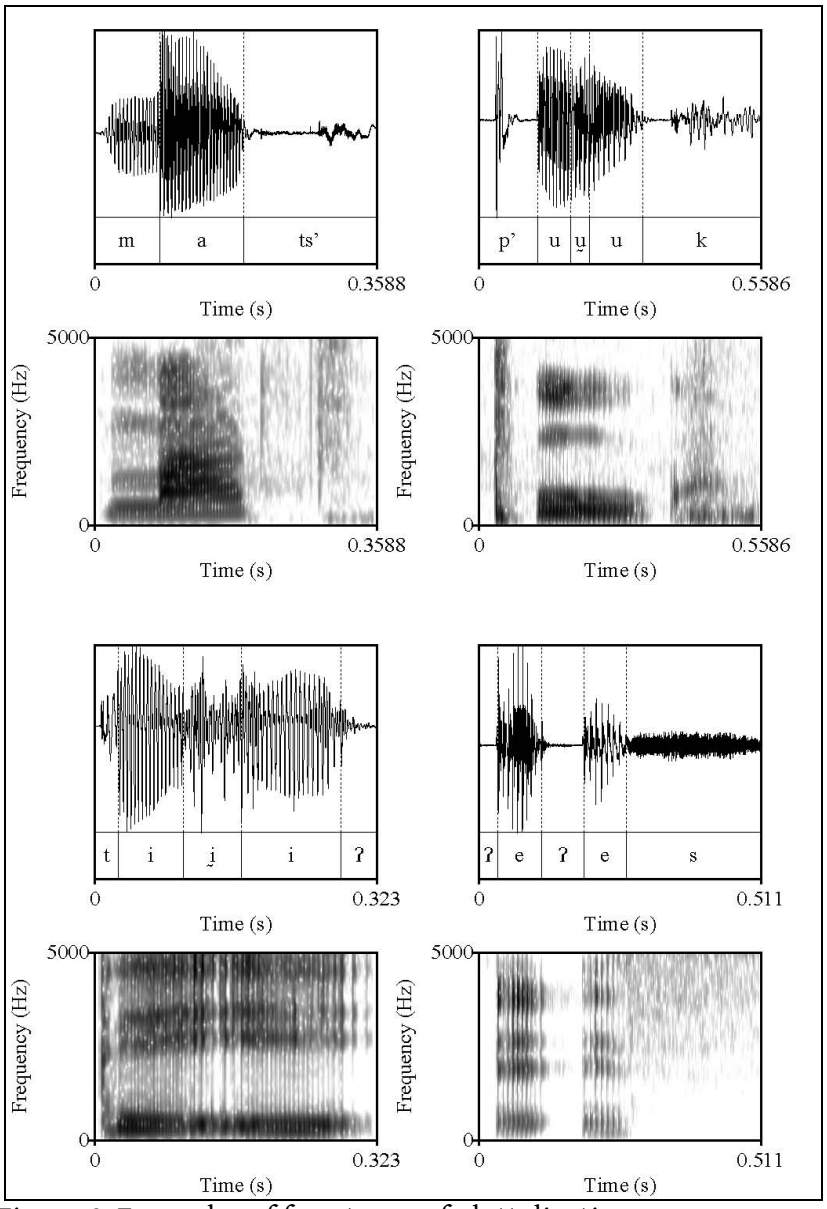

Figure 2: Examples of four types of glottalization
Top row: modal voice, *ma'ats'* /máa̯ts'/ 'hull corn'; weak glottalization, *p'u'uk* /p'úu̯k/ 'cheek'
Bottom row: creaky voice, *ti'i'* /tíi̯ʔ/ 'there'; full glottal stop, *e'es* /ʔée̯s/ 'show'

2.1.2 Measuring Pitch

As explained above, pitch measurements (in Hz) were extracted at 10 ms intervals for the duration of the vowel using PRAAT. Seven pitch values are referred to in this paper: the maximum pitch value produced during vowel production, the minimum pitch value produced during vowel production, and five pitch values produced at normalized time points (beginning of vowel, 25%, 50%, and 75% of vowel production, and end of vowel). These five pitch values are used to define pitch contours for each vowel, and the maximum pitch value minus the minimum pitch value defines the *pitch span* of the vowel. Maximum and minimum pitch values were obtainable for every token, but pitch values at the five normalized time points were not

always available (due to aperiodicity generally caused by creak or full glottal closure).[8] These missing values were recorded as such, but did not result in throwing out the obtainable pitch values for a given vowel.

In order to make meaningful cross-speaker comparisons of pitch values, pitch measurements in Hertz are transformed into *semitones over the baseline* (s/b). This transform is based on the work of Nolan (2003), who found that pitch spans (of intonational contours) showed the least inter-speaker variation when measured in semitones, and the work of Pierrehumbert (1980), who showed that a speaker-specific baseline could be used to scale pitch measurements and, again, minimize inter-speaker variation. As shown in (2), s/b is calculated by using the standard equation for deriving semitones from Hz – $12*\log_2$(Hz/reference Hz) – but instead of using a constant value for the "reference Hz", a speaker- and context-specific baseline is used. The baseline is (somewhat arbitrarily) defined as the average pitch value produced at the mid point of LOW TONE vowels in a given context (i.e. isolation or phrase-final) by a given speaker. Thus, for example, a pitch measurement of 2 s/b indicates a value that is 2 semitones above that particular speaker's baseline (in that particular context). As shown in Frazier (2009), this transform is successful at minimizing inter-speaker variation in YM.

(2) equation used to transform Hertz into semitones over the baseline

$s/b = 12*\log_2(Hz/\text{baseline Hz})$
where baseline Hz is the average pitch value produced at the mid point of LOW TONE vowels for a given speaker in a given context

2.2 Results

2.2.1 Vowel Length

GLOTTALIZED vowels are slightly longer than HIGH TONE vowels, though this difference is only statistically significant for production experiment 1:

Table 1: Mean vowel length (ms) for HIGH TONE and GLOTTALIZED vowels [9]

	HIGH TONE	GLOTTALIZED	t	p	df
exp. 1 (isolation)	200	208	2.25	.02	625
exp. 2 (phrase-final)	186	188	0.31	.76	288

2.2.2 Glottalization

For GLOTTALIZED vowels, modal voice and creaky voice are the two most common types of glottalization, while HIGH TONE vowels are (unsurprisingly) almost always produced with modal voice (see Table 2). While it is clear that these two vowel shapes differ in terms of glottalization, the fact that so many GLOTTALIZED vowels are produced with modal voice suggests that glottalization alone does not signal a contrast.

[8] Vocal fold vibration during the production of creaky voice was not always aperiodic, and so there are many pitch measurements that come from portions of vowels where creaky voice occurs. See §2.2.4 for further discussion of pitch and creaky voice.
[9] All t statistics in this paper are calculated using a mixed linear regression model to account for multiple observations within subjects.

Table 2: Distribution of glottalization types

	exp. 1 (isolation)				exp. 2 (phrase-final)			
	mod.	w.g.	cr.	g.s.	mod.	w.g.	cr.	g.s.
GLOTTALIZED	35%	13%	44%	8%	59%	9%	29%	3%
HIGH TONE	94%	2%	4%	0%	96%	1%	3%	0%

2.2.3 Pitch

The average pitch contours of HIGH TONE and GLOTTALIZED vowels as spoken in isolation are shown in Fig. 3. GLOTTALIZED vowels start with higher pitch, which then drops more rapidly, than HIGH TONE vowels. Both vowel shapes end with relatively low pitch. GLOTTALIZED vowels have a larger pitch span than HIGH TONE vowels ($\bar{x}_{GLOT}$ = 6.1 semitones; $\bar{x}_{HI}$ = 4.2 semitones; $t(624) = 7.82$, $p < .01$).

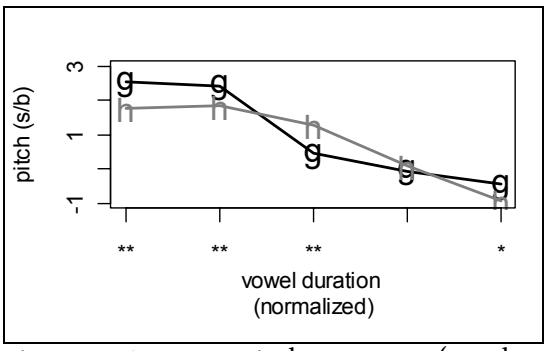

Figure 3: Average pitch contours (production experiment 1; isolation)
g = GLOTTALIZED; h = HIGH TONE; asterisks indicate time points with statistically significant differences in pitch values for each vowel shape (** ($p < .01$); * ($p < .05$))

There are some gender-based differences in the production of pitch, as shown in Fig. 4. Namely, GLOTTALIZED vowels are produced with higher pitch (relative to the speaker's baseline) by males than by females. This means that the initial pitch of GLOTTALIZED vowels is significantly higher than the initial pitch of HIGH TONE vowels for males ($t(355) = 5.40$, $p < .01$) but not for females ($t = 0.71(288)$, $p = .48$). This result is discussed in more detail in §2.2.4. Even though productions by the two genders differ in terms of initial pitch, GLOTTALIZED vowels have a larger pitch span for both genders ($\bar{x}_{GLOT,male}$ = 5.9 semitones; $\bar{x}_{HI,male}$ = 4.8 semitones; $\bar{x}_{GLOT,female}$ = 6.5 semitones; $\bar{x}_{HI,female}$ = 3.4 semitones).

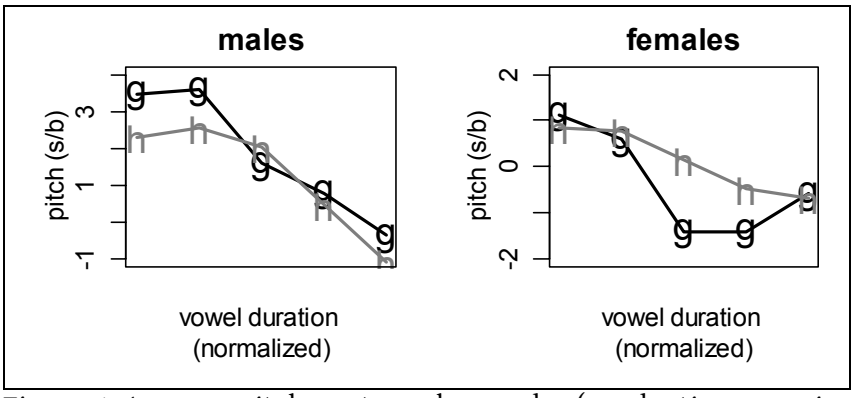

Figure 4: Average pitch contours by gender (production experiment 1; isolation)
g = GLOTTALIZED; h = HIGH TONE

The pitch contours produced in production study 2, as shown in Fig. 5, show that both vowel shapes begin and end with the same pitch values in this context, though the pitch of GLOTTALIZED vowels drops earlier in vowel production. Statistical analysis shows that differences between the two contours are mostly nonsignificant.

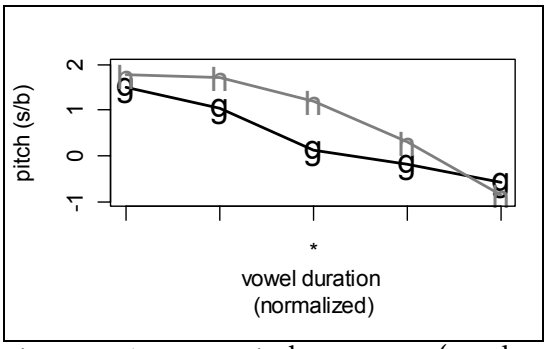

Figure 5: Average pitch contours (production experiment 2; phrase-final)
g = GLOTTALIZED; h = HIGH TONE; asterisks indicate time points with statistically significant differences in pitch values for each vowel shape (** ($p < .01$); * ($p < .05$))

Again, the pitch contours of GLOTTALIZED vowels are produced differently by the two genders (see Fig. 6). GLOTTALIZED vowels do have higher initial pitch than HIGH TONE vowels in phrase-final context when spoken by males ($t(118) = 2.61$, $p = .01$), whereas HIGH TONE vowels have higher pitch throughout vowel production (until the end of the vowel) for females (for initial pitch, $t(162) = -2.36$, $p = .02$).

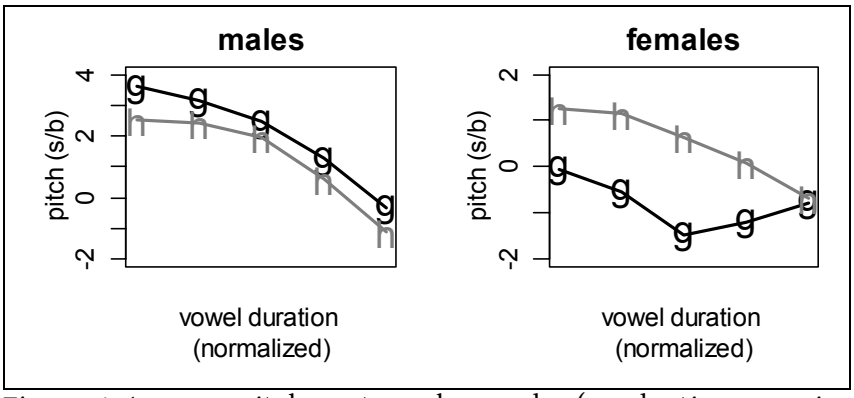

Figure 6: Average pitch contours by gender (production experiment 2; phrase-final)
g = GLOTTALIZED; h = HIGH TONE

Pitch spans are larger for GLOTTALIZED vowels, whether the data is averaged across all participants ($\bar{x}_{GLOT}$ = 6.1 semitones; $\bar{x}_{HI}$ = 4.3 semitones; t (616) = 3.18, p < .01) or across each gender separately ($\bar{x}_{GLOT,male}$ = 6.5 semitones; $\bar{x}_{HI,male}$ = 4.7 semitones; $\bar{x}_{GLOT,female}$ = 5.9 semitones; $\bar{x}_{HI,female}$ = 4.1 semitones).

2.2.4 The Interaction of Pitch, Glottalization, and Gender

The results presented in the previous section indicate that GLOTTALIZED vowels do not have the same pitch contours for both males and females, even when using the semitones over the baseline transform, which is designed to minimize cross-speaker variation. This result can be better understood by looking at the pitch contours that are produced with different glottalization types. In Fig. 7 we see the average pitch contours for GLOTTALIZED vowels (as produced in production experiment 1) by gender and glottalization type (modal voice, weak glottalization, and creaky voice).[10] When GLOTTALIZED vowels are produced without glottalization (modal voice only), they have roughly the same contour for males and females and are produced with higher pitch than that of HIGH TONE vowels. However, the pitch contours of those vowels produced with either weak glottalization or creaky voice are strikingly different between the two genders. For males, creaky voice and weak glottalization are correlated with even higher initial pitch, while, for females, glottalization is correlated with lower initial pitch and dramatic decreases in pitch during the middle portion of the vowel (where glottalization is canonically produced). The cause of this distinction is the fact that females and males are producing creaky voice with similar f0 values (in Hz).[11] For example, the average pitch value produced at the middle time point of GLOTTALIZED vowels produced with either weak glottalization or creaky voice is 147 Hz for males and 167 Hz for females, as compared to the middle time point of GLOTTALIZED vowels produced with modal voice, which is 156 Hz for males and 212 Hz for females. This indicates that pitch produced during creaky voice is not a function of a speaker's natural pitch range in the same way that pitch produced

[10] Those vowels produced with a full glottal stop are excluded here because of how few tokens had this glottalization type and because no pitch values can be extracted from the portion of the vowel with a glottal stop.
[11] A similar result was found for speakers of English. Blomgren et al. (1998) measured f0 during productions of "modal register" and "vocal fry" and found that the average f0 for females was much higher than males during modal register (211.0 Hz for females as compared to 117.5 Hz for males) but not during vocal fry (48.1 Hz for females and 49.1 Hz for males).

during modal voice is. Pitch produced during creaky voice is far below the baseline for females but near, or even slightly above, the baseline for males. Thus, when these pitch values are transformed into s/b, a measurement that is sensitive to a speaker's natural pitch range, we get the gender-based differences documented in Fig. 7. Because males more consistently produce GLOTTALIZED vowels with higher pitch than HIGH TONE vowels, a male's voice is used for the stimuli of perception experiment 2 (see §4).

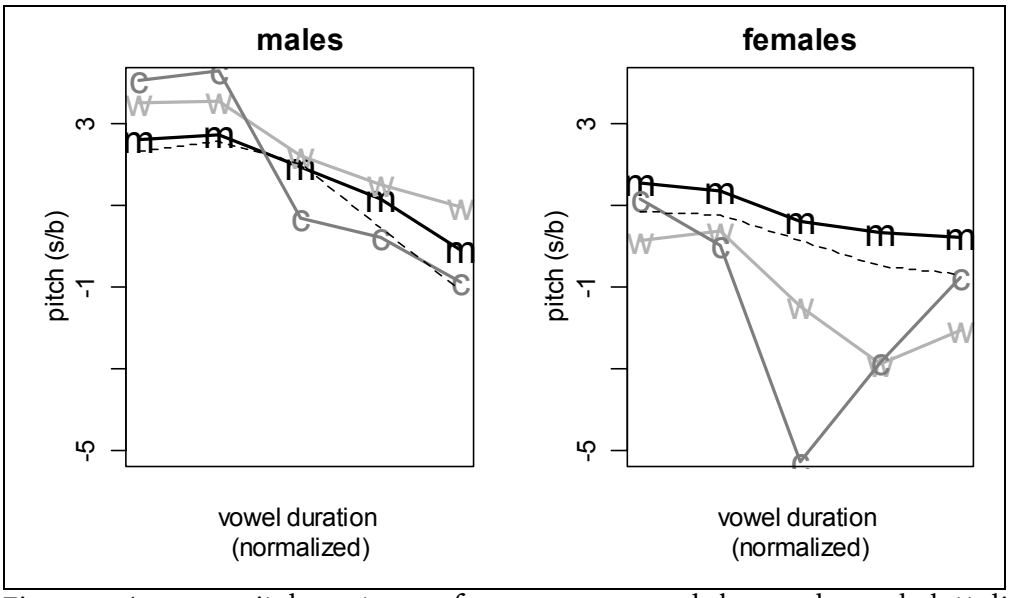

Figure 7: Average pitch contours of GLOTTALIZED vowels by gender and glottalization type (production experiment 1; isolation)
m = modal voice; w = weak glottalization, c = creaky voice
The thin dashed line shows the average pitch contour of HIGH TONE vowels.

2.3 Summary

The above description of HIGH TONE and GLOTTALIZED vowels has shown that length, pitch, and glottalization all contribute to this contrast (in increasing order of importance). GLOTTALIZED vowels are more likely to be produced with glottalization, with a larger pitch span, with higher initial pitch (especially by males), and with a slightly longer duration (though this last characteristic is not very robust). The perception experiments presented in §§3-4 are designed to determine which of these cues are attended to by listeners when determining whether they heard a GLOTTALIZED or a HIGH TONE vowel.

3. Perception Experiment 1: Natural Stimuli

As we saw in §2, the phonetic dimensions of length, pitch, and glottalization systematically differ in productions of HIGH TONE and GLOTTALIZED vowels in YM. However, there is also a high degree of overlap between the permissible values of a given parameter for both vowel shapes. In other words, many productions of HIGH TONE vowels are legal productions of GLOTTALIZED vowels, and vice versa (with the one notable exception of a full glottal stop, which is only produced with GLOTTALIZED vowels).

There are thus two goals of this perception experiment. The first is to determine if listeners can indeed correctly identify tokens with HIGH TONE and GLOTTALIZED vowels when given no other contextual information. The second is to determine which of the various phonetic dimensions influence the listeners' decisions.

3.1 Methods

3.1.1 Participants

Sixteen speakers who participated in the two perception experiments are reported on here (2 males from Mérida and 9 females and 5 males from Santa Elena). All had also participated in production experiment 2, and some had participated in production experiment 1. All participants are fluent in Spanish in addition to YM; all speak YM in the home and in daily life. Two participants from Santa Elena are also proficient in English. The two participants from Mérida are originally from smaller towns on the western side of the peninsula.

3.1.2 Procedure and Stimuli

Participants heard a stimulus and were asked to choose whether they heard a word with a HIGH TONE vowel or a word with GLOTTALIZED vowel. Experiments were run with PRAAT (Boersma & Weenink 2006). The participant listened to each stimulus through headphones and then used a mouse to select the appropriate word on the computer screen. For each choice on the computer screen, a word was displayed in both YM and Spanish. The Spanish translation was given to ensure that there were no orthography misunderstandings. Additionally, there was a repeat button on the screen that participants could use up to three times if they wished to hear the stimulus again.

Before the experiment began, participants went through a practice perception task. This was done to ensure both that the participant understood the instructions and that they were comfortable using a mouse to interact with the computer screen. If necessary, the participant was shown how to use a mouse, and this usually required the practice session to be done more than once. There was one participant who did not like using the mouse and did not want to learn. She instead used her finger to point to a box on the screen, and I clicked on the box for her. Participants were encouraged to ask questions if necessary and were compensated for their time.

The stimuli used for this experiment were natural, unmanipulated recordings of the words *k'a'an* /k'áa̰n/ 'strong' and *k'áan* /k'áan/ 'hammock'. The Spanish translations given for these words were *fuerte* and *hamaca*, respectively. Each word was spoken by 24 different speakers of YM (from Mérida, Santa Elena, and Sisbicchén); the words were recorded during production experiment 1.[12] Participants heard each of the 48 stimuli once and were asked to choose between *k'a'an* and *k'áan* as the word they thought they heard, as detailed above.

[12] As discussed in §2, speakers from Sisbicchén produce pitch and length differently than speakers from Santa Elena and Mérida. Participants were not told that they were listening to speakers from different locations, and there is no evidence in the data to suggest that participants responded differently to stimuli from different dialect groups.

3.2 Results

The first goal of this experiment is to determine to what extent listeners can correctly identify HIGH TONE and GLOTTALIZED vowels without any contextual clues. As shown in Table 3, participants performed better than chance (Rao-Scott $\chi^2(1) = 19.2$, $p < .01$),[13] though it is doubtful that one should consider accuracy rates of 63-64% as highly successful. This data suggests that, at the very least, participants are not merely guessing and, to a limited extent, can identify HIGH TONE and GLOTTALIZED vowels on the basis of phonetic production alone without the aid of any semantic or other discourse cues.

Table 3: Confusion matrix (perception experiment 1)

	stimulus	
response	*k'a'an* (GLOT.)	*k'áan* (H.T.)
k'a'an (GLOT.)	240 (62.5%)	138 (35.9%)
k'áan (H.T.)	144 (37.5%)	246 (64.1%)

The second goal of the experiment is to determine which phonetic cues influence the listeners' decisions in discriminating between HIGH TONE and GLOTTALIZED vowels. The results (see Fig. 8) show a statistically significant effect of all measured phonetic dimensions: glottalization type (Rao-Scott $\chi^2(3) = 55.2$, $p < .01$), vowel length ($z = 3.97$, $p < .01$), initial pitch ($z = 4.31$, $p < .01$), and pitch span ($z = 7.26$, $p < .01$).[14] More glottalization, longer vowel duration, higher initial pitch, or a larger pitch span triggers more GLOTTALIZED vowel responses (as opposed to HIGH TONE vowel responses).

[13] The Rao-Scott χ^2 is analogous to Pearson's χ^2 and is adjusted for multiple observations within subjects.
[14] The tests for vowel length, initial pitch, and pitch span use a mixed logistic regression model to account for multiple observations within subjects and a null hypothesis of no linear association between the independent variable (vowel length, initial pitch, or pitch span) and the logit of the dependent variable (response).

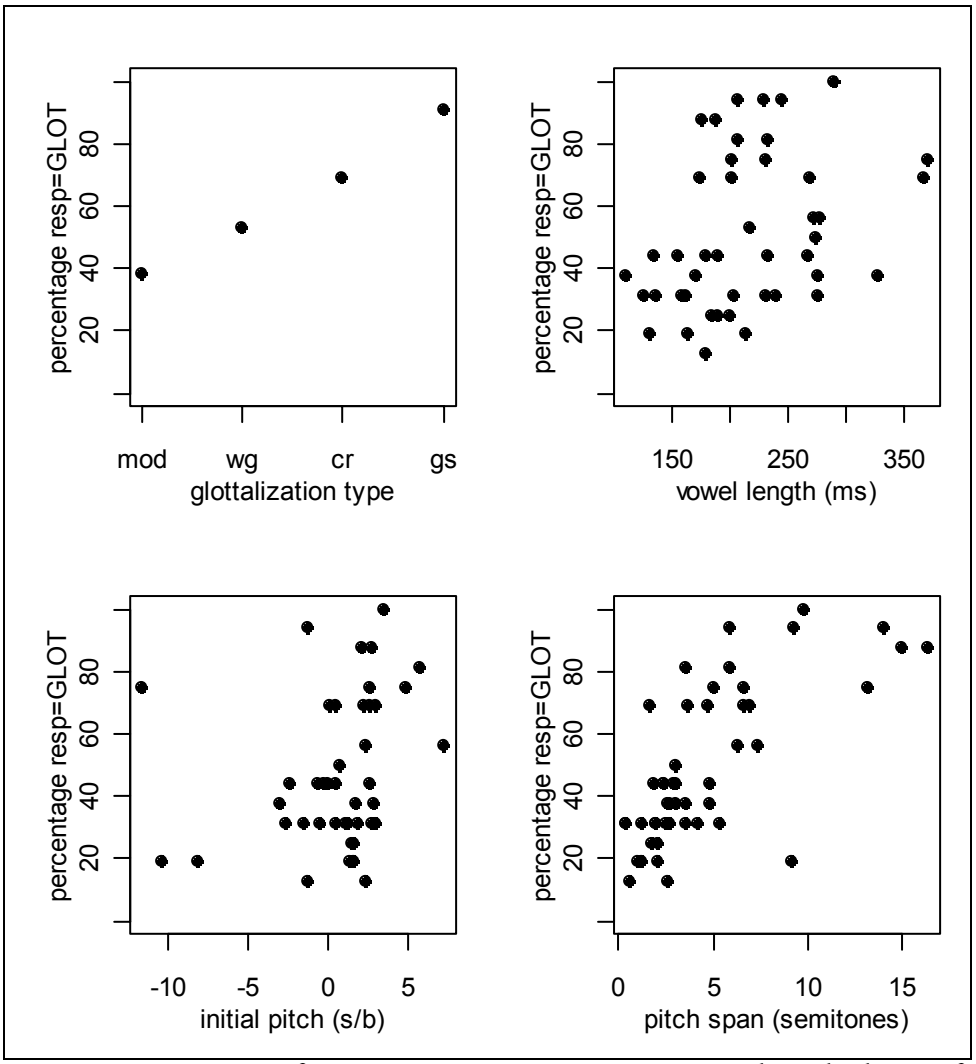

Figure 8: Percentage of times response = GLOTTALIZED vowel on the basis of glottalization type, vowel length, initial pitch, and pitch span of the stimulus

3.3 Discussion

The results of perception experiment 1 indicate that listeners make use of all available phonetic cues when discriminating between HIGH TONE and GLOTTALIZED vowels. In fact, they even attend to vowel length, which is only loosely correlated with phonemic category (see Table 1). This result provides support for bidirectional models of phonetics and phonology, such as Boersma's (2007) Bidirectional Stochastic OT. In this model, phonetic values are related to phonological categories via constraints that have the same mean ranking values in both the production and the perception grammar (and variation is accounted for through stochastic evaluation). Thus the production grammar tells the speaker how to map a phonological category like /v̋ṽ/ onto particular values for pitch, glottalization, etc. The perception grammar takes these phonetic values as the input and tells the listener how to map them onto a phonological category. This model predicts that whatever phonetic values are deemed optimal for some phonological category by the production grammar will be also be mapped onto that phonological category by the perception grammar.

For vowel length, initial pitch values, and pitch spans, this prediction is borne out. The production grammar is more likely to map a GLOTTALIZED vowel – /v́v̰/ – onto phonetic forms with longer vowel lengths, higher initial pitch values, and larger pitch spans, while it is more likely to map a HIGH TONE vowel – /v́v/ – onto phonetic forms with shorter vowel lengths, lower initial pitch values, and smaller pitch spans. (Though, of course, there is variation in the phonetic forms that can be generated by the production grammar, as predicted by stochastic evaluation.) The perception grammar is then more likely to map a phonetic form with a longer vowel length, higher initial pitch value, and larger pitch span onto a GLOTTALIZED vowel.

This correlation is not as straightforward with glottalization type. In production (see Table 2), HIGH TONE vowels are almost always produced with modal voice, whereas GLOTTALIZED vowels are produced with all four glottalization types, with modal voice and creaky voice being the most common. Because a glottal stop is so rarely produced with GLOTTALIZED vowels, it may be surprising that it is such a good perceptual cue. However, this result makes sense if we look at how often each glottalization type occurs as a production of a particular vowel shape. As shown in Table 4, the vast majority of productions with weak glottalization, creaky voice, or a glottal stop are GLOTTALIZED vowels, whereas the majority of productions with modal voice are HIGH TONE vowels. Thus, it is understandable for the perception grammar to map any phonetic form with glottalization onto a GLOTTALIZED vowel.

Table 4: Percentage of times each glottalization type occurs as a production of each vowel shape (production experiment 1; isolation)

	mod.	w.g.	cr.	g.s.
GLOTTALIZED	26%	87%	91%	100%
HIGH TONE	74%	13%	9%	0%

The fact that glottal stops are such good perceptual cues but are so rarely produced is an instance of the *prototype effect* (see Boersma 2006 and references therein). Listeners prefer for GLOTTALIZED vowels to have glottal stops because this makes them easy to identify, but speakers prefer not to produce a glottal stop because of the effort involved. This production-perception mismatch can be accounted for with the articulatory constraints of the Bidirectional Model.[15] Articulatory constraints penalize effortful phonetic forms (with no reference to the input) and hence function similarly to the markedness constraints of classic OT. Thus an articulatory constraint such as *[ʔ] ("don't produce a glottal stop") can rule out the production of a glottal stop even though the rest of the production grammar may prefer the mapping of /v́v̰/ → [v́ʔv]. This articulatory constraint plays no role in the perception grammar because phonetic forms (denoted by brackets instead of slashes) are not the candidates of evaluation, and so the perception grammar prefers the mapping [v́ʔv] → /v́v̰/.

4. Perception Experiment 2: Manipulated Stimuli

The results of perception experiment 1 indicate that listeners make use of a variety of phonetic cues in discriminating between HIGH TONE and GLOTTALIZED vowel in YM. One

[15] Another method of accounting for the prototype effect is to add a *hyperarticulated form* to a model of phonetics and phonology, such that the prototypicality task involves the mapping of categorical representations onto hyperarticulated targets, whereas normal production goes a step further and maps the hyperarticulated form onto a "normal speech" phonetic form (Johnson et al. 1993). As Boersma (2006) argues, Bidirectional Stochastic OT allows for the analysis of the prototype effect without positing an additional hyperarticulated form.

shortcoming with the design of this experiment is that, because natural productions were used, the stimuli represent only those productions that happened to be produced by each of the 24 speakers. There is thus no systematic relation among the relevant phonetic dimensions, and hence it is hard to test how listeners are affected by the interactions of these phonetic dimensions. For example, in the production data presented in §2, we saw that, for females, GLOTTALIZED vowels with modal voice tend to have higher pitch (than both GLOTTALIZED vowels produced with glottalization and HIGH TONE vowels). This could mean that listeners are more attentive to pitch when given a stimulus with modal voice, since both HIGH TONE and GLOTTALIZED vowels may be produced with modal voice. Perception experiment 2 was designed to test how the interaction of pitch and glottalization influence listeners in a discrimination task. The effect of vowel length is not investigated in this experiment as it is the least correlated with vowel shape in production and in order to simplify the experimental design and analysis.

The results of this experiment indicate that listeners reacted differently than expected to these stimuli. Listeners made use of glottalization alone when making their decisions, while completely ignoring pitch. The implications of these results are discussed in §4.3.

4.1 Methods

4.1.1 Participants and Procedures

Perception experiment 2 took place immediately after perception experiment 1, and so the participants are the same for both experiments. In this task, participants were again forced to choose between a word with a HIGH TONE vowel and a word with a GLOTTALIZED vowel. Two minimal pairs were used: *k'áan* /k'áan/ 'hammock' vs. *k'a'an* /k'aa̰n/ 'strong' and *cháak* /tʃáak/ 'rain' vs. *cha'ak* /tʃaa̰k/ 'starch'.[16] Each stimulus was played in the context of the frame sentence *Tin wa'alaj* ___ 'I said ___'. A frame sentence was used in this task because of the concern that the combination of manipulated stimuli and no context would make the participants unable to do anything but guess. With the frame sentence, the listener becomes familiar with the pitch range of the speaker and can use that information.[17] The procedure of this experiment is the same as in perception experiment 1, except that the computer screen showed whichever minimal pair was relevant for the particular stimulus being heard.

Participants heard each of the 32 stimuli (16 manipulated stimuli for each minimal pair) three times, for 96 trials. All participants completed all 96 trials, but some of the data was rejected. Three participants from Santa Elena always selected *cháak* when given the choice of *cháak* vs. *cha'ak*, and so these 144 trials (48 for each participant) were discarded and are not included in the results that follow. I assume that these participants were simply not familiar with the word *cha'ak* (see footnote 16).

[16] The minimal pair *k'a'an* vs. *k'áan* is robust; both words are common lexical items in everyday usage. The minimal pair *cha'ak* and *cháak* is not as robust in that *cha'ak* 'starch' is not a common lexical item. I had first tried to manipulate stimuli with another robust minimal pair (*ku'uk* 'squirrel' and *kúuk* 'elbow'), but I found that the high vowel [u] sounded much less natural after resynthesis with different pitch contours. The *cha'ak/cháak* minimal pair was used because it was the best minimal pair with a low vowel that I could find. (Exact minimal pairs for the contrast of HIGH TONE and GLOTTALIZED vowels are actually quite rare in YM.) The Spanish translations provided for the YM words were *fuerte* for 'k'a'an', *hamaca* for 'k'áan', *sagú* for 'cha'ak', and *lluvia* for 'cháak'.

[17] The frame sentence used with the perception study (Tin wa'alaj __. 'I said __') is slightly different from the frame sentence used in production study 2 (Tu ya'alaj __. 'S/he said __'); a different subject pronoun is used. This difference does not affect the prosodic or syntactic constituency of the sentence.

4.1.2 Manipulation of Stimuli

Stimuli were manipulated from one natural production of *k'an* 'ripe' and *chak* 'red' (both are current lexical items with a SHORT vowel) as spoken in isolation by a male from Mérida who did not participate in the perception experiments. This male did participate in both production experiments, and so the data obtained there will be used to calculate his baseline in phrase-final position. The frame sentence (*Tin wa'alaj __.* 'I said __.') was an unmanipulated recording spoken by the same male.

Using each original recording (*k'an* and *chak*), 16 stimuli were manipulated to fit each combination of four types of glottalization and four values for initial pitch (see details below). All manipulations were done with PRAAT. Because the original stimulus has a SHORT vowel, the vowel was lengthened by copying and pasting whole pitch periods in the central portion of the vowel, so that the resulting vowel (with modal voice) was about 200 ms long (about the mean length of long vowels). There are minor deviations in the vowel lengths of each stimulus due to the necessity of cutting/pasting whole pitch periods; the average vowel length of the 32 stimuli is 199 ms (standard deviation = 3.8). See Table 5 for precise measurements of each stimulus.

4.1.2.1 Manipulation of Pitch

Four different pitch contours were created using PRAAT and were defined by three points – the beginning of the vowel, 75 ms after the beginning of the vowel, and the end of the vowel. The first two points had the same value for a given contour and had different values for each of the four pitch contours (125, 140, 155, and 170 Hz), while the last point had a constant value for all stimuli (110 Hz, which continued through the coda [n] for *k'áan/k'a'an*). The four different pitch contours will henceforth be abbreviated as L (low), ML (mid-low), MH (mid-high), and H (high).[18] In semitones over the baseline (calculated by using this speaker's baseline as obtained from production experiment 2), the different initial pitch values are -0.1, 1.9, 3.6, and 5.2 s/b. When a new stimulus is resynthesized with each pitch contour, minor fluctuations in pitch can occur (see Table 5). Fig. 9 shows the four synthesized pitch contours and where they fall relative to this speaker's natural average productions of these vowel shapes.

There is some variation in the pitch spans of different stimuli with the same initial pitch value but with different glottalization types, due to the fact that creaky voice is synthesized by inserting extremely low pitch values (see below). After synthesizing a stimulus with weak glottalization or creaky voice, f0 measurements are sometimes obtainable from the creaky portions of the vowel, and sometimes not (due to aperiodicity). The measured pitch span for each stimulus is given in Table 5. In general, as initial pitch increases (for a given lexical item and glottalization type) so does the pitch span.[19]

---

[18] I use abbreviations that look like phonological tone markers so that they will be easily recognizable by the reader. It should be kept in mind that these are not tonal markers, as both HIGH TONE and GLOTTALIZED vowels are marked by high tone in their phonological form. These abbreviations are meant to signify where the exact acoustic value is in the range of the acoustic values under consideration. The L marker, for example, does not indicate low tone or even low pitch overall; it indicates the lowest pitch value of the four pitch values used in this experiment.

[19] The exception to this generalization is that, for each glottalization type, the pitch category of MH has a smaller pitch span than the pitch category of ML for the *k'áan/k'a'an* stimuli.

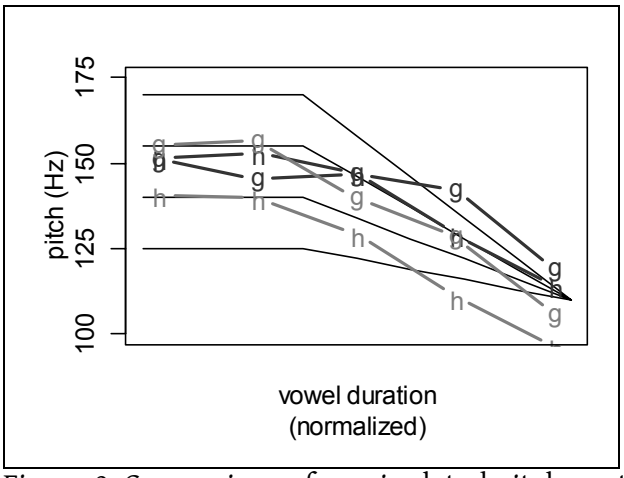

Figure 9: Comparison of manipulated pitch contours with speaker's natural productions The thick gray lines show the average pitch contours for HIGH TONE (h) and GLOTTALIZED (g) vowels (from production experiment 1 (lighter gray) and production experiment 2 (darker gray)) as spoken by the producer of the original tokens used to manipulate the stimuli for perception task 2. The thin black lines show the four manipulated pitch contours.

Table 5: Length and pitch measurements of manipulated stimuli

		cha'ak/cháak			k'a'an/k'áan		
glot. type	pitch cat.	vowel length (ms)	initial pitch (s/b)	pitch span (semi-tones)	vowel length (ms)	initial pitch (s/b)	pitch span (semi-tones)
mod.	L	196	-0.7	1.8	201	-0.7	2
	ML	196	1.2	3.3	201	1.2	5.8
	MH	196	3	4.6	201	3	4.1
	H	196	4.6	5.9	201	4.6	7
w.g.	L	201	-0.7	1.8	202	-0.7	2.1
	ML	198	1.2	3.9	200	1.2	5.9
	MH	194	3	5.5	200	3	4.1
	H	196	4.6	7.2	204	4.6	7.6
cr.	L	195	-0.7	8.2	202	-0.7	7.5
	ML	195	1.2	10.3	203	1.2	11.7
	MH	194	2.9	12.1	200	3	10.7
	H	190	4.6	12	197	4.6	13.5
g.s.	L	207	-0.7	8.2	204	-0.7	4.9
	ML	203	1.2	10.3	205	1.2	11.7
	MH	200	2.9	10.3	201	3	6.6
	H	197	4.6	12	197	4.6	12

4.1.2.2 Manipulation of Glottalization

Each token with modal voice (and one of the four pitch contours) was then modified to create a new token with each of the other three glottalization types: weak glottalization, creaky voice, and a full glottal stop. In order to create a token with "weak glottalization", each pitch tier (Fig. 9) was altered by adding a pitch value of 35 Hz at 10 ms after the second point (85 ms after the start of the vowel). When a token is resynthesized with such a pitch contour, the resulting token contains a very brief portion of what sounds like creaky voice and resembles the pattern of weak glottalization as produced by native speakers of YM (see Fig. 2).

Tokens with "creaky voice" were manipulated by starting with the weak glottalization pitch tier (with one pitch point at 35 Hz). This pitch tier was modified by adding another point at 35 Hz at 10 ms after the first one. In this manner, the impression of creaky voice is synthesized through the creation of extremely low pitch values. In order to make the stimuli with creaky voice sound more natural, the peak intensity of the portion of the vowel with creaky voice was scaled down by 30%.

In order to synthesize a glottal stop, the stimuli with creaky voice were used as a starting point. About 75 ms were deleted from the vowel in order to insert 75 ms of silence and maintain a vowel of the same length. The most stable portions of the vowel were deleted so that formant transitions were not affected. Because, in YM, a glottal stop is generally followed (and sometimes preceded) by creaky voice, most of the deleted portions of the vowel were those with modal voice so that the synthesized creaky voice remains. After the portion of silence was inserted into the middle of the remaining vowel, the peak intensity of a couple of creaky pulses surrounding this silence were again lowered by 30% The resulting token contains a glottal stop (represented by silence) surrounded by creaky pulses. Spectrograms and waveforms showing the four different glottalization types are displayed in Fig. 10.

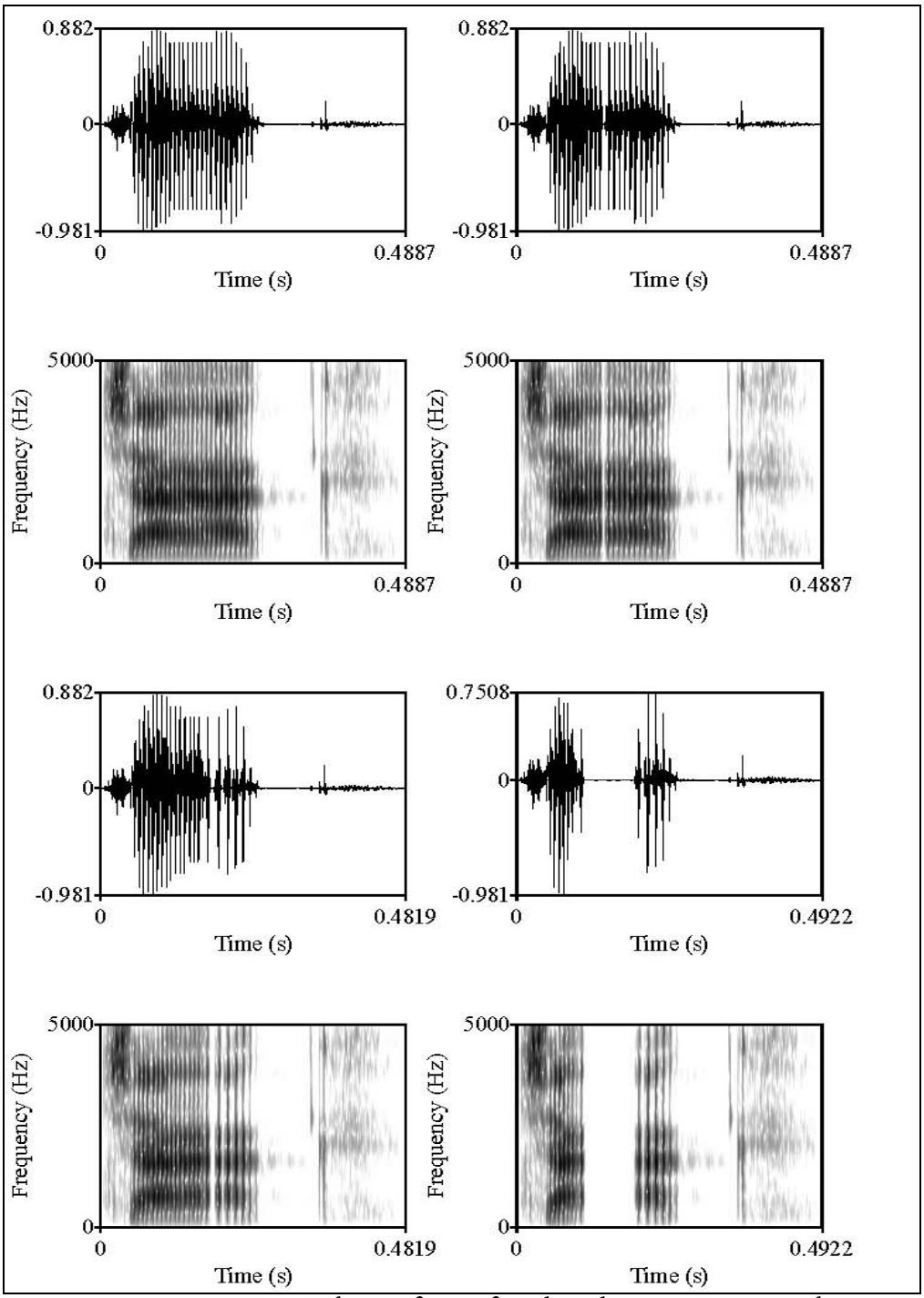

Figure 10: Spectrograms and waveforms for glottalization types used in perception exp. 2 These stimuli represent *cháak/cha'ak* with an initial pitch of 140 Hz (the ML category).

4.2 Results

The results of this experiment, as presented in Table 6, show that glottalization alone influenced the participants' decisions. As glottalization increases, listeners are more likely to select a token with a GLOTTALIZED vowel, regardless of the pitch contours of the stimuli. This

means that neither initial pitch nor pitch span was attended to by the listener. There are no right or wrong responses to this task as all of the stimuli were manipulated from tokens that are not possible responses.

Table 6: Percentage of times a GLOTTALIZED vowel was chosen for each stimulus

	pitch categories			
glottalization type	L	ML	MH	H
glottal stop	80	83	77	82
creaky voice	63	62	68	68
weak glottalization	44	43	42	43
modal	29	26	26	23

significant effect of glottalization: Wald $\chi^2(3) = 32.95$, $p < .01$
nonsignificant effect of pitch: Wald $\chi^2(3) = 1.26$, $p = .74$
nonsignificant interaction: Wald $\chi^2(9) = 3.46$, $p = .94$
(standard errors are adjusted for multiple observations within subjects)

4.3 Discussion

Given the results of perception experiment 1, where listeners made use of every available cue, no matter how loosely the cue is correlated with vowel shape in production, the results of perception experiment 2 are surprising. In this task, the participants completely ignored the phonetic dimension of pitch and focused solely on glottalization. The two perception experiments differed in two significant ways: natural vs. manipulated stimuli and isolation vs. phrase-final context. In this section I argue that there is little reason to expect that stimulus context can account for the different results and that the literature on categorical perception provides reasons to believe that stimulus quality does account for the differences.

The production of HIGH TONE and GLOTTALIZED vowels does differ by context, as documented in §2. Initial pitch is not as high for GLOTTALIZED vowels (relative to the initial pitch of HIGH TONE vowels) in words in phrase-final context (as opposed to words spoken in isolation). In fact, as shown in Fig. 6, HIGH TONE vowels actually have higher pitch for females in phrase-final context. Additionally, the speaker of the frame sentence and of the words that were manipulated to create the stimuli for this experiment produced HIGH TONE and GLOTTALIZED vowels with the same initial pitch values in phrase-final context (see Fig. 9). It could thus be the case that pitch is not a good cue to this contrast in phrase-final position.

There are three reasons why this is not a satisfactory explanation of the different results of the perception experiments. First, with regard to initial pitch, it was shown that, on average, males do produce higher initial pitch values for GLOTTALIZED vowels even in phrase-final context. Second, vowel length is only marginally correlated with vowel shape but was readily used by listeners when responding to natural stimuli. This shows that listeners are willing to use phonetic dimensions that are not the most robust indicators of contrast. Third, pitch span is highly correlated with vowel shape regardless of context and gender of speaker; GLOTTALIZED vowels tend to have larger pitch spans than HIGH TONE vowels. The pitch categories of L, ML, MH, and H represent both increasingly higher initial pitch values and increasingly larger pitch spans (with the exception of some stimuli with MH initial pitch having a smaller pitch span than those with ML initial pitch, as discussed in §4.1.2.1).

I believe that the different results of the perception experiments cannot be attributed

to a difference of context (isolation vs. phrase-final) and must therefore be attributed to a difference of stimulus quality. One effect of stimulus quality on perception has been documented in the literature on *categorical perception*, or the ability to distinguish between phonemic categories coupled with an inability to distinguish among phonetically different members of a phonemic category (see Liberman et al. 1957). Van Hessen and Schouten (1999: 58) show that categorical perception increases as stimulus quality increases, and they conclude that "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."

If we consider this quote in light of the different results of the perception experiments, it seems clear that participants in experiment 1 (with natural stimuli) "had to listen to the full spectrum and all its subtle interacting cues", while participants in experiment 2 did not. So why did the participants in this experiment listen for glottalization and not pitch? What was the basis of this decision: (universal) cognitive processes or a (language-specific) grammar? Or were they just guessing?

It is unlikely that participants were simply guessing because all participants behaved similarly; for each participant individually, there is a nonsignificant effect of pitch, and for all but four participants, there is a significant effect of glottalization. This means that perhaps these four participants were confused by the manipulated stimuli and hence just guessed, but it is unlikely that if all participants were guessing that their decisions would be correlated with one phonetic dimension and never the other.

The question that remains is whether or not the perceptual behavior exhibited in perception experiment 2 is determined by the (language-specific) grammar or some universal bias. It is possible that some aspect of human anatomy and/or cognitive abilities pushed the speakers to listen for glottalization and not other cues, i.e. it is possible that somehow glottalization is naturally a more salient feature or that it was the only readily available cue in the manipulated stimuli. This is unlikely as, to my ear and to the ear of other English speakers who have listened to the stimuli, the pitch differences among the four categories are quite striking (see Fig. 9). Further experimentation would need to be done in order to determine how non-YM speakers might react to the stimuli of perception experiment 2, but in the absence of such experimentation I find these "universal" explanations unsatisfying. Furthermore, the patterns of production in YM suggest that, in this language, glottalization is a more important cue than pitch in distinguishing GLOTTALIZED and HIGH TONE vowels, and so this fact may explain why listeners focus on glottalization alone when responding to manipulated stimuli.

In Table 4 we saw the percentage of times each glottalization type was produced as either a GLOTTALIZED or HIGH TONE vowel. This table is repeated in table 7, along with information on how often different categories of initial pitch and pitch span were produced as either a GLOTTALIZED or HIGH TONE vowel. What we see in this table is that glottalization is the only cue that, if used by itself, is likely to lead the listener to the correct decision. If the listener decides to draw a line and categorically perceive all modal voiced vowels as HIGH TONE vowels and all vowels with weak glottalization, creaky voice, or a glottal stop as GLOTTALIZED vowels, they will be right a high percentage of the time (74% of time for modal voiced tokens, and 93%, 95%, and 100% of the time for tokens with weak glottalization, creaky voice, or a full glottal stop, respectively). Initial pitch and pitch span, on the other hand, will only allow the listener to perform slightly better than chance. If the listener decides to interpret all tokens with an initial pitch of 2 s/b or less as a HIGH TONE vowel and all tokens with

an initial pitch of more than 2 s/b as a GLOTTALIZED vowel, they will only be right about half the time. If the listener decides that the demands of the linguistic situation are such that all cues are not available and hence decides to focus on only one cue, glottalization is the cue that will lead to the highest rate of success in discriminating HIGH TONE from GLOTTALIZED vowels in YM.

Table 7: Percentage of times each phonetic category is produced as a GLOTTALIZED vowel or a HIGH TONE vowel

	glottalization type			
	mod.	w.g.	cr.	g.s.
GLOTTALIZED	26%	87%	91%	100%
HIGH TONE	74%	13%	9%	0%
	initial pitch category (s/b)			
	< 0	0 – 2	2 – 4	> 4
GLOTTALIZED	42%	38%	54%	65%
HIGH TONE	58%	62%	46%	35%
	pitch span category (semitones)			
	< 2	2 – 4	4 – 6	> 6
GLOTTALIZED	37%	35%	49%	66%
HIGH TONE	63%	65%	51%	34%

In YM, the production of glottalization is more closely correlated with underlying vowel shape than the other cues are. The native language user will have learned this in the language acquisition process, and I propose that this knowledge is used to adapt the grammar for degraded language situations. In the case of perception experiment 2, participants assessed the quality of the stimuli, concluded that there were not enough available cues for following the standard perception grammar, and thus focused on the one cue that is most likely to lead to success – glottalization. This was an informed choice based on language-specific knowledge.

Nittrouer (2002: 719) documents how English speakers learn to use the cues that are the most informative in distinguishing among fricatives and concludes that "children learn what information in the signal must be extracted in order to apprehend phonetic structure in the language they are acquiring." The cues that will help the listener in "apprehending phonetic structure" are of course language-specific and can be influenced by phoneme inventory (Wagner et al. 2006). It is thus uncontroversial that the YM listener will naturally attend to those phonetic dimensions that have been determined to be helpful in identifying what the speaker said (given the phoneme inventory and phonetic properties of such phonemes in YM). The interesting result here is that, not only have native speakers learned to use certain phonetic cues, but they have also learned to focus on the single most reliable cue when necessary.

It should be clear that this altered perception grammar is useful for more than just laboratory tasks. Real language users communicate in non-ideal settings. They shout across long distances, talk over extraneous noise, or maybe even talk with their mouths full of food. It seems natural that listeners would want to adapt to such situations with a perception grammar that will facilitate comprehension in the face of these obstacles.

While it is beyond the scope of this paper to work out exactly what mechanisms alter the grammar in such degraded situations, the idea of a perception grammar that is specific to degraded linguistic situations has implications for production in the context of the

Bidirectional Model. If glottalization is the preferred cue to perceiving the contrast between HIGH TONE and GLOTTALIZED vowels, it should also be the preferred to cue for emphasizing this contrast in production. In other words, perhaps when YM speakers try to disambiguate a word with a GLOTTALIZED vowel from a word with HIGH TONE vowel in a noisy room, they will emphasize glottalization (and not pitch). To generalize, we should be able to find symmetries between the cues used by listeners in degraded language settings and those emphasized by speakers in the same types of situations. Such experiments have not been conducted to my knowledge, and this is an open area of research that is motivated by the idea that listeners purposefully alter their perception grammar in the face of less than ideal stimuli.

5. Conclusions

The production experiments reported in §2 showed that multiple phonetic properties are used to distinguish HIGH TONE from GLOTTALIZED vowels in YM. In increasing order of importance, GLOTTALIZED vowels tend to be longer, to have higher initial pitch, to have a larger pitch span, and to be produced with more glottalization. There are gender-based differences in the production of GLOTTALIZED vowel related to the interaction of pitch and creaky voice. Creaky voice causes a dramatic dip in f0 for females relative to their natural pitch range, while the f0 produced during creaky voice is within the natural pitch range of males.

In perception experiment 1 (§3), participants responded to natural stimuli and their decisions were influenced by length, pitch (both initial pitch and pitch span), and glottalization. This means that the phonetic values that distinguish HIGH TONE form GLOTTALIZED vowels in production are used by the listener as cues to the contrast. This result is predicted by bidirectional models of phonetics and phonology (such as Bidirectional Stochastic OT (Boersma 2007)) which state that the production and perception grammars relate phonetic values to phonological forms via constraints with the same ranking in both grammars. Another advantage of the Bidirectional Model is that it can account for the prototype effect. In YM, a full glottal stop is an excellent perceptual cue but is rarely produced, meaning that speakers are avoiding the production of peripheral tokens even when such tokens are the most likely to be accurately perceived by the listener.

The results of perception experiment 2 (§4) showed that listeners attended to glottalization alone when responding to manipulated stimuli. I propose that this was a language-specific choice (and hence controlled by the perception grammar) such that, in the face of a degraded linguistic situation, listeners focused on the one cue that is most likely to lead to success. This means that the grammar of a language must encode information on the reliability of certain phonetic dimensions in signaling contrast.

References

Avelino, Heriberto, Eurie Shin, Sam Tilsen, Reiko Kataoka, and Jeff Pynes. 'The phonetics of laryngealization in Yucatec Maya.' Paper presented at the LSA annual meeting in Anaheim, California, 2007.

Blair, Roberto W., and Refugio Vermont Salas. *Spoken Yucatec Maya,* Book I: Lessons 1-6, with revisions by Refugio Vermont Salas (1968, 1994), Norman A. McQuown (1994); 1995 version. Chapel Hill, NC: Duke University-University of North Carolina Program in Latin American Studies, 1965.

Blankenship, Barbara. "The time course of breathiness and laryngealization in vowels". PhD diss, UCLA, 1997.

Blomgren, Michael, Yang Chen, Manwa L. Ng, and Harvey R. Gilbert. "Acoustic, aerodynamic, physiologic, and perceptual properties of modal and vocal fry registers." *Journal of the Acoustical Society of America* 103 (1998): 2649-2658.

Boersma, Paul. "Prototypicality judgements as inverted perception." In *Gradience in Grammar*, edited by Gisbert Fanselow, Caroline Féry, Matthias Schlesewsky, and Ralf Vogel, 167-184. Oxford: Oxford University Press, 2006.

Boersma, Paul. "Some listener-oriented accounts of *h*-aspiré in French." *Lingua* 117 (2007): 1989-2054.

Boersma, Paul, and David Weenink. Praat: doing phonetics by computer (Version 4.4.05), 2006. [Computer program]. Retrieved from http://www.praat.org/.

Bricker, Victoria, Eleuterio Poʔot Yah, and Ofelia Dzul Poʔot. *A dictionary of the Maya language as spoken in Hocabá, Yucatán*. Salt Lake City: University of Utah Press, 1998.

Cho, Taehong, and Peter Ladefoged. "Variation and universals in VOT: evidence from 18 languages." *Journal of Phonetics* 27 (1999): 207-229.

Escudero, Paola, and Paul Boersma. "Bridging the gap between L2 speech perception research and phonological theory." *Studies in Second Language Acquisition* 26 (2004): 551-585.

Frazier, Melissa. "Tonal dialects and consonant pitch interaction in Yucatec Maya." In *New Perspectives in Mayan Linguistics*, edited byHeriberto Avelino. Cambridge: Cambridge University Press, In Press.

Frazier, Melissa. "The production and perception of pitch and glottalization in Yucatec Maya." PhD diss., University of North Carolina, Chapel Hill, 2009.

Gordon, Matthew, and Peter Ladefoged. "Phonation types: a cross-linguistic overview." *Journal of Phonetics* 29 (2001): 383-406.

Gordon, Raymond G., Jr., ed. *Ethnologue: languages of the world*, 15th ed. Dallas: SIL International, 2005. Online version: http://www.ethnologue.com/.

Johnson, Kieth, Edward Flemming, and Richard Wright. "The hyperspace effect: phonetic targets are hyperarticulated." *Language* 69 (1993): 505–528.

Liberman, Alvin M., Katherine Safford Harris, and Belver C. Griffith. "The discrimination of speech sounds within and across phoneme boundaries." *Journal of Experimental Psychology* 54 (1957): 358-368.

Liberman, A. M., P. C. Delattre, and F. S. Cooper. "Some cues for the distinction between voiced and voiceless stops in initial position." *Language and Speech* 1 (1958), 153-167.

Lisker, Leigh, and Arthur S. Abramson. "Cross-language study of voicing in initial stops: acoustical measurements." *Word* 20 (1964): 384-422.

Lisker, Leigh, and Arthur S. Abramson. "The voicing dimension: some experiments in comparative phonetics." In *Proceedings of the Sixth International Congress of Phonetic*

Sciences, edited by Bohuslav Hála, Milan Romportl, & Přemysl Janota, 563-567. Prague: Academia, 1970.
Nittrouer, Susan. "Learning to perceive speech: how fricative perception changes, and how it stays the same." *Journal of the Acoustical Society of America* 112 (2002), 711-719.
Nolan, Francis. "Intonational equivalence: an experimental evaluation of pitch scales." *Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona* (2003): 771-774.
Peterson, Gordon E., and Harold L. Barney. Control methods used in a study of the vowels." *Journal of the Acoustical Society of America* 24 (1952): 175–184.
Peterson, Gordon E., and Ilse Lehiste. "Duration of syllable nuclei in English." *Journal of the Acoustical Society of America* 32 (1960): 693–703.
Pierrehumbert, Janet. "The phonology and phonetics of English intonation." PhD diss., MIT, 1980.
Redi, Laura, and Stefanie Shattuck-Hufnagel. "Variation in the realization of glottalization in normal speakers." *Journal of Phonetics* 29 (2001): 407-429.
van Hessen, A. J, and M. E. H. Schouten. "Categorical perception as a function of stimulus quality." *Phonetica* 56 (1999): 56-72.
Wagner, Anita, Mirjam Ernestus, and Anne Cutler. "Formant transitions in fricative identification: the role of native fricative inventory." *Journal of the Acoustical Society of America 120* (2006): 2267-2277.